ADVANCED THEORY OF STATISTICS

1. Distribution Theory
2. Estimation
3. Tests of Hypotheses

Lectures delivered by D. G. Chapman
(Stat 611-612)
1958-1959
North Carolina State College
Notes recorded by R. C. Taeuber

Institute of Statistics
Mimeo Series No. 241
September, 1959
# TABLE OF CONTENTS

## I. PRELIMINARIES

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concepts of set theory</td>
<td>1</td>
</tr>
<tr>
<td>Probability measure</td>
<td>3</td>
</tr>
<tr>
<td>Distribution function (Th. 1)</td>
<td>5</td>
</tr>
<tr>
<td>Density function (def. 8)</td>
<td>8</td>
</tr>
<tr>
<td>Random variable (def. 9)</td>
<td>10</td>
</tr>
<tr>
<td>Conditional distribution</td>
<td>10</td>
</tr>
<tr>
<td>Functions of random variables</td>
<td>12</td>
</tr>
<tr>
<td>Transformations</td>
<td>12</td>
</tr>
<tr>
<td>Riemann – Stieljes Integrals</td>
<td>15</td>
</tr>
</tbody>
</table>

## II. PROPERTIES OF UNIVARIATE DISTRIBUTIONS — CHARACTERISTIC FUNCTIONS

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard distributions</td>
<td>20</td>
</tr>
<tr>
<td>Expectation and moments</td>
<td>21</td>
</tr>
<tr>
<td>Tchebycheff's Inequality (Th. 8 + corollary)</td>
<td>24</td>
</tr>
<tr>
<td>Characteristic functions</td>
<td>25</td>
</tr>
<tr>
<td>Cumulant generating functions (def. 17)</td>
<td>28</td>
</tr>
<tr>
<td>Inversion theorem for characteristic functions (Th. 12)</td>
<td>30</td>
</tr>
<tr>
<td>Central Limit Theorem (Th. 14)</td>
<td>33</td>
</tr>
<tr>
<td>Laplace Transform</td>
<td>36</td>
</tr>
<tr>
<td>Fourier Transform</td>
<td>36</td>
</tr>
<tr>
<td>Derived distributions (from Normal)</td>
<td>36</td>
</tr>
</tbody>
</table>

## III. CONVERGENCE

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergence of distributions</td>
<td>39</td>
</tr>
<tr>
<td>Convergence in probability (def. 18)</td>
<td>41</td>
</tr>
<tr>
<td>Weak law of large numbers (Th. 16)</td>
<td>41</td>
</tr>
<tr>
<td>A Convergence Theorem (Th. 18)</td>
<td>44</td>
</tr>
<tr>
<td>Convergence almost everywhere (def. 20)</td>
<td>49</td>
</tr>
<tr>
<td>Strong law of large numbers (Th. 20)</td>
<td>51</td>
</tr>
<tr>
<td>Other types of convergence</td>
<td>53</td>
</tr>
</tbody>
</table>

## IV. POINT ESTIMATION

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preliminaries</td>
<td>55</td>
</tr>
<tr>
<td>Ways to formulate the problem of obtaining best estimates</td>
<td>56</td>
</tr>
<tr>
<td>Restricting the class of estimates</td>
<td>56</td>
</tr>
<tr>
<td>Optimum properties in the large</td>
<td>58</td>
</tr>
<tr>
<td>Dealing only with large sample (asymptotic) properties</td>
<td>60</td>
</tr>
<tr>
<td>Methods of Estimation</td>
<td>61</td>
</tr>
<tr>
<td>Method of moments</td>
<td>61</td>
</tr>
<tr>
<td>Method of least squares</td>
<td>61</td>
</tr>
<tr>
<td>Gauss-Markov Theorem (Th. 21)</td>
<td>61</td>
</tr>
<tr>
<td>Maximum Likelihood Estimates</td>
<td>64</td>
</tr>
<tr>
<td>Single parameter estimation (Th. 22)</td>
<td>65</td>
</tr>
<tr>
<td>Multiparameter estimation (Th. 23)</td>
<td>70</td>
</tr>
</tbody>
</table>
### Unbiased Estimation
- Information Theorem (Cramer-Rao) (Th. 24) ........................................ 76
- Multiparameter extension of Cramer-Rao (Th. 25) ...................................... 80
- Summary of usages and properties of m.l.e. ................................................ 85
- Sufficient statistics (def. 31) .......................................................... 86
- Improving unbiased estimates (Th. 26) ...................................................... 88
- Complete sufficient statistics (Th. 27) ...................................................... 92
- Non-parametric estimation (m.v.u.e.) ....................................................... 95

### \( \chi^2 \)-estimation .................................................................
- Maximum likelihood estimation ......................................................... 102
- Minimum \( \chi^2 \) .............................................................. 102
- Modified minimum \( \chi^2 \) ......................................................... 103
- Transformed minimum \( \chi^2 \) ......................................................... 103
- Summary .............................................................. 104
- Examples ............................................................ 105
- Minimax estimation ........................................................... 108
- Wolfowitz's minimum distance estimation .............................................. 110

### V. TESTING OF HYPOTHESES -- DISTRIBUTION FREE TESTS

- Basic concepts .......................................................... 113
- Distribution free tests ..................................................... 115
- One sample tests ........................................................ 117
- Probability transformation .................................................. 121
- Tests for one sample problems ............................................. 121
- Convergence of sample distribution function (Th. 31) .......................... 124
- Approaches to combining probabilities ........................................... 129
- Two sample tests ......................................................... 133
- Pitman's theorem on asymptotic relative efficiency (Th. 32) ................. 140

### VI. PARAMETRIC TESTS -- POWER

- Optimum tests .......................................................... 148
- Probability ratio test (Neyman-Pearson) (Th. 33) ................................. 149
- U.M.P.U. test (Th. 35) ..................................................... 156
- Invariant tests ........................................................ 161
- Distribution of non-central \( t \) (\( \tau \)) ........................................... 163
- Tests for variances ......................................................... 168
- Summary of normal tests .................................................... 169
- Maximum likelihood ratio tests ................................................. 173
- General Linear Hypothesis ...................................................... 175
- Power of analysis of variance test .............................................. 182
- Multiple Comparisons ....................................................... 186
- Tukey's procedure ......................................................... 186
- Scheffe's test for all linear contrasts (Th. 38) .................................. 187
- \( \chi^2 \)-tests ............................................................. 191
- Power of a simple \( \chi^2 \)-test ................................................ 191
- One sided \( \chi^2 \)-test ...................................................... 192
VII. MISCELLANEOUS: CONFIDENCE REGIONS

Correspondence of confidence regions and tests (Th. 40) 199
Confidence interval for the ratio of two means 200

List of Theorems 202
List of Definitions 204
List of Problems 206
I: Preliminaries:

Set: A collection of points in $\mathbb{R}^k$ (Euclidean $k$-dimensional space) = $S$

Def. 1/ $S_1 + S_2$ is the set of points in either or both sets.

$S_1 + S_2$ is the set of points in both $S_1$ and $S_2$.

if $S_1$ is contained in $S_2$ ($S_1 \subset S_2$) then $S_2 - S_1$ is the set of points in $S_2$ but not in $S_1$.

Exercise 1/ Show that $S_1 + S_2 = S_2 + S_1$

$S_1 S_2 = S_2 S_1$

There is also the obvious extension of these definitions of set addition and multiplication to 3 or more sets.

$$\sum_{n=1}^{\infty} S_n = \text{the set of points in at least one of the } S_n$$

$$\prod_{n=1}^{\infty} S_n = \text{the set of points common to all the } S_n$$

$S^*$ is defined as the complement of $S$ and is the same as $\mathbb{R}^k - S$.

Lemma 1/ $(S_1 + S_2)^* = S_1^* S_2^*$

Proof: Let $\epsilon$ denote "is an element of"

$x \in (S_1 + S_2)^*$ means that $x$ is not a member of either $S_1$ or $S_2$,

i.e. $x \in S_1^*, x \in S_2^*$

therefore $x \in S_1^*, x \in S_2^*$

since $x$ is common to both $S_1^*, S_2^*$, $x \in S_1^* S_2^*$.
To complete the proof

\[ x \in S_1^* S_2^* \quad \Rightarrow \quad x \in S_1^* \text{ and } x \in S_2^* \]
\[ \Rightarrow x \notin S_1^* \text{ and } x \notin S_2^* \]
\[ \Rightarrow x \notin (S_1 + S_2)^* \]
\[ \Rightarrow x \in (S_1 + S_2)^* \]

**Exercise 2**
Show that \( S_2 \cap S_1 = S_2 S_1^* \).

**Exercise 3**
In \( \mathbb{R}^2 \) define \( S_1 = \{x, y : x^2 + y^2 \leq 1\} \)

i.e., the set of points \( x, y \) subject to the restriction \( x^2 + y^2 \leq 1 \)

\[ S_2 = \{x, y : |x| \leq 0.8, |y| \leq 0.8\} \]

\[ S_3 = \{x, y : x = 0\} \]

Represent graphically \( S_1 + S_2, S_1^* S_2^*, S_2 S_1^* S_1 S_2 = S_1 S_3^* \).

**Def. 2**
If \( S_1 < S_2 < S_3 \ldots \ldots \) (an exploding family)

We define:
\[ \lim_{n \to \infty} S_n = \sum_{n=1}^{\infty} S_n \]

And if \( S_1 \supset S_2 \supset S_3 \ldots \ldots \) (a nested family)

We define:
\[ \lim_{n \to \infty} S_n = \prod_{n=1}^{\infty} S_n \]

Such sequences of sets are called monotone sets.

**Exercise 4**

(a) Show that the closed interval in \( \mathbb{R}^2 \) \( \{x, y : |x| \leq 1, |y| \leq 1\} \) may be represented as an infinite product of a set of open intervals.

**Ans:** \( S_n = \{x, y : |x| < 1 + \frac{1}{n}, |y| < 1 + \frac{1}{n}\} \)

(b) Show that the open interval in \( \mathbb{R}^2 \) \( \{x, y : |x| < 1, |y| < 1\} \) can be represented as an infinite sum of closed intervals.
Ans.: \( S_n = \{ x, y : |x| \leq 1 - \frac{1}{n}, |y| \leq 1 - \frac{1}{n} \} \)

Probability is generally thought of in terms of sets, which is why we study sets.

Def. 3/ Borel Sets --- the family of sets which can be obtained from the family of intervals in \( \mathbb{R}^k \) by a finite or enumerable sequence of operations of set addition, multiplication, or complementation are called Borel Sets.

The word multiplication could be deleted, since multiplication can be performed by complementation, e.g.:

\[
(S_1 + S_2)^* = S_1^* S_2^* \quad (a_1^* + S_2^*)^* = S_1^* S_2^* (S^*)^* = S
\]

Def. 4/ \( A(S) \) is an additive set function if

1/ for each Borel Set \( A(S) \) is a real number, and
2/ if \( S_1, S_2, \ldots \) are a sequence of disjoint sets

\[
A \left( \bigcup_{n=1}^{\infty} S_n \right) = \sum_{n=1}^{\infty} A(S_n).
\]

Examples:

--- area is a set function

--- in \( \mathbb{R}^1 \) \( A(S) = \int f(x) \, dx \) \( A_1(S) = \int x \, dx \)

Def. 5/ \( P(S) \) is a probability measure on \( \mathbb{R}^k \) if

1/ \( P \) is an additive set function
2/ \( P \) is non-negative
3/ \( P(\mathbb{R}^k) = 1 \)

\( \emptyset \) will denote the empty set which contains no points, i.e. \( \emptyset = \mathbb{R}^k^*; \emptyset + \mathbb{R}^k = \mathbb{R}^k \).

Ex. 5/ \( P(\emptyset) = 0 \)

Ex. 6/ if \( S_1 \subseteq S_2 \) then \( P(S_1) \leq P(S_2) \)

Lemma 2/ \( P(S_1 + S_2 + \ldots) \leq P(S_1) + P(S_2) + \ldots \)

Problem: Prove lemma 2.
Lemma 3: \( P(\lim_{n \to \infty} S_n) = \lim_{n \to \infty} P(S_n) \) if \( S_n \) is a monotone sequence.

Proof: case 1: \( S_1 \subset S_2 \subset S_n \ldots \ldots \)

Define: \( S_1' = S_1 \)
\( S_2' = S_2 - S_1 \)
\( S_3' = S_3 - S_2 \) etc.

These sets \( S_n' \) are disjoint; also \( \sum_{n=1}^{\infty} S_n = \sum_{n=1}^{\infty} S_n' \)

\[
P(\lim_{n \to \infty} S_n) = P\left(\sum_{n=1}^{\infty} S_n\right)
\]
\[
= P\left(\sum_{n=1}^{\infty} S_n'\right)
\]
\[
= \sum_{n=1}^{\infty} P(S_n')
\]
the additive property of \( P \)

\[
= P(S_1') + P(S_2') + P(S_3') + \ldots
\]
\[
= P(S_1) + P(S_2 - S_1) + P(S_3 - S_2) \ldots
\]
\[
= P(S_1) + P(S_2) - P(S_2) + P(S_3)
\]
etc.
\[
= P(S_n) \text{ after n steps}
\]
\[
= \lim_{n \to \infty} P(S_n)
\]

Case 2: \( S_1 \supset S_2 \supset S_3 \ldots \ldots \)

Problem 2: Prove lemma 3 for case 2.

Def. 6: Associated with any probability measure \( P(S) \) on \( R \) there is a point function \( F(x) \) defined by

\[
F(x) = P(-\infty, x].
\]

\( F(x) \) is called a distribution function --- d.f.
Theorem 1: Any distribution function, \( F(x) \), has the following properties:
1. It is a monotone, non-decreasing sequence.
2. \( F(-\infty) = 0 \), \( F(+\infty) = 1 \)
3. \( F(x) \) is continuous on the right.

Proof: 1/ For \( x_1 < x_2 \) we have to show that \( F(x_1) \leq F(x_2) \)

\[
F(x_1) = P(-\infty, x_1) \quad F(x_2) = P(-\infty, x_2)
\]

The interval \((-\infty, x_1) \subseteq \text{the interval } (-\infty, x_2)\)

From exercise 6 we have that \( P(I_1) \leq P(I_2) \)

Therefore \( F(x_1) \leq F(x_2) \)

2/a/ If we define \( G_n \) as the interval \((-\infty, -n) \) \( n=1,2,3, \ldots \)

Then \( G_1 \supset G_2 \supset G_3 \ldots \)

\( G = \lim_{n \to \infty} G_n = \emptyset \) (the empty set)

\[
\lim_{n \to \infty} F(n) = \lim_{n \to \infty} P(G_n) = P(\lim_{n \to \infty} G_n)
\]

From lemma 3

\[
= P(G) = P(\emptyset) = 0
\]

b/ Follows in a similar fashion by defining \( G_n = (-\infty, n) \)

3/ Pick a point \( a \) -- for this point we want to show

\[
\lim_{x \to a} F(x) = F(a).
\]

Consider a nested sequence \( \varepsilon_n \to 0, \varepsilon_n > 0 \).

Then \( \lim_{n \to \infty} F(a + \varepsilon_n) = F(a) \) is the property to be shown.

If we define \( H_n = (-\infty, a + \varepsilon_n) \) \( n=1,2,3, \ldots \)

\[
\lim_{n \to \infty} H_n = H = \text{interval } (-\infty, a)
\]

\[
\lim_{n \to \infty} P(H_n) = P(\lim H_n) = P(H)
\]

lemma 3

Therefore \( \lim_{n \to \infty} F(a + \varepsilon_n) = F(a) \)
Problem 3: Show that

\[ F(a) = \lim_{x \to a} F(x) + P\{[a]\} \quad \text{where } [a] \text{ is the set whose only point is } a. \]

Or in familiar terms \( F(a) = F(x = 0) + Pr(x = a) \)

where \( F(x = 0) \) is the limit from the left.

Theorem 2: To any point function \( F(x) \) satisfying properties 1, 2, and 3 of theorem 1, there corresponds a probability measure \( P(S) \) defined for all Borel sets such that for any interval \( (-\infty, x) \)

\[ P(-\infty, x) = F(x), \]

Proof omitted — see Cramer p. 53 referring to p. 22.

Theorem 3: A distribution function \( F(x) \) has at most a countable number of discontinuities.

Proof: Let \( v_n \) be the number of points of discontinuity with a jump \( \frac{1}{n} \)
then \( v_n \leq n \) which is what we have to show.

Suppose the contrary holds, i.e., \( v_n > n \).

Then if we let \( S_n \) be the set of such discontinuities, we have

\[ 1 - F(r_1) \geq P(S_n) > \frac{1}{n} (n) \geq 1 \]

which is a contradiction.

Therefore, the total number of discontinuities \( \sum_{n=1}^{\infty} v_n < \sum_{n=1}^{\infty} n \)

where \( \sum_{n=1}^{\infty} n \) is the sum of the integers which is a countable sum.

Notation:

[ ] square brackets — means the end point is not included in the interval — i.e., an open interval.

( ) round brackets — mean the end points are included in the interval, i.e., a closed interval.

\( (a, b] \) is the interval \( a \) to \( b \), including \( a \) but not \( b \).

Def. 7: In \( R_k \) to each probability measure \( P(S) \) there corresponds a unique distribution function

\[ F(x) = F(x_1, x_2, \ldots, x_k) \]

\[ = F\left[ \text{interval } (-\infty, -\infty, \ldots; x_1, x_2, \ldots, x_k) \right] \]
The interval is the set of points in $\mathbb{R}_k$

$$-\infty < x_1 \leq x_i \quad i = 1, 2, \ldots, k$$

**Theorem 4:** $F(x_1, x_2, \ldots, x_k)$ has the following properties:

1. It is continuous from the right in each variable
2. $F(-\infty, x_2, \ldots, x_k) = F(x_1, -\infty, x_3, \ldots, x_k) = \ldots = 0$
   
   $F(\infty, +\infty, \ldots, +\infty) = 1$

3. $\Delta_k F(a_1, a_2, \ldots, a_k) \geq 0$ 
   
   i.e., the $P$ measure of any interval is non-negative.

Conversely, if $F(x_1, x_2, \ldots, x_k)$ has these properties, then there is a unique $P$-measure defined by

$$F(I) = F(x_1, x_2, \ldots, x_k).$$

That is, $I$ is the interval $[-\infty, -\infty, \ldots, x_1, x_2, \ldots, x_k]$.

In $\mathbb{R}_2$

$$F(I) = F(a_1 + h_1, a_2 + h_2) - F(a_1, a_2 + h_2) - F(a_1 + h_1, a_2) + F(a_1, a_2)$$

$$= \Delta_2 F(a_1, a_2)$$
\[ F(I) = F(a_1 + h_1, a_2 + h_2, a_3 + h_3) \]

\[ F(a_1, a_2 + h_2, a_3 + h_3) = F(a_1 + h_1, a_2, a_3 + h_3) \]

\[ F(a_1 + h_1, a_2 + h_2, a_3) \]

\[ F(a_1, a_2 + h_2, a_3) + F(a_1 + h_1, a_2, a_3) \]

\[ F(a_1, a_2, a_3) \]

\[ \Delta_3 F(a_1, a_2, a_3) \]

The proof of theorem 4 is by analogy with the linear case (theorem 1),
\[ F(x_1, + \infty, \ldots, + \infty) = F([- \infty, x_1]) \]

\[ = F_1(x_1) \]

\[ = \text{the marginal distribution of } x_1 \]

(similarly for other dimensions)

\text{Def. 8:}

If \( F \) is continuous and differentiable in all variables, then

\[ \frac{\partial^k F}{\partial x_1 \partial x_2 \ldots \partial x_k} = f(x_1, x_2, \ldots x_k) \]

is the density function of \( x_1, x_2, \ldots, x_k \).

\text{Exercise 7:}

In \( \mathbb{R}_2 \)
\[ F(x, y) = 0 \text{ if } x \leq 0 \text{ or } y \leq 0 \]
\[ = \frac{(x+y)}{2} \text{ for } 0 < x \leq 1 \]
\[ 0 < y \leq 1 \]
\[ = 1 \text{ for } x > 1, y > 1 \]

Can this be a distribution function in \( \mathbb{R}_2 \)?

How can the definitions be completed?
Solution -- consider the marginal distribution of $x$

$$F_1(x) = F(x, +\infty) \geq F(x, 1) = \frac{x + 1}{2}$$

$$F_1(0) = \frac{(0 + 1)}{2} = \frac{1}{2}$$

But in fact $F(0, 0) = 0$

$$F(0, y) = 0 \text{ for all } y$$

$$F_1(0) = 0$$

Therefore there is a contradiction. $F$ cannot be a proper distribution function.

If $F(x, y)$ is a proper distribution function, then the two marginal distributions

$$F_1(x) = F(x, +\infty)$$

$$F_2(y) = F(+\infty, y)$$

must be proper and in this case they break down.

**Problem 4:** If we define $f(x, y) = x + y$

$$0 \leq x \leq 1$$

$$0 \leq y \leq 1$$

$$= 0 \text{ elsewhere}$$

Find $F(x, y); F_1(x); \text{ and } F_2(y)$

Show that $F(x, y)$ satisfies the properties of a distribution function.
Def. 9: Random Variable

We assume we have experiments which yield a vector valued set of observations

\[ \mathbf{X} = (X_1, X_2, \ldots, X_k) \] with the properties:

1. For each Borel set \( S \) in \( \mathbb{R}_k \) there is a probability measure \( P(S) \) which is the probability that the whole vector \( \mathbf{X} \) falls in the set in \( S \)

\( \text{\( (P(S) \) is non-negative, additive, and \( P(R_k) = 1 \).} \)

Cramer p. 152-4
axioms 1 and 2

2. If \( X_1, \ldots, X_n \) are random variables in \( \mathbb{R}_{k_1}, \mathbb{R}_{k_2}, \ldots, \mathbb{R}_{k_n} \)

then the combined vector \( (X_1, X_2, \ldots, X_n) \) is also a random variable

in \( \mathbb{R}_{k_1 + k_2 + k_3 \ldots + k_n} \).

Conditional Distribution

\[ (X, Y) \] are random variables in \( \mathbb{R}_{k_1}, \mathbb{R}_{k_2} \).

Let \( S \) and \( T \) be sets in \( \mathbb{R}_{k_1}, \mathbb{R}_{k_2} \).

Def. 10: If \( P(\ X \text{ belongs to } S) > 0 \), then we define conditional probability

\[ P(Y \subseteq T \mid X \subseteq S) = \frac{P(Y \subseteq T, X \subseteq S)}{P(X \subseteq S)} \].

We show that \( P(Y \subseteq T \mid X \subseteq S) \) does satisfy the requirements of a probability measure

1. It is non-negative since \( P(Y \subseteq T, X \subseteq S) \) is non-negative.
2. It is additive since \( P(Y \subseteq T, X \subseteq S) \) is additive in \( \mathbb{R}_{k_2} \).

\[ P(Y \subseteq T_1 \mid X \subseteq S) + P(Y \subseteq T_2 \mid X \subseteq S) = P(Y \subseteq T_1 + T_2 \mid X \subseteq S) \]

3. \[ \frac{P(Y \subseteq R_k, X \subseteq S)}{P(X \subseteq S)} = \frac{P(X \subseteq S)}{P(X \subseteq S)} = 1 \]

If \( P(Y \subseteq T) > 0 \) we could also define

\[ P(X \subseteq S \mid Y \subseteq T) = \frac{P(X \subseteq S, Y \subseteq T)}{P(Y \subseteq T)} \]
In familiar terminology what we are saying is that

\[ \Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} \quad \text{or} \quad \Pr(A, B) = \Pr(A \cap B) \Pr(B). \]

If we have the corresponding distribution functions

\[ F(x, y); F_1(x) = F(x, +\infty); \text{ and } F_2(y) = F(+\infty, y) \]

**Def. 11:** \( X \) and \( Y \) are independent random variables if the joint distribution function \( F(x, y) \) factors into \( F_1(x) F_2(y) \).

See p. 160 Cramer — he goes first to probability measures, then to d.f.

---

**notation:**

--- Capital Latin letters used for random variables in general.
--- Small Latin letters used for observations or specific values of the random variables.

\[ \Pr(X \leq x) = F(x) \]

**Def. 11 — extension:**

In the case of \( n \) random variables, \( X_1, X_2, \ldots, X_n \) these are independent if

\[ F(x_1, x_2, \ldots, x_n) = F_1(x_1) F_2(x_2) \cdots F_n(x_n). \]

**Note:** Three variables may be pairwise independent, but may not be (mutually) independent — see the example on p. 162 of Cramer.

If density functions exist, then \( X_1, X_2, \ldots, X_n \) independent means that

\[ f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n). \]

**Note:** The fact that the density functions factor does not necessarily mean independence since dependence may be brought in through the limits.

e.g., \( X \) and \( Y \) are distributed uniformly on \( OAB \)

\[ f(x, y) = \begin{cases} 2 & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases} \]

**Exercise 8:** Find \( F(x, y); F_1(x); \) and \( F_2(y). \)
Functions of Random Variables:

$g$ is a Borel measurable function if for each Borel set $S$, there is a set $S'$ such that

$x \in S' \implies g(x) \in S$

is also a Borel set.

A notation sometimes used is that $S' = g^{-1}(S)$ --- that $S'$ is the inverse image of $S$ under the mapping $g$.

Now consider $Y = g(X)$ where $X$ is a random variable.

$$\Pr(y \in S) = \Pr(x \in S') = \Pr(S')$$

Therefore, any Borel measurable function, $Y = g(X)$ of a random variable, $X$, is itself a random variable.

$$\Pr(y \in S) = \Pr(g^{-1}(S))$$

This extends readily to $k$ dimensions.

Transformations:

We have $X$ and $Y$ which are random variables with distribution function $F(x, y)$ and a density function $f(x, y)$.

Let $\alpha = \beta_1(X, Y)$  $\beta = \beta_2(X, Y)$

Where $\beta_1$ and $\beta_2$ are 1 to 1 with $\alpha$, $\beta$ are continuous, differentiable, and the Jacobian of the transformation is non-vanishing.

$$J = \begin{vmatrix} \frac{\partial x}{\partial \alpha} & \frac{\partial x}{\partial \beta} \\ \frac{\partial y}{\partial \alpha} & \frac{\partial y}{\partial \beta} \end{vmatrix}$$

We then have the inverse functions

$$X = \psi_1(\alpha, \beta)$$

$$Y = \psi_2(\alpha, \beta)$$

The density function $f(a, b)$ of the random variables $\alpha$, $\beta$ is

$$f[\psi_1(a, b), \psi_2(a, b)] \left| J \right|$$
However under the transformation the limits of the variables will be changed and these have to be worked out in each individual case. (See Anderson and Bancroft.)

**Problem 5**: X, Y are uniformly and independently distributed on (0, 1).

Find the distribution of Z = XY and -2 ln XY.

**Example**: For X, Y as in problem 5, find the distribution function of Z = X + Y.

**Solution**: Consider also W = X - Y

Consider the joint distribution of (Z, W) (0, 1)

\[
\begin{align*}
Z &= X + Y \\
W &= X - Y \\
\therefore \frac{Z + W}{2} &= X \quad \frac{Z - W}{2} = Y
\end{align*}
\]

\[
J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2}
\]

Density of Z, W = f(x, y) |J| = 1 \cdot \frac{1}{2}

The limits of Z and W are dependent

\[
\begin{align*}
Z &= X + Y \\
W &= X - Y
\end{align*}
\]

If Z = z, then W takes on values from (0 - z) thru 0 to z = 0, so that

\[
\text{for } Z = z \leq 1 \quad -z \leq W \leq z
\]

\[
f(z, w) = \frac{1}{2} \quad \text{with limits } -z \leq W \leq +z
\]

\[
F_1(z) = \int_{-z}^{+z} |J| \, dw = \int_{-z}^{+z} \left( \frac{1}{2} \right) \, dw = \frac{1}{2} \left. w \right|_{-z}^{+z} = z
\]

\[
F(z) = \int_{0}^{z} F(t) \, dt = \frac{z^2}{2} \quad 0 \leq z \leq 1
\]
If \( X = z > 1 \), then \( z \) takes on values from \((z - 1) = 1\) to \(1 - (z - 1)\) or from \((z-2)\) to \((2-z)\)

\[
f_1(z) = \sum_{z=2}^{2-z} \frac{1}{2} \, dz = 2 - z
\]

\[
F(z) = \frac{1}{2} + \int_{1}^{z} (2-z) \, dz = \frac{1}{2} - \frac{(2-z)^2}{2}
\]

\[
= \frac{1}{2} - \frac{(2-z)^2}{2} + \frac{1}{2} = 1 - \frac{(2-z)^2}{2}
\]

\[
\therefore F(z) = \frac{z^2}{2} \quad 0 \leq z \leq 1
\]

\[
= 1 - \frac{(2-z)^2}{2} \quad 1 \leq z \leq 2
\]

\[f(z) = z \quad 0 \leq z \leq 1\]

\[= 2 - z \quad 1 \leq z \leq 2\]

See p. 245, 6 in Cramer

Joint density of \( Z, W = \frac{1}{2} \)

\[
f(z, w) = \begin{cases} \frac{1}{2} & 0 \leq z \leq 1 \\ -z & z \leq w \leq z \\ \frac{1}{2} & 1 \leq z \leq 2 \\ z - 2 & 2 \leq w \leq 2 - z \end{cases}
\]
If the transformation is not 1 to 1 (that is $J = 0$) then the usual device to avoid the difficulties that may arise is to divide $R_k$ into regions in each of which the transformation is 1 to 1, and then work separately in each region.

i.e., consider in $R_1$ $Y = X^2$ We should consider 2 separate cases:

$$
\begin{align*}
X &< 0 \\
X &\geq 0
\end{align*}
$$

Cramer P. 167

Riemann-Stieltjes Integral:

Let $F(x)$ be a d.f., with at most a finite number of discontinuities in $(a, b)$ and let $g(x)$ be a continuous function, then we can define

$$
\int_a^b g(x) \, dF(x) \quad \text{as follows:}
$$

Cramer P. 71-74

Divide $(a, b)$ into $n$ sub-intervals $x_1, x_2, \ldots, x_n$ of length $\Delta$

Let

$$
S_n = \sum_{i=1}^{n} \left[ \inf_{x_{i-1} \leq x \leq x_i} g(x) \right] \left[ F(x_i) - F(x_{i-1}) \right]
$$

Also define

$$
\int_{-\infty}^{+\infty} g(x) \, dF(x) = \lim_{a \to -\infty} \lim_{b \to +\infty} \int_a^b g(x) \, dF(x)
$$

provided the limit exists

and in general

$$
\int_a^b g(x) \, dF(x) \text{ has all the usual properties of the familiar Riemann integral.}
$$

They can be shown to have a common limit.

So the common limit is called the R-S integral,

$$
\frac{S_n}{n} < \frac{S^-_n}{n} \quad \text{but as } n \to \infty, \Delta \to 0 \quad \frac{S_n}{n} \text{ is increasing, } \frac{S^-_n}{n} \text{ is decreasing}
$$
If \( F(x) \) has density \( f(x) \) which is continuous except at a finite number of points, then
\[
F(x_i) - F(x_{i-1}) = f(x') (x_i - x_{i-1}) \quad x_{i-1} < x' < x_i
\]
\[
= f(x') \Delta_i(x)
\]
\[
\hat{S}_n = \sum \left[ \inf g(x) \right] \left[ f(x') \right] \Delta
\]
\[
\lim_{n \to \infty} \hat{S}_n = \int_a^b g(x) \, dF(x) = \int_a^b g(x) \hat{f}(x) \, dx = \text{ordinary Riemann integral.}
\]

If \( F(x) \) has only jumps at \( x_1, x_2, \ldots, x_n \) and elsewhere is constant
\[
\int_a^b g(x) \, dF(x) = \sum g(x_i) \Pr [x = x_i]
\]
\[\text{the jump}\]

If \( g(x) \) is continuous then this limit (the R-S Integral) exists. Also, if \( g(x) \) has at most a finite number of discontinuities and so does \( F(x) \) and they don't coincide, then the R-S integral exists.

**Def. 12:** \( X \) is a discrete random variable if there exists a countable set of points, \( x_1, x_2, \ldots, x_n \) with \( \Pr(X = x_i) = p_i \) and \( \sum p_i = 1 \) \( \left[ \text{elsewhere } F(x) \text{ is constant, i.e., } F'(x) = 0 \right] \).

for such a discrete random variable, the R-S integral reduces to a sum:
\[
\int_a^b g(x) \, dF(x) = \lim_{n \to \infty} \sum g(x'_i) \left[ F(x'_i) - F(x'_{i-1}) \right]
\]
where the $x_i'$ are points of division of $(a, b)$ and $x_i''$ is an intermediate point in the $i^{th}$ interval

$$
= \lim_{n \to \infty} \sum g(x_i') p_i
$$

$$
= \sum g(x_i') p_i
$$

summed over the set of points $x_i$ in $(a, b)$ -- the points where there is some probability.

**Def. 13:** $X$ is a continuous random variable if $F(x)$ is continuous and has a derivative $f(x)$ continuous except at a countable number of points.

$$
F(x_i') = F(x_{i-1}) = f(x_i') \left[ x_i' = x_{i-1} \right] \quad \text{(the theorem of the mean)}
$$

$$
x_{i-1} \leq x_i' \leq x_i \quad \text{with}
$$

$$
\int_a^b g(x) \, dF(x) = \lim_{n \to \infty} \sum g(x_i') \left[ F(x_i') - F(x_{i-1}) \right]
$$

$$
= \lim_{n \to \infty} \sum g(x_i') f(x_i') \Delta_i
$$

$$
= \int_a^b g(x) \, f(x) \, dx
$$

We can extend this definition to $k$-dimensions readily by writing:

$$
\int_a^b g(x_1', \ldots, x_k') \, d_{x_1'} \ldots d_{x_k'} \, F(x_1', \ldots, x_k')
$$

$$
= \lim_{n \to \infty} \sum g(x_1', \ldots, x_k') \, \Delta_k \, F(x_1', \ldots, x_k')
$$

For a def. of $\Delta_k$ see p. 8
If \( F(x_1, x_2, \ldots, x_k) \) is continuous and the density \( f(x_1, x_2, \ldots, x_k) \) exists and is continuous, then
\[
\left\{ \begin{array}{c}
g(x_1, \ldots, x_k) \int_{x_1}^{x} \cdots \int_{x_k}^{x} F(x_1, \ldots, x_k) \\
a_1 & b_1 \\
a_k & b_k
\end{array} \right. 
\]

In \( \mathbb{R} \)
\[
\int_{-\infty}^{x} dF(x) = F(x) - F(-\infty) = F(x)
\]

\[
\int_{a}^{b} dF(x) = F(b) - F(a)
\]

If we let \( b \to +\infty \), \( a \to -\infty \)
\[
\int_{-\infty}^{+\infty} dF(x) = 1
\]

and this extends easily to the \( k \)-dimensional case, so that we have:
\[
\left\{ \begin{array}{c}
g(x_1, \ldots, x_k) \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} F(x_1, \ldots, x_k) \\
a_1 & b_1 \\
a_k & b_k
\end{array} \right. = 1
\]

Consider \( k = 2 \), and the marginal distributions
\[
F_1(x_1) = F(x_1, +\infty) = \int_{-\infty}^{+\infty} dF_{x_2}(x_1, x_2)
\]
\[
= \lim_{\begin{array}{c}
a \to -\infty \\
b \to +\infty
\end{array}} \int_{a}^{b} dF_{x_2}(x_1, x_2)
\]
\[
\lim_{a \to -\infty, b \to +\infty} \left[ F_1(x_1, b) - F_1(x_1, a) \right] = F_1(x_1, +\infty) - F_1(x_1, -\infty) = F_1(x_1, +\infty) - 0
\]

This also extends readily to \( R_k \)

\[
F_1(x_1) = F(x_1, +\infty, \ldots, +\infty)
\]

\[
= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} F(x_1, x_2, \ldots, x_k) \, dx_2 \cdots dx_k
\]

with \( x_1 \) held fixed.

If the density function exists, then this reduces to a \( k-1 \) integral

\[
f_1(x_1) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \ldots, x_k) \, dx_2 \cdots dx_k
\]

**Problem 6**

If \( X_1, \ldots, X_n \) have independent, uniform on \((0, 1)\), distributions,

show that

\[
-2 \sum_{i=1}^{n} \ln X_i
\]

has a \( \chi^2 \) distribution with \( 2n \) d.f.

and indicate the statistical application of this.

**see** Snedecor ch. 9

Fisher -- about p. 100

Anderson + Bancroft -- last chapter of section 1

From problem 5b

\[
Z = -2 \ln X \cdot Y = -2 \ln X + -2 \ln Y
\]

or is the sum of 2 \( \chi^2 \) with 2 d.f. each

**References on integrals**

-- Cramer pp. 39-40

-- Sokolnikoff -- Advanced Calculus -- ch. 4
Chapter II

Properties of Univariate Distributions; Characteristic Functions

Standard Distributions:

A. Trivial or point mass (discrete)

\[ \Pr [X = a] = 1 \quad F(x) = 0 \quad x < a \]
\[ F(x) = 1 \quad x \geq a \]

B. Uniform (continuous)

\[ F(x) = 0 \quad x < 0 \]
\[ F(x) = x \quad 0 \leq x \leq 1 \]
\[ F(x) = 1 \quad x > 1 \]

C. Binomial (discrete)

\[ \Pr [X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 1, 2, \ldots, n \quad 0 \leq p \leq 1 \]

\[ \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = (1-p + p)^n = 1^n = 1 \]

\[ \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = 1 \text{ is an identity in } p, n \]

D. Poisson (discrete)

\[ \Pr [X = k] = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 1, 2, \ldots, \infty \quad \lambda > 0 \]

\[ \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1 \]

E. Negative binomial (discrete)

\[ \Pr [X = k] = \binom{r+k-1}{r-1} p^r (1 - p)^k \quad k = 1, 2, \ldots, \infty \]
\[ 0 < p < 1 \]
\[ r \text{ is an integer} \]

Example: Draw from an urn with proportion p of red balls, with replacement, until we get r reds out. The random variable in this situation is the number of non-reds drawn in the process; to have k black balls means that in the
first \( r + k - 1 \) trials, we got \( r - 1 \) reds and \( k \) blacks, and then on the last trial got a red; the probability of this is

\[
\binom{r+k-1}{r-1} p^{r-1} (1 - p)^k \times p.
\]

This is what is referred to as inverse sampling in that the number of defectives is specified rather than specifying the sample size which is then scrutinized for the number of defectives.

F. Normal distribution (continuous)

\[
f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

\( \sigma > 0 \)

\(-\infty < \mu < \infty \)

\(-\infty < x < \infty \)

**Problem 7:** Prove that

\[
\sum_{k=0}^{\infty} \binom{r+k-1}{r-1} p^r (1 - p)^k = 1
\]

i.e., is an identity in \( r, p \)

**Hint:** is in the name - express \((a+b)^n\) in an infinite series.

**Def. 1:** If \( X \) is a random variable with distribution \( F(x) \) and if \( \int_{-\infty}^{\infty} g(x) \, dF(x) \) exists, then we define the expectation of \( g(X) \) as

\[
E[g(X)] = \int_{-\infty}^{\infty} g(x) \, dF(x)
\]

this being the R - S integral,

if \( X \) is continuous \( E[g(X)] = \int_{-\infty}^{\infty} g(x) \, f(x) \, dx \)

if \( X \) is discrete \( E[g(X)] = \sum_{v=0}^{\infty} g(x_v) \, p_v \)

**Problem 8:** Given \( F(x) = 0 \) \( x < 0 \)

\( = \frac{1}{2} \) \( x = 0 \)

\( = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X} e^{-t^2/2} \, dt \) \( x > 0 \)

(This is a censored normal distribution -- i.e., all the negative values are concentrated at the origin)
Find: \( E(X) \)

Def. 15: If \( E \left[ X - E(X) \right]^k \) exists it is defined to be the \( k^{th} \) central moment and is denoted \( \mu_k \).

If \( E(X)^k \) exists it is defined as the \( k^{th} \) moment about the origin, and is denoted \( \mu_k \).

Exercise 9: Find \( E(X) \) for each of the standard distributions.

Theorem 5: \( E \left[ g(X) + h(Y) \right] = E \left[ g(X) \right] + E \left[ h(Y) \right] \)

Proof: Let \( F(x, y) \) be the joint d.f. of \( X, Y \) and \( F_1 \) and \( F_2 \) the marginal d.f.

\[
E \left[ g(X) + h(Y) \right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[ g(x) + h(y) \right] d_{xy} F(x, y) = \int \int \left[ g(x) d_{xy} F(x, y) + \int h(y) d_{xy} F(x, y) \right]
\]

Theorem 6: If \( X \) and \( Y \) are independent random variables, then

\( E \left[ g(X) h(Y) \right] = E \left[ g(X) \right] E \left[ h(Y) \right] \)


Corollary: If \( X \) and \( Y \) are independent random variables, then

\( \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \)

Moments: \( \mu_1 = E(X) = \text{mean} \)
\( \mu_2 = E(X - \mu)^2 = \text{variance} \)
\( \mu_3 = E(X - \mu)^3 \)
\( \mu_4 = E(X - \mu)^4 \)

etc.

for the normal distribution \(- N(0, 1)\)

\[
f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}
\]

\[
E(X^k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx
\]

all odd moments \((k \text{ odd}) = 0\) by a "symmetry" argument.
\[ E(x^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \, dx = 1 \] from integration by parts

\[ E(x^4) = 3 \]

or in general

\[ E(x^{2n}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2n} e^{-x^2/2} \, dx = 1 \cdot 3 \cdot 5 \cdots (2n - 1) \]

which can be shown by induction.

**Theorem 7:** Let \( \alpha_0, \alpha_1, \alpha_2, \ldots \) be the moments of a distribution function \( F(x) \), i.e.,

\[
\alpha_k = \frac{1}{\infty} \int_{-\infty}^{\infty} x^k \, dF(x)
\]

then if for some \( r > 0 \),

\[
\sum_{k=0}^{\infty} \frac{\alpha_k x^k}{k!}
\]

converges absolutely then \( F(x) \) is the only distribution with these moments.

**Proof:** See Cramer, p. 176.

**Example:** \( N(0, 1) \)

\[
\alpha_{2k+1} = 0
\]

\[
\alpha_{2k} = \frac{(2k - 1)!!}{2^{k-1} (k-1)!} = 1 \cdot 3 \cdot 5 \cdots (2k - 1)
\]

\[
\sum_{k=0}^{\infty} \frac{\alpha_k x^k}{k!} = \sum_{k=0}^{\infty} \frac{(2k - 1)!!}{2^k (k-1)!} \frac{x^{2k}}{(2k)!}
\]

since odd terms drop out.

\[
= \sum_{k=0}^{\infty} \frac{r^{2k}}{2^{k-1} (k-1)! \cdot 2k} = \sum_{k=0}^{\infty} \frac{(r^2)^k}{2^k k!}
\]

\[
= \sum \frac{1}{k!} \left( \frac{r^2}{2} \right)^k \quad \text{now} \quad \sum_{k=0}^{\infty} \frac{k^k}{k!} = e^x
\]

\[ = e^{r^2/2} \]

= exponential series
\[ \sum_{k=0}^{\infty} \frac{r^k}{k!} \] converges absolutely for all \( r \), therefore the only distribution with these moments is the d.f. with the density
\[ f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \] i.e., the normal

**Problem 9:** Find the moments of the uniform distribution and show that this is the only d.f. with these moments.

**Theorem 8:** (Tchebycheff's)

If \( g(x) \) is a non-negative function of \( X \) then for every \( K > 0 \)

\[ \Pr \left[ g(X) \geq K \right] \leq \frac{E \left[ g(X) \right]}{K} \]

**Proof:**

\[ E \left[ g(X) \right] = \int_{-\infty}^{\infty} g(x) \, dF(x) \]

Let \( S \) be the set of values of \( X \) where \( g(X) \geq K \)

\[ \Pr \left[ g(X) \geq K \right] \geq \int_{S} g(x) \, dF(x) \quad \text{since the smallest value of } g(X) \text{ in } S \text{ is } K \]

\[ \geq K \int_{S} dF(x) = K \Pr \left[ g(X) \geq K \right] \]

\[ \therefore \Pr \left[ g(X) \geq K \right] \leq \frac{E \left[ g(X) \right]}{K} \]

**Corollary:** The above (th. 8) converts readily into the more familiar form

\[ \Pr \left[ |X - \mu| \geq K \sigma \right] \leq \frac{1}{K^2} \]

**Proofs:** (See p. 182 in Cramer) setting

\[ g(X) = (X - \mu)^2 \quad K = k^2 \sigma^2 \quad E \left[ g(X) \right] = \sigma^2 \]

\[ \Pr \left[ (X - \mu)^2 \geq (k \sigma)^2 \right] \leq \frac{\sigma^2}{k^2 \sigma^2} \]

taking the square root of the left hand side
\[
\Pr \left[ \left| X - \mu \right| \geq k \sigma \right] \leq \frac{1}{k^2}
\]

**Theorem 9:** If \( X_n \) is a sequence of binomial random variables with parameters \( n, p \), then given any \( \varepsilon > 0, \delta > 0 \) there exists an \( n_0 \) such that for \( n > n_0 \)

\[
\Pr \left[ \left| \frac{X_n}{n} - p \right| \geq \varepsilon \right] \leq \delta
\]

(which says that if you take larger and larger samples, then the observed ratio \( X/n \) approaches the true value)

\( X_n \) = number of successes in \( n \) independent trials with a probability of success in each trial = \( p \)

**Proof:**

\[\sigma^2_{X_n} = np(1-p)\]

\[\sigma^2(X_n/n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \leq \frac{1}{4n}\]

From corollary to theorem 8

\[\Pr \left[ \left| \frac{X_n}{n} - p \right| \geq k \sigma \right] \leq \frac{1}{k^2}\]

Choose: \( \frac{1}{k^2} = \delta \) or \( k = \frac{1}{\sqrt{\delta}} \)

\[k \sqrt{\frac{p(1-p)}{n}} = \varepsilon \text{ or } n = \frac{p(1-p)}{\delta \varepsilon^2}\]

Hence if \( n \) is chosen this large, from the corollary to theorem 8 the stated probability inequality follows.

**Note:** Theorem 9 could be rewritten

\[\Pr \left[ \left| \frac{X_n}{n} - p \right| \geq \varepsilon \right] \leq \frac{1}{\ln \varepsilon^2} \quad n = \frac{p(1-p)}{\delta \varepsilon^2}\]

\[\delta = \frac{p(1-p)}{n \varepsilon^2} \leq \frac{1}{\ln \varepsilon^2}\]

**Characteristic Functions:**

**Def. 16:** Characteristic function:

\[\phi_X(t) = \int_{-\infty}^{\infty} e^{ixt} dF(x)\]

or since \( e^{ixt} = \cos xt + i \sin xt \)

\[\phi_X(t) = \int_{-\infty}^{\infty} \cos xt dF(x) + i \int_{-\infty}^{\infty} \sin xt dF(x)\]
the successive derivatives of \( \phi(t) \) when evaluated at \( t = 0 \) yield the moments of \( F(x) \) except for a factor of a power of "i".

\[
\frac{d}{dt} \phi(t) = \phi'(t) = \int_{-\infty}^{\infty} ix e^{ixt} dF(x)
\]

\[
\phi'(0) = i \int_{-\infty}^{\infty} x e^{0} dF(x) = i \mu
\]

\[
\phi''(t) = \int_{-\infty}^{\infty} (ix)^2 e^{ixt} dF(x)
\]

\[
\phi''(0) = i^2 \int_{-\infty}^{\infty} x^2 dF(x) = i^2 \kappa_2
\]

in general

\[
\phi^k(0) = i^k E(X^k) = i^k \kappa_k
\]

the moment generating function operates in the same manner, except it does not include the factor "i", and is therefore not as general in application.

\[
MGF = M(t) = E(e^{Xt}) = \int_{-\infty}^{\infty} e^{xt} dF(x) \quad \text{if this integral exists}
\]

and operates by evaluating successive derivatives with respect to \( t \) at \( t = 0 \)

Lemma: if \( E(X^k) \) exists, then \( \phi^k(t) \) exists and is continuous. The converse is also true.

Examples:

1. Trivial distribution: \( Pr[X = a] = 1 \)
   \[
   \phi(t) = \int_{-\infty}^{\infty} e^{ixt} dF(x) = e^{iat} x 1 = e^{iat}
   \]

2. Binomial:
   \[
   \phi(t) = \sum_{k=0}^{n} e^{ikt} \binom{n}{k} p^k (1-p)^{n-k}
   \]
   \[
   = \sum \binom{n}{k} \left( pe^{it} \right)^k (1-p)^{n-k} = \left[ pe^{it} + (1-p) \right]^n
   \]
3. Poisson: \( \phi(t) = \sum_{k=0}^{\infty} e^{itk} \frac{e^{-\lambda} \lambda^k}{k!} \)

\[ = \sum_{k=0}^{\infty} \frac{e^{-\lambda} (\lambda e^{it})^k}{k!} \]

\[ = e^{-\lambda} e^{\lambda e^{it}} \]

\[ = e^{\lambda (e^{it} - 1)} \]

4. Normal \((0, 1)\):

\[ \phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} \, dx = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2 - 2itx}{2}} \, dx \]

\[ = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{x^2 - 2itx + (it)^2}{2} + \frac{(it)^2}{2}} \, dx \]

Setting \( y = x - it \):

\[ = \left\{ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-y^2/2} \, dy \right\} e^{\frac{-t^2}{2}} \]

\[ = e^{\frac{-t^2}{2}} \]

(Note the term in curly brackets is the integral of a normal density and equals one.)

* The validity of this complex transformation has to be justified. See Problem II.

If \( X \) is \( N(0, 1) \)

\( \phi(t) = e^{\frac{-t^2}{2}} \)

\( \phi'(t) = -\frac{2t}{2} e^{\frac{-t^2}{2}} \)

\( \phi''(t) = -e^{\frac{-t^2}{2}} + t^2 e^{\frac{-t^2}{2}} \)

\( \phi''(0) = -1 \)

\( E(X^2) = i^2 \phi'' = 1 \)
Problem 10: Prove
\[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} \, dx = 1 \]

Problem 11: Show that
\[ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-x^2/2} \, dx = e^{-t^2/2} \]
without using the transformation used in class

Hint: \( e^{itx} = \cos tx + i \sin tx \)

Theorem 10:
\[ \phi_{AX + B}(t) = e^{iBt} \phi_X(At) \]
where \( X \) is a random variable with a d.f. \( F(x) \) and a characteristic function \( \phi (t) \)

Proof: \( Y = AX + B \) where \( A, B \) are constants

\[ \phi_Y(t) = E(e^{itY}) = E\left[ e^{it(AX + B)} \right] \]
\[ = E \left( e^{itAX} e^{itB} \right) = e^{itB} E(e^{it(AX)}) \]
\[ = e^{iBt} \phi_X(At) \]

if \( X \) is \( N(0, 1) \) then \( Y = \sigma X + \mu \) is \( N(\mu, \sigma^2) \)

\[ \phi_Y(t) = e^{it\mu} \phi_X(\sigma t) \]
\[ = e^{it\mu} e^{-(\sigma t)^2/2} \]
\[ = e^{it\mu} e^{-\sigma^2 t^2/2} \]

Def. 17: The cumulant generating function, \( K(t) \), is defined to be:

\[ K(t) = \ln \phi(t) \]

Example: For \( Y \) which is \( N(\mu, \sigma^2) \)

\[ K(t) = it\mu - \frac{\sigma^2 t^2}{2} \]
Note: For further discussion of cumulants see Cramer or Kendall.

Note: Originally cumulants = semi-invariants (British school name = Scandinavian school name) — however, semi-invariants have been extended so that now cumulants are a special case.

**Theorem 11:** If $X_1, X_2, \ldots, X_n$ are independent random variables, then

$$\phi(t) = \prod_{i=1}^{n} \phi_{X_i}(t)$$

(the c.f. of a sum = the product of the individual c.f.s.)

**Proof:**

$$E\left[e^{it \sum X_i}\right] = E\left[e^{it X_1} e^{it X_2} \ldots e^{it X_n}\right]$$

$$= E\left[e^{it X_1}\right] E\left[e^{it X_2}\right] \ldots E\left[e^{it X_n}\right]$$

by independence

$$= \phi(t) \phi(t) \ldots \phi(t) = \prod_{i=1}^{n} \phi(t)$$

**Example:** If $X_i$ are NID $(\mu_i, \sigma_i^2)$, then the c.f. of $Y = \sum_{i=1}^{n} X_i$

$$E\left[e^{it \sum X_i}\right] = \prod_{i=1}^{n} \left(e^{it \mu_i - \frac{t^2 \sigma_i^2}{2}}\right)$$

$$= e^{it \sum \mu_i - \frac{t^2 \sum \sigma_i^2}{2}}$$

therefore we could say that $Y$ is $N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$

To justify this last step we need to show the converse of $F(X) \rightarrow \phi_X(t)$ i.e., that

$\phi_X(t) \rightarrow F(X)$. Therefore, we need the following lemma and theorem.

**Lemma:**

$$\lim_{T \to \infty} \frac{2}{\pi} \int_{-T}^{T} \frac{\sin h t}{t} dt = \begin{cases} -1 & h < 0 \\ 0 & h = 0 \\ +1 & h > 0 \end{cases}$$

**Proof:**

$$J(\alpha, \beta) = \int_{0}^{\infty} e^{-\alpha u} \frac{\sin \beta u}{u} du \quad \alpha > 0$$

$$\frac{\partial J}{\partial \beta} = \int_{0}^{\infty} e^{-\alpha u} \cos \beta u du \quad \text{Notes: Differentiation under the integral can be justified.}$$
\[ \frac{\alpha}{\kappa^2} \quad \text{see tables or integrate by parts twice} \]

\[ \int \frac{\partial J}{\partial \beta} \, d\beta = \int \frac{\alpha}{\kappa^2 + \beta^2} \, d\beta = \arctan \frac{\beta}{\kappa} + C \]

Let \( \beta \to 0 \) then \( J(\alpha, 0) = e^{\alpha t}, \quad 0 \cdot du = 0 \)

\[ \arctan 0 + C = 0 \]
\[ 0 + C = 0 \]
\[ \therefore C = 0 \]

\( J(\alpha, \beta) = \arctan \frac{\beta}{\alpha} \)

Let \( \alpha \to 0, \) and put \( \beta = h, \) then

\[ \lim_{\alpha \to 0} J(\alpha, h) = \arctan \frac{h}{\alpha} \text{ depending on } h + \text{ or } - \]

\[ \frac{\pi}{2} \quad h > 0 \]
\[ -\frac{\pi}{2} \quad h < 0 \]

**Theorem 12:** If \( F(x) \) is continuous at \( a - h, \) \( a + h, \) then

\[ F(a + h) - F(a - h) = \lim_{\alpha \to \infty} \int_{-\infty}^{\infty} \frac{\sin \alpha t}{t} e^{-ita} \phi(t) \, dt \]

If \( \int_{-\infty}^{\infty} |\phi(t)| \, dt < \infty \) then \( f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) \, dt \)

**Note:** Recall that \( \phi(t) = \int_{-\infty}^{\infty} e^{-itx} f(x) \, dx \) \( \text{[def. 16]} \)

Combining this with the above theorem means that given \( \phi(t) \) or \( f(x) \) we can determine the other.

**Proof:**

Define \( J = \frac{1}{\pi} \int_{-T}^{T} \frac{\sin \alpha t}{t} e^{-ita} \phi(t) \, dt \)

\[ = \frac{1}{\pi} \int_{-T}^{T} \frac{\sin \alpha t}{t} e^{-ita} \left[ \int_{-\infty}^{\infty} e^{-itx} \, dF(x) \right] \, dt \]
Interchanging integrals (reversing the order of integration) which can be justified this becomes

\[
\frac{1}{n} \int_{-\infty}^{\infty} \left\{ \int_{-T}^{T} \frac{\sin \frac{ht}{t}}{\sin \frac{hx}{x}} \left( \frac{\varepsilon^{-ita} \varepsilon^{itx}}{\varepsilon^{-it(x-a)}} \right) \, dt \right\} \, dF(x)
\]

\[
= \frac{1}{n} \int_{-\infty}^{\infty} \left\{ \int_{-T}^{T} \frac{\sin \frac{ht}{t}}{\sin \frac{hx}{x}} \left[ \cos t(x-a) + i \sin t(x-a) \right] \, dt \right\} \, dF(x)
\]

Note: The \( \frac{\sin}{t} \) sin term is an odd function \( \Rightarrow \int_{-T}^{T} f(t) \, dt = 0 \)

The \( \frac{\sin}{t} \) cos term is an even function \( \Rightarrow \int_{-T}^{T} f(t) \, dt = 2\int_{0}^{T} f(t) \, dt \)

\[
= \frac{1}{n} \int_{-\infty}^{\infty} 2 \int_{0}^{T} \frac{\sin \frac{ht}{t}}{\sin \frac{hx}{x}} \cos t(x-a) \, dt \, dF(x)
\]

Note: \( 2 \cos A \sin B = \sin (A+B) - \sin (A - B) \)

\[
J = \frac{1}{n} \int_{-\infty}^{\infty} \left\{ \int_{0}^{T} \frac{\sin \frac{t(x-a+h)}{t} - \sin \frac{t(x-a-h)}{t}}{t} \, dt \right\} \, dF(x)
\]

now take the limit as \( T \to \infty \)

using the lemma just proved, with that \( h = (x-a+h) \) here

<table>
<thead>
<tr>
<th>( x=a-h )</th>
<th>( x=a+h )</th>
</tr>
</thead>
<tbody>
<tr>
<td>For ( \sin t(x-a+h) )</td>
<td>( x=a+h&lt;0 )</td>
</tr>
<tr>
<td>For ( \sin t(x-a-h) )</td>
<td>( x-a-h&lt;0 )</td>
</tr>
<tr>
<td>For ( x ) in each region</td>
<td>both = ( -\frac{\pi}{2} )</td>
</tr>
<tr>
<td>2nd part = ( \frac{\pi}{2} )</td>
<td></td>
</tr>
<tr>
<td>Whole integral (in brackets)</td>
<td>0</td>
</tr>
</tbody>
</table>
i.e., in the region \( a - h \leq x \leq a + h \)

\[
\int_0^T \frac{1}{t} \sin t(x - a + h) \, dt - \int_0^T \frac{1}{t} \sin t(x - a - h) \, dt = \pi
\]

elsewhere \( = 0 \)

There

\[
J = \frac{1}{\pi} \left( \int_{a-h}^{a+h} \pi \, dF(x) + \int_{-\infty}^{0} \, dF(x) + \int_{a+h}^{0} \, dF(x) \right)
\]

\[= F(a + h) - F(a - h)\]

Proof of the second statement in the theorem:

\[
\frac{F(a + h) - F(a - h)}{2h} = \frac{1}{2h \pi} \int_{-\infty}^{\infty} \frac{\sin ht \, e^{-\text{ita}}}{t} \, \phi(t) \, dt
\]

taking the limit of both sides as \( h \to 0 \)

\[
f(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \lim_{h \to 0} \frac{\sin ht \, e^{-\text{ita}}}{ht} \, \phi(t) \, dt
\]

therefore

\[
f(a) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\text{ita}} \, \phi(t) \, dt
\]

**Problem 12:** Let \( X_i \ (i = 1, 2, \ldots, n) \) have a density function given by

\[
f(x) = \begin{cases} a-1 & x > 0 \quad \text{for} \quad 0 \leq x \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

Find the density of \( Y = \frac{1}{n} \sum_{i=1}^{n} X_i \)

\(X_i\) are independent

(may need the result of Cramer p. 126)

**Problem 13:** Define a factorial moment

\[
E(x(x-1) \ldots (x-r+1))
\]

Define

\[
F^*(t) = \int_{-\infty}^{\infty} (1 + t)^x \, dF(x) \quad \text{as the factorial moment generating function.}
\]
Find
\[ F^*(t) \] for the binomial and use this to get the factorial moments.

Problem 11. If \( \phi(t) = e^{-|t|} \) find the density of \( f(x) \) corresponding to \( \phi \). Find the distribution of the mean of \( n \) independent variables with this d.f.

Theorem 13: A necessary and sufficient condition that a sequence of distribution functions \( F_n \) tend to \( F \) at every point of continuity of \( F \) is that \( \phi_n \), the characteristic function corresponding to \( F_n \), tend to a function, \( \phi(t) \) which is continuous at \( t = 0 \) [or tends to a function \( \phi(t) \) which is itself a characteristic function].

Proof: Omitted — see Cramer p. 96-98

This theorem says, if we have
\[
F_1 \quad F_2 \quad F_3 \quad \ldots \ldots \quad F
\]
\[
\phi_1 \quad \phi_2 \quad \phi_3 \quad \ldots \ldots \quad \phi
\]
we go from the \( F_1 \) to the \( \phi_1 \) — observe that the \( \phi_1 \) tend to a limit [for example the normal approximation to the binomial] — observe that the limit is itself a characteristic function [of the normal] — then go to \( F \) by previously discussed techniques.

Theorem 14: (Central Limit Theorem)

If \( X_1, X_2, X_3 \ldots \ldots \) are a sequence of independently and identically distributed random variables with a distribution function \( F(x) \) with finite first and second moments [say mean \( \mu \) and variance \( \sigma^2 \)] then:

1— the distribution of \( Y_n = \sqrt{n} \left( \frac{1}{n} \sum X_i - \mu \right) \) tends, as \( n \to \infty \), to the normal distribution with mean 0 and variance \( \sigma^2 \)

2— for any interval \((a, b)\)
\[
\lim_{n \to \infty} \frac{\Pr \left[ a < Y_n < b \right]}{\frac{1}{\sqrt{2\pi} \sigma}} = \int_a^b e^{-t^2/2\sigma^2} \, dt
\]

3— the sequence \([Y_n]\) is asymptotically \( N(0, \sigma^2) \)

Proof: Denote the c.f. of \( X_1 \) as \( \phi(t) \)
to get the c.f. of \( Y_n = \sqrt{n} \frac{\sum (X_i - \mu)}{n} = \frac{\sum (X_i - \mu)}{\sqrt{n}} \)
\[
\phi_{Y_n}(t) = \left[ e^{-\frac{\mu}{\sqrt{n}}} \phi \left( \frac{t}{\sqrt{n}} \right) \right]^n
\]
if we expand $\varphi(t)$ in a Taylor series, we have

$$\varphi(t) = 1 + i \kappa_1 t + i^2 \kappa_2 t^2/2 + \text{remainder} \quad \kappa_1 = \mu, \quad \kappa_2 = \mu^2 + \sigma^2$$

$$= 1 + i \mu t - (\mu^2 + \sigma^2) t^2/2 + R(t)$$

where $\frac{R(t)}{t^2} \to 0$ as $t \to 0$

$$\ln \varphi_n(t) = n \left[ \ln e^{-i \frac{\mu}{\sqrt{n}} t} + \ln \varphi \left( \frac{t}{\sqrt{n}} \right) \right]$$

$$= i \sqrt{n} \mu t + n \ln \varphi \left( \frac{t}{\sqrt{n}} \right)$$

$$= i \mu \sqrt{n} t + n \ln \left[ 1 + \frac{i \mu t}{\sqrt{n}} - \frac{(\mu^2 + \sigma^2)t^2}{2n} + R \left( \frac{t}{\sqrt{n}} \right) \right]$$

\textbf{Note:} $\ln (1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \ldots$,

$$= x - \frac{x^2}{2} + R'$$

where $\frac{R'}{x} \to 0$ as $x \to 0$

setting $x = i \mu \sqrt{n} t = (\mu^2 + \sigma^2) \frac{t^2}{2n} + R \left( \frac{t}{\sqrt{n}} \right)$ we get

$$\ln \varphi_n(t) = -i \sqrt{n} \mu t + n \left[ i \mu \frac{t}{\sqrt{n}} - (\mu^2 + \sigma^2) \frac{t^2}{2n} + R \left( \frac{t}{\sqrt{n}} \right) \right.$$

$$\left. - \frac{t^2}{2} \left( i \mu \frac{t}{\sqrt{n}} \right)^2 + \frac{t^3}{n^{3/2}} \text{An} + \frac{R''(t)}{n^2} \right]$$

$$= -\frac{\sigma^2 t^2}{2} + n R \left( \frac{t}{\sqrt{n}} \right) + \frac{R''}{n}$$

$$\lim_{n \to \infty} \ln \varphi_n(t) = -\frac{\sigma^2 t^2}{2} \quad \text{since} \quad \lim_{n \to \infty} \frac{R''}{n^{3/2}} = 0$$

$$\Rightarrow \lim_{n \to \infty} \varphi_n(t) = e^{-\frac{\sigma^2 t^2}{2}} \quad \text{since} \quad \lim_{n \to \infty} \frac{R(t)}{n^{3/2}} \to 0$$

which is the characteristic function of the normal distribution $N(0, \sigma^2)$. 
The theorem says, for any given $\varepsilon$ there is some sufficiently large $n(a, b, c, \sigma)$ such that
\[
\left| \Pr \left[ a < Y_n < b \right] - \frac{1}{\sqrt{2\pi} \sigma} \int_a^b e^{-t^2/2\sigma^2} \, dt \right| < \varepsilon
\]

**Problem 15:** Let $X$ be a Poisson random variable with parameter $\lambda$. Show that for any $(a, b)$
\[
\lim_{\lambda \to \infty} \Pr \left[ X < \frac{X - \lambda}{\sqrt{\lambda}} < \frac{b - \lambda}{\sqrt{\lambda}} \right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} \, dt
\]

a. directly -- use Sterling's approximation
b. by characteristic functions

see Feller ch. VII

Problem 15 -- says that the standardized Poisson random variable is asymptotically normally distributed $N(0, 1)$

Cauchy Distribution -- see problem 14 -- has no mean or variance since $e^{-|t|}$ is not differentiable at the origin

\[
\lim_{b \to \infty} \lim_{a \to \infty} \frac{1}{\pi} \int_a^b \frac{x}{1+x^2} \, dx = \frac{1}{2\pi} \ln(1+x^2) \bigg|_a^b = \infty - \infty
\]

therefore $E(x)$ does not exist for a Cauchy random variable, even though the distribution is symmetric about the origin.

**Theorem 15:** (Liapounoff)

If $X_1, X_2, \ldots, X_n$ are independent random variables with means $\mu_1$ and variances $\sigma_1^2$ and with
\[
\rho_1^3 = \frac{E|X_1 - \mu_1|^3}{\sigma_1^3} < \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{\sum x_i^3}{\left( \sum \sigma_1^2 \right)^{3/2}} = 0
\]

then
\[
T = \frac{\sum X_i - \sum \mu_1}{\left( \sum \sigma_1^2 \right)^{1/2}} \quad \text{is asymptotically} \quad N(0, 1)
\]

Problem 16: Define $M^*(t) = E(x^t)$ as the Mellin Transform.

If $X$ has density $f(x) = k x^{k-1}$, $0 \leq x \leq 1$
then $M^*(t) = \frac{k}{k+t}$

If $X$ has density $f(x) = -k^2 x^{k-1} \ln x$, $0 \leq x \leq 1$
then $M^*(t) = \left(\frac{k}{k+t}\right)^2$

Use this to find the density of $Y = X_1 X_2$ where the $X_i$ are independent and have density $f(x) = k x^{k-1}$.

State a theorem necessary to validate this approach.

_Impulse Transform:_

$$E\left[ e^{-sx} \right] = f(s) = \int_0^\infty e^{-sx} f(x) \, dx \quad \text{if } f(x) = 0 \quad x < 0$$

--- extensive uses in differential equations
--- extensive tables of the I.T. in the literature --- tables for passing from the transform to the function and vice versa --- see Doetsch: Tables of the Laplace Transform.

_Notes:_ Replacing $(s)$ by $(-it)$ gives the mgf, or by $(+it)$ gives the c.f.

_Fourier Transform — the mathematical name for the characteristic function_:

$$E\left[ e^{itx} \right]$$

We have previously noted that if $X, Y$ are NID with means $\mu_1$ and variances $\sigma_1^2$
then $Z = X + Y$ is normal $N(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2)$.

There is a converse to this "addition theorem for normal variates" to the effect that:

if $Z = X + Y$ where $X, Y$ are independent and $Z$ is normal, then $X, Y$ are both normal --- see Cramer p. 213.

_Derived Distributions (from the Normal and others):_ 

<table>
<thead>
<tr>
<th>name</th>
<th>density</th>
<th>$c.f.$</th>
<th>$E(x)$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_n^2$</td>
<td>$\frac{1}{n/2} e^{-x/2} x^{n/2 - 1}$</td>
<td>$(1 - 2it)^{-n/2}$</td>
<td>$n$</td>
<td>$2n$</td>
</tr>
</tbody>
</table>
-- if $X_1, \ldots, X_n$ are NID(0, 1), $\sum_{1}^{n} X_i^2$ is $\chi_n^2$

-- it has the additive property $\chi_{m+n}^2 = \chi_m^2 + \chi_n^2$

**Gamma**

$$\frac{\lambda}{\Gamma(\lambda)} x^{\lambda-1} \frac{1}{\gamma^\lambda (1+it)^{-\lambda}}$$

$x > 0$

-- $\chi_1^2$ is a special case of the gamma

-- Pearson's Type III distribution with starting point at the origin.

**Student's t**

$$\frac{1}{\sqrt{mn}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \frac{1}{\left(1+\frac{x}{n}\right)^{m/2}}$$

$n > 1$

-- $t = \frac{X\sqrt{n}}{\sqrt{m}}$ where $X, Y$ are independent, $X$ is N(0, 1), $Y$ is $\chi_n^2$

-- $t$ with 1 d.f. is a Cauchy distribution

**F**

$$\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{n/2} \frac{x^{m/2} - 1}{(1+\frac{m}{n}x)^{m+n/2}}$$

$x > 0$

-- if $X, Y$ are independent, $X$ is $\chi_m^2$, $Y$ is $\chi_n^2$, then $F = \frac{X/m}{Y/n}$

Fisher's $z$ is defined by $F = e^{2z}$ see Cramer p. 243.

**Beta**

$$\frac{\Gamma(p+q)}{\Gamma(p) \Gamma(q)} x^{p-1} (1-x)^{q-1}$$

$$\beta(p, q)$$

$p/q = \frac{(p+q)^2(p+q+1)}{m/n \times \frac{F}{1 + m/n \times F}}$

$0 < x < 1$

-- putting $p = \frac{m}{2}, q = \frac{n}{2}$, we get $\beta = \frac{m F}{1 + m/n \times F}$

-- see Kendall for the relation to the Binomial.
\[ \Gamma(p) = \int_0^\infty e^{-x} x^{n-1} \, dx \quad n > 0 \]

\[ \Gamma(p) = (p-1) \Gamma(p-1) \]

if \( p \) is a positive integer \( \Gamma(p) = (p-1)! \)

\[ \lim_{p \to \infty} \frac{\Gamma(p+1)}{p^p e^{-p} \sqrt{2\pi p}} = 1 \quad \text{Stirling's Formula} \]

\[ \lim_{p \to \infty} \frac{\Gamma(p+h)}{p^h \Gamma(p)} = 1 \quad h \text{ Fixed} \]

**Problem 17:** \( X \) and \( Y \) are NID \((0, \sigma^2)\). Find the marginal distribution of

1. \( r = \sqrt{x^2 + y^2} \)

2. \( \theta = \arctan \frac{Y}{X} \)

**Problem 18:**
Cramer p. 319 No. 10.
Chapter III

Convergence of Distributions:

a. Central Limit Theorem (theorem 14)

b. Poisson distribution  (problem 15)  if X is a Poisson r.v., then

\[ \lim_{\lambda \to \infty} \Pr \left[ a \leq \frac{X - \lambda}{\sqrt{\lambda}} < b \right] = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-t^2/2} \, dt \]

Proof: 1. By c.f. is straightforward  see sol. to problem 15, or p. 250.
2. Direct

\[ P(a, b) = \sum_{x=a}^{b} e^{-\lambda} \frac{\lambda^x}{x!} \]

Let \( K = \lambda + \sqrt{\lambda} \, x \)

\[ x = \frac{K - \lambda}{\sqrt{\lambda}} \]

\[ x + \Delta x = \frac{K + 1 - \lambda}{\sqrt{\lambda}} \]

\[ \Delta x = \frac{1}{\sqrt{\lambda}} \]

\[ P(a, b) = \sum_{x=a}^{b} e^{-\lambda} \frac{\lambda^x + x/\sqrt{\lambda}}{(\lambda + x/\sqrt{\lambda})!} \]

using Stirling's Formula:

\[ \frac{n!}{n^{n} e^{-n} (2\pi n)^{1/2}} = 1 + O(n) \]

where \( O(n) \to 0 \) as \( n \to \infty \)

\[ = \sum_{a}^{b} e^{-\lambda} \frac{\lambda^x + x/\sqrt{\lambda}}{(2\pi)^{1/2}(\lambda + x/\sqrt{\lambda})^\lambda + x/\sqrt{\lambda} + 1/2 e^{-(\lambda + x/\sqrt{\lambda})}} \]

\[ = \frac{1}{\sqrt{2\pi}} \sum_{a}^{b} \frac{e^{x/\sqrt{\lambda}}}{(1 + \frac{x}{\sqrt{\lambda}})^{\lambda}} \cdot \frac{1}{(1 + \frac{x}{\sqrt{\lambda}})^{x/\sqrt{\lambda}}} \cdot \frac{1}{\sqrt{\lambda}} \cdot \frac{1}{(1 + \frac{x}{\sqrt{\lambda}})^{1/2}} \left[ 1 + O(\lambda) \right] \]
Notes: \[ \lim_{\lambda \to \infty} \frac{1}{(1 + \frac{x}{\sqrt{\lambda}})^\lambda} = \lim_{x/\sqrt{\lambda} = u \to \infty} \frac{1}{(1 + \frac{x}{u})^u} = e^{-x^2} \]

\[ \lim_{\lambda \to \infty} \ln \frac{e^{x/\sqrt{\lambda}}}{(1 + x/\sqrt{\lambda})^\lambda} = \lim_{\lambda \to \infty} \left( \frac{x}{\sqrt{\lambda}} - \lambda \ln \left( 1 + \frac{x}{\sqrt{\lambda}} \right) \right) \]

\[ = \lim_{\lambda \to \infty} \left( \frac{x^2}{2\lambda} - \frac{x}{\sqrt{\lambda}} + R \left( \frac{1}{\lambda^{3/2}} \right) \right) \]

\[ = \frac{x^2}{2} \]

\[ P_{a,b} = \frac{1}{\sqrt{2\pi}} \int_a^b \frac{e^{x^2/2}}{x^{2/2}} \Delta x \left( 1 + \theta'(\lambda) \right) \]

where \[ \theta'(\lambda) \to 0 \]

hence the limit \[ \lim_{\lambda \to \infty} P_{a,b} = \lim_{\lambda \to \infty} \left( \text{Riemann sum} \frac{1}{\sqrt{2\pi}} \sum_{a} e^{-x^2/2} \Delta x \right) \]

\[ = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} \, dx \]

Note: For a similar proof for the binomial case, see J. Neyman; First Course in Statistics, Chapter 4.

c. If \( X \) has the \( \chi^2 \) distribution, then

\[ \frac{X - n}{\sqrt{2n}} \] is \( N(0,1) \) as \( n \to \infty \)


d. If \( X \) has the Student's distribution with \( n \) d.f., then

\[ \frac{X - \bar{X}}{\sqrt{\frac{n}{n-2}}} \] is \( N(0,1) \) as \( n \to \infty \)

Proof: Deferred for now; a proof working with densities is given by Cramer, p. 250.

e. If \( X \) has the Beta distribution with parameters \( p, q \), then the standardized variate is \( N(0,1) \) as \( p \to \infty, q \to \infty \), and \( p/q \) remains finite

Note: \( X \) has mean \( \frac{p}{p+q} = \frac{1}{1 + q/p} \)

Proof: Omitted
Problem 19: \( X \) is a negative binomial r.v. \((r, p)\)

- Find the limiting d.f. of \( X \) as \( p \to 1; \ q \to 0; \ r \to \lambda \) (finite).

- State and prove a theorem that shows that under certain conditions a linear function of \( X \) is \( \mathcal{N}(0, 1) \).

Problem 20: \( X, Y \) have a joint density \( f(x, y) \).

Define \( E \left[ g(X) \mid Y \right] = \int_{-\infty}^{\infty} g(x) f(x \mid y) dx \)

where \( f(x \mid y) = f(x, y) / f_1(y) \). \( f_1(y) \) is the marginal density of \( y \).

Show that \( E \left[ g(X) \right] = \int_{-\infty}^{\infty} \int_Y g(x \mid y) f_1(y) dy dx \)

Use this to find the unconditional distribution of \( X \) if \( X \mid Y \) is \( \mathcal{B}(Y, p) \) while \( Y \) is Poisson \((\lambda)\).

Convergence in Probability

**Def. 10:** A sequence of r.v.s, \( X_1, X_2, \ldots, X_n \) is said to converge in probability to a r.v. \( X \) if given \( \delta > 0 \) there exists a number \( N(\varepsilon, \delta) \) such that for \( n > N(\varepsilon, \delta) \)

\[
\Pr \left[ \left| X_n - X \right| > \varepsilon \right] < \delta
\]

which is written \( X_n \xrightarrow{p} X \)

"\( \xrightarrow{p} \)" indicates converging in probability.

As a special case of this we have \( X_n \xrightarrow{p} c \) if given \( \delta > 0 \) there exists \( N \) such that for \( n > N \)

\[
\Pr \left[ \left| X_n - c \right| > \varepsilon \right] < \delta
\]

Further, in this notation, theorem 9 may be written: if \( X_n \) is a binomial r.v. with parameter \( n \), then

\[
\frac{X_n}{n} \xrightarrow{p} p
\]

**Theorem 16:** If \( X_1, X_2, \ldots, X_n \) are a sequence of independent random variables with means \( \mu_i \) and variances \( \sigma_i^2 \) then if

\[
\sum_{i=1}^{n} \frac{\sigma_i^2}{i^2} \to 0
\]

then

\[
Y_n = \frac{1}{n} \sum_{i=1}^{n} (X_i = \mu_i) \xrightarrow{p} 0
\]
Proof: By the corollary to the Tchebycheff Theorem (thm. 8)

\[ \Pr \left[ \left| \frac{1}{n} \sum_{i=1}^{n} (X_1 - \mu_1) \right| > K \sigma_{\frac{1}{n}} \right] \leq \frac{1}{K^2} \]

given \( n \), \( \delta \) choose \( \frac{1}{K^2} = \frac{\delta^2}{2} \leq \delta \)

Now \( \sigma_{\frac{1}{n}}^2 = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 \to 0 \) as \( n \to \infty \)

Put \( K = \sqrt{\frac{2}{\delta}} \) and take \( n \) sufficiently large so that

\[ \frac{\sqrt{2}}{\sqrt{\delta}} \left( \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2 \right)^{1/2} \leq \varepsilon \]

with this choice of \( n \)

\[ \Pr \left[ \left| \frac{1}{n} \sum_{i=1}^{n} (X_1 - \mu_1) \right| > \varepsilon \right] < \delta \]

If \( X_1, X_2, X_3, \ldots \) have the same distribution, then \( \sigma_{\frac{1}{n}}^2 = \frac{\sigma_2^2}{n} \)

and \( \sigma_{\frac{1}{n}}^2 = \frac{\sigma_2^2}{n} \to 0 \) as \( n \to \infty \)

so that in this particular case, \( \bar{X} \xrightarrow{p} \mu \).

Theorem 17: (Khintchine) (Weak Law of Large Numbers)

If \( X_1, X_2, \ldots, X_n \) are independently and identically distributed r.v. with
mean \( \mu \), then \( \bar{X} \xrightarrow{p} \mu \).

Proof: (See Cramer, p. 253-4)

\[ \phi_X(t) = E(e^{it\bar{X}}) = E(e^{i(\frac{1}{n} \sum X_1)t}) \]

\[ \phi_{\frac{X}{\bar{X}}}(t) = E(e^{i(\frac{X}{\bar{X}})^t}) = \left[ \phi_X(\frac{t}{n}) \right]^n \]

\[ \ln \phi_{\frac{X}{\bar{X}}}(t) = n \ln \phi_X(\frac{t}{n}) = n \ln \left[ 1 + i\mu \frac{t}{n} + R \right] \]

where \( \frac{R}{n} \to 0 \) as \( t \to 0 \)
\[ \phi_{\overline{X} - \mu}(t) = e^{-i\mu t} \phi(t) \]

\[ \ln \phi_{\overline{X} - \mu}(t) = -i\mu t + n \ln(1 + i\mu \frac{t}{n} + R) \]

Note: \( \ln(1 + x) = x + R(x) \)

where \( \frac{R(x)}{x} \to 0 \) as \( x \to 0 \)

\[ = -i\mu t + n \left[ i\mu \frac{t}{n} + R\left(\frac{t}{n}\right) + R' \right] \]

\[ = \frac{R_n(t/n)}{t/n} \text{ t which as } n \to \infty \quad \frac{t}{n} \to 0 \]

so \( \frac{R_n(t/n)}{t/n} \to 0 \)

hence \( \phi_{\overline{X} - \mu}(t) \to 1 \) as \( n \to \infty \)

but if \( \phi(t) = 1 \quad F(x) = 0 \quad x < 0 \)

\[ = 1 \quad x \geq 0 \]

or, in other words, the limiting distribution of \( \overline{X} - \mu \) is the trivial d.f. which takes the value 0 with probability 1.

Questions:

If \( X_1, X_2, X_3, \ldots, X_n \xrightarrow{p} X \)

Do \( \mu_1, \mu_2, \mu_3, \ldots, \mu_n \xrightarrow{p} \mu \) ?

Do \( \sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots, \sigma_n^2 \xrightarrow{p} \sigma^2 \) ?

i.e., does \( X_n \xrightarrow{p} X \) imply \( \mu_n \xrightarrow{p} \mu \) ?

Not necessarily, as shown by the following example.

Example:

Let \( X_n \) be defined as follows:

\[ X_n = n^2 \quad \text{with probability} \quad \frac{1}{n} \]

\[ = 0 \quad \text{with probability} \quad 1 - \frac{1}{n} \]
\[ X_n \xrightarrow{p} 0 \]
\[ E \left[ X_n \right] = 0(1 - \frac{1}{n}) + n^2 \left( \frac{1}{n} \right) = n \]

as \( n \to \infty \quad E \left[ X_n \right] \to \infty \)

**Problem 21:** Let \( X \) be the r.v. defined as the length of the first run in a series of binomial trials with \( \Pr[\text{success}] = p \). Find the distribution of \( X \), \( E \left[ X \right] \), and \( \sigma^2(X) \).

**Problem 22:** Let \( X_1, X_2, ..., X_n \) be independently uniformly distributed on \( 0,1 \). Let \( Y = \min (X_1, X_2, ..., X_n) \). Find the d.f. of \( Y \), \( E \left[ Y \right] \), and \( \sigma^2(Y) \).

Find the asymptotic d.f. of \( nY \) and of \( \frac{Y - E(Y)}{\sigma_Y} \).

Is \( \frac{Y - E(Y)}{\sigma_Y} \) a \( N(0,1) \) as \( n \to \infty \)?

**Theorem 18:** Let \( X_1, X_2, ..., X_n \) be a sequence of random variables with distribution functions \( F_1(x), F_2(x), ..., F_n(x) \to F(x) \). Let \( Y_1, Y_2, ..., Y_n \) be a sequence of random variables tending in probability to \( c \).

Define: \( U_n = X_n + Y_n \); \( V_n = X_n Y_n \); and \( W_n = X_n / Y_n \).

a) the d.f. of \( U_n \to F(x-c) \); and if \( c > 0 \)

b) the d.f. of \( V_n \to F(x/c) \)

c) the d.f. of \( W_n \to F(x/c) \)

**Proof:** All three parts are similar --- see Cramer p. 254-5 for proof of the third part.

Proof of the first statement (a):

Assume \( x = c \) is a point of continuity of \( F \). Let \( \varepsilon \) be small so that \( x = c \pm \varepsilon \) is an interval of continuity.

Let \( S_1 = \) the set of points such that
\[ X_n + Y_n \leq x; \quad |Y_n - c| \leq \varepsilon \]

\( S_2 = \) the set of points such that
\[ X_n + Y_n \leq x; \quad |Y_n - c| > \varepsilon \]

\( S = S_1 + S_2 = \) the set of points such that
\[ X_n + Y_n \leq x \]
\[ P(S_2) = \Pr \left[ X_n + Y_n \leq x, \ | Y_n - c| > \varepsilon \right] \leq \Pr \left[ | Y_n - c| > \varepsilon \right] \]

which tends to 0 as \( n \to \infty \), therefore we can choose \( n_1 \) so that \( n > n_1 \) implies \( P(S_2) < \frac{\delta}{3} \).

In \( S_1 \): \( c - \varepsilon \leq Y_n \leq c + \varepsilon \), thus

\[ F(x - c - \varepsilon) - \frac{\delta}{3} \leq F(x - c - \varepsilon) = P \left[ X_n \leq x - (c + \varepsilon) \right] \leq P(X_n + Y_n \leq x) = P(X_n = x - Y_n) \]

\[ \leq P \left[ X_n \leq x - (c - \varepsilon) \right] = F_{n_2} \left[ x - c - \varepsilon \right] \leq F(x - c + \varepsilon) - \frac{\delta}{3} \]

where \( n_2 \) is chosen so that when \( n > n_2 \), \( F_n(x) - F(x) \leq \frac{\delta}{3} \) in the vicinity of \( c \).

Therefore, in \( S_1 \)

\[ F(x - c - \varepsilon) - \frac{\delta}{3} \leq F(X_n + Y_n \leq x) \leq F(x - c + \varepsilon) + \frac{\delta}{3} \]

\( \varepsilon \) can be chosen so that \( F(x - c + \varepsilon) - F(x - c - \varepsilon) \leq \frac{\delta}{3} \) for \( n > \max(n_1, n_2) \).

Noting that \( \Pr \left[ X_n \leq x \right] = \Pr \left[ X_n + Y_n \leq x \right] = P(S_1) + P(S_2) \), we can write:

\[ -\delta \leq \frac{\delta}{3} - \frac{\delta}{3} \leq F(x - c - \varepsilon) - \frac{\delta}{3} + 0 - F(x - c) \leq \]

\[ P(S_1) + P(S_2) - F(x - c) = \Pr \left[ X_n \leq x \right] - F(x - c) \]

\[ = F(x - c + \varepsilon) + \frac{\delta}{3} + \frac{\delta}{3} - F(x - c) \leq \frac{\delta}{3} + \frac{\delta}{3} + \frac{\delta}{3} \leq \delta \]

which makes use of \( F(x - c + \varepsilon) - F(x - c) = \frac{\delta}{3} \)

\[ F(x - c - \varepsilon) - F(x - c) = -\frac{\delta}{3} \]

This then reduces to \( -\delta \leq \Pr \left[ X_n \leq x \right] - F(x - c) \leq \delta \)

hence \( \left| \Pr \left[ X_n \leq x \right] - F(x - c) \right| \leq \delta \) for \( n > \max(n_1, n_2) \)

which is what we set out to prove in the first place.
Theorem 19: If \( X_n \xrightarrow{p} c \) and if \( g(x) \) is a function continuous at \( x = c \),
\[
g(X_n) \xrightarrow{p} g(c)
\]

Problem 23: Prove Theorem 19. (work with fact that \( g(x) \) is continuous)

**Example of Theorem 18:**

Show \( t_n \) is \( A \mathcal{N}(0,1) \)

\[
t_n = \frac{\sqrt{n}(\overline{x} - \mu)}{s_n} \quad \text{where } x_1, x_2, \ldots \text{ are independent with mean } \mu, \text{ var } \sigma^2
\]

\[
s_n^2 = \frac{\sum (x_i - \overline{x})^2}{n-1}
\]

\[
t_n = \frac{\sqrt{n}(\overline{x} - \mu)}{\sigma} \cdot \frac{s_n}{\sigma} = \frac{\sqrt{n}(\overline{x} - \mu)}{\sigma} \quad \frac{s_n}{\sigma} \quad \text{which is in the } W_n \text{ form}
\]

\[
x_n = \frac{\sqrt{n}(\overline{x} - \mu)}{\sigma} \quad y_n = \frac{s_n}{\sigma}
\]

\( X_n \) is \( A \mathcal{N}(0,1) \) by the central limit theorem.

Hence, if we show that
\[
y_n = \frac{s_n}{\sigma} \xrightarrow{p} 1
\]

then the statement that \( t \) is \( A \mathcal{N}(0,1) \) follows from theorem 18(c).

\[
\frac{s_n^2}{\sigma^2} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{(n-1)\sigma^2} = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^2(n-1)}
\]

\[
= \frac{n}{n-1} \frac{\sum (x_i - \mu)^2}{\sigma^2} - \frac{n}{n-1} \frac{(\overline{x} - \mu)^2}{\sigma^2}
\]

\[
\frac{s_n^2}{\sigma^2} = \frac{n}{n-1} \left[ \frac{1/n \sum (x_i - \mu)^2}{\sigma^2} - \frac{(\overline{x} - \mu)^2}{\sigma^2} \right]
\]
by Khintchine's theorem (No. 17) this sample mean tends to $\sigma^2$

i.e., \( \frac{1}{n} \sum \frac{(\bar{X}_i - \mu)^2}{\sigma^2} \to \sigma^2 \)

therefore, the first term \( \frac{n}{n-1} \to 1 \)

\( \frac{n}{n-1} \to 1 \) as \( n \to \infty \)

hence, it remains to show that

\[ \frac{(\bar{X} - \mu)^2}{\sigma^2} \to 0 \]

but we know that \[ \frac{\bar{X} - \mu}{\sigma} \to 0 \] as \( n \to \infty \) by Khintchine's theorem (No. 17).

Therefore, \[ \frac{(\bar{X} - \mu)^2}{\sigma^2} \to 0 \] and \[ \frac{s_n^2}{\sigma^2} \to 1 \]

and thus by another application of Theorem 19 \[ \frac{s_n^2}{\sigma^2} \to 1 \]

Note: The \( t_n \) in this case is Student's distribution if the \( x \) are independently and normally distributed --- however, this is for general \( t_n \).

Re Theorem 19, see: Mann and Wald; Annals of Math. Stat., 1943;
"On Stochastic Order Relationships" --- for extending ordinary limit properties to probability limit properties.

Misc. remarks: On the Taylor series remainder term as used in the proofs of theorems 14, 17, and on p. 37 --- see Cramer p. 122.

If \( f(x) \) is continuous and a derivative exists, then we can write

\[ f(x) = f(a) + (x - a)f'(a + \theta (x - a)) \quad 0 \leq \theta \leq 1 \]

\[ = f(a) + (x - a)[f'(a) + f'(a + \theta(x - a))] - f'(a) \]

\[ = f(a) + (x - a)f'(a) + R \]

where \( R = (x - a)[f'(a + \theta(x - a)) - f'(a)] \)
\[
\frac{R}{x - a} = f^\prime \left[ a + \Theta(x - a) \right] - f^\prime(a) \\
\frac{f^\prime(x)}{f^\prime(a)} \to 0 \\
\text{then if } f^\prime \text{ is continuous, as } x \to a \\
\lim_{x \to a} \frac{R}{x - a} = 0 \quad \text{i.e., } R \text{ converges to } 0 \text{ faster than } x - a \\

\text{Remark: If there exists an } A \text{ such that } \int_{-\infty}^{A} g(x) \, dF_n(x) < \epsilon \\
\text{and } \int_{A}^{\infty} g(x) \, dF_n(x) < \epsilon \\
\text{for } n = 1, 2, 3, \ldots \text{ then if } \frac{F_n}{F} \xrightarrow{p} F \\
\int_{-\infty}^{\infty} g(x) \, dF_n(x) \xrightarrow{p} \int_{-\infty}^{\infty} g(x) \, dF(x) \\
\text{Ref. Cramer, p. 74}
\]

\text{Question: Under what conditions does } E(t_n) = 0 \text{ for all } n \\
or does } E(t_n) \xrightarrow{a} 0 \text{ as } n \to \infty? \\
(t_n \text{ is defined as in the example illustrating theorem 18)}

\text{Counter-example:}

Define 
\[ p_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1} e^{-x^2/2} \]
\[ X_n \text{ is normal } (0,1) \text{ except on the interval } (1 - \frac{1}{n}, 1) \]
\text{and } X_n = 1 \text{ with probability } p_n.

Then with probability \((p_n)^n\) \(X_1, X_2, \ldots, X_n = 1\) 

in which case 
\[ t_n \sqrt{n(1 - \mu)} = \infty \]

therefore \(E(t_n)\) is not defined.

\text{Problem 24:}

\(X\) is Poisson \(\lambda\), then 
\[ Y = \frac{(X - \lambda)^2}{X} \]
\text{is asymptotically } \chi^2(1) \text{ as } \lambda \to \infty
Convergence Almost Everywhere:

Def. 19: A sequence of r.v.s, $X_1, X_2, \ldots \rightarrow X$ a.e. if given $\varepsilon, \delta$ there exists as $N$ such that

$$\Pr \left[ |X_j - X| < \varepsilon \quad j = N, N+1, N+2, \ldots \right] \geq 1 - \delta$$

Ref.: Feller, Chapter 9.

Note: Convergence almost everywhere is sometimes called "strong convergence," convergence in probability is sometimes called "weak convergence".

Example: (of a case when $X_n \overset{p}{\rightarrow} c$ but $X_n \not\rightarrow c$ a.e.)

- $X_n = 0$ with probability $1 - \frac{1}{n}$
- $X_n = 1$ with probability $\frac{1}{n}$

the $X$'s are independent.

(1) To show $X_n \overset{p}{\rightarrow} 0$

$$\Pr \left[ \bigcup_{n=N}^{\infty} \{X_n = 0 \} \right] = \int_{N}^{\infty} \Pr[X_n < \varepsilon] \, \text{d}x = \int_{N}^{\infty} \left(1 - \frac{1}{N+j}\right) \, \text{d}x$$

for any $\varepsilon < 1$ and $\frac{1}{n}$ can be made arbitrarily small by increasing $n$.

(2) $X_n \not\rightarrow 0$ a.e.

$$\Pr \left[ \bigcup_{n=N}^{\infty} \{X_n = 0 \} \right] < \varepsilon \quad n = N, N+1, N+2, \ldots$$

Note: $1 - x \leq e^{-x}$ for $0 \leq x \leq 1$

$$\leq \sum_{j=0}^{\infty} e^{-\frac{1}{N+j}} = \exp \left\{ -\sum_{j=0}^{\infty} \frac{1}{N+j} \right\}$$

which series is divergent

$$\leq e^{-\infty} = 0$$
therefore \( \Pr \left[ \left| X_n - 0 \right| < \epsilon \quad n = N, N+1, N+2, \ldots \right] = 0 \)

due to

therefore \( X_n \not\rightarrow 0 \) a.e.

**Problem 25:**

If \( X_n = 0 \) with probability \( 1 - \frac{1}{2^n} \)

\( X_n = 1 \) with probability \( \frac{1}{2^n} \)

then \( X_n \not\rightarrow 0 \) a.e.

**Problem 26:**

\( \varphi_X(t) = \cos t a \)

1. What is the d.f. of \( X \)?
2. Is \( \overline{X} \) A. N. as \( n \to \infty \) (suitably normalized)?
3. To what does \( \overline{X} \) converge as \( a \to 0 \)?

**Problem 27:**

\( X_1 \) and \( Y_1 \) are independent and identically distributed random variables with means \( \mu, \nu \) and variances \( \sigma_1^2, \sigma_2^2 \). Find and prove the asymptotic d.f. of \( \overline{X}_n \overline{Y}_n \) (suitably normalized).

**Kolmogorov inequality:** Let \( \{X_i\} \) be a sequence of r.v. with means \( \mu_i \) and variances \( \sigma_i^2 \).

Then

\[
\Pr \left[ \left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu_1 \right| \leq K \quad K = 1, 2, \ldots, m \right] > 1 - \frac{1}{K^2}
\]

**Examples:**

\[
\begin{align*}
X_1 & \quad 0 \quad 1 \\
X_2 & \quad 0 \quad 1 \\
\Pr \left[ \left| X_1 \right| < K, \quad \left| X_1 + X_2 \right| < \sqrt{2K} \right] & > 1 - \frac{1}{K^2}
\end{align*}
\]
Theorem 20:

If \( x_1, x_2, x_3, \ldots \) are independent random variables with means \( \mu_1 \)

and variances \( \sigma_1^2 \), then, if \( \sum_{1}^{\infty} \frac{\sigma_1^2}{i^2} \) converges

\[
\left( \bar{X}_n - \bar{\mu}_n \right) \rightarrow 0 \text{ a.e.}
\]

or we say that \( \bar{X} \) obeys the strong law of large numbers

where \( \bar{\mu} = \frac{1}{n} \sum_{1}^{n} \mu_i \)

Proof of Theorem 20:

Let \( A_j \) be the event that for some \( n \) in the interval \( 2^{j-1} < n < 2^j \)

\[
|\bar{X}_n - \mu_n| > \varepsilon \quad \text{(violating the definition of convergence a.e.)}
\]
\[ Pr \left( A_j \right) = Pr \left( \left| Y_n - \mu \right| > \varepsilon \text{ for some } n \right) \]
\[ = Pr \left( \left| \sum_{i=1}^{n} (X_i - \mu_i) \right| > n\varepsilon \right) \]
\[ \leq Pr \left( \left| \sum_{i=1}^{n} (X_i - \mu_i) \right| > 2^{j-1}\varepsilon \right) \text{ inserting the lower bound on } n \]
\[ \leq Pr \left[ \left( \sum \sigma_i^2 \right)^{1/2} > \frac{2^{j-1}\varepsilon}{\left( \sum \sigma_i^2 \right)^{1/2}} \right] \]

from the Kolmogorov inequality:
\[ Pr \left[ \frac{X_1 - \mu_1}{\sigma_1} < K; \frac{(X_1 + X_2) - (\mu_1 + \mu_2)}{(\sigma_1^2 + \sigma_2^2)^{1/2}} < K; \ldots \right] \]
\[ \frac{\left| \sum (X_i - \mu_i) \right|}{\left( \sum \sigma_i^2 \right)^{1/2}} < K \]
\[ > 1 - \frac{1}{K^2} \]

hence for the event \( A \)
\[ Pr \left[ A_j \right] \leq \frac{\sum_{i=1}^{2^j} \sigma_i^2}{2^{2(j-1)} \varepsilon^2} \]
\[ \leq \frac{h \sum_{i=1}^{2^j} \sigma_i^2}{2^{2j} \varepsilon^2} \]
\[ \sum_{j=1}^{\infty} Pr \left[ A_j \right] \leq \frac{h}{2} \sum_{j=1}^{\infty} \frac{1}{2^j} \sum_{i=1}^{2^j} \sigma_i^2 \]
Interchanging the order of summation

\[
\leq \frac{4}{e^2} \sum_{i=1}^{\infty} \sigma_i^2 \sum_{j=1}^{\infty} \frac{1}{2^{2j+1}}
\]

Note that \(\sum_{j=1}^{\infty} \frac{1}{2^{2j+1}} = \frac{1}{1 - \frac{1}{2^2}} \) since it is a geometric series.

\[
\leq \frac{4}{e^2} \sum_{i=1}^{\infty} \sigma_i^2 \cdot \frac{4}{3} = \frac{16}{3e^2} \sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2}
\]

Now this sum is finite by hypothesis — hence \(\sum_j \Pr[A_j]\) converges and we can choose \(N\) so that

\[
\sum_{j=1}^{\infty} \Pr[A_j] < \delta
\]

hence definition 19 for convergence a.e. is satisfied.

**Corollary:** (to theorem 20)

If \(X_i\) are independent and identically distributed \((i = 1, 2, 3, \ldots)\) with mean \(\mu\) and variance \(\sigma^2\) then

\(\mathbf{X} \rightarrow \mu\) a.e.

Proof: is immediate since \(\sum \frac{\sigma_i^2}{i^2} = \sigma^2 \sum \frac{1}{i^2}\) which converges, i.e. < \(\infty\).

**Other Types of Convergence:**

1. Convergence in the mean

   \[ l.i.m. \quad X_n = X \quad \text{if} \quad \lim_{n \to \infty} \mathbb{E} \left[ (X_n - X)^2 \right] = 0 \]

   Note: \(l.i.m. = \text{limit in the mean}\)

   Implies convergence in probability but not convergence a.e.

2. Law of the iterated logarithm

Ref.: Feller, Chapter 8

3. St. Petersburg paradox

\[ X = 2^n \text{ with probability } \frac{1}{2^n}, \quad n = 1, 2, 3, \ldots \]

\[ \sum_{1}^{\infty} \frac{1}{2^n} = 1 \]

Note:

E.g. Toss a coin until heads comes up -- count the total number of tosses required \((= n)\) -- bank pays \(2^n\) to the player.

\[ E(x) = \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \sum_{1}^{\infty} 1 = \infty \]

For a fair game, the entry fee should be equal to the expected gain; therefore, this game presents a problem in the determination of the "fair" entry fee.

Ref.: Feller, p. 235-7 -- he shows

\[ \Pr \left[ \frac{\sum_{1}^{n} X_i}{n \ln n} \right. \left. = 1 \mid > \delta \right] < \delta \]

That is, the game "becomes fair" if the entry fee is \(n \ln n\).
ESTIMATION (Point):

Cramer --- ch. 32-3

\( F(x, \Theta) \) --- a family of distributions

\( x, \Theta \) may be vector valued, in which case they will be denoted \( \bar{x}, \Theta \)
\( x \in \mathbb{R}_k \)
\( \Theta \in \Omega \) --- parameter space

Example: 1: if \( x_1, x_2, \ldots, x_n \) are NID(\( \mu, \sigma^2 \)) then

\( \Omega \) consists of all possible \( \mu \)
and all positive \( \sigma^2 \)

\( \infty \) \( \mu \) \( \sigma^2 \) \( \infty \)

2: \( F(x) \) is \( F \) the family of all continuous distributions

\( \Omega \) is then the space of all continuous d.f.

--- this is the non-parametric case
--- might wish to estimate \( E_F(x) = \int_{-\infty}^{\infty} x \, dF(x) \)

provided that we add the restriction that \( E_F(X) < \infty \)

Estimate of \( g(\Theta) \) is some function of \( X \) from \( \mathbb{R}_k \) to \( \Omega \) which in some sense comes close to \( g(\Theta) \)

Or in the "general decision theory" point of view (Wald)

an estimate of \( g(\Theta) \) is a decision function \( d(x) \) and we have associated with each decision function a loss function \( W \left[ d(x), \Theta \right] \) with \( W = 0 \) whenever \( d(x) = g(\Theta) \)

The choice of loss function is arbitrary, but we frequently choose

\[ W \left[ d(x), \Theta \right] = \left[ d(x) - g(\Theta) \right]^2 \]

Def. 20: Risk Function is defined as

\[ R(d, \Theta) = E \left\{ W \left[ d(x), \Theta \right] \right\} = \int_{-\infty}^{\infty} W \left[ d(x), \Theta \right] \, dF(x, \Theta) \]
Example: $X_1, X_2, \ldots, X_n$ are NID($\mu, 1$)

$$d^*(X) = \bar{X}$$

$$R(d^*, \mu) = E(\bar{X} - \mu)^2 = \frac{1}{n}$$

A "best" estimate might be defined as one which minimizes $R(d, \theta)$ with respect to $d$ uniformly in $\theta$.

$$R(d, \theta) \leq R(d^*, \theta) \text{ for all } \theta \text{ with } d^* \text{ any other estimator}.$$  

consider the estimate $d(x)$ of $g(\theta)$ defined as $d(x) = g(\theta_0) \quad R(d, \theta_0) = 0$

Hence a uniformly (supposedly) minimum risk estimate can be found only if there exists a $d(x)$ such that $R[d(x), \theta] = 0$

An example would be similar to asking which is better for estimating time -- a stopped clock which is right twice a day, or one that runs five minutes slow.

Since a uniformly best estimate is virtually impossible to find, we want to consider alternative

**WAYS to formulate the problem of obtaining best estimates:**

I. by restricting the class of estimates

_1. unbiased estimates

**Def. 21:** $d(X)$ is unbiased if $E[d(X)] = g(\theta)$

$d(X)$ is a minimum variance unbiased estimate (m.v.u.e.) if

$$E[(d(X) - g(\theta))^2]$$

is minimized over unbiased estimates $d$

_2. invariant estimates

Let $h(X)$ be the transformation of a real line into itself which induces a transformation $h$ on parameterspace. If $d[h(X)] = R[d(X)]$ then $d(X)$ is invariant under this transformation.

Example: family of $d_x \phi$ with $E(X) < \infty$

Problem is to estimate $E(X)$

$$h(x) = ax + b \quad R[E(X)] = aE(X) + b$$

An estimate $d(X)$ of $\mu$ is invariant if $d[h(X)] = R[d(X)]$

$$d(ax + b) = a d(X) + b$$

Therefore $\bar{X}$ is an invariant estimate of $\mu$ under this transformation.

Note that $d(X) = \mu_0$ is not invariant.
3. Best linear unbiased estimates (b.l.u.e.)

Def. 22: Estimates of \( g(\theta) \) which are unbiased, linear in the \( X_i \) and which among such estimates have minimum variance are b.l.u.e.

**Problem 28:** \( X_1, X_2, \ldots, X_n \) are independent random variables with mean \( \mu \) and and variance \( \sigma^2 \)

show that \( \bar{X} \) is the b.l.u.e. of \( \mu \).

**Problem 29:** \( X_1, X_2, \ldots, X_n \) are NID (\( \mu, \sigma^2 \))

\[
W(d(X), \theta) = \begin{cases} 
  b[d(X) - \mu] & \text{if } d(X) > \mu \\
  c[d(X) - \mu] & \text{if } d(X) < \mu 
\end{cases}
\]

a. \( d(X) = \bar{X} \) find \( R(\bar{X}, \mu, \sigma) \)

(note: the answer depends on the loss function constants only, not \( \mu \))

b. \( d(X) = \bar{X} + a \) — show how to determine \( a \) such that \( R(d, \mu, \sigma) \) is minimized

(note: the answer involves \( \phi(z) \) which is defined as

\[
\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt
\]

Comment on this problem:
An orthogonal transformation
— is a rotation or reflection
— \( Y = AX \) where \( A \) is orthogonal
— \( |J| = 1 \)

\[ \sum y_i^2 = \sum x_i^2 \]

— For \( y_i = a_{i1} x_1 + a_{i2} x_2 + \cdots + a_{in} x_n \); \( \sum_{j=1}^{n} a_{ij}^2 = 1; \sum_{j} a_{ij} a_{kj} = 0 \) \( i \neq k \)

In general if \( d(X) \) is a function of \( T(X_1, X_2, \ldots, X_n) \)

\[
R(d, \theta) = E[W(d, \theta)]
= \int \cdots \int W[T(x), \theta] dF(x)
\]
Making the transformation 

\[ y = T(x) \]

\[ y = \cdots \{ \text{n-1 functions independent of the first one} \} \]

Then

\[ R \left[ d(T), \theta \right] = \int W \left[ T(x), \theta \right] dF_1(y_1) \cdots \left( \begin{array}{c} \text{marginal distribution} \\ \text{of } y_1 = T(x) \end{array} \right) \left( \begin{array}{c} \text{conditional distribution} \\ \text{of } y_2, \ldots, y_n \end{array} \right) \]

\[ = \int W \left[ y_1, \theta \right] dF(y_1) \]

\[ = \int W \left[ T(x), \theta \right] dF(T(x)) \]

II. Optimum Properties in the Large

1. Bayes Estimates

**Def. 23:** If \( \theta \) has a known "a priori" distribution \( H(\theta) \) then the Bayes estimate of \( g(\theta) \) is that \( d(\theta) \) which minimizes

\[ \int R(d, \theta) \, dH(\theta) \]

with respect to \( d(x) \)

**Example:** \( X \) is \( B(n, p) \) and \( p \) is uniform on \((0, 1)\)

Let \( W \left[ d(X), \theta \right] = [d(X) - p]^2 \) and minimize this with respect to \( d \)

\[ R(d, \theta) = \sum_{x=0}^{n} [d(x) - p]^2 \binom{n}{x} p^x (1 - p)^{n-x} \]

Average risk = risk function averaged with respect to \( p \)

\[ \frac{1}{0} \int R(d, \theta) \, dp \]
\[
\sum_{x=0}^{n} \frac{1}{x! (n-x)!} \left[ d^2(x) - 2pd(x) + p^2 \right] \frac{\binom{n}{l}}{x! (n-x)!} p^x (1-p)^{n-x} \, dp
\]

Note: \( \int_0^1 p^a (1-p)^b \, dp = \frac{a! \cdot b!}{(a+b+1)!} \)

Using this evaluation on each part separately we have

\[
\sum_{x=0}^{n} \left[ d^2(x) - 2d(x) \frac{\binom{n}{l}}{x! (n-x)!} \frac{\binom{n+l}{x} (n-x)!}{(n+1)!} \frac{(x+1)! (n-x)!}{(n+2)!} \right] + \frac{\binom{n}{l}}{x! (n-x)!} \frac{(x+1)! (n-x)!}{(n+3)!} \]

\[
= \sum_{x=0}^{n} \left[ \frac{d^2(x)}{n+1} - 2d(x) \frac{(x+1)(x+2)}{(n+1)(n+2)(n+3)} \right]
\]

\[
= \frac{1}{n+1} \sum_{x=0}^{n} \left[ \frac{d^2(x)}{n+2} - \frac{2d(x)(x+1)}{(n+2)(n+3)} + \left( \frac{x+1}{n+2} \right)^2 + \frac{(x+1)(x+2)}{(n+2)(n+3)} - \left( \frac{x+1}{n+2} \right)^2 \right]
\]

\[
= \frac{1}{n+1} \sum_{x=0}^{n} \left[ \left( \frac{d(x)}{n+2} - \frac{x+1}{n+2} \right)^2 + \frac{x+1}{n+2} \left( \frac{n+1-x}{(n+2)(n+3)} \right) \right]
\]

This is certainly minimised with respect to \( d(x) \) if each term in the first summation is zero — i.e. if

\[ d(x) = \frac{x+1}{n+2} \]

**Problem 30**: if \( d(X) = \frac{X}{n} \) find \( R(d, p) \) as a function of \( p \) and also

\[
\text{average } R = \int_0^1 R(d, p) \, dp
\]

- if \( p \) is uniformly distributed on \((0, 1)\)
- \( R(d, p) = E \left[ (d(X) - p)^2 \right] \)
2. Minimax Estimate

**Def. 24:** \( d(X) \) is a minimax estimate if \( d(X) \) minimizes \( \sup_{\theta} R(d, \theta) \) in comparison to any other estimate \( d^*(X) \).

\[
R(\theta)
\]

\( d_1(\theta) \) is minimax estimate since it has a minimum maximum point

\[
d_2
\]

\[ R(d_1, \theta) \quad R(d_2, \theta) \]

i.e., we get the min (with respect to \( d \)) of the max (with respect to \( \theta \)) of \( R(d, \theta) \)

\[
\text{-- or we take the inf (}d\sup(\theta)\text{) of } R(d, \theta)
\]

3. Constant Risk Estimates

**Def. 25:** A constant risk estimate is one for which \( R(d, \theta) \) is constant with respect to \( \theta \).

**Problem 31:** Find a constant risk estimate among linear estimates of \( p \) if \( X \) is \( B(n, p) \).

III -- By dealing only with large sample (asymptotic) properties of estimates

**Def. 26:** Consistent Estimates \( \Rightarrow d_n(X) \) is consistent if:

\[
d_n(X) \overset{P}{\rightarrow} \theta
\]

(it does not necessarily follow that \( \mathbb{E}(d_n) \rightarrow \theta \) or that \( \sigma(d_n)^2 \rightarrow 0 \))

**Problem 32:** If \( \mathbb{E}(d_n) \rightarrow \theta \) and \( \sigma(d_n)^2 \rightarrow 0 \) then \( d_n(X) \) is consistent.

(These are the sufficient conditions for consistency)

**Def. 27:** Best Asymptotically Normal Estimates (B.A.N.E.) \( \Rightarrow d(X) \) is a B.A.N. estimate if:

1. \[
\frac{d_n - \mathbb{E}(d_n)}{\sigma(d_n)} \text{ is } \mathcal{N}(0, 1)
\]

2. if \( d_n^* \) is any other A. N. estimate, then

\[
\lim_{n \to \infty} \frac{\sigma^2(d_n)}{\sigma^2(d_n)} \leq 1
\]
METHODS OF ESTIMATION

A -- Methods of Moments (K. Pearson)

\[ \Theta = \Theta_1, \Theta_2, \ldots, \Theta_k \] -- equate the first \( k \) sample moments to \( k \) population moments (expressed as functions of \( \Theta_1, \Theta_2, \ldots, \Theta_k \)). Solve these equations for \( \Theta_1, \Theta_2, \ldots, \Theta_k \) and these are the moment estimates.

**Example:** \( X_1, X_2, \ldots, X_n \) are NID(\( \mu, \sigma^2 \))

- First population moment = \( \mu \)
- Second population moment = \( \sigma^2 \)

\[ \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

then \( \bar{X} = \hat{\mu}; \quad \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{\sigma}^2 \) ("n" being the divisor used by K. Pearson)

**Note:** This method yields poor estimates in many cases -- has very few optimum properties

B -- Method of Least Squares (Gauss -- Markov)

Let \( \mathbf{X} = (X_1, X_2, \ldots, X_n) \) be a r.v. with \( \mathbb{E}(\mathbf{X}) = \mathbf{A\Theta} \quad \Theta = (\Theta_1, \Theta_2, \ldots, \Theta_s) \)

\[ \mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1s} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{ns} \end{pmatrix} s \leq n \]

\( \mathbf{A} \) is of rank \( s \)

or we have \( \mathbb{E}(X_i) = \sum_{j=1}^{n} a_{ij} \Theta_j \quad i = 1, 2, \ldots, n \)

also \( \sigma^2(x_i, x_j) = \delta_{ij} \sigma^2 \) i.e., the covariances = 0

**Def. 28:** \( \mathbf{\Theta}^* \) is a least squares estimate of \( \Theta \) if \( \mathbf{\Theta}^* \) minimizes

\[ (\mathbf{X} - \mathbf{A\Theta})' (\mathbf{X} - \mathbf{A\Theta}) = \sum_{i=1}^{n} (X_i - a_{11}\Theta_1 - \cdots - a_{is}\Theta_s)^2 \]

**Theorem 21:** (Gauss -- Markov)

With the given conditions on \( X_1, X_2, \ldots, X_n \) the least squares estimate \( \mathbf{\Theta}^* \) is a best linear unbiased estimate (b.l.u.e.) of \( \Theta \).

Lehman

we first show that

\[ \Theta^* = C^{-1} A' X \]

where \( C = A'A \) \( C' = C \) since it is symmetric.

if we write \( \Theta = C^{-1} A' X + Y \) then we are trying to minimize

with respect to \( y \),

\[ \Phi = (X - A \Theta)'(X - A \Theta) \]

\[ = (X - AC^{-1} A' X - AY)'(X - AC^{-1} A' X - AY) \]

\[ = (X - AC^{-1} A' X)'(X - AC^{-1} A' X) - (X - AC^{-1} A' X)'AY \]

\[ - (AY)'(X - AC^{-1} A' X) + (AY)'(AY) \]

now the cross-product terms equal zero

\[ e.g. - X'AY + X'A(C^{-1})'A'AY = - X'AY + X'AY = 0 \]

since \( (C^{-1})' = C^{-1} \) \( C = A'A \)

\( (C^{-1})'A'A = C^{-1}C = I \)

similarly the second cross-product also = 0

hence \( \Phi \) is minimized with respect to \( y \) if \( (AY)'(AY) \) is minimized which will happen if \( AY = 0 \) since \( A \) has maximum rank \( s \) this will happen only if \( y = 0 \)

or writing \( \Phi = \sum_{i=1}^{n} (x_i - a_{il} \Theta_l - \cdots - a_{ij} \Theta_j - \cdots - a_{is} \Theta_s)^2 \)

formally minimizing \( \Phi \) in the usual fashion by differentiating with respect to the \( \Theta_j \)

\[ \frac{\partial \Phi}{\partial \Theta_j} = -2 \sum_{i=1}^{n} a_{ij}(x_i - a_{il} \Theta_l - \cdots - a_{ij} \Theta_j - \cdots - a_{is} \Theta_s) = 0 \]

\[ j = 1, 2, \ldots, s \]

solving these equations

\[ \sum_{i=1}^{n} a_{ij}x_j = (\sum_{i=1}^{n} a_{ij}a_{il}) \Theta_l + \cdots + (\sum_{i=1}^{n} a_{ij}a_{is}) \Theta_s \]

or \( A'X = (A'A) \Theta \)
To show that $\hat{\theta}^*$ so defined (i.e., $= \theta^{-1} A' X$) is B.L.U.E.,

consider a linear estimate $B_{x_n X}$

$$E[EX] = \theta \quad B E[X] = \theta \quad BA = \theta \quad \text{thus} \quad BA = I$$

note that $\theta^2 BB'$ is the covariance matrix of $EX$ -- the elements in the

diagonal are the variances of the estimates $\theta$ -- we thus wish to minimize
these diagonal elements (by proper choice of $B$) subject to the

restriction that $BA = I$

$$(B - \theta^{-1}A') (B - \theta^{-1}A')' = BB' - B(A')' (\theta^{-1})' - \theta^{-1} A' B' + \theta^{-1} A' (A')' (\theta^{-1})$$

using the relationships that $BA = I \quad A'A = \theta \quad C = \theta \quad (\theta^{-1})' = \theta^{-1}$

$$(B - \theta^{-1}A') (B - \theta^{-1}A')' = BB' - (\theta^{-1})' - \theta^{-1} + \theta^{-1} = BB' - \theta^{-1}$$

thus $BB' = \theta^{-1} + (B - \theta^{-1}A') (B - \theta^{-1}A')'$

minimization will occur if $(B - \theta^{-1}A') = 0$ or if $B = \theta^{-1}A'$ ($\theta = \theta^{-1}A' X$)

hence $\hat{\theta}^*$ are B.L.U.E.

example: if $X_1, X_2, \ldots, X_n$ are uncorrelated with $E(X_1) = \mu$ and

common variance $\sigma^2$ then the least squares estimate of $\mu$ is $\bar{X}$

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

let $A = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ here $s = 1$ (we have a $1 \times n$ matrix of 1's)

$$C = A' A = n \quad \theta^{-1} = \frac{1}{n}$$

$$\hat{\mu}^* = \theta^{-1} A' X = \frac{1}{n} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$$

Problem 33: $E(X_1) = \alpha + \beta t_i$ \quad $i = 1, 2, \ldots, n$

t_i are known constants and assume $\sum_{i=1}^{n} t_i = 0$

find B.L.U.E. of $\alpha, \beta$ using theorem 21,
Aiken extended this result in 1934 to the case where the \( X_i \) are correlated and we know the correlation matrix \( V \) (up to an arbitrary multiplier) -- b.l.u.e. are also least squares estimates which are obtained by minimizing

\[
(X - A'\theta)'V^{-1}(X - A'\theta)
\]

ref: Plackett: Biometrika, 1949, p.45

C. Maximum Likelihood Estimates (ref. Cramer ch. 33)

Def. 29: Likelihood function

\[
L = f(X_1, X_2, \ldots, X_n, \theta) \text{ if the } X_i's \text{ are continuous}
\]

\[
= p(X_1, X_2, \ldots, X_n, \theta) \text{ where the } p's \text{ are discrete probabilities}
\]

if the \( X_i \) are continuous, independent, and identically distributed

\[
L = \prod_{i=1}^{n} f(X_i, \theta) \text{ or } \ln L = \sum_{i=1}^{n} \ln f(X_i, \theta)
\]

Def. 30: A Maximum Likelihood Estimate (M.L.E.) is that value of \( \theta \) (denoted \( \hat{\theta} \)) which maximizes the function \( L \) (or \( \ln L \))

It may happen (from the third case in def. 29) that \( \hat{\theta} \) is the solution of the set of equations

\[
\frac{\partial \ln L}{\partial \theta_i} = 0 \quad i = 1, 2, \ldots, s \quad \theta = (\theta_1, \theta_2, \ldots, \theta_n)
\]

Regularity conditions needed in the maximum likelihood derivations (ref: Cramer 500-504)

1. The \( X_i \) are continuous, independent, and identically distributed.

   We will assume first that \( \theta \) is a scalar.

2. \( \frac{\partial \ln f(X_i, \theta)}{\partial \theta} \) is a function of \( X_i \) and hence is a random variable.

   we assume that \( \frac{\partial^k \ln f(X_i, \theta)}{\partial \theta^k} \) exists for \( k = 1, 2, 3 \)

3. \( \Omega \) is an interval and \( \theta_0 \) (the true value of \( \theta \)) is an interior point

4. \( \frac{\partial^k \ln f(X_i, \theta)}{\partial \theta^k} < F_k(X_i) \) which is integrable over \((-\infty, \infty)\)
\[ E \left[ f_3(x_1) \right] < M \text{ for all } \theta \quad \text{--- i.e., it is bounded} \]

5. \[ E \left[ \frac{\partial \ln f_1}{\partial \theta} \right]^2 = \int_{-\infty}^{\infty} \left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \right)^2 f(x_1, \theta) \, dx_1 = k^2 \quad 0 < k^2 < \infty \]

**Theorem 22:** If \( f_1 \) satisfy the regularity conditions, and if \( L \), or \( \ln L \), has a unique maximum, then

1. the maximum likelihood estimate \( \hat{\theta} \) is the solution of the equation

\[ \frac{\partial \ln L}{\partial \theta} = 0 \]

2. \( \hat{\theta} \) is consistent

3. \( \sqrt{n} \left( \hat{\theta} - \theta_0 \right) \) is asymptotically normal with mean 0 and variance

\[ \frac{1}{E \left[ \frac{\partial \ln f_1}{\partial \theta} \right]^2} \]

**Proof:**

1. since \( \frac{\partial \ln L}{\partial \theta} \) is continuous with a continuous derivative, if

\[ \ln L = \sum_{1}^{n} \ln f(x_1, \theta) \]

has a maximum, \( \frac{\partial \ln L}{\partial \theta} = 0 \) at this max.

2. to show that \( \hat{\theta} \) is consistent

\[ f_1 = f(x_1, \theta) \]

\[ \frac{\partial \ln f_1}{\partial \theta} = \left. \frac{\partial \ln f_1}{\partial \theta} \right|_{\theta = \theta_0} + \left. \frac{\partial^2 \ln f_1}{\partial \theta^2} \right|_{\theta = \theta_0} (\theta - \theta_0) + \left. \frac{\partial^3 \ln f_1}{\partial \theta^3} \right|_{\theta = \theta_0} (\theta - \theta_0)^2 \]

summing each term on both sides of the equality, dividing by \( n \), and doing some substituting

\[ \frac{1}{n} \frac{\partial \ln L}{\partial \theta} = \frac{1}{n} \sum_{i=1}^{n} \left. \frac{\partial \ln f_1}{\partial \theta} \right|_{\theta = \theta_0} + \frac{1}{n} \sum_{i=1}^{n} \left( \left. \frac{\partial^2 \ln f_1}{\partial \theta^2} \right|_{\theta = \theta_0} (\theta - \theta_0) \right) + \frac{z}{n} \sum_{i=1}^{n} f_3(x_1) (\theta - \theta_0)^2 \]

the term in the third derivative is replaced by the term in regularity condition \( 4a \) --- and multiplied by the factor \( z(0 \leq z \leq 1) \) to restore the equality

this equation can be written:

\[ \frac{1}{n} \frac{\partial \ln L}{\partial \theta} = B_0 + B_1 \left( \theta - \theta_0 \right) + z \frac{B_2 (\theta - \theta_0)^2}{2} \]
note that we have: \[
\int_{-\infty}^{\infty} f(x_i, \theta) \, dx_i = 1
\]

from which we can show:

\[
\frac{\partial}{\partial \theta} \int f(x_i, \theta) \, dx_i = 0 \iff \int \frac{\partial f(x_i, \theta)}{\partial \theta} \, dx_i = 0 \iff \int \left( \frac{1}{f(x_i, \theta)} \frac{\partial f_i}{\partial \theta} \right) f(x_i, \theta) \, dx_i = 0
\]

therefore:

\[
E \left[ \frac{\partial \ln f_i}{\partial \theta} \right] = 0
\]

also, differentiating a second time we have:

\[
\frac{\partial^2}{\partial \theta^2} \int f(x_i, \theta) \, dx_i = 0
\]

or

\[
\int \frac{\partial^2 f(x_i, \theta)}{\partial \theta^2} = 0
\]

\[
E \left[ \frac{\partial^2 \ln f_i}{\partial \theta^2} \right] = \int_{-\infty}^{\infty} \frac{\partial^2 \ln f_i}{\partial \theta^2} f(x_i, \theta) \, dx_i = \int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \left[ \frac{1}{f(x_i, \theta)} \frac{\partial f_i}{\partial \theta} \right] f(x_i, \theta) \, dx_i
\]

\[
= \int_{-\infty}^{\infty} \left\{ \frac{1}{f(x_i, \theta)} \frac{\partial^2 f_i}{\partial \theta^2} f(x_i, \theta) - \frac{1}{f_i} \left( \frac{\partial f_i}{\partial \theta} \right)^2 f(x_i, \theta) \right\} \, dx_i
\]

... = 0 by [2] above

\[
= - \int_{-\infty}^{\infty} \left( \frac{1}{f_i} \frac{\partial f_i}{\partial \theta} \right)^2 f_i \, dx_i = - \int_{-\infty}^{\infty} \left( \frac{\partial \ln f_i}{\partial \theta} \right)^2 f_i \, dx_i
\]

\[
= - E \left[ \frac{\partial \ln f_i}{\partial \theta} \right]^2 = - k^2
\]
or \[-E\left(\frac{\partial^2 \ln f(x_1, \theta)}{\partial \theta^2}\right) = E\left(\frac{\partial \ln f(x_1, \theta)}{\partial \theta}\right)^2 = k^2\]

Thus, by Khintchine's theorem (no. 17)

\[B_0 \xrightarrow{p} 0\]
\[B_1 \xrightarrow{p} -k^2\]
\[B_2 \xrightarrow{p} E[F_3(x_1)] < M\]

Let \(S\) be the set of points where

\[|B_0| < \delta^2, \quad B_1 \leq \frac{1}{2} k^2, \quad |B_2| < 2M.\]

We can find an \(n\), given \(\varepsilon\), \(\delta\), such that \(\Pr[S] > 1 - \varepsilon\).

In \(S\) the right hand side (r.h.s.) of the Taylor series expansion is to be considered.

Consider: \(\theta = \theta_0 + \delta\)

\[r.h.s. = B_0 + B_1 \delta + \frac{1}{2} B_2 \delta^2\]

\[\leq \delta^2(1 + M) - \frac{1}{2} k^2 \delta\]

letting \(z = 1\)

So that, if \(\delta < \frac{k^2}{2(1+M)}\) the r.h.s. \(\leq 0\).

Considering: \(\theta = \theta_0 - \delta\)

\[r.h.s. = B_0 - B_1 \delta + \frac{1}{2} z B_2 \delta^2\]

\[\geq - \delta^2 + \frac{1}{2} k^2 \delta^2 = -(M + 1) \delta^2 + \frac{1}{2} k^2 \delta\]

\[\geq 0\]

for the same \(\delta < \frac{k^2}{2(1+M)}\)

Hence, in \(S\), which occurs with probability \((1 - \varepsilon)\), \(\frac{1}{n} \ln \frac{L}{\theta} = 0\) has a root in the interval \((\theta_0 - \delta, \theta_0 + \delta)\) and \(\ln L\) has a maximum in the interval (at the root).
\( \hat{\theta} \) is then the maximum likelihood estimate and the solution of the equation

\[
\frac{1}{n} \frac{\partial \ln L}{\partial \theta} = B_0 + B_1 (\hat{\theta} - \theta_0) + \frac{1}{2} z B_2 (\hat{\theta} - \theta_0)^2 = 0
\]

which yields

\[
\hat{\theta} - \theta_0 = -\frac{B_0}{B_1 + \frac{z}{2} B_2 (\hat{\theta} - \theta_0)}
\]

Multiplying both sides by \( k \sqrt{n} \)

\[
(\hat{\theta} - \theta_0) k \sqrt{n} = -\frac{B_0 \sqrt{n}}{k} \frac{1}{B_1 + \frac{z}{2} \frac{B_2}{k} (\hat{\theta} - \theta_0)}
\]

We know that:

\[
B_0 = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln f_i}{\partial \theta} \bigg|_{\theta_0}
\]

\[
E[B_0] = 0 \quad \text{Var}[B_0] = \frac{1}{n} \left( \frac{\partial \ln f_i}{\partial \theta} \right)^2 = \frac{k^2}{n}
\]

Therefore

\[
\frac{\sqrt{n} B_0}{k}
\]

is \( \mathcal{N}(0, 1) \)

\[
B_1 \xrightarrow{p} -k^2 \quad B_2 < M \text{ (i.e., it is bounded)}
\]

\[
\hat{\theta} - \theta_0 \xrightarrow{p} 0
\]

Thus, by the relationships we have just stated, and by use of theorem 18

\[
k \sqrt{n} (\hat{\theta} - \theta_0) \text{ is } \mathcal{N}(0, 1)
\]

or:

\[
\frac{1}{k^2}\sqrt{n} (\hat{\theta} - \theta_0) \text{ is } \mathcal{N}(0, \frac{1}{k^2} E \left[ \left( \frac{\partial \ln f_i}{\partial \theta} \right)^2 \right])
\]

Example:

\[
f(x) = a e^{-ax} \quad x \geq 0
\]

Find the m.l.e., of \( a \), and its asymptotic distribution.

\[
L = a^n e^{-a} \sum_{i=1}^{n} x_i
\]
\[ \ln L = n \ln(a) = a \sum_{i=1}^{n} x_i \]

\[ \frac{\partial \ln L}{\partial a} = \frac{n}{a} - \sum_{i=1}^{n} x_i = 0 \]

therefore \[ \hat{a} = \frac{n}{\sum x_i} = \frac{1}{\bar{x}} \]

We can easily verify that \( E(x) = \frac{1}{\hat{a}} \); \( E(\bar{x}) = \frac{1}{\hat{a}} \) or \( \hat{a} = \frac{1}{\bar{x}} \)

by the method of moments.

\[ \frac{\partial^2 \ln L}{\partial a^2} = -\frac{n}{a^2} \quad \frac{n \partial^2 \ln f_i}{\partial a^2} \]

\[ = -\frac{2}{a^2} \ln f_i = \left( \frac{\partial \ln f_i}{\partial a} \right)^2 = \frac{1}{a^2} \]

Variance = \( \frac{1}{E \left[ \left( \frac{\partial \ln f_i}{\partial a} \right)^2 \right]} = \frac{1}{1/a^2} = a^2 \)

Hence: \( \sqrt{n} \left( \frac{1}{\bar{x}} - \hat{a} \right) \) is \( A N(0, a^2) \)

or \( \frac{\sqrt{n} (\frac{1}{\bar{x}} - \hat{a})}{a} \) is \( A N(0, 1) \).

---

Problem 34: \( f(x) = a^2 e^{-a^2 x} \quad x > 0 \)

-- find the m.l.e. of \( a \) and its asymptotic distribution.

-- verify that the same result could be obtained by a Taylor Series expansion.

Problem 35: \( X_i \) is Poisson \( \lambda \) \( (i = 1, 2, \ldots, n) \)

-- find the m.l.e. of \( \lambda \) and its asymptotic distribution.
Problem 36: X is uniform on the interval (0, a).

-- Find the m.l.e. of a [not by differentiating] (\hat{\theta}).

-- Correct it for bias and find the asymptotic distribution of the unbiased estimate (\hat{a}^*).

-- Another unbiased estimate is \( a^* = 2 \bar{X} \).

Compare the actual variances of \( \hat{a}^*; a^* \).

Note: \( a^* \) is a moment estimate -- for another comparison of the method of moments with the method of maximum likelihood, see Cramer p. 505.

Note: The m.l.e. is invariant under single valued functional transformations i.e., the m.l.e. of \( g(\theta) \) is \( g(\hat{\theta}) \).

Remark: If \( d_n(X) \) is \( A N(\mu, \sigma^2/n) \) and \( g[d] \) is continuous with continuous first and second derivatives and the first derivative \( \neq 0 \) at \( x=\mu \)
then \( \sqrt{n} \left[ g(d_n) - g(\mu) \right] \) is \( A N(0, \left[ g'(\mu) \right]^2 \sigma^2) \).

Proof: by use of a Taylor series expansion:
\[
\begin{align*}
\hat{g(d_n(x))} &= g(\mu) + g'(\mu) \left[ d_n - \mu \right] + g''(\mu) \frac{(d_n - \mu)^2}{2}
\end{align*}
\]
From which we can get
\[
\begin{align*}
\sqrt{n} \left[ \frac{g(d_n) - g(\mu)}{\sigma} \right] &= g'(\mu) \sqrt{n} \frac{(d_n - \mu)}{\sigma} + \frac{g''(\mu)}{2 \sigma} \frac{(d_n - \mu)^2}{p} \xrightarrow{p \to 0} 0 \\
\xrightarrow{\text{bounded}} A N(0, 1)
\end{align*}
\]

Hence by use of theorem 18
\[
\begin{align*}
\sqrt{n} \left[ \frac{g(d_n) - g(\mu)}{\sigma g'(\mu)} \right] \text{ is } A N(0, 1).
\end{align*}
\]

---

Multiparameter case:

Theorem 23: Under generalized regularity conditions of theorem 22, if
\[ \Theta' = (\theta_1, \theta_2, \ldots, \theta_n) \]
then for sufficiently large \( n \) and with
probability $1 - \varepsilon$, the m.l.e. of $\theta$ is given as the solution of the equations

$$\frac{\partial \ln L}{\partial \theta_i} = 0 \quad i = 1, 2, \ldots, s$$

and further: $\hat{\theta}_1^0, \hat{\theta}_2^0, \ldots, \hat{\theta}_s^0$ has asymptotically a joint normal distribution with means $\theta_1, \theta_2, \ldots, \theta_s$ and variance-covariance matrix $V^{-1}$

$$V = n \begin{pmatrix}
E \left( \frac{\partial^2 \ln f}{\partial \theta_1 \partial \theta_1} \right) & E \left( \frac{\partial^2 \ln f}{\partial \theta_1 \partial \theta_2} \right) & \cdots & E \left( \frac{\partial^2 \ln f}{\partial \theta_1 \partial \theta_s} \right) \\
E \left( \frac{\partial^2 \ln f}{\partial \theta_2 \partial \theta_1} \right) & E \left( \frac{\partial^2 \ln f}{\partial \theta_2 \partial \theta_2} \right) & \cdots & E \left( \frac{\partial^2 \ln f}{\partial \theta_2 \partial \theta_s} \right) \\
\vdots & \vdots & \ddots & \vdots \\
E \left( \frac{\partial^2 \ln f}{\partial \theta_s \partial \theta_1} \right) & E \left( \frac{\partial^2 \ln f}{\partial \theta_s \partial \theta_2} \right) & \cdots & E \left( \frac{\partial^2 \ln f}{\partial \theta_s \partial \theta_s} \right)
\end{pmatrix}$$

This is the so-called information matrix used in multiparameter estimation.

Sketch of part of the proof:

$$0 = \frac{1}{n} \frac{\partial \ln L}{\partial \theta_i} = \frac{1}{n} \frac{\partial \ln L}{\partial \theta_i} \bigg|_{\theta^0} + \frac{1}{n} \sum_{j=1}^{s} (\theta_j - \theta_j^0) \frac{\partial^2 \ln L}{\partial \theta_j \partial \theta_i} \bigg|_{\theta^0}$$

+ second and higher degree terms $i = 1, 2, \ldots, s$

(Note: $\ln L$ terms can be replaced by $\sum \ln f_i = n \ln f_i$ terms.)

From theorem 22:

$$\ln L = \sum_{1}^{n} \ln f(x_i, \theta)$$

$$E \left( \frac{\partial \ln L}{\partial \theta_j} \right) = E \left( \sum \left[ \frac{\partial \ln f_i}{\partial \theta_j} \right] \right) = 0$$

$$- E \left( \frac{\partial^2 \ln f_i}{\partial \theta_j \partial \theta_j} \right) = E \left( \frac{\partial \ln f_i}{\partial \theta_j} \right)^2$$

We also need a set of covariance terms:

$$- E \left( \frac{\partial^2 \ln f_i}{\partial \theta_j \partial \theta_k} \right) = E \left[ \left( \frac{\partial \ln f_i}{\partial \theta_j} \right) \left( \frac{\partial \ln f_i}{\partial \theta_k} \right) \right]$$

which follows in a manner similar to the derivation of the variance expression in theorem 22.
Now we can write:

\[ \beta_{jo} = \frac{1}{n} \sum_{i=1}^{n} \frac{\ln f(x_i, \theta)}{\theta_{jo}} \quad \text{as} \quad p \to 0 \]

\[ b_{jk} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \ln f(x_i, \theta)}{\theta_{j} \theta_{k}} \quad \text{as} \quad p \to 0 \]

The maximum likelihood equations can be written (ignoring the second degree terms in the expansions):

\[ -\beta_0 = B (\hat{\theta} - \theta^0) \quad \text{in matrix form or completely written out as:} \]

\[ -\beta_{1o} = (\hat{\theta}_1 - \theta^0) b_{11} + \cdots + (\hat{\theta}_s - \theta^0) b_{1s} \]

\[ \vdots \]

\[ -\beta_{so} = (\hat{\theta}_1 - \theta^0) b_{s1} + \cdots + (\hat{\theta}_s - \theta^0) b_{ss} \]

\[ B \xrightarrow[p \to \infty]{\text{in}} B[E(B)]^{-1} \quad \text{(i.e., each element } b_{jk} \text{ replaced by } E[b_{jk}] \text{)} \]

For large \( n \): \( B[E(B)]^{-1} \xrightarrow[p \to \infty]{\text{in}} \)

\[ -B_0 = B[E(B)]^{-1} [E(B)] (\hat{\theta} - \theta^0) = [E(B)][(\hat{\theta} - \theta^0) \]

or:

\[ \hat{\theta} - \theta^0 = [-E(B)]^{-1} B_0 \]

\[ [E(B)] = V \quad \text{and is non-singular.} \]

Variance-covariance matrix of \( \hat{\theta} - \theta^0 = E(B)^{-1} \left[ \begin{array}{c} \text{var-cov} \ B_0 \end{array} \right] \left[ E(B)^{-1} \right] \)

\[ V^{-1} \]

\[ V^{-1} \]

**Example** of the information matrix used in the multiparameter estimation:

\[ f(x; a, b) = \frac{ab}{\Gamma(b)} e^{-ax} x^{b-1} \quad x > 0 \quad a > 0 \quad b > 0 \]

\[ \ln L = n b \ln a - n \ln \Gamma(b) - a \sum_{i=1}^{n} x_i - (b - 1) \bar{x}_L \]

\[ n \]

\[ \text{defining } \bar{x}_L = \frac{\sum \ln x_i}{n} \]
\[
\frac{1}{n} \frac{\partial \ln L}{\partial a} = \frac{b}{a} - \bar{x} = 0
\]

\[
\frac{1}{n} \frac{\partial \ln L}{\partial b} = \ln a - \frac{\Gamma'(b)}{\Gamma(b)} + \bar{x}_L = 0
\]

From the first equation \( \hat{a} = \frac{b}{\bar{x}} \)

Substituting this in the second equation we get:

\[
\ln \frac{b}{a} = F_2(b) + \bar{x}_L = 0
\]

defining \( F_2(b) = \frac{\Gamma'(b)}{\Gamma(b)} = \frac{d}{db} \left( \ln \Gamma(b) \right) \)

\[
\ln b = F_2(b) = \ln \frac{b}{a} - \bar{x}_L
\]

Note: Pearson has compiled tables for \( F_2(b) \) which he called the di-gamma function

Thus we have:

\[
\frac{\partial^2 \ln L}{\partial a^2} = n \left( -\frac{b}{a^2} \right)
\]

\[
\frac{\partial^2 \ln L}{\partial a \partial b} = n \left( \frac{1}{a} \right)
\]

\[
\frac{\partial^2 \ln L}{\partial b^2} = n \left[ \frac{d^2}{db^2} - \ln \Gamma(b) \right] = n \left[ -F_2'(b) \right]
\]

\[
V = n \begin{bmatrix}
\frac{b}{a^2} & \frac{1}{a} \\
\frac{1}{a} & F_2'(b)
\end{bmatrix}
\]

\[
|V| = n \left[ \frac{b}{a^2} F_2'(b) - \frac{1}{a^2} \right]
\]

note: \( F_2'(b) \) is called the tri-gamma

The asymptotic variances or covariances are thus:

\[
\text{of } \hat{a} = \frac{F_2'(b)}{|V|} = \frac{F_2'(b) a^2}{n \left[ bF_2'(b) - 1 \right]}
\]

\[
\text{of } \hat{a}, \hat{b} = \frac{1/\bar{x}}{|V|} = \frac{a}{n \left[ bF_2'(b) - 1 \right]}
\]
of \( \hat{b} = \frac{b/a^2}{|V|} = \frac{b}{n \left[bF_2(b) - 1 \right]} \)

\( \sqrt{n} \left( \hat{b} - b \right); \sqrt{n} \left( \hat{a} - a \right) \) thus have a joint normal asymptotic distribution with means 0 and variance-covariance matrix:

\[
\left( \frac{V}{n} \right)^{-1} = \begin{bmatrix}
\frac{a^2 F_2'(b)}{bF_2(b) - 1} & a \\
\frac{a}{bF_2(b) - 1} & \frac{b}{bF_2(b) - 1}
\end{bmatrix}
\]

Exercise: \( f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \)

- Find the m.l.e. of \( \mu \) and \( \sigma^2 \) and find the information matrix.

- The m.l.e. of \( \hat{\mu} \) and \( \hat{\sigma}^2 \) are \( \bar{x}, \frac{\sum (x_i - \bar{x})^2}{n} \)

- \( \hat{\mu}, \hat{\sigma}^2 \) are independent, therefore the covariance terms in the information matrix are 0.

Problem 37: \( f(x) = \frac{1}{2} e^{-|x-\mu|} \quad -\infty < x < \infty \)

Note: This is the so-called Laplace distribution. It is an example of finite theory -- the m.l.e. is not a minimum-variance estimate for any finite sample size.

a) Find the m.l.e. of \( \mu \) (based on \( n \) independent observations).

Does it satisfy the conditions of theorem 22? why not?

b) Is \( \bar{x} \) an unbiased estimate of \( \mu \)? find \( \sigma^2 \).

Problem 38: \( X \) is Poisson \( \lambda \).

\( Y \) is Poisson \( \lambda \mu \).

Find the m.l.e. of \( \lambda, \mu \) and also their information matrix.
Check that the same results are obtained by expanding $\mu$ in a Taylor series about $\mu$, hence this result is asymptotic as $\lambda \to \infty$.

$\mu \lambda \to \infty$.

Problem 390: $\frac{X_i}{a}$ is $\chi^2$; $\frac{Y_j}{a + b}$ is $\chi^2$; $X, Y$ are independent

$i = 1, 2, \ldots, m$ \quad $j = 1, 2, \ldots, n$

$a > 0$ \quad $b > 0$

Find the $m_{\lambda\varepsilon\varepsilon}$ of $a$, $b$ and also the asymptotic variance-covariance matrix.
D. UNBIASED ESTIMATION

Theorem 24: Information Theorem (Cramer-Rao)

If \( d(X) \) is a regular estimate of \( \theta \) and \( E \left[ d(X) \right] = \theta + b(\theta) \) (where \( b(\theta) \) is a possible bias factor) then

\[
\sigma_d^2 \geq \frac{\left( 1 + b'(\theta) \right)^2}{n \ k^2}
\]

for all \( n \)

where

\[
k^2 = E \left[ \frac{\partial \ln f(x)}{\partial \theta} \right]^2 = -E \left[ \frac{\partial^2 \ln f(x)}{\partial \theta^2} \right]
\]

the equality holds if and only if \( d(X) \) is a linear function of \( \frac{\partial \ln f}{\partial \theta} \).

Regularity conditions for the Information Theorem:

1. \( \theta \) is a scalar in open interval.
2. \( \frac{\partial f}{\partial \theta} \) exists for almost all \( x \) (the set where \( \frac{\partial f}{\partial \theta} \) does not exist must not depend on \( \theta \)).

(Note: Problem 36 where \( f(x) = \frac{1}{a} \) (i.e. uniform on \( (0, a) \) ) would be an exception to this condition).
3. \( \int f(x, \theta) \, dx \) can be differentiated under the integral sign with respect to \( \theta \).
4. \( \int d(x) \ f(x, \theta) \, dx \) can be differentiated under the integral sign with respect to \( \theta \).
5. \( 0 < k^2 < \infty \)

Proof: From the proof of theorem 22 we remember that

\[
E\left( \frac{\partial \ln f}{\partial \theta} \right) = 0 \quad -E\left( \frac{\partial^2 \ln f}{\partial \theta^2} \right) = E\left( \frac{\partial \ln f}{\partial \theta} \right)^2 = k^2
\]

Differentiating both sides of this equation, we get

\[
\int d(x) \frac{1}{f} \frac{\partial f}{\partial \theta} f(x) \, dx = 1 + b'(\theta)
\]
which can also be written 
\[ E \left[ d(x) \left. \frac{\partial \ln f}{\partial \theta} \right| \right] = 1 + b^i(\theta). \]

The correlation coefficient of \( S(x) = \frac{\partial \ln f(x)}{\partial \theta} \) and \( d(x) \leq 1 \)

\[
\left( \frac{E \left[ S(x) d(x) \right]}{\sigma_d(x) \cdot \sigma_S(x)} \right)^2 \leq 1
\]

Note: \( E \left[ S(x) \right] \cdot E \left[ d(x) \right] \) term vanishes since \( E \left[ S(x) \right] = 0 \)

or

\[
\sigma_d^2 \geq \frac{\left[ 1 + b^i(\theta) \right]^2}{\sigma_S^2(x)}
\]

now

\[
\sigma_S^2(x) = \sum_1^n E \left( \frac{\partial \ln f(x)}{\partial \theta} \right)^2 = n k^2
\]

therefore

\[
\sigma_d^2(x) \geq \frac{\left[ 1 + b^i(\theta) \right]^2}{n k^2} = \frac{\left[ 1 + b^i(\theta) \right]^2}{n E \left[ \frac{\partial \ln f(x)}{\partial \theta} \right]^2}
\]

Since \( r^2 = 1 \) if and only if the random variables are linearly related it follows that the equality in this result holds if and only if \( d(x) \) is a linear function of \( S(x) \).

Example: \( X_1, X_2, \ldots, X_n \) are NID(\( \mu, \sigma^2 \)), \( \sigma^2 \) known.

Find the C-R lower bound for the variance of unbiased estimates of \( \mu \).

\[
L = \frac{1}{(2 \pi \sigma)^{n/2}} e^{-\sum_1^n (x_i - \mu)^2 / 2 \sigma^2}
\]

\[
\ln L = K - \frac{\sum_1^n (x_i - \mu)^2}{2 \sigma^2}
\]

\[
\frac{\partial \ln L}{\partial \mu} = \frac{\sum_1^n (x_i - \mu)}{\sigma^2}
\]
\[- \frac{\partial^2 \ln L}{\partial \mu^2} = \frac{n}{\sigma^2} \]

\[k^2 = \frac{1}{\sigma^2} \]

Hence any regular unbiased estimate \( d(\bar{x}) \) of \( \mu \) must satisfy \( \sigma_d^2 \geq \frac{\sigma^2}{n} \)

but \( \sigma^2 \bar{x} = \frac{\sigma^2}{n} \) hence \( \bar{x} \) is a minimum variance unbiased estimate (m.v.u.e.).

**Example (Discrete): Binomial**

\( x \) is \( B(n, p) \) -- find the C-R lower bound for the unbiased estimates of \( p \).

\[ L = \binom{n}{x} p^x (1 - p)^{n-x} \]

\[ \ln L = x \ln p + (n-x) \ln (1 - p) \]

\[ \frac{\partial \ln L}{\partial p} = \frac{x}{p} - \frac{n-x}{1-p} = 0 \]

\[ \frac{\partial^2 \ln L}{\partial p^2} = \frac{x}{p^2} + \frac{n-x}{(1-p)^2} \]

\[ -E \left[ \frac{\partial^2 \ln L}{\partial p^2} \right] = \frac{np}{p^2} + \frac{n(1-p)}{(1-p)^2} = \frac{n}{p(1-p)} = n k^2 \]

Hence \( \sigma_d^2 \geq \frac{p(1-p)}{n} \) but \( \sigma^2 \frac{\bar{x}}{n} \) so that \( \frac{x}{n} \) is m.v.u.e.

**Note:** Recall that the C-R bound is achieved if and only if \( d(\bar{x}) \) is a linear function of \( \frac{\partial \ln f}{\partial \theta} \) -- if \( \bar{x} \) is not a linear function of \( \frac{\partial \ln f}{\partial \theta} \) then \( \bar{x} \) will not achieve the C-R lower bound.

**Problem 1.5:** Find the C-R lower bound for the unbiased estimates of the Poisson parameter \( \lambda \) based on independent observations \( X_1, X_2, \ldots, X_n \).

**Example:** Negative Binomial (Ref. Lehman Ch. 2, p. 2-21, 22)

\[ L = Pr(x) = pq^x \]

i.e., sample until a single success occurs

\[ \ln L = \ln p + x \ln (1 - p) \]

\[ \frac{\partial \ln L}{\partial p} = \frac{1}{p} - \frac{x}{1-p} \]
\[
- \frac{\frac{\partial^2 \ln L}{\partial p^2}}{p^2} = \frac{1}{p^2} + \frac{x}{(1-p)^2} \quad E(x) = \frac{1-p}{p}
\]

\[
- E \left[ \frac{\frac{\partial^2 \ln L}{\partial p^2}}{p^2} \right] = \frac{1}{p^2} + \frac{1}{p(1-p)} = \frac{1}{p^2 (1-p)} = n k^2
\]

thus \( \sigma_d^2 \geq p^2 (1 - p) \)

To find the M.L.E. of \( p \):

\[
\frac{\partial \ln L}{\partial p} = \frac{1}{p} - \frac{x}{1-p} \quad \text{or} \quad \hat{p} = \frac{1}{1+x}
\]

--- it can be shown that this estimate of \( p \) is biased.

Can we find an unbiased estimate? Yes --- by solving the equation \( E \left[ d(X) \right] = p \).

To do this we observe \( X: 0 \quad 1 \quad 2 \quad 3 \quad \ldots \quad n \quad \ldots \ldots \)

\[
\begin{align*}
\text{Prob: } & p \quad pq \quad pq^2 \quad pq^3 \quad \ldots \quad pq^n \quad \ldots \ldots \\
E \left[ d(X) \right] &= d(0)p + d(1)pq + d(2)pq^2 + \ldots + d(n)pq^n + \ldots = p \\
\text{or} \quad d_0 + d_1q + d_2q^2 + \ldots + d_nq^n + \ldots &= 1
\end{align*}
\]

If this power series is to be an identity in \( q \) we can equate the coefficients on the left and on the right

\[
d_0 = 1 \quad d_1 = 0 \quad d_2 = 0 \quad \ldots \ldots
\]

The unbiased estimate is thus \( \hat{p} = \frac{1}{1+x} = 1 \) if \( x = 0 \)

\[
= 0 \quad \text{if} \quad x \geq 1
\]

i.e., the decision as to the status of the whole lot is based on the first observation.

This is unbiased, but \( \ldots \ldots \)

(this example is used to show why we don't always want an unbiased estimate)

\[
\sigma^2 = pq \geq p^2 \quad \text{(the C-R lower bound)} \quad \text{--- thus the C-R lower bound can't be met}
\]

\( \hat{p} \)

... since this estimate is the only unbiased estimate by the uniqueness of power series.
To find an unbiased estimator we can try to solve the equation
\[ \int_{-\infty}^{\infty} d(x) f(x, \theta) \, dx = \theta \] for the continuous case.

or
\[ \sum d(n) p_n(\theta) = \theta \] for the discrete case.

These equations, however, are in general rather messy and difficult to handle.

**Problem 1.** \( f(x) = k \, x^{k-1} \quad 0 \leq x \leq 1 \quad k > 0 \)

1. Find the C-R lower bound.

2. Is it attained by the M.L.E.?

**Multiparameter extension of the Information (C-R) Theorem**

We have \( f(x; \theta_1, \theta_2, \ldots, \theta_k) \).

Assume the same regularity conditions as for Theorem 24 in all the \( \theta \)'s.

Denote:
\[ S_i = \frac{\partial \ln L}{\partial \theta_i} \quad \text{and} \quad E \left[ S_i S_j \right] = \lambda_{ij} \]

\[ \Lambda = \begin{pmatrix} \lambda_{11} & \cdots & \lambda_{1k} \\ \vdots & \ddots & \vdots \\ \lambda_{k1} & \cdots & \lambda_{kk} \end{pmatrix} \]

\( \Lambda \) is non-singular.

\[ E \left[ S_i \right] = 0 \quad \text{as in theorems 22, 24} \]

Let \( d(X) \) be an unbiased estimate of \( \theta_1 \)

so that \( \int d(x) f(x; \theta_1, \theta_2, \ldots, \theta_k) \, dx = \theta_1 \).

By differentiation with respect to \( \theta_1 \)
\[ E \left[ d \, S_1 \right] = 1. \]

By differentiation with respect to \( \theta_j \)
\[ E \left[ d \, S_j \right] = 0. \]
Theorem 25: Under the regularity conditions stated
\[ \sigma_d^2 \geq I_1' \Lambda^{1} I_1 \]
where \( I_1' = (1, 0, \ldots, 0) \) \( k \) components
or
\[ \sigma_d^2 \geq \begin{vmatrix} \lambda_{22} & \cdots & \lambda_{2k} \\ \vdots & \ddots & \vdots \\ \lambda_{k2} & \cdots & \lambda_{kk} \end{vmatrix} \]

Proof: Observe that \( E\left[ (d_1 - \theta_1 - \sum_{i=1}^{k} a_i S_i)^2 \right] > 0 \) where the \( a_i \)'s are arbitrary constants to be determined.

\[ E\left[ (d_1 - \theta_1)^2 \right] = 2 E\left[ (d_1 - \theta_1) \sum a_i S_i \right] + E\left[ \sum a_i S_i \right]^2 > 0 \]

From the above relationships this becomes
\[ \sigma_d^2 = 2a_1 + \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j E[S_i S_j] > 0 \]

which then becomes
\[ \sigma_d^2 \geq 2a_1 + \sum_{i=1}^{k} \sum_{j=1}^{k} a_i a_j \lambda_{ij} \quad \text{for all real } a_1, a_2, \ldots, a_k. \]

Call the right hand side of this inequality \( \varphi(a) \)

We wish to maximise \( \varphi(a) \) with respect to \( a \) to get the best possible statement about the bound of the variance.

\[ \varphi(a) = 2a_1 - \sum_{i=1}^{k} a_i^2 \lambda_{ii} - \sum_{i \neq j} a_i a_j \lambda_{ij} \]
\[ \frac{\partial \phi}{\partial a_1} = 2 - 2a_1 \lambda_{11} - 2 \sum_{j=2}^{k} a_j \lambda_{ij} = 0 \]

or
\[ \sum_{j=1}^{k} a_j \lambda_{ij} = 1 \]

\[ \frac{\partial \phi}{\partial a_i} = -2a_1 \lambda_{ii} - 2 \sum_{j=1, j \neq i}^{k} a_j \lambda_{ij} = 0 \]

or
\[ \sum_{j=1}^{k} a_j \lambda_{ij} = 0 \quad i = 2, 3, \ldots, k \]

Hence the equations in \( a \) to maximize \( \phi(a) \) are
\[ \sum_{j=1}^{k} a_j^0 \lambda_{ij} = 1 \]
\[ \sum_{j=1}^{k} a_j^0 \lambda_{ij} = 0 \quad i = 2, 3, \ldots, k \]

where \( a_j^0 \) is a maximizing value of \( a_j \).

How, multiply the \( i \)th equation by \( a_i^0 \) and add all the equations.

The left hand side of the sum so obtained is
\[ \sum_{i=1}^{k} a_i^0 \sum_{j=1}^{k} a_j^0 \lambda_{ij} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_i^0 a_j^0 \lambda_{ij} \]

The right hand side is \( a_1^0 \).

Hence
\[ \phi_{\text{max}}(a) = 2a_1^0 - a_1^0 = a_1^0 \]

Therefore
\[ c_{d1}^2 \geq a_1^0 \]

where:
\[ \Lambda \otimes I = I_1 = \begin{pmatrix} 1 \\ \vdots \\ 0 \end{pmatrix} \]
\[ a_i^0 = \Lambda^{-1} I_1 \]
\[ a_1^0 = I_1' \Lambda^{-1} I_1 \]
or, as stated in the theorem:

\[ \sigma_d^2 \geq I_1 \wedge I_1^{-1} = \begin{vmatrix}
\lambda_{22} & \cdots & \lambda_{2k} \\
\vdots & \ddots & \vdots \\
\lambda_{k2} & \cdots & \lambda_{kk}
\end{vmatrix}
\]

Remember: \( \lambda_{ij} = E[\{ S_i S_j \}] \quad S_i = \frac{\partial \ln L}{\partial a} \)

**Example:** \( f(x) = \frac{ab}{\Gamma(b)} x^{b-1} e^{-ax} \quad x > 0 \)

--- the generalized Gamma or Pearson's Type III

--- \( a, b \) are the unknown parameters

Find the lower bound for the variance of the unbiased estimate of \( a \).

\[ \ln L = n \ln a - n \ln \Gamma(b) - ax + n(b-1) \ln x \]

\[ S_1 = \frac{\partial \ln L}{\partial a} = \frac{nb}{a} - \bar{x} \]

\[ S_2 = \frac{\partial \ln L}{\partial b} = n \ln a - n \frac{\partial}{\partial b} \left[ \ln \Gamma(b) \right] + n \ln x \]

denote \( \frac{\partial}{\partial b} \left[ \ln \Gamma(b) \right] = F_2(b) \)

\[ \lambda_{11} = E[S_1^2] = - \frac{\partial^2 \ln L}{\partial a^2} = \frac{nb}{a^2} \]

\[ \lambda_{22} = E[S_2^2] = - \frac{\partial^2 \ln L}{\partial b^2} = n F_2'(b) \]

\[ \lambda_{12} = \lambda_{21} = E[S_1 S_2] = - \frac{\partial^2 \ln L}{\partial a \partial b} = - \frac{n}{a} \]
\[ \Lambda = n \left( \begin{array}{cc} \frac{b}{a^2} & -\frac{1}{a} \\ -\frac{1}{a} & F_2^i(b) \end{array} \right) \]

\[ \sigma^2_a \geq \frac{F_2^i(b)}{n \left( \frac{b F_2^i(b)}{a^2} - \frac{1}{a} \right)} = \frac{a^2 F_2^i(b)}{n \left[ b F_2^i(b) - 1 \right]} = \frac{a^2}{n \left[ b - \frac{1}{F_2^i(b)} \right]} \]

Note: See the example for Theorem 23.

Note: We started the proof with \( E \left[ d_1 - e_1 - \sum_{i=1}^{k} a_i S_i \right]^2 \geq 0 \).

The strict inequality will hold unless \( \left[ d_1 - e_1 - \sum_{i=1}^{k} a_i S_i \right] \) is essentially constant.

Therefore: If the multiparameter C-R lower bound is attained, \( d_1 \) is a linear function of \( \sum_{i=1}^{k} a_i S_i \)

and, as before, the M.L.E. (corrected for bias if necessary) attains the lower bound (if there is any estimate that attains that lower bound).

**Problem 42:** \( X_1, \ldots, X_n \) are negative binomial

\[ \Pr \left[ X = x \right] = \binom{r + x - 1}{r - 1} p^r q^x \quad p + q = 1 \quad x = 0, 1, 2, \ldots \]

Find the C-R lower bounds for the unbiased estimates of \( p \)

1. If \( r \) is known
2. If \( r \) is unknown
Problem 42 (b): Show that if $X \mid \lambda$ is Poisson $(\lambda)$ and $\lambda$ is distributed with density

$$f(\lambda) = \frac{\lambda^{b-1} e^{-\lambda}}{(b-1)!}$$

then $X$ is negative binomial. Identify functions of $a$, $b$ with $p$, $r$.

---

Summary on usages re m.l.e.:

We have the following criteria for estimates:

1. Efficiency (Cramer) -- attains the C-R lower bound.
2. (Asymptotic) Efficiency (Fisher) -- $(n)$ (variance) $\rightarrow \frac{1}{k^2}$ in the limit.
4. Best asymptotic normal (b.a.n.) -- among A.N. estimates, no other has a smaller asymptotic variance.

Properties of m.l.e.

1. If there is an efficient (Cramer) estimate (i.e., if the estimate is a linear function of $\frac{\partial \ln L}{\partial \theta}$) then the m.l.e. is efficient (Cramer).
2. If the m.l.e. is efficient (Cramer) and unbiased it is m.v.u.e. Other m.l.e. may or may not be m.v.u.e.
3. Under the general regularity conditions, the m.l.e. is asymptotically efficient (Fisher).
4. Among the class of unbiased estimates, then the m.l.e. are b.a.n., otherwise (i.e., if the m.l.e. are biased) we can not say that the m.l.e. are necessary b.a.n.

ref: Le Cam; Univ. of Calif. Publications in Statistics; Vol. I, No. 11, 1953

-- The class of efficient (Cramer) estimates is contained in the class of m.v.u.e. -- which is the real reason we are interested in attaining the C-R lower bound.

-- 1, 2 are finite results, i.e. for any $n$.
-- 3, 4 are finite (asymptotic) results -- for large $n$ only.
Let $X$ be a random variable with density $f(x, \theta)$, $\theta \in \Theta$.

$T(X)$ is a statistic, i.e., a function of the random variable $(X)$.

Assume that the conditional d.f. of $X$, given $T = t$, exists.

**Def. 31:** $T(X)$ is a sufficient statistic for $\theta$ if the distribution of $X$ given $T = t$ is (mathematically) independent of $\theta$, that is, if

$$f(x, \theta) = g(T, \theta) \cdot h(x | T)$$

$h$ is independent of $\theta$.

**Examples:**

**No. 1** suppose $X_1, X_2, \ldots, X_n$ are $N(\mu, 1)$

$$f(x, \mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum (x_i - \mu)^2}{2}}$$

$$= \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2} \sum x_i^2} e^{\frac{n\mu^2}{2}}$$

$$T(x) = \sum_{1}^{n} x_i = n \bar{x}$$

$$f(x, \mu) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{\sum (x_i - \bar{x})^2}{2} - \frac{n(\bar{x} - \mu)^2}{2}}$$

$\bar{x}$ is thus sufficient for $\mu$.

**No. 2** Poisson:

$$f(x_1, x_2, \ldots, x_n) = e^{-n\lambda} \frac{\lambda^{\sum x_i}}{\prod x_i !}$$

$$f(x) = e^{-n\lambda} \left( \frac{\lambda^{\sum x_i}}{\prod x_i !} \right) \left( \frac{1}{n} \sum x_i \right) \left( \frac{1}{n} x_1 \right) \ldots \left( \frac{1}{n} x_n \right)$$

$$g(T, \lambda) \quad h(x | T)$$
\[ \sum x_i \] is Poisson \((n\lambda)\); given \(\sum x_i\), the individual observations are multinomial and \(T = \sum x_i = n \bar{x}\) is sufficient for \(\lambda\).

**No. 3 -- Normal**

\[
f(x) = \frac{1}{(\sqrt{2\pi \sigma})^n} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}
\]

\[
= \frac{1}{(\sqrt{2\pi \sigma})^n} e^{-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}} e^{-\frac{n(\bar{x} - \mu)^2}{2\sigma^2}}
\]

Here sufficient statistics for \(\mu, \sigma^2\) are \(\bar{x}, \sum_{1}^{n} (x_i - \bar{x})^2\).

**Problem 43:** Find sufficient statistics (if any) for the parameters of the following distributions:

- **a-- Gamma** \(f(x) = \frac{a^b}{\Gamma(b)} e^{-ax} x^{b-1} \quad x > 0\)
- **b-- power** \(f(x) = k x^{k-1} \quad 0 \leq x \leq 1\)
- **c-- Beta** \(f(x) = \frac{1}{\beta(a, b)} x^{a-1} (1 - x)^{b-1} \quad 0 \leq x \leq 1\)
- **d-- Cauchy** \(f(x) = \frac{1}{\pi} \left( \frac{1}{1 + (x - \mu)^2} \right)\)
- **e-- neg. binomial** \(p(x) = \binom{x + r - 1}{r - 1} p^r q^x \quad x = 0, 1, 2, \ldots\)
- **f-- normal** mean: \(\mu + \beta t_i\) \(t_i\) known variance: \(\sigma^2\)

**Lemma:** for any real number \(c\)

\[
E\left[ x - c \right]^2 \geq E_T \left[ E_x(x_i|T) - c \right]^2
\]

- **Remark:**
Proof: \( E[X^2 | t] \geq \left( E[X | t] \right)^2 \) where \( t \) is any particular value of the statistic \( T \).

\[
E[X] = E_T[E_x(X|T)]
\]

\[
E[X^2] = E_T[E_x(X^2|T)] \geq E_T[E(X|T)]^2
\]

replacing \( X \) by \((X - c)\) we get

\[
E[(X - c)^2] \geq E_T[E_x(X - c|T)]^2
\]

\[
\geq E_T[E_x(X|T) - c]^2
\]

Remark: the equality holds only if \( X \) is a function of \( T \) since the equality holds at step 1 if \( X|t \) is a constant.

Theorem 26: Let \( X \) have density \( f(x, \theta) \).

Let \( T \) be a sufficient statistic for \( \theta \).

If \( d(X) \) is any unbiased estimate of \( g(\theta) \), then \( \Psi(T) = E_x[d(X)|T] \) is an unbiased estimate of \( g(\theta) \) and \( \sigma^2_{\Psi} \leq \sigma^2_d \) with the equality holding only if \( \Psi(T) = d(x) \).

Proof: Since \( T \) is sufficient then the distribution of \( d(X)|T \) is independent of \( \theta \) and hence

\[
\Psi(T) = E_x[d(X)|T] \quad \text{is independent of} \quad \theta.
\]

1. \( \quad E\left[\Psi(T)\right] = E_T\left[E_x[d(X)|T]\right] = E[d(x)] = g(\theta) \)

2. In the result of the lemma put \( d(X) = X \quad g(\theta) = c \) then it follows immediately that

\[
\sigma^2_d = E[d(X) - g(\theta)]^2 \geq E_T[\Psi(T) - g(\theta)]^2
\]

\[
\geq \sigma^2_{\Psi}
\]
Examples:

1. Binomial $X_i$ is $B(1, p)$ for $i = 1, 2, \ldots, n$

$$f(x_1, x_2, \ldots, x_n) = p^x(1-p)^{n-x}$$

where $X = \sum_{i=1}^{n} x_i$

(since $f(x)$ is ordered, the coefficient term is omitted)

$X$ is a sufficient statistic for $p$.

$E(X_1) = p$ so that $X_1$ is an unbiased estimate of $p$.

$\sigma^2_{X_1} = p(1-p)$

Define $\psi(X) = E[X_1 | X]$

-- by the theorem $\psi(X)$ will be unbiased and have smaller variance.

$X_1 = 1$ if success on trial 1

$= 0$ if failure on trial 1

in $n$ trials we have $X$ successes

$\Pr[X_1 = \text{success} | X] = X/n$

$\Pr[X_1 = 0 | X] = 1 - \frac{X}{n}$

$\Pr[X_1 = 1 | X] = \frac{X}{n}$

therefore

$$\psi(X) = E[X_1 | X] = \frac{X}{n}$$

2. Problem \#2: Negative binomial; $r$ known

show that

1. $X$ is sufficient for $p$.

2. $(1 - X_1)$ is an unbiased estimate of $p$.

note: $X_i = 1$ if failure on trial $i$

$= 0$ otherwise

find $E[X_1 | X]$.

Example No. 3: Uniform distribution on $(0, \Theta)$

$X = \sum X_i$

$Z = \max(X_1, X_2, \ldots, X_n)$ is sufficient for $\Theta$.

$E[X_1] = \frac{\Theta}{2}$ so that $2X_1$ is an unbiased estimate of $\Theta$.

$\psi(Z) = E[2X_1 | Z]$ will be unbiased and have smaller variance.
proof: If \( X_i \) is one of the observations less than \( Z \) then \( X_i \) is uniform on \((0, Z)\),
then \( E[2X_1|X_1 < Z] = Z \cdot \left( \frac{Z}{Z} \right) = Z \).

If \( X_i \) is the maximum (i.e., = \( Z \)) then \( E[2X_1|Z] = 2Z \).

Thus:
\[
E[2X_1|Z] = \frac{n-1}{n} Z + \frac{1}{n} 2Z = (1 - \frac{1}{n})Z + \frac{2}{n} Z
\]
\[
= (1 + \frac{1}{n}) Z = (\frac{n+1}{n}) Z
\]

b- Given that \( X \) is a Poisson random variable,
estimate \( e^{-\lambda} \) (i.e., estimate the probability of getting a zero observation).

Given \( X_1, X_2, \ldots, X_n \),
\[ p(X_1, \ldots, X_n) = e^{-\lambda n} \frac{(\lambda)^{\sum X_i}}{\prod X_i} \]

we know that \( T = \sum X_i \) is sufficient for \( \lambda \) and
\[ p(T) = e^{-n\lambda} \frac{(n\lambda)^T}{T!} \]
(see example No. 2 on p. 86)

(also \( \bar{x} \) is m.l.e. for \( \lambda \) and \( e^{-\bar{x}} \) is m.l.e. of \( e^{-\lambda} \).

What is an unbiased estimate of \( e^{-\lambda} \)?

Let \( n_0 = \) the number of zeros among \( X_1, X_2, \ldots, X_n \).
\[
E\left[ \frac{n_0}{n} \right] = e^{-\lambda} = Pr[X = 0]
\]
so that \( \frac{n_0}{n} \) is an unbiased estimate of \( e^{-\lambda} \).

Define the estimate of \( e^{-\lambda} \)
\[
\psi'(T) = E\left[ \frac{n_0}{n} | T \right]
\]
Let:
\[
Y_1 = 1 \text{ if } X_1 = 0 \\
\quad \quad = 0 \text{ if } X_1 > 0
\]
then \( n_0 = \sum_{i=1}^{n} Y_i \)
\[ E[Y_1 | T] = Pr[Y_1 = 1 | T] = Pr[X_1 = 0 | T] = (1 - \frac{1}{n})^T \]

\[ E \left[ \frac{Y_0}{n} | T \right] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i | T] = \frac{1}{n} \left[ n \left( 1 - \frac{1}{n} \right)^T \right] \]

To show that \( \Psi(T) = (1 - \frac{1}{n})^T \) is unbiased:

\[
E \left[ \Psi(T) \right] = \sum_{T=0}^{\infty} \left( 1 - \frac{1}{n} \right)^T e^{-n\lambda} \frac{(n\lambda)^T}{T!} = \sum e^{n\lambda} \frac{\left( 1 - \frac{1}{n} \right)(n\lambda)^T}{T!} = e^{n\lambda} e^{n\lambda - \lambda} = e^{-\lambda}
\]

Problem 4.5: Find the variance of \( \Psi(T) = (1 - \frac{1}{n})^T \)

Compare it with the C.R. lower bound for unbiased estimates of \( e^{-\lambda} \).

Sufficient statistics are not unique:

Example 1: Poisson \( T = \sum X_i \) is sufficient, but \( \frac{T}{n} = \bar{X} \) is equally good.

Example 2: \( X_1, X_2, \ldots, X_n \) are \( N(\mu, 1) \)

\[ f(x) = \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{\sum(x_i - \mu)^2}{2}} = \left( \frac{1}{\sqrt{2\pi}} \right)^n e^{-\frac{\sum x_i^2 + 2\mu \sum x_i - n \mu^2}{2}} \]

\[ \sum x_i^2, \bar{X} \] are sufficient for \( \mu \), however \( \sum x_i^2 \) is not necessary and gives no more information about \( \mu \) than does \( \bar{X} \)

Example 3: \( f(x; \mu, \sigma^2) = \left( \frac{1}{\sqrt{2\pi} \sigma} \right)^n e^{-\frac{\sum(x_i - \mu)^2}{2\sigma^2}} \)
\[ T_1 = \bar{x} \quad T_2 = \sum (x_i - \bar{x})^2 \]

or also \( T_1 = \bar{x} \quad T_2 = \sum x_i^2 \) — this set gives the same information (need to combine them to estimate \( \sigma^2 \))

Example 4: Same as No. 3, but \( \mu \) is known,

then \( T = \sum (x_i - \mu)^2 \) is sufficient for \( \sigma^2 \).

Remark: (aimed at problem 13, but holds in general)

The m.l.e. is a function of the sufficient statistic(s)

since \( L = g(T, \theta) h(X, T) \)

\[ \ln L = \ln g + \ln h \]

\[ \frac{\partial \ln L}{\partial \theta_i} = \frac{1}{g(T, \theta)} \frac{\partial g}{\partial \theta_i} = 0 \quad \frac{\partial \ln L}{\partial \theta_0} = 0 \text{ since it is independent of } \theta \]

The solution of this equation (which gives the m.l.e.) obviously depends only on \( T \).

Def. 32: A sufficient statistic is called complete if

\[ E_\theta[h(T)] = 0 \text{ implies } h(T) = 0, \]

\[ E_\theta \text{ denotes identically equal in } \theta \]

i.e., \[ \int h(T) f(T, \theta) dT = 0 \]

Theorem 27: If \( T \) is a complete sufficient statistic and \( d(T) \) is an unbiased estimate of \( g(\theta) \) then \( d(T) \) is an essentially unique minimum variance unbiased estimate (m.v.u.e.).

Proof: Let \( d_1(T) \) be any other unbiased estimate of \( g(\theta) \); then we know that

\[ E_\theta[d(T)] = g(\theta) \]
\[ E_\theta [d_1(T)] = g(\theta) \]

thus by subtraction \[ E_\theta [d(T) - d_1(T)] = 0 \] for all \( \theta \).

The completeness of \( T \) implies that \( d(T) - d_1(T) = 0 \) (see def. 32)
or that \( d(T) = d_1(T) \).

Further -- let \( d'(X) \) be any other unbiased estimator. We know from theorem 26
that \[ E[d'(X) \mid T] = \psi(T) \] is unbiased and \( \sigma^2_\psi \leq \sigma^2_d \), with the equality
holding only if \( d' \) is a function of \( T \).

But, by the first part of the proof \[ \psi(T) = d(T) \]

Hence, if we start with an unbiased estimator not a function of \( T \), we can improve it. If we start with an unbiased estimator that is a function of \( T \), it is \( d(T) \). This is the contention of the theorem.

Example No. 1: Binomial \( X \) is \( B(n, p) \)
\[ p(x) = \binom{n}{x} p^x (1-p)^{n-x} \]

For completeness of \( X \), which has previously been shown to be a sufficient statistic, we need
\[ \sum_{x=0}^{n} h(x) \binom{n}{x} p^x (1-p)^{n-x} \overset{p}{\Rightarrow} 0 \implies h(x) = 0 \cdot \]

The left hand side is a polynomial in \( p \) of degree \( n \).

\[ P_n(p) \overset{p}{\equiv} 0 \quad i.e. \quad a_n p^n + a_{n-1} p^{n-1} + \ldots + a_0 \overset{p}{\equiv} 0 \]

For this to be identically zero in \( p \) implies that all coefficients are zero, which means \( h(x) = 0 \) for every \( x = 0, 1, 2, \ldots, n \),
therefore \( h(x) \overset{p}{\equiv} 0 \) and \( X \) is a complete sufficient statistic,

\[ \frac{X}{n} \text{ is m.v.u.e.} \]

Example No. 2: Poisson \( X_1, X_2, \ldots, X_n \) are Poisson \((\lambda)\).

\[ T = \sum X_i \text{ is sufficient and is Poisson } (n\lambda) \]

For completeness, we must have
\[ \sum_{T=0}^{\infty} h(T) e^{-n\lambda} \frac{(n\lambda)^T}{T!} \overset{\lambda}{\equiv} 0 \text{ implying } h(T) = 0 \text{ on the integers} \]
\[ \sum_{T=0}^{\infty} h(T) \left( \frac{n\lambda}{T} \right)^T = 0 \]

\[ h_0 + h_1 \frac{n\lambda}{1} + h_2 \frac{(n\lambda)^2}{2!} + \cdots = 0 \]

Such a power series identically zero means each coefficient = 0 or \( h(T) = 0 \), so that \( T \) is complete -- therefore \( \frac{T}{n} \) is m.v.u.e. of \( \lambda \),

\[ (1 - \frac{T}{n})^T \text{ is m.v.u.e. of } e^{-\lambda}. \]

Example No. 3: Normal \( X_1, X_2, \ldots, X_n \) are \( N(\mu, 1) \).

\( \overline{X} \) is sufficient and is \( N(\mu, \frac{1}{n}) \).

\[ E\left[ h(\overline{X}) \right] = \frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} h(\overline{x}) e^{-\frac{(\overline{x} - \mu)^2}{2}} \, d\overline{x} = \frac{\sqrt{n}}{\sqrt{2\pi}} \int h(\overline{x}) e^{-\frac{\overline{x}^2}{2}} e^{-n\overline{x}\mu} \, d\overline{x} = \frac{\sqrt{n}}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}} \int h(\overline{x}) e^{-\frac{\overline{x}^2}{2}} e^{-n\overline{x}\mu} \, d\overline{x} \equiv \mu 0 \]

This is a bilateral Laplace transform.

By the theory of Laplace transforms (ref: Widder's book) if the Laplace transform \( \equiv 0 \) identically, then the function = 0 or

\[ h(\overline{x}) e^{-\frac{\overline{x}^2}{2}} = 0 \]

which implies that \( h(\overline{x}) = 0 \), therefore \( \overline{X} \) is complete.

If \( X_1, X_2, \ldots, X_n \) are \( N(\mu, \sigma^2) \) then by a similar argument \( (\overline{X}, s^2) \) are complete for \( \mu, \sigma^2 \).

Remark: Any estimate which is unbiased and a function of \( \overline{X}, s^2 \) is m.v.u.e. of \( g(\mu, \sigma^2) \).

In particular, if \( g(\mu, \sigma^2) = \mu^2 + \sigma^2 = E[X^2] \),

\[ \frac{1}{n} \sum X_i^2 \text{ is an unbiased estimate of } \mu^2 + \sigma^2. \]

\[ \frac{1}{n} \sum X_i^2 = \frac{(n-1)s^2 + \overline{x}^2}{n} \text{ so } \frac{1}{n} \sum X_i^2 \text{ is m.v.u.e. of } \mu^2 + \sigma^2. \]
Problem 46: Find m.v.u.e. of $\lambda_p$ where $Pr[x < \lambda_p] = p$.

and $X_1, X_2, \ldots, X_n$ are NID($\mu, \sigma^2$)

Example: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$

$Y_1, Y_2, \ldots, Y_n$ are $N(\nu, \sigma^2)$

Sufficient statistics are $(\bar{X}, \bar{Y}, s_x^2, s_y^2)$

(i.e., we have four sufficient statistics for three parameters),

thus this sufficient statistic is not complete.

Remark: for completeness, the sufficient statistic vector must have the same number of components as the parameter vector.

Remark: for an example of what can happen by forcing the criteria of unbiasedness on an estimate, see Lehmann p. 3-13, 114.

Non-parametric Estimation (m.v.u.e.)

Let $X_1, X_2, \ldots, X_n$ be independent random variables with continuous d.f. $F(x)$.

The problem is to estimate some function $g(F)$, e.g.

$$g(F) = \int_{-\infty}^{\infty} x \, dF(x) = EX $$

provided $EX < \infty$

$$g(F) = \sigma^2_x$$

$$g(F) = F(x)$$ i.e., we want to estimate the density function

$$g(F) = F(a) = Pr[X \leq a]$$

$$g(F) = F(b) - F(a)$$

$$g(F) = (X_1, X_n)$$ such that $F(X_n) - F(X_1) \geq 1 - \alpha$

(two-sided tolerance limit problem)
Theorem 26: If \( d(X_1, X_2, \ldots, X_n) \) is symmetric in \( X_1, X_2, \ldots, X_n \) and 
\[ E[d(X)] = g(F) \] then \( d(X) \) is m.v.u.e. of \( g(F) \).

Proof: Sufficient statistic is 
\[ T = (\sum x_1, \sum x_1^2, \ldots, \sum x_1^n). \]

Consider the \( n \) equations:
\[ \sum x_1 = t_1 \]
\[ \sum x_1^2 = t_2 \]
\[ \ldots \]
\[ \sum x_1^n = t_n \]

These equations have at most \( nj \) solutions.

Assume (as is true with probability 1) that all the \( X \)'s are distinct — if 
\( X_1, X_2, \ldots, X_n \) is a solution, so is any permutation of the \( X \)'s.

There are \( nj \) permutations of the \( X \)'s so these give a complete set of solutions.

Since the sufficient statistic may be regarded as a set of observations,
order disregarded, any function of \( T \) is symmetric in \( X_1, X_2, \ldots, X_n \).

To show completeness, we must show
\[
\begin{align*}
\int g(T) \, dF(T) &\equiv 0 \quad \longrightarrow \quad g(T) = 0.
\end{align*}
\]

Consider the sub-family of densities
\[ C(\theta_1, \theta_2, \ldots, \theta_n) e^{-x^{2n} - \theta_1 x - \theta_2 x^2 - \ldots - \theta_n x^n} \]

\[ f(x) = C_n\sum x_1^{2n} - \theta_1 \sum x_1 - \theta_2 \sum x_1^2 - \ldots - \theta_n \sum x_1^n \]
\[ = C_n\sum x_1^{2n} - \theta_1 t_1 - \theta_2 t_2 - \ldots - \theta_n t_n \]

\[ f(x) = h(x) \exp(- \theta_1 t_1 - \theta_2 t_2 - \ldots - \theta_n t_n) \]

\[ T = (t_1, t_2, \ldots, t_n) \]
\[
 f(T) = f(t_1, t_2, \ldots, t_n) = C \exp(-\sum_{i=1}^{n} \theta_i t_i - \sum_{i=1}^{n} \theta_i t_i) p(t_1, \ldots, t_n)
\]

where \( p(t_1, t_2, \ldots, t_n) \) is polynomial in the \( t \)'s obtained from the Jacobian --- it does not involve the \( \theta \)'s and is non-negative.

Hence, if
\[
 g(T) \sum_{j=1}^{n} \theta_j t_j
\]

by the same type of theory as in the normal case (uniqueness of the bilateral Laplace transform) this implies

\[
 g(T) p(T) = 0 \quad \text{or} \quad g(T) = 0
\]

Example 1: Estimate \( E(X) \).

\( X \) is symmetric function of \( X_1, X_2, \ldots, X_n \), further, it is unbiased so that by theorem 28 it is m.v.u.e.

Example 2: Estimate \( F(x) \).

We have, from the sample,

\[
 F_n(x) = \frac{1}{n} \left( \# x_i \leq x \right) = \frac{1}{n} \sum_{i=1}^{n} \psi_i(x) \quad \text{where} \quad \psi_i(x) = 1 \text{ if } x_i \leq x \\
 = 0 \text{ if } x_i > x
\]

\[ F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \psi_i(x) \] indicates symmetry in the \( x \)'s.

It then remains to show that it is unbiased, i.e., for each \( x \): \( E[F_n(x)] = F(x) \).

For each fixed \( x \) the number of \( x_i \leq x \) is a binomial random variable with parameters \( [n, F(x)] \) --- and hence, since the sample frequency is an unbiased estimate of the binomial parameter, we have the required unbiasedness.

Remark: \( F_n(x) \xrightarrow{p} F(x) \) for all \( x \) \hspace{1cm} [Proof given later]
Problem 47: Find the m.v.u.e. of \( \Pr[X \leq a] \) (a known) given observations \( X_1, X_2, \ldots, X_n \) with d.f. \( F(x) \).

Problem 48: Given \( X_1, X_2, \ldots, X_n \) are NID(0, \( \sigma^2 \)) and

\[
S = \sum X_i^2
\]

is sufficient for \( \sigma^2 \),

a) find the minimum risk estimate of \( \sigma^2 \) among functions of the form \( aS \).

b) is there a constant risk estimator of \( \sigma^2 \) of the form \( aS + b \)?

---

\( F \). \( \chi^2 \)-estimation

-- most generally applied to multinomial situations.

-- usually associated with Karl Pearson.


**Multinomial Distribution:**

Given a series of trials with:

possible outcomes of each independent trial \( E_1, E_2, \ldots, E_k \)

probabilities for a given outcome: \( p_1, p_2, \ldots, p_k \)

and \( n \) experiments result in:

\( v_1, v_2, \ldots, v_k \)

with the restrictions that \( \sum p_i = 1 \) \( \sum v_i = n \)

\[
\Pr[v_1, v_2, \ldots, v_k] = \frac{n!}{v_1! v_2! \cdots v_k!} p_1^{v_1} p_2^{v_2} \cdots p_k^{v_k}
\]

[This is a term of the multinomial expansion of \( (p_1 + p_2 + \ldots + p_k)^n \)].

**Characteristic function:**

\[
\Phi_{v_1, v_2, \ldots, v_k}(t_1, t_2, \ldots, t_k) = \sum \frac{n!}{v_1! v_2! \cdots v_k!} (p_1 e^{t_1})^{v_1} \cdots (p_k e^{t_k})^{v_k}
\]

\[
= (p_1 e^{t_1} + \ldots + p_k e^{t_k})^n
\]
\[ \varphi_{v_1, v_2, \ldots, v_k}(t_1, 0, \ldots, 0) = (p_1 e^{t_1} + p_2 + \cdots + p_k)^n \]

\[ = [p_1 e^{t_1} + (1 - p_1)]^n \]

i.e., any one variable in a multinomial situation is binomial \( B(n, p_1) \).

Moments:

\[ E[v_i] = np_i \]

\[ \sigma_{v_i}^2 = np_i(1 - p_i) \]

\[ \sigma^2_{v_i v_j} = -np_ip_j \frac{\partial^2 \varphi}{\partial t_i \partial t_j} \bigg|_{t=0} = E[v_i] E[v_j] \]

Asymptotic distributions:

a) \[ z_i = \frac{v_i - np_i}{\sqrt{np_i(1-p_i)}} \]
   \( \) is \( \mathcal{N}(0, 1) \) as \( n \rightarrow \infty \).

b) \[ \sum_{i=1}^{k} z_i^2 = \sum_{i=1}^{k} \frac{(v_i - np_i)^2}{np_i} \]
   \( \) has a \( \chi^2 \)-distribution with \( k-1 \) d.f. as \( n \rightarrow \infty \)
   (see Cramer for the transformation from \( k \) dependent to \( k-1 \) independent variables)

c) as \( n \rightarrow \infty \), \( np_i \rightarrow \lambda_i \)
   \( i = 1, 2, \ldots, k-1 \)

\( v_1, v_2, \ldots, v_{k-1} \) have a limiting multi-Poisson distribution with

parameters \( \lambda_1, \lambda_2, \ldots, \lambda_{k-1} \)
\( (v_1, v_2, \ldots, v_{k-1} \) are independent in the limit).

Example: (Homogeneity)

We have 3 sets of multinomial trials with possible outcomes \( E_1, E_2, E_3 \)

with probabilities \( \Theta_1, \Theta_2, 1 - \Theta_1 - \Theta_2 \) for each set of trials.
Therefore we have the following outcomes

\[
\begin{array}{ccc}
v_{11} & v_{12} & v_{13} \\
v_{21} & v_{22} & v_{23} \\
v_{31} & v_{32} & v_{33} \\
v_{*1} & v_{*2} & v_{*3} \\
\end{array}
\quad n_1 \quad n_2 \quad n_3 \quad N
\]

\[
\ln L = \ln \left[ \frac{v_{11}}{p_{11}} \cdot \frac{v_{12}}{p_{12}} \cdot \frac{v_{13}}{p_{13}} \cdot \frac{v_{21}}{p_{21}} \cdot \frac{v_{22}}{p_{22}} \cdot \frac{v_{23}}{p_{23}} \cdot \frac{v_{31}}{p_{31}} \cdot \frac{v_{32}}{p_{32}} \cdot \frac{v_{33}}{p_{33}} \right] + \ln K
\]

\[= v_{*1} \ln \theta_1 + v_{*2} \ln \theta_2 + v_{*3} \ln (1 - \theta_1 - \theta_2) + \ln K\]

\[
\frac{\partial \ln L}{\partial \theta_1} = \frac{v_{*1}}{\theta_1} - \frac{v_{*3}}{1 - \theta_1 - \theta_2} = 0
\]

\[
\frac{\partial \ln L}{\partial \theta_2} = \frac{v_{*2}}{\theta_2} - \frac{v_{*3}}{1 - \theta_1 - \theta_2} = 0
\]

These give the two equations:

1. \( (v_{*1} + v_{*3}) \theta_1 + v_{*1} \theta_2 = v_{*1} \)

2. \( v_{*2} \theta_1 + (v_{*2} + v_{*3}) \theta_2 = v_{*2} \)

which when added together yield

\( N \theta_1 + N \theta_2 = v_{*1} + v_{*2} = N - v_{*3} \)

Multiplying (3) by \( \frac{v_{*2}}{N} \) we get

\( v_{*2} \theta_1 + v_{*2} \theta_2 = \frac{v_{*2} (v_{*1} + v_{*2})}{N} \)

(2) - (4) yields \( v_{*3} \theta_2 = \frac{v_{*2} v_{*3}}{N} \)

or \( \hat{\theta}_2 = \frac{v_{*2}}{N} \)
Similarly:
\[ \hat{\theta}_1 = \frac{v_1}{N} \]

It can also be found that
\[ v(\hat{\theta}_2) = \frac{\theta_2(1-\theta_2)}{N} \]

\[ \text{COV}(\hat{\theta}_1, \hat{\theta}_2) = -\frac{\theta_1 \theta_2}{N} \]

---

**Problem 9:** (Independence)

Consider one sequence of \( N \) trials that result in one of the \( rc \) events

\[ E_{11}, E_{12}, \ldots, E_{rc} \]

with probabilities \( p_{ij} = \rho_i \tau_j \)

where \( \sum_{i=1}^{r} \rho_i = 1 \); \( \sum_{j=1}^{c} \tau_j = 1 \).

Find the m.l.e. of \( \rho_i, \tau_j \) and also their variances and covariances.

---

**\(\chi^2\)-estimation; general case**

**Given:** \( s \) series of \( n_i \) trials, each trial resulting in one of the events \( E_1, \ldots, E_k \)

- The probability of \( E_j \) occurring on any trial in the \( i^{th} \) series is \( p_{ij} \);

\[ \sum_{j=1}^{k} p_{ij} = 1 \]

- The random variable in the problem is the number of occurrences of each event:

\[ v_{i1}, v_{i2}, \ldots, v_{ik}, \sum_{j=1}^{k} v_{ij} = n_i \]

- The \( p_{ij} \) are functions of \( \theta \) (continuous with continuous first and second derivatives).

- Expectation is with respect to "\( j \)" i.e., \( E[v_{ij}] = n_i p_{ij} \) -- the use of the "\( i \)" subscript is a convenience, we could consider the \( s \) trials as one big trial.
The following methods of estimating \( \theta \) are asymptotically equivalent, i.e.,

(i) they are consistent
(ii) they are asymptotically normal
(iii) they have equal asymptotic variances

1. maximum likelihood
2. minimum \( \chi^2 \)
3. modified minimum \( \chi^2 \)
4. transformed minimum \( \chi^2 \)

1. **Maximum Likelihood Estimation**:

Maximize

\[
\prod_{i=1}^{s} \prod_{j=1}^{k} \left[ p_{ij}(\theta) \right]^{v_{ij}} \quad \text{or} \quad \sum_{i=1}^{s} \sum_{j=1}^{k} v_{ij} \ln p_{ij}(\theta)
\]

subject to \( \sum_{j=1}^{k} p_{ij}(\theta) = 1 \) for each \( i, \) with respect to \( \theta; \)

or actually solve the equations:

\[
\sum_{i} \sum_{j} \frac{v_{ij}}{p_{ij}(\theta)} \frac{\partial p_{ij}}{\partial \theta} = 0 \tag{1}
\]

(provided suitable regularity conditions are imposed as given in theorems 22, 23).

2. **Minimum \( \chi^2 \)**

\[
\sum_{i} \sum_{j} \frac{(v_{ij} - n_{ij} p_{ij})^2}{n_{ij} p_{ij}} \quad \text{is asymptotically distributed as} \quad \chi^2 \quad \text{with} \ s(k-1) \ d.f.
\]

The method is to minimize this expression with respect to \( \theta \) or to solve the equations:

\[
-2 \sum_{i} \sum_{j} \frac{n_{ij} (v_{ij} - n_{ij} p_{ij})}{n_{ij} p_{ij}} \frac{\partial p_{ij}}{\partial \theta} + \sum_{i} \sum_{j} \frac{(v_{ij} - n_{ij} p_{ij})^2}{n_{ij} p_{ij}^2} \frac{\partial p_{ij}}{\partial \theta} = 0 \tag{2}
\]

This set of equations could be expressed as:

\[
\sum_{i} \sum_{j} \frac{v_{ij}}{p_{ij}(\theta)} \frac{\partial p_{ij}}{\partial \theta} - \sum_{i} n_{ij} \sum_{j} \frac{\partial p_{ij}}{\partial \theta} + \frac{1}{2} \sum_{i} \sum_{j} \frac{(v_{ij} - n_{ij} p_{ij})^2}{n_{ij} p_{ij}^2} \frac{\partial p_{ij}}{\partial \theta} = 0
\]
\[ \sum_{j} p_{ij} = 1 \quad \sum_{j} \frac{\partial p_{ij}}{\partial \theta} = 0 \]

therefore the second term equals 0

\[ R = \text{the third term} \xrightarrow{p} 0 \quad \text{as} \quad N \to \infty \]

Hence, equation [2] is the same as equation [1] except for \( R \) and thus it is seen that under suitable regularity conditions the solutions of equation [2] tend in probability to the solutions of equation [1].

3. Modified Minimum \( \chi^2 \)

Replace \( p_{ij} \) in the denominator of the regular minimum \( \chi^2 \) equation with its estimated value, and then minimize with respect to \( \theta \):

\[ \chi^2_m = \sum_{i} \sum_{j} \frac{(v_{ij} - n_{i}p_{ij})^2}{v_{ij}} \]

subject to \( \sum_{j} p_{ij} = 1 \) for \( i = 1, 2, \ldots, s. \)

Comparing the two \( \chi^2 \) methods

\[
\begin{align*}
\chi^2 - \chi^2_m &= \sum_{i} \sum_{j} (v_{ij} - n_{i}p_{ij})^2 \left( \frac{1}{n_{i}p_{ij}} - \frac{1}{v_{ij}} \right) \\
&= \sum_{i} \sum_{j} (v_{ij} - n_{i}p_{ij})^2 \left( \frac{v_{ij} - n_{i}p_{ij}}{n_{i}p_{ij}v_{ij}} \right) \\
&= \sum_{i} \sum_{j} \frac{(v_{ij} - n_{i}p_{ij})^3}{n_{i}p_{ij}v_{ij}} = R
\end{align*}
\]

\[ R \xrightarrow{p} 0 \quad \text{faster than does} \quad \chi^2 \quad \text{therefore methods (2) and (3)} \]

are asymptotically equivalent.

The equations to be solved in this case are:

\[ \sum_{i} \sum_{j} \frac{n_{i}(v_{ij} - n_{i}p_{ij})}{v_{ij}} \frac{\partial p_{ij}}{\partial \theta} = 0 \quad [3] \]

4. Transformed Minimum \( \chi^2 \)

Recall that if \( \frac{\sqrt{N}(X_n - \mu)}{\sigma_n} \) is asymptotically normal, then \( \frac{g(X_n) - g(\mu)}{\sigma g'(\mu)} \) is \( \text{A} \). \( \text{N} \).

and the asymptotic variance of \( g(X_n) \) is \( \sigma^2 \left[ g'(\mu) \right]^2 \).
Before writing the necessary equation, write

\[ q_{ij} = \frac{v_{ij}}{n_1} \quad \text{thus the } q_{ij} \text{ are random variables.} \]

\[ q_{ij} \xrightarrow{p} p_{ij} \quad \mathbb{E}[q_{ij}] = p_{ij} \]

Thus the modified minimum \( \chi^2 \) equation can be written:

\[ \chi^2_m = \sum_i \sum_j \frac{n_1 (q_{ij} - p_{ij})^2}{q_{ij}} \]

The equivalent estimation procedure is thus to minimize

\[ \sum_i \sum_j \frac{n_1 \left[ g(q_{ij}) - g(p_{ij}) \right]^2}{q_{ij} \left[ g'(q_{ij}) \right]^2} \]

(the numerator should be \( p_{ij} \left[ g'(p_{ij}) \right]^2 \) but the difference occasioned by the modification \( \to 0 \))

for \( g \)'s which are continuous with continuous first and second derivatives, and with \( g'(p_{ij}) \) bounded away from zero.

This method will be useful if the transformation, \( g(p_{ij}(\Theta)) \) is simple.

The equations to be solved are thus:

\[ \sum_i \sum_j \frac{n_1 \left[ g(q_{ij}) - g(p_{ij}) \right]}{q_{ij} \left[ g'(q_{ij}) \right]^2} \quad g'(p_{ij}) \frac{\partial p_{ij}}{\partial \Theta} = 0 \quad [4] \]

Summary:

Method (1) \([\text{m.l.e.}]\) -- is most useful if the \( p_{ij} \) can be expressed as products, i.e., \( p_{ij} = p_i p_j \) as in problem 49.

Method (2) \([\text{minimum } \chi^2]\) -- most useful if \( p_{ij} \) can be expressed as a sum of probabilities, i.e., \( p_{ij} = x_i + d_j \).

Method (3) \([\text{modified } \chi^2]\) -- same as (2).

Method (4) \([\text{transformed } \chi^2]\) -- may be most useful in some special cases.
Examples of minimum - $\chi^2$ procedures:

1. Linear Trend: $n_1 = 2$, $p = 3$

<table>
<thead>
<tr>
<th>obs.</th>
<th>prob.</th>
<th>obs.</th>
<th>prob.</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{11}$</td>
<td>$p_{11} = p - \Delta$</td>
<td>$v_{12}$</td>
<td>$p_{12} = 1 - p + \Delta$</td>
<td>$n_1$</td>
</tr>
<tr>
<td>$v_{21}$</td>
<td>$p_{21} = p$</td>
<td>$v_{22}$</td>
<td>$p_{22} = 1 - p$</td>
<td>$n_2$</td>
</tr>
<tr>
<td>$v_{31}$</td>
<td>$p_{31} = p + \Delta$</td>
<td>$v_{32}$</td>
<td>$p_{32} = 1 - p - \Delta$</td>
<td>$n_3$</td>
</tr>
</tbody>
</table>

Problem is to estimate both $p$ and $\Delta$.

Using the modified minimum - $\chi^2$ procedure:

$$
\chi^2 = \left[ v_{11} - n_1(p-\Delta) \right]^2 \left[ \frac{1}{v_{11}} + \frac{1}{v_{12}} \right] + \left[ v_{21} - n_2p \right]^2 \left[ \frac{1}{v_{21}} + \frac{1}{v_{22}} \right] + \left[ v_{31} - n_3(p+\Delta) \right]^2 \left[ \frac{1}{v_{31}} + \frac{1}{v_{32}} \right]
$$

$$
- \frac{1}{2} \frac{\partial^2 \chi^2}{\partial p^2} = a_1 \left[ v_{11} - n_1(p-\Delta) \right] + a_2 \left[ v_{21} - n_2p \right] + a_3 \left[ v_{31} - n_3(p+\Delta) \right] = 0
$$

where $a_1 = n_1 \left( \frac{1}{v_{11}} + \frac{1}{v_{12}} \right)$

$$
- \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \Delta^2} = -a_1 \left[ v_{11} - n_1(p-\Delta) \right] + a_3 \left[ v_{31} - n_3(p+\Delta) \right] = 0
$$

These two equations yield the following two equations which can be solved simultaneously for $p$ and $\Delta$:

$$
p \sum n_1 a_1 + (a_3 n_3 - a_1 n_1) \Delta = \sum a_1 v_{11}
$$

$$
(a_3 n_3 - a_1 n_1) p + (a_1 n_1 + a_3 n_3) \Delta = a_3 v_{31} - a_1 v_{11}
$$

Note: remember that this procedure is asymptotically equivalent to the m.l.e. -- therefore the asymptotic variance-covariance matrix can be found by evaluating

$$
\frac{1}{2} \frac{\partial^2 \chi^2}{\partial p^2} \quad \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \Delta^2} \quad \frac{1}{2} \frac{\partial^2 \chi^2}{\partial p \partial \Delta}
$$

at the expectations since $-\chi^2/2$ is the exponent of the asymptotic normal distribution.

$$
\frac{1}{2} \frac{\partial^2 \chi^2}{\partial p^2} = \sum a_1 n_1 \quad \frac{1}{2} \frac{\partial^2 \chi^2}{\partial p \partial \Delta} = a_3 n_3 - a_1 n_1 \quad \frac{1}{2} \frac{\partial^2 \chi^2}{\partial \Delta^2} = a_1 n_1 + a_3 n_3
$$
Recall \[ a_1 = n_1 \left( \frac{1}{v_{11}} + \frac{1}{v_{12}} \right) \]

\[ E[v_{11}] = n_1(\rho - \Delta) \quad E[v_{12}] = n_1(1-\rho + \Delta) \]

Replacing the random variables in \( a_1 \) by their expectations yields

\[ a_1 \xrightarrow{p} \left( \frac{1}{\rho - \Delta} + \frac{1}{1-\rho + \Delta} \right) = \frac{1}{(\rho - \Delta)(1-\rho + \Delta)} \]

Similarly,

\[ a_2 \xrightarrow{p} \frac{1}{p(1-p)} \]

\[ a_3 \xrightarrow{p} \frac{1}{(\rho + \Delta)(1-p - \Delta)} \]

Thus the inverse of the \( A \cdot V \cdot COV \) matrix is

\[ V^{-1} = \begin{pmatrix}
  a_1 n_1 + a_2 n_2 + a_3 n_3 & = a_1 n_1 + a_3 n_3 \\
  a_1 n_1 + a_2 n_2 + a_3 n_3 & = a_1 n_1 + a_3 n_3 \\
  a_1 n_1 + a_2 n_2 + a_3 n_3 & = a_1 n_1 + a_3 n_3 \\
  a_1 n_1 + a_2 n_2 + a_3 n_3 & = a_1 n_1 + a_3 n_3 \\
\end{pmatrix} \]

---

2. Logistic (Bio-assay) problem (Bergson)

--- applying greater dosages produces greater kill (or reaction)

\[ s = \text{general} \quad n_1 = 2 \]

\( x_1, x_2, \ldots, x_n \) dosages

\( n_1, n_2, \ldots, n_s \) receive dosages

\( v_1, v_2, \ldots, v_s \) die (or react)

\( n_1 = v_1, \ldots, n_s = v_s \) survive

\[ P_i = \frac{1}{1 + e^{-\left(\alpha + \beta x_i \right)}} \]

\[ 1 - P_i = \frac{e^{-\left(\alpha + \beta x_i \right)}}{1 + e^{-\left(\alpha + \beta x_i \right)}} \]

\[ \frac{1 - P_i}{P_i} = e^{-\left(\alpha + \beta x_i \right)} \]

\[ q_i = \frac{v_i}{n_1} = \text{proportion dying or reacting} \]

\[ -ln \left( \frac{1 - P_i}{P_i} \right) = \alpha + \beta x_i \]
Applying the transformed $\chi^2$ method to estimate $\alpha$, $\beta$ and using the transformation

$$g(x) = - \ln \left( \frac{1-x}{x} \right) = \ln \left( \frac{x}{1-x} \right)$$

$$g'(x) = \frac{1-x}{x} \frac{(1+x)-(x)}{(1-x)^2} = \frac{1}{x(1-x)}$$

The transformed $\chi^2$ is given by:

$$\chi_T^2 = \sum_{i=1}^{s} n_i \left[ \frac{[g(q_{i1}) - g(p_{i1})]^2}{q_i [g'(q_{i1})]^2} + \frac{[g(l=q_{i1}) - g(l=p_{i1})]^2}{(1-q_i) [g'(1-q_{i1})]^2} \right]$$

$$= \sum_{i=1}^{s} n_i \left[ [g(q_{i1}) - g(p_{i1})]^2 \left( \frac{1}{q_i [g'(q_{i1})]^2} + \frac{1}{(1-q_i) [g'(1-q_{i1})]^2} \right) \right]$$

Putting in the values of $g$, $g'$, we have

$$\chi_T^2 = \sum_{i=1}^{s} n_i \left[ \ln \frac{q_{i1}}{1-q_{i1}} - (\alpha + \beta x_{i1}) \right]^2 \left[ \frac{[q_{i1}(1-q_{i1})]^2}{q_{i1}} + \frac{[q_{i1}(1-q_{i1})]^2}{1-q_{i1}} \right]$$

$$= \sum_{i=1}^{s} n_i \left[ \ln \frac{q_{i1}}{1-q_{i1}} - (\alpha + \beta x_{i1}) \right]^2 \left[ q_{i1}(1-q_{i1}) \right]$$

**Problem 50:**

Consider $r \times c$ sequences of $n$ trials which may result in $E$ or $\bar{E}$ where in trial $(ij)$

$$\Pr[E] = \alpha_i + \beta_j \quad \Pr[\bar{E}] = 1 - \alpha_i = \beta_j \quad i = 1, 2, \ldots, r \quad j = 1, 2, \ldots, c$$

Set up equations to estimate $\alpha_i$, $\beta_j$ by

1. maximum likelihood.
2. minimum modified $\chi^2$

based on observations $v_{ij}$, $n = v_{ij}$.  

- 107 -
Problem as stated produces $r+c$ equations in $r+c$ unknowns, but their matrix is singular -- i.e., the equations are not independent by virtue of the fact that the sum of the $r$ equations for the $\alpha_i$ equals the sum of the $c$ equations for the $\beta_j$.

Therefore, reduce the number of sequences to $h_i$

and add the restriction that $p_{22} = p_{12} + p_{21} - p_{11}$

i.e.: $p_{11} = \alpha_1 + \beta_1$

$p_{12} = \alpha_1 + \beta_2$

$p_{21} = \alpha_2 + \beta_1$

$p_{22} = \alpha_2 + \beta_2 = p_{12} + p_{21} - p_{11}$

i.e., reparameterize and find the equations necessary to solve for the new parameters.

Problem 51:

Given $s$ sequences of $n_1$ trials may result in $E$ or $\tilde{E}$ where in sequence $i$

$$Pr[E] = e^{-\lambda}$$

$$Pr[\tilde{E}] = 1 - e^{-\lambda}$$

($\alpha_i$ known; $i = 1, 2, \ldots, s$)

(let the number of $E$'s observed be denoted by $v_1$).

Set up the equations to estimate $\lambda$ by

1. maximum likelihood.

2. minimum modified $x^2$.

and find the asymptotic variance of $\hat{\lambda}$.

(note: Ferguson discusses this problem in his Dec. 58 article in the Annals)

G. Minimax estimation:

Recall that $d(X)$ is a minimax estimate of $g(\theta)$ if

$$\sup_{\theta} R[d, \theta]$$

is minimized by $d(X)$.

$$R[d, \theta] = E[d(X) - g(\theta)]^2 = \int [d(x) - g(\theta)]^2 f(x) \, dx$$

[see def. 24 on p. 54]
Methods to try to find minimax estimates:

1. Cramer-Rao

\[ \sigma_d^2 \geq \frac{[1+b'(\theta)]^2}{nE\left[ \frac{\partial \ln f(x)}{\partial \theta} \right]^2} = \frac{[1 + b'(\theta)]^2}{nE\left[ \frac{\partial^2 \ln f(x)}{\partial \theta^2} \right]} \]

\[ E[\hat{d}(X) - g(\theta)]^2 = E[\hat{d}(X) - E[\hat{d}(X)] + E[\hat{d}(X)] - g(\theta)]^2 = \sigma_d^2 + [b(\theta)]^2 \]

\[ R[\hat{d}, \theta] \geq \frac{[1+b'(\theta)]^2}{nE\left[ \frac{\partial \ln f(x)}{\partial \theta} \right]^2} + b^2(\theta) = k(\theta) \]

If we have an estimate \( \hat{d}(X) \) which actually has the risk \( k(\theta) \) then any other estimator with risk less than \( k(\theta) \) leads to a contradiction.

Example: \( X_1, X_2, \ldots, X_n \) are \( N(\mu, \sigma^2) \)

and we know \( R(\bar{X}) = \frac{\sigma^2}{n} \)

Lehmann shows that \( \frac{[1+b'(\theta)]^2}{n\sigma^2} + b^2(\theta) \leq \frac{\sigma^2}{n} \)

which implies that \( b(\theta) = 0 \)

Therefore \( \bar{X} \) is a minimax estimator of \( \mu \).

2. This method based on the following theorem which will be stated without proof.

**Theorem 29:** If a Bayes estimator \[ \text{def. 23} \] has constant risk \[ \text{def. 25} \] then it is minimax.

\( \hat{d}(X) \) is constant risk if \( E[\hat{d}(X) - g(\theta)]^2 \) is constant.

\( \hat{d}(X) \) is Bayes if \( \hat{d}(X) \) minimizes \( \int [d(x) - g(\theta)] f(x, \theta) df(\theta) \)

where \( G \) is the "a priori" distribution of \( \theta \).

ref: Lehmann -- section 4, p. 19-21

**Example:** Binomial -- \( X \) is binomial \( B(n, p) \).

Recall that we found that \( d(X) = \frac{\sqrt{n}}{1+\sqrt{n}} \frac{x}{n} + \frac{1}{2(1+\sqrt{n})} \)

is a constant risk estimator of \( p \) [see problem 31, p. 54].
If \( p \) has a prior Beta distribution
\[
f(p) = K p^{\frac{\sqrt{n}}{2}} (1 - p)^{\frac{\sqrt{2}}{2} - 1}
\]
then \( d(X) \) is Bayes;

hence the minimax estimate of the binomial parameter \( p \) is
\[
\frac{\sqrt{n}}{1+\sqrt{n}} \frac{X}{n} + \frac{1}{2(1+\sqrt{n})}.
\]

Graphically we have the following risk functions:

---

Problem 52: Compare \( R_{d_m}(p) \) and \( R_{d_m}(p) \) for \( n = 25 \).

\[
(\hat{d} = \frac{X}{n}; \quad d_m = \text{minimax estimate})
\]

H. Wolfowitz's Minimum Distance Estimation

\( X_1, X_2, \ldots, X_n \) are observations from the distribution \( F(x, \theta) \).

We are also given some measure of distance between 2 distributions \( \rho(F, G) \).

\[
e.g. \quad \rho(F, G) = \sup_{-\infty < x < \infty} |F(x) - G(x)|
\]

\[
\rho_1(F, G) = \int [F(x) - G(x)]^2 \frac{dF(x) + G(x)}{2}
\]

Also we have that the sample \( d.s.f., \quad F_n(x) = \frac{\text{no. of } X_i's < x}{n} \) (see p. 97).
\( \hat{\theta} \) is a minimum distance estimate of \( \theta \) if \( \rho[F(x, \theta), F_n(x)] \) is minimized by choosing \( \theta = \hat{\theta} \).

Note: We want the whole sample d.f. to agree with the whole theoretical distribution -- not just the means or the variances agreeing.

In particular, if we use the "sup-distance" then \( \hat{\theta} \) is that estimate of \( \theta \) which yields

\[
\min_{\theta} \sup_{-\infty < x < \infty} \left| F(x, \theta) - F_n(x) \right|.
\]

Remark: \( \hat{\theta} \) is a consistent estimate of \( \theta \).

Proof: Suppose for all sufficiently large \( n \), \( \hat{\theta} \) differs from \( \theta \) by more than \( \delta \).

Then for some \( \delta \)

\[
\sup_{-\infty < x < \infty} \left| F(x, \hat{\theta}) - F(x, \theta_0) \right| > \delta
\]

and for some \( n_1 \)

\[
\sup_{-\infty < x < \infty} \left| F_{n_1}(x) - F(x, \theta_0) \right| < \frac{\delta}{3}
\]

We want to show that

\[
\sup_{-\infty < x < \infty} \left| F_n(x) - F(x, \theta_0) \right| < \sup_{-\infty < x < \infty} \left| F_{n_1}(x) - F(x, \hat{\theta}) \right|.
\]

Let \( x' \) be the \( x \) such that

\[
F(x', \hat{\theta}) - F(x', \theta_0) = \delta
\]

\[
F_{n_1}(x') - F(x', \theta_0) < \frac{\delta}{3}
\]

Therefore \( F(x', \hat{\theta}) - F_{n_1}(x') > \frac{2}{3} \delta \)

thus \( \sup_{-\infty < x < \infty} \left| F(x', \theta) - F_{n_1}(x') \right| > \frac{2}{3} \delta \)

\[
> \sup_{-\infty < x < \infty} \left| F_n(x) - F(x, \theta_0) \right|
\]

but this contradicts the fact that \( \theta \) minimizes \( \sup_{-\infty < x < \infty} \left| F_n(x) - F(x, \theta) \right| \)

therefore \( \hat{\theta} \) is consistent.
Example: Find the minimum distance estimate for $\mu$ given $X_1, X_2, X_3$ are $N(\mu, 1$).

$X_1 = 1$  $X_2 = 2$  $X_3 = 4$

$\Phi(z_1) = \begin{array}{c}
0.081 \\
0.345 \\
0.915 \\
0.325 (0.67 - 0.345)
\end{array}$

$\Phi(z_1) = \begin{array}{c}
0.097 \\
0.382 \\
0.955 \\
0.288
\end{array}$

$\Phi(z_1) = \begin{array}{c}
- \\
0.371 \\
0.952 \\
0.299
\end{array}$

$\Phi(z_1) = \begin{array}{c}
- \\
0.386 \\
0.956 \\
0.284 (0.67 - 0.386)
\end{array}$

Other values of $\mu$ were tried, but the min sup-distance = 0.284, therefore $\hat{\mu} = 2.29$. ($\bar{x} = 2.33$)

Problem 53: Take 4 observations from a table of normal deviates (add an arbitrary factor if desired) and find the minimum distance estimate of $\mu$. 
Chapter V: TESTING OF HYPOTHESES -- DISTRIBUTION FREE TESTS

I. Basic concepts.

\[ X_1, X_2, \ldots, X_n \] have distribution function \( F(x) \) -- continuous

-- absolute continuous

\([\text{density } f(x)]\)

The hypothesis \( H \) specifies that \( F \sim F_0 \) (some family of d.f.)

Alternative \( H \) specifies that \( F \not\sim F_0 \)

Def. 33: A test is a function \( \varphi(x) \) taking on values between 0 and 1

i.e., if \( \varphi(x) = 1 \) reject \( H \) with probability 1

\[ \begin{align*}
    k & \quad \text{for } \varphi(x) = k \\
    0 & \quad \text{for } \varphi(x) = 0
\end{align*} \]

(note that this considers a test as a function instead of the usual consideration of regions)

Def. 34: A test is of size \( \alpha \) if \( E[\varphi(x)] \) for \( F \sim F_0 \leq \alpha \)

\[ E[\varphi(x)] = \int_{-\infty}^{\infty} \varphi(x) \, df(x) \]

(this says that the probability of rejecting \( H \) when it is true \( \leq \alpha \))

Def. 35: Power of a test is \( E[\varphi(x)] \) for \( F \not\sim F_0 \) and is denoted \( \beta_{\varphi}(F) \)

Def. 36: \( \{ \varphi_n \} \) is a consistent sequence of tests if for \( F \not\sim F_0 \)

\[ \beta_{\varphi_n}(F) \rightarrow 1 \quad \text{as } n \rightarrow \infty \]

Def. 37: Index of a sequence of tests:

\( N(\varphi_n, \varphi, \alpha, \beta) \) is the least integer \( n \) such that, given that

\( \varphi \) is of size \( \alpha \), \( \beta_{\varphi_n}(F | F \not\sim \varphi') \geq 1 - \beta \)
Def. 38: Asymptotic Relative Efficiency (A.R.E.)

Let $p(\theta, \theta_0)$ be some measure of the distance of $\theta$ from $\theta_0$.

Let $q_n$ and $\hat{q}_n$ be two sequences of consistent tests.

then the A.R.E. of $\hat{q}$ to $q_n$ is defined to be

$$\lim_{p \to 0} \frac{N(q_n, \theta, \alpha, \beta)}{N(\hat{q}_n, \theta, \alpha, \beta)}$$ provided the limit exists.

Problem 51: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$ (\(\sigma\) known)

$H_0$: $\mu = 0$  alternative: $\mu > 0$

Consider two tests of $H_0$:

1. the mean test: reject $H$ if $\bar{x} > \frac{z_{1-\alpha}}{\sqrt{n}} \sigma$

   $[\phi(\bar{x}) = 1$ if $\bar{x} > \frac{z_{1-\alpha}}{\sqrt{n}} \sigma$

   $= 0$ if $\bar{x} \leq \frac{z_{1-\alpha}}{\sqrt{n}} \sigma$

2. the median test: [use the large sample distribution of the median, i.e., the sample median is asymptotically normal with mean = the population median and variance = $n\sigma^2/2n$]

\[\hat{\phi}\] test: reject $H$ if the sample median $> z_{1-\alpha} \frac{\sigma \sqrt{n}}{\sqrt{2n}}$

a) find the index for each test for the alternative $\mu$.
b) find the $A_{R,E}$, i.e.

$$\lim_{\mu^1 \to 0} \frac{N(\mathcal{F}_p \mu^1, \lambda, \beta)}{N(\mathcal{F}_p \mu^1, \lambda, \beta)}$$

II. Distribution Free Tests

refs: Siegel -- Non-parametric Statistics
Fraser -- Non-parametric Methods in Statistics
Savage -- Bibliography in the Dec. 1953 J.A.S.A.

A. Quantile and Median Tests:

Def. 39: Let the solution of the equation $F(\lambda_p) = p$ define the quantile $\lambda_p$.

If the solution is not unique, define $\lambda_p = \inf_x [F(x) = p]$.

Given the problem to test $H_0: \lambda_p = \lambda_0$ against the alternative $H_1: \lambda_q = \lambda_0$

note: the alternative could be stated $\lambda_p \neq \lambda_0$ but this statement precludes consideration of the power of the test since the power in this case would be undetermined by virtue of the unknown behavior of $F(x)$.

Test: $X = \text{the number of } x_1, x_2, \ldots, x_n \leq \lambda_0$

under $H_0: X$ is $B(n, p)$

for the two sided test ($q \neq p$) at level $\alpha$ using the normal approximation

$$\text{reject } H \text{ if: } \frac{|x - np|}{\sqrt{np(1-p)}} \geq z_{1-\alpha/2}$$

Problem 55: Compute the power of the test (using the normal approximation) as a function of $q$ for $n = 100$ and $p = 0.5$

Example: for paired comparisons, to test if the median is equal to zero set up the series of observations $d_1, d_2, \ldots, d_n$, $d_i = x_i - y_i$

where $x_i \leq y_i \iff d_i \leq 0$

$x_i > y_i \iff d_i > 0$
This test that the median is zero is often referred to as the sign test, since \( X \) in this case is merely the number of \( d_i \) with a negative sign.

**Definition:** \( \hat{\lambda}_p \), the sample quantile, is defined as the solution of \( F_n(\hat{\lambda}_p) = p \)

**Theorem 30:** \( \hat{\lambda}_p \) has density function

\[
f(\hat{\lambda}_p) = h(x) = \frac{n!}{\mu!(n-\mu-1)!} \left[F(x)\right]^\mu \left[1 - F(x)\right]^{n-\mu-1} f(x)
\]

where \( \mu = \lfloor np \rfloor \) which denotes the greatest integer in \( np \)

**Proof:** ref: Cramer p. 368

\[
\Pr[\hat{\lambda}_p \leq x] = \Pr[\mu + 1 \text{ or more observations} \leq x]
\]

\[
= \sum_{j=\mu+1}^{n} \binom{n}{j} [F(x)]^j [1 - F(x)]^{n-j} = F(\hat{\lambda}_p)
\]

To get the density \( h(x) \) [assuming that \( F(x) \) has a density \( f(x) \)] we differentiate the summation, getting the following two summations [from each part of the product that occurs in each term summed]

\[
h(x) = \sum_{j=\mu+1}^{n} \binom{n}{j} j[F(x)]^{j-1} [1 - F(x)]^{n-j} f(x)
\]

\[
- \sum_{j=\mu+1}^{n} \binom{n}{j} (n - j) [F(x)]^j [1 - F(x)]^{n-j-1} f(x)
\]

The corresponding terms [in \( F^2(1 - F)^b \)] cancel each other except for the first \( j = \mu + 1 \) term in the first summation which has no corresponding term in the second sum.

Thus:

\[
h(x) = \frac{n!}{\mu!(n-\mu-1)!} \left[F(x)\right]^\mu \left[1 - F(x)\right]^{n-\mu-1} f(x)
\]

By the usual limiting process applied to density functions, \( \hat{\lambda}_p \) is asymptotically normal, with mean \( \lambda_p \) and variance \( \frac{1}{f^2(\lambda_p)} \frac{p(1-p)}{n} \)

i.e. \( \lambda_p \) is \( A \sim N(\lambda_p, \frac{p(1-p)}{n f^2(\lambda_p)}) \)
Problem 56: Let \( x \) have density \( \frac{1}{c} f(\frac{x-\mu}{c}) \)

where \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \), \( \int_{-\infty}^{\infty} x \, f(x) \, dx = 0 \), \( \int_{-\infty}^{\infty} x^2 f(x) \, dx = 1 \)

and \( f(x) \) is symmetric about 0.

\( H_0: \mu = 0 \quad H_1: \mu = \mu_1 > 0 \)

Find the condition on \( f(x) \) such that the A.R.E. of the median to the mean is greater than 1. Use large sample normal approximations for both tests.

Specialize this result to \( f(x) = \frac{1}{2c} e^{-\frac{|x-\mu_1|}{c}} \), \( -\infty < x < \infty \)

---

III. One Sample Tests

\( H_0: F = F_0 \) completely specified \( H_1: F = F_1 < F_0 \)

\( H_1^*: F = F_1 \neq F_0 \)

problems: 1) goodness of fit (usually a wider problem since \( F_0 \) is usually not completely specified)

2) "slippage test"

3) combination of tests -- ref: A Birnbaum, JASA, Sept. 1954

4) randomness in time or space -- ref: Bartholomew, Barton, and David; Biometrika; circa 1955-6

Randomness in Time:

Assume: \( \Pr[\text{n events in } (0,t) \text{ (the time interval } 0 \text{ to } t) ] = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \)

i.e., is Poisson with parameter \( \lambda t \)

Probability of event occurring in one time interval is independent of an occurrence in any other time interval.

\[ 0 \quad T_1 \quad T_2 \quad T_3 \quad \cdots \quad T_n \quad T \]
Let $T_i$ be the time at which the $i^{th}$ event occurred.

$$t_i = T_i - T_{i-1} \quad T_i = \sum_{j=1}^{i} t_j$$

$$\Pr[t_i > t] = \Pr[\text{no event occurs in time } (0,t)] = e^{-\lambda t}$$

$f(t)$ is the density of the $t_i = \lambda e^{-\lambda t}$ (the exponential distribution)

$$f(t_1, t_2, \ldots, t_n) = \lambda^n e^{-\lambda \sum t_i}$$ since the $t_i$'s are independent

Distribution of $T_1, T_2, \ldots, T_n$ is obtained by transformation

$$T_1 = t_1$$
$$T_2 = t_1 + t_2$$
$$\quad \quad \quad \quad \quad \quad \quad |J| = 1$$
$$\quad \quad \quad \quad \quad \quad \quad \frac{\partial}{\partial t_i}$$
$$T_n = t_1 + t_2 + \ldots + t_n$$

$$f(T_1, T_2, \ldots, T_n) = \lambda^n e^{-\lambda T_n}$$

but $0 \leq T_1 \leq T_2 \leq \ldots \leq T_n$

Therefore, let us find the conditional distribution of $T_1, T_2, \ldots, T_n$ given that $n$ events occurred in the fixed time interval $(0, T)$.

$$f(T_1, T_2, \ldots, T_n; n) = \text{density of } T_1, T_2, \ldots, T_n \text{ and also the probability that no events occur in the time interval } (T_n, T)$$

$$= \lambda^n e^{-\lambda T_n} e^{\lambda(T - T_n)}$$

$$= \lambda^n e^{-\lambda T}$$

$$f(n) = e^{\lambda T} \frac{(\lambda T)^n}{n!}$$

therefore $f(T_1, T_2, \ldots, T_n; n) = n!(\frac{1}{T})^n$

This distribution is the distribution of $n$ ordered uniform independent random variables on $(0, T)$. 
Ordered Observations:

\[ X_1, X_2, \ldots, X_n \text{ have density } f(x) \text{ and are independent.} \]
\[ Y_1, Y_2, \ldots, Y_n \text{ are the ordered } X\text{'s.} \]

Joint density of the \( Y \)'s is \( n! \int f(y_1) f(y_2) \ldots f(y_n) \) \(-\infty < y_1 \leq y_2 \leq \ldots \leq y_n < \infty\)

If the \( X \)'s have a continuous distribution then we may disregard any equalities between the \( X \)'s or the \( Y \)'s.

The marginal distribution of the \( Y_1 \) can be obtained by three methods, i.e.

1. Integration:

\[
= n! \int_{-\infty}^{\infty} \int_{-\infty}^{y_1} \int_{-\infty}^{y_i} \int_{-\infty}^{y_3} \int_{-\infty}^{y_2} \int_{-\infty}^{y_1} f(y_1) \text{ dy}_1 \text{ dy}_2 \text{ dy}_3 \text{ dy}_{i-1} \text{ dy}_i \text{ dy}_{i+1} \text{ dy}_n \cdots
\]

note: -- the observations above \( y_1 \) are constrained by the next lower observation, since this is an ordered sequence

-- the observations below \( y_1 \) are constrained from above

Integrating this we get

\[
g_1(y) = \text{ the marginal density of } Y_1
\]

\[
= \frac{n!}{(n-1)! (i-1)!} \left[ 1 - F(y_1) \right]^{n-1} \left[ F(y_i) \right]^{i-1} f(y_i)
\]

2. By differentiation:

\[
\Pr[Y_1 < y] = \Pr[\text{ i or more observations fall to the left of } y]
\]

\[
= \sum_{j=1}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}
\]

\[
g_1(y) = \frac{\partial}{\partial y} \sum \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}
\]
3. Heuristic Method (Wilks):

\[
\frac{\text{dy}}{Y_i} = \frac{n!}{(i-1)! \cdot (n-i)!} \left[F(y)\right]^{i-1} \left[1 - F(y)\right]^{n-i} f(y)
\]

(i-1) observations in the interval \((\infty, y)\)

(n-i) observations in the interval \((y, +\infty)\)

The probability of this is essentially a trinomial distribution, therefore

\[
g_i(y) = \frac{n!}{(i-1)! \cdot (n-i)!} \left[F(y)\right]^{i-1} \left[1 - F(y)\right]^{n-i} = \Pr[Y_i \text{ is in the interval } dy]
\]

Example: In particular, if we have the uniform distribution:

\[
f(y) = 1 \quad F(y) = y
\]

\[
g_i(y) = \frac{n!}{(n-1)! \cdot (i-1)!} y^{i-1} (1 - y)^{n-i} \quad \text{(which is a Beta distribution)}
\]

and

\[
E[Y] = \frac{i}{n+1}
\]

We could also define the spacings:

\[
S_i = Y_i - Y_{i-1} \quad i = 1, 2, \ldots, n+1
\]

\[
E[S_1] = \frac{1}{n+1} \quad \sum_{i=1}^{n+1} S_i = 1
\]

put \(Y_{n+1} = 1\) \quad \(Y_0 = 0\)

Problem 57: Show that each \(S_1\) has the same density.

a) find this density.

b) find i) \(E\left[\sum S_1^2\right]\)  
ii) \(E\left[\sum_{i=1}^{n+1} \left|S_i - \frac{1}{n+1}\right|\right]\)
Lemma: Probability Transformation:

If \( X \) has distribution \( F(x) \), \( F(X) \) has a uniform distribution on \((0,1)\).

Proof:

\[
\Pr[F(X) \leq u] = \Pr[X \leq F^{-1}(u)] = F[F^{-1}(u)] = u \quad 0 \leq u \leq 1
\]

One Sample Problem can be put in the following form by use of the probability transformation:

Given \( X_1, X_2, \ldots, X_n \); let \( X_1 < X_2 < \cdots < X_n \) be the \( X \)'s ordered increasingly and define \( U_i = F_0(X_i) \quad i = 1, 2, \ldots, n \)

we have a set of ordered observations \( U_1, U_2, \ldots, U_n \) on the interval \((0,1)\) which under

- \( H_0: \) has uniform distribution (with density \( n:\) )

  [or equivalently, that the original \( X \)'s had d.f. \( F_0(x) \); i.e., that the correct probability transformation was used]

and under \( H_1: \) has distribution \( F_1 \left[ F_0^{-1} \right] = G_0 \).

two-sided problem  one-sided problem

\[
\begin{array}{c}
\text{H}_0 \text{ true} \\
G
\end{array}
\quad
\begin{array}{c}
\text{H}_0 \text{ true} \\
G
\end{array}
\]

G\( u \leq u \) -- with the strict inequality for some interval

Some of the Tests for One Sample Problems

1. \( \bar{U} \) which is \( \text{As}\text{N}(\frac{1}{2}, \frac{1}{12n}) \)

   -- for the one-sided situation

   -- consistent

   -- reject \( H \) if \( \bar{U} \) is "too large", i.e., if \( \bar{U} > \frac{1}{2} + z_{1-\alpha} \left( \frac{1}{12n} \right) \)

   (for the above pictured situation)

   or if \( \bar{U} \) is "too small" in the converse situation, \([i.e., G(u) > u]\).
2. \( -2 \sum_{i=1}^{n} \ln U_i \) which is \( \chi^2 \) with \( 2n \) d.f. (ref. problem 6)

- \( \chi^2 \) is exact, not asymptotic
- consistent
- for the one-sided problem
- used in combination problems
- reject H if \( \chi^2 \) is "too large"

3. \( -2 \sum_{i=1}^{n} \ln (1 - U_i) \) also is \( \chi^2 \) with \( 2n \) d.f.

- for the one-sided problem
- Pearson's counter to Fisher's advocating No. 2
- consistent
- reject H if \( \chi^2 \) is "too small"

4. Distance Type Problem

Kolmogorov Statistic defined as follows

\[
\begin{align*}
D^+ &= \sup_{u \leq u_1} \left[ F_n(u) - u \right] \\
D^- &= \sup_{u \geq u_1} \left[ u - F_n(u) \right] \\
D &= \sup_{u \leq u_1} \left| F_n(u) - u \right| \\
\end{align*}
\]

for the one-sided problem

5. Related to the Kolmogorov statistic is

\[
\begin{align*}
R^+ &= \sup_{u \leq u_1} \frac{F_n(u) - u}{u} & \text{one-sided problem} \\
R &= \sup_{u \leq u_1} \frac{|F_n(u) - u|}{u} & \text{two-sided problem} \\
\end{align*}
\]

- not necessarily consistent
- \( a_n \) is arbitrary but positive
- derived by a Hungarian, Renyi
- not of very great merit
6. \( \omega_n^2 = \frac{1}{n} \int_0^1 \left[ F_n(u) - u \right]^2 \, du \)

--- two-sided or one-sided problem
--- sort of a continuous analogue of \( \chi^2 \)
--- consistent
--- due to Von Mises and Smirnov

7. \( \omega_n = \frac{1}{2} \sum_{i=1}^{n+1} |S_i - \frac{1}{n+1}| \quad S_i = u_i - u_{i-1} \)

--- ref: Sherman, Annals, 1950
--- one-sided or two-sided problem
--- consistent
--- is \( A_n \) \( \left( \frac{1}{4}, \frac{2e^{-5}}{10e^2} \right) \)

8. \( \beta = \sum_{i=1}^{n+1} S_i^2 \)

--- one-sided or two-sided problem
--- due to Moran
--- is consistent
--- \( \frac{n\beta}{2} = 1 \) is \( A_n(0, 1) \)

9. \( U_1 \) (Wilkinson Combination Procedure)
--- one-sided problem
--- generally not consistent

**Problem 58:**

a) Find the test based on \( U_1 \) for the set of alternatives \( G(u) > u \) with \( g(u) = G'(u) \)

b) Find the power of the test for the alternative \( G(u) = u^k \) \( 0 < k < 1 \)

c) Find the limiting power as \( n \to \infty \) (if the limit is 1, the test is consistent)

10. \( \chi^2 \)

11. Neyman's smooth tests
--- discovered by Neyman about 1937, but never generally used
--- ref: Neyman, Skandinavisk Aktuarietidskrift; 1937
PEARSON - Biometrika; 1938
David - Biometrika; 1938
Problem 59: \( X_1, X_2, \ldots, X_n \) are independent with d.f. \( F(x) \), density \( f(x) \)

\[ R = \max_{1 \leq i \leq n} X_i \]

a) Find the distribution of \( R \).

b) Suppose \( F(x) \) has density of the form \( \frac{1}{\sigma} f \left( \frac{x-\mu}{\sigma} \right) \)

where \( \int f(x) \, dx = 1 \)
\( \int x^2 f(x) \, dx = 1 \)

show that \( \mathbb{E}[R] = k(\sigma, n) \sigma \).

---

**Theorem 31:**

If \( F \) is continuous, \( F_n \xrightarrow{p} F \) for all \( x \).

**Proof:** We want to show that given \( \epsilon, \delta \) we can find \( N \) such that for \( n > N \)

\[ \Pr \left[ \left| F_n(x) - F(x) \right| < \epsilon \text{ for all } x \right] > 1 - \delta \]

Let \( A, B \) be such that

\[ F(A) < \frac{\epsilon}{2} \]
\[ 1 - F(B) < \frac{\epsilon}{2} \]

Pick out points \( X_{(1)}, X_{(2)}, \ldots, X_{(k-1)} \) in \( A, B \) such that \( F[X_{(1)}] = F[X_{(k-1)}] \)

which we can do because of uniform continuity.

Now set \( A = X_{(0)} \quad B = X_{(k)} \)

Consider a sample of size \( n \). Let \( n_i = f[X_j] \) that fall in the \( i \)th interval \( (X_{(i-1)}, X_{(i)}) \).

\( n_0, n_1, \ldots, n_{k+1} \) are multinomial with probabilities \( p_i = F[X_{(i)}] - F[X_{(i-1)}] \)

Now, \( \sum_{i=0}^{k+1} \frac{(n_i - np_i)^2}{np_i} \)

has a \( \chi^2 \) distribution with \( k+1 \) d.f.

We can find an \( M \) such that \( \Pr \left[ X_{k+1}^2 > M \right] < \delta \)
Since the above sum is less than $M$, each term and also its square root are certainly less than $M$, therefore:

$$
\Pr \left[ \frac{1}{\sqrt{n P_i}} \left| \frac{n_i}{n} - P_i \right| < M \text{ for each } i = 0, 1, \ldots, k+1 \right] > 1 - \delta
$$

which could be written

$$
\Pr \left[ \frac{1}{\sqrt{n P_i}} \left| \frac{n_i}{n} - P_i \right| < \frac{M}{\sqrt{n}} \right. \text{ for each } i = 0, 1, \ldots, k+1 \right] > 1 - \delta
$$

Recall that

$$
F_n[x(i)] = \frac{\# \{ x_j \leq x(i) \}}{n} = \sum_{j=0}^{i} \frac{n_j}{n}
$$

Now

$$
\left| F_n[x(i)] - F[x(i)] \right| = \left| \sum_{j=0}^{i} \frac{n_j}{n} - \sum_{j=0}^{i} P_j \right| = \left| \sum_{j=0}^{i} \left( \frac{n_j}{n} - p_j \right) \right| \leq \sum_{j=0}^{i} \left| \frac{n_j}{n} - p_j \right|
$$

Choose $n$ so large that

$$
\frac{(k+2)M/P_i}{\sqrt{n}} < \frac{\delta}{2}
$$

for all $i$.

Hence, if $n$ is chosen this large, with probability $1 - \delta$ the following relationship will hold:

$$
\left| F_n[x(i)] - F[x(i)] \right| < \frac{(i+1)M/P_i}{\sqrt{n}} < \frac{\delta}{2}
$$

Consider $x$ lying between $x(i-1)$ and $x(i)$

$$
-\varepsilon = -\frac{\delta}{2} - \frac{\varepsilon}{2} \quad \leq \quad F[x(i)] - F[x(1)] + F[x(i-1)] - F_n[x(i)] =
$$

$$
F[x(i-1)] - F_n[x(i)] \leq F(x) - F_n(x) \leq F[x(i)] - F_n[x(i-1)]
$$

$$
\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
$$

with probability $1 - \delta$.

---

*utilizing the following relationships:*

**making use of the following relationship:**
- 126 -

\[ -\frac{\varepsilon}{2} < F[x(1)] - F_n[x(1)] < \frac{\varepsilon}{2} \]

\[ F[x(1)] - F[x_{i-1}] \leq \frac{\varepsilon}{2} \]

\[ -\frac{\varepsilon}{2} < F[x_{i-1}] - F[x(1)] < 0 \]

Therefore the theorem holds.

---

Example: Referring to test No. 4 given on page 122.

\[ F_n \xrightarrow{p} F \]

If \( H \) is true, \( F_n \rightarrow \) uniform, thus for \( n \) sufficiently large \( |F_n(u) - u| < \varepsilon \).

If \( G \) is true, \( |F_n(u) - G(u)| < \varepsilon \) and thus \( |F_n(u) - u| > \varepsilon \).

Therefore, the test based on \( D = \sup |F_n(u) - u| \) will reject \( H \) with probability tending to 1

hence the test based on \( D \) is consistent.

Problem 59 (Addendum) (see p. 124):

\( c) \) Find the distribution of \( R \) for

1) \( X_1, X_2, \ldots, X_n \) uniform on \( (0, 1) \)

2) \( X_1, X_2, \ldots, X_n \) with exponential density \( f(x) = a e^{-ax} \)

Note: in the uniform case \( U_1 \) and \( R \) are dependent, in the exponential case they are independent "since the upper limit of the observations is \( \infty \)

---

Remark: The Kolmogorov statistics, \( D_n, D^+, \) and \( D^- \) are in fact invariant under the probability integral transform.

Proof: we have to show that

\[ \sup_{-\infty < x < \infty} |F_n(x) - F(x)| = \sup_{0 \leq u \leq 1} |F_n(u) - u| \]

\[ \sup_{0 \leq u \leq 1} |F_n(u) - u| = \sup_{0 \leq F(x) \leq 1} \left| F_n[F(x)] - F(x) \right| \]
Now \( F_n(x) = \frac{1}{n} \) \( Y_1 \leq x \leq Y_{i+1} \) where the \( Y \)'s are the ordered observations \( X_1, X_2, \ldots, X_n \)

\[
F_n(F(x)) = \frac{1}{n} \quad \text{where} \quad F(Y_i) \leq F(x) \leq F(Y_{i+1})
\]

Therefore:

\[
\sup_{0 \leq u \leq 1} | F_n(u) - u | = \sup_{-\infty < x < \infty} | F_n(x) - F(x) |
\]

**Distributions of \( D_n \):**

When \( H \) is true, and \( U_1, U_2, \ldots, U_n \) are uniform, then

\( a \) \( \lim_{n \to \infty} \Pr \left[ \sqrt{n} D_n^+ < z \right] = 1 - e^{-2z^2} \) (due to Smirnov)

- Finite distribution of \( D_n^+ \) given by Z. Birnbaum + Tingey in the Annals of Math. Stat.,
- Tabled by Miller in JASA, 1956, pp. 113-115

\( b \) \( L(z) = \lim_{n \to \infty} \Pr \left[ \sqrt{n} D_n^- < z \right] = 2 \sum_{m=1}^{\infty} (-1)^{m+1} e^{-2m^2 z^2} \) (due to Kolmogorov)


The simplest proof of both results is due to Doob in the Annals of Math. Stat., 1949

Some tables on the finite distribution of \( D_n \) are given by Massey in JASA, 1949, pp. 68-77.

Consider now that \( H \) is true, i.e., that \( U \) has d.f. \( G(u) \):

\( D_n^- \) test is to reject \( H \) if \( D_n^- > \varepsilon_n \) where

\[
\Pr \left[ D_n^- > \varepsilon_n \right] = \alpha
\]

\[
\Pr \left[ \sqrt{n} D_n^- > \sqrt{n} \varepsilon_n \right] = e^{-2\varepsilon_n^2} = \alpha
\]

thus \( \varepsilon_n = \sqrt{\frac{\ln \alpha}{2n}} \)

\[\Delta = \max \text{ difference between } u, G(u)\]

\[D_n^- = \sup_{0 \leq u \leq 1} \left[ u - F_n(u) \right]\]

Reject \( H \) if \( D_n^- > \varepsilon_n \)
Suppose \( H_1 \) is true so that \( F_n \) is the sample d.f. from \( G(u) \) with maximum \( |u - G(u)| = \Delta \) at \( u = u_0 \).

Then:
\[
\Pr[D_n^- > \varepsilon_n] \geq \Pr[u_0 - F_n(u_0) > \varepsilon_n]
\geq \Pr[F_n(u_0) - u_0 < -\varepsilon_n]
\geq \Pr[n F_n(u_0) < n(u_0 - \varepsilon_n)]
\]

But \( F_n(u_0) \) is a binomial random variable with expectation \( u_0 + \Delta \) and
\[
\text{Binomial probability} = \sum_{k=0}^\infty \binom{n(u_0 + \Delta)}{k} \left( \frac{u_0}{u_0 + \Delta} \right)^k \left( \frac{\Delta}{u_0 + \Delta} \right)^{n-k}.
\]

Problem 60: Find the bound on the power of the \( D_n^- \) test where \( G(u) = u^2 \).

Using the normal approximation given an explicit form for this bound in terms of
\[
\phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt
\]

Test No. 5: Renyi Statistic
\[
R^+ = \sup_{a \leq u \leq l} \frac{F_n(u)-u}{u}, \quad R = \sup_{a \leq u \leq l} \frac{|F_n(u)-u|}{u}
\]

Limiting distribution:
\[
\lim_{n \to \infty} \Pr \left[ \frac{1}{\sqrt{n}} R^+ < z \right] = \sqrt{\frac{2}{\pi}} \int_0^z e^{-t^2/2} dt
\]

For the distribution of \( R \) and further discussion see an article by Renyi in Acta Mathematica (Magyar), 1953.

Test No. 6: \( \omega_n^2 = \int_0^1 \left[ F_n(u) - u \right]^2 du = \int_0^\infty \left[ F_n(x) - F(x) \right]^2 dF(x)
\]

\[
= \int_0^1 \left[ F_n^2(u) - 2uF_n(u) + u^2 \right] du
\]
\[ F_n(u) = \frac{1}{n} \quad u_i \leq u \leq u_{i+1} \]

e.g., \( F_n(u) \) is flat for an interval

\[
\sum_{i=1}^{n} \left\{ \left[ \left( \frac{i}{n} \right)^2 - 2u \frac{i}{n} \right] du \right\} + \int_0^1 u^2 \ du
\]

recall: \( u_{n+1} = 1 \quad u_0 = 0 \)

\[
= \sum_{i=1}^{n+1} \left( \frac{i}{n} \right)^2 (u_{i+1} - u_i) - \frac{2}{n} \sum_{i=1}^{n} \frac{i}{n} (u_{i+1} - u_i) + \frac{1}{3}
\]

\[
= 1 + \frac{1}{n^2} \sum_{i=1}^{n} (1 - 2i)u_i - \frac{1}{n} \left( \sum_{i=1}^{n} u_i^2 \right) - 1 + \frac{1}{3}
\]

\[
\frac{\omega^2_n}{n} = \sum_{i=1}^{n} \frac{u_i^2}{n} = 2 \sum_{i=1}^{n} \frac{u_i}{n} + \sum_{i=1}^{n} \frac{u_i^2}{n^2} + \frac{1}{3}
\]

Cramer shows that:

\[
\frac{\omega^2_n}{n} = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left[ u_i - \frac{2i-1}{2n} \right]^2
\]

Tables of \( \lim_{n \to \infty} \text{Pr} \left[ n\omega^2 < z \right] \) have been given by T. W. Anderson and Darling in the Annals of Math. Stat., 1952, p. 206.

Problem 61: Find \( E \left[ \omega^2_n \right] \)

Approaches to Combining Probabilities:

Example: The following probabilities are from a one-sided t-test:

<table>
<thead>
<tr>
<th>P</th>
<th>F</th>
<th>P</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>.026</td>
<td>.1</td>
<td>.76</td>
<td>.6</td>
</tr>
<tr>
<td>.115</td>
<td>.2</td>
<td>.61</td>
<td>.7</td>
</tr>
<tr>
<td>.27</td>
<td>.3</td>
<td>.89</td>
<td>.8</td>
</tr>
<tr>
<td>.36</td>
<td>.4</td>
<td>.92</td>
<td>.9</td>
</tr>
<tr>
<td>.75</td>
<td>.5</td>
<td>.98</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Under the basic hypothesis the \( p \)'s are uniform on \((0, 1)\)

**Alternative tests:**

1. \( x^2_{(20)} = -2 \sum \ln U_i = 18.274 \quad \text{P} = 0.57 \quad \text{accept } H \)

2. To test alternatives of the form \( G(u) > u \) the Kolmogorov-Smirnov test statistic is \( D^+ = F_n - u \).

\[ \sup D^+ = 0.085 \]

From Miller's tables: \( \Pr \left[ D^+_{10} > 0.342 \right] = 0.15 \)

3. \( \bar{u} = 0.588 \)

\[ z = \frac{\bar{u} - 0.5}{\frac{1}{12n}} = \frac{0.588}{\sqrt{\frac{1}{120}}} = 0.964 \]

\[ \Pr \left[ z < 0.964 \right] = 0.67 \]

Since \( E[\bar{U}] \) under \( H_1 < E[\bar{U}] \) under \( H_0 \) we will reject \( H \) if \( z < z_{\alpha} \), therefore we cannot reject \( H_0 \).

4. Against two-sided alternatives

\( D = 0.35 \quad (0.75 - 0.40) \)

See Massey's tables for small sample size probabilities.

Using the large sample approximation:

\[ \Pr \left[ \sqrt{n} D > z \right] = 2 \sum_{m=1}^{\infty} (-1)^m e^{-2m^2z^2} = L(z) \]

\[ \Pr \left[ D > \frac{0.35}{\sqrt{10}} \right] = L(0.35) \]

**Example:**

In the following table:

- the \( X_i \) are taken from a table of \( N(0, 1) \) normal deviates.
- the \( X_{(i)} \) are the ordered \( X_i \)
- the \( U_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X_{(i)}} e^{-t^2/2} \, dt = \Pr \left[ z < X_{(i)} \right] \)
- 131 -

The $S_i$ are the spacings between the $U_i$

<table>
<thead>
<tr>
<th>$X^{(1)}$</th>
<th>$X^{(1)}$</th>
<th>$U_i$</th>
<th>$S_i$</th>
<th>$\frac{1}{n+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.42</td>
<td>-0.38</td>
<td>0.35</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>-0.02</td>
<td>-0.02</td>
<td>0.49</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>0.33</td>
<td>0.08</td>
<td>0.53</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>0.40</td>
<td>0.09</td>
<td>0.54</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>1.76</td>
<td>0.37</td>
<td>0.64</td>
<td>0.10</td>
<td>0.09</td>
</tr>
<tr>
<td>0.09</td>
<td>0.40</td>
<td>0.66</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>0.08</td>
<td>0.42</td>
<td>0.66</td>
<td>0.05</td>
<td>0.09</td>
</tr>
<tr>
<td>1.12</td>
<td>0.88</td>
<td>0.71</td>
<td>0.16</td>
<td>0.09</td>
</tr>
<tr>
<td>-0.38</td>
<td>1.12</td>
<td>0.87</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>0.37</td>
<td>1.76</td>
<td>0.96</td>
<td>0.04</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Under $H_0^{*}$ the $X$'s are $N(0, 1)$.

The test statistic is:

$$\omega_n = \frac{1}{2} \sum_{i=0}^{10} \left| S_i - \frac{1}{n+1} \right| = 0.38$$

Ignoring the slight negative correlation between the $S_i$, one could use a normal approximation with:

$$E[\omega_n] = \frac{1}{6} = 0.17 \quad V(\omega_n) = \frac{2e^{-5}}{10e^2} = (0.07)^2$$

Example:

If you want to test $H_1^{*}$ the $X$'s are $X(5)$, then you should use the probability transformation:

$$U = \frac{1}{2^{5/2}\sqrt{\pi}} \int_{0}^{\infty} e^{-t/2} \frac{5}{t^{5/2}} dt$$
Combination of Test Probabilities:

$H_0^2$: $U$ is uniform

$H_1^2$: $U$ has distribution $G(u) > u$ (i.e., the observations tend to be smaller)

ref: A. Birnbaum, JASA, September 1954

He considers a comparison of the following tests:

1- Fisher: $-2 \sum \ln U_i$

2- Pearson: $-2 \sum \ln (1 - U_i)$

--- found unsatisfactory for most applications.

3- $U_1$ -- not consistent, but is this important?

Birnbaum concluded that $-2 \sum \ln U_i$ was better for the one particular case of testing normal means.

Other possible tests: $D, \omega^2, \bar{U}$; have not been studied in this light.

Goodness of Fit Tests:

$H_0^2$: $F = F_0$ completely specified

--- The test usually involves estimation of parameters.

--- The only completely worked out theory is for the $\chi^2$-test.

For other suggestions see:

--- F. David, Biometrika, 1938-9


They present a Monte Carlo derived distribution of $\omega^2$ and $D$ for the case of testing $H_0^2$ X's are $N(\mu, \sigma^2)$ where $\mu^2_n$ and $\sigma^2_n$ are estimated by $\bar{x}$, $s^2$ working with $n=25$, $n=100$.


--- The $D^+_n$ test is a "minimax" test (among Fisher, Pearson, $D^+_n$, $\omega^2_n$, $U$) of the one-sided hypothesis $H_0^2$: $F = F_0$ versus $H_1^2$: $F = F_1 < F_0$.

i.e., "minimax" in the sense that it has minimum power to pick up easy-to-detect alternatives, maximum power to pick up hard-to-detect alternatives.
IV. Two Sample Problems:

\[ X_1, X_2, \ldots, X_m \text{ have d.f. } F(x) \]
\[ Y_1, Y_2, \ldots, Y_n \text{ have d.f. } G(y) \]
\[ H_0: F = G \quad H_1: F < G \text{ or } F > G \]
\[ H_1': F \neq G \]

In the parametric case one could use the normal approximation and the two-sample t-test on the means, but these hypotheses are somewhat wider.

Partial list of tests:

1. Median Test
2. Runs Test
3. Wilcoxon's Test (also called the Mann-Whitney test)
4. Kolmogorov-Smirnov D-test
5. Ven der Waerden's X-test (or Terry's C-test)

All we need for these tests is to be able to order the observations; magnitudes are not important; e.g.,

\[
\begin{array}{ccccccc}
  & X & X & X & X & Y & Y & Y \\
\end{array}
\]

Test 1: For the median test set up a 2x2 table classifying the observations as above or below the median of the combined sample. For example:

<table>
<thead>
<tr>
<th></th>
<th>Below</th>
<th>Above</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X )</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>( Y )</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

| Total | \( \frac{m+n}{2} \) | \( \frac{m+n}{2} \) | \( m+n \) |

Test Statistics is the usual \( \chi^2 \) for 2x2 contingency tables with one d.f.

Test 2: For the runs test, set

\[ r = \text{number of runs of } X \text{'s and of } Y \text{'s in the combined sample (in our example } r = 4) \]

\( H_0 \) is rejected if \( r < r_0 \).

If \( H_0 \) is true, then the \( X \)’s and \( Y \)’s are intermingled and the value of \( r \) will be "large"; if \( H_1 \) is true, then \( r \) will be "small".
Test 3: For the Wilcoxon (Mann-Whitney) test, we define:

\[ u_{ij} = \begin{cases} 1 & \text{if } X_i < Y_j \\ 0 & \text{otherwise} \end{cases} \]

\[ U = \sum_{i=1}^{m} \sum_{j=1}^{n} u_{ij} \text{ in our example } U = 22 \text{ (4+4+4+5+5)} \]

Wilcoxon's original test was based on \( R_i^x \)

where \( R_i^x \) = the sum of the ranks of \( X_i \) in the combined sample ordering from the smallest.

Similarly \( R_j^y \) = the sum of the ranks of \( Y_j \).

Problem 62: Prove: \( U = mn + \frac{m(m+1)}{2} - \sum_{i=1}^{m} R_i^x - \sum_{j=1}^{n} R_j^y - \frac{n(n+1)}{2} \)

Test 4: Kolmogorov-Smirnov define \( D \) for the two-sample problem as:

\[ D_{mn} = \sup \left| F_m(x) - G_n(x) \right| \]

\[ -\infty < x < \infty \]

Test 5: For Van der Waerden's \( X \)-test we

let \( \phi (u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt \)

and define: \( \Upsilon(u) = \phi^{-1}(u) \)

then the test statistic is: \( X = \sum_{i=1}^{m} \Upsilon \left( \frac{R_i^x}{m + n - 1} \right) \)

Problem 63: Let \( Q \) = number of \( Y \)'s which exceed \( \max(X_1, X_2, \ldots, X_m) \).

a) Find the distribution of \( Q \) in the general case.

b) Specialize the result in (a) when \( H_0 \) is true.

c) Find the limiting distribution of \( Q \) for case (b) as \( m \to \infty, n \to \infty, \frac{m}{n} \to \lambda \).
Runs Test (Test No. 2):

Ref.: Mood Chap. 16

\( r \) = no. of runs of X's or Y's in the ordered combined sample.

There are two cases to be considered— that of an even and that of an odd number of runs.

**Even case:** The number of arrangements of \( m \) X's and of \( n \) Y's that have the property of giving rise to \( 2k \) runs can be found by a simple generating function device.

Again, there are two possibilities; i.e., starting with a run of X's or of Y's, so starting with either, we must multiply the end result by 2 since the two starts are symmetric.

Starting with the X's, they are divided into \( k \) groups, all non-zero. To find the number of ways of doing this, consider the coefficient of \( t^m \) in the expansion of

\[
(t + t^2 + t^3 + \cdots)^k = \left( \frac{t}{1-t} \right)^k = t^k (1 - t)^{-k}
\]

\[
= t^k \left( 1 + kt + \frac{(-k)(-k-1)}{2} t^2 + \cdots \right)
\]

\[
= t^k \sum_{j=0}^{\infty} \binom{k+j-1}{k-1} t^j \frac{(k+j-1)!}{j! (k-1)!}
\]

The coefficient of \( t^m \) is found when \( j = m - k \), and

\[
\frac{(k + m - k - 1)!}{(m - k)! (k - 1)!} = \binom{m - 1}{k - 1}
\]

Similarly the number of arrangements of the \( n \) Y's into \( k \) non-zero groups is

\[
\binom{n - 1}{k - 1}
\]

Therefore, the number of arrangements of X's and Y's with 2\( k \) runs is

\[
2 \binom{m - 1}{k - 1} \binom{n - 1}{k - 1}
\]

The total number of arrangements possible with \( m \) X's and \( n \) Y's is \( \frac{(m + n)!}{m! n!} \).
Therefore:

$$\Pr [r = 2k] = \frac{2^m}{(m+n)} \binom{m-1}{k-1} \binom{n-1}{k-1}$$

**Odd Case:** \( (r = 2k + 1) \) For the case when the number of runs is odd, i.e., to determine \( \Pr [r = 2k + 1] \), the argument is similar, but we start and end with either X or Y (instead of starting with one end and ending with the other), therefore

$$\Pr [r = 2k + 1] = \frac{(m-1)(n-1)}{(m+n)} \binom{m-1}{k-1} \binom{n-1}{k}$$

$$E(r) = \frac{2mn}{m+n} + 1 \approx 2 N \alpha \beta$$

$$V(r) = \frac{2mn(2mn - m - n)}{(m+n)(m+n-1)} \approx 4 N \alpha^2 \beta^2$$

Where: \( N = m+n \), \( m = N\alpha \), \( n = N\beta \), \( \alpha + \beta = 1 \)

By Sterling's approximation methods, we can show that

$$\lim_{N \to \infty} \frac{r - 2N\alpha\beta}{2\sqrt{N}} \text{ is } \mathcal{N}(0,1)$$

**Test:** \( H_0 \) is rejected if \( r < r_0 \)

\( r_0 \) can be determined from the left-hand tail of the normal approximation or from tables in

2. Siegel, Table F
3. Dixon + Massey, Table 11

**Wilcoxon (Mann-Whitney) Test (Test No. 3):**

$$U = \sum_i \sum_j u_{ij} \quad \text{where} \quad u_{ij} = 1 \text{ if } X_i < Y_j$$

$$= 0 \text{ otherwise}$$

If \( H \) is true, \( E(U) = \sum_i \sum_j E(u_{ij}) = \frac{mn}{2} \)
For the general case, assume \( \Pr[Y > X] = p \)

\[
= \sum_{x=0}^{\infty} \Pr[Y > x | X = x] \ dF(x)
\]

\[
= \sum_{x=0}^{\infty} \left[1 - G(x)\right] \ dF(x)
\]

\[
= E \left[ \{1 - G(x)\} \ | \ F(x) \right]
\]

In this general case, \( E(U) = mn \ p \)

\[
E(U^2) = \sum_i \sum_j E(u_{ij}^2)
\]

\[
+ \sum_i \sum_{j 
eq b} E(u_{ij}, u_{ib})
\]

\[
+ \sum_{i 
eq a} \sum_j E(u_{ij}, u_{aj})
\]

\[
+ \sum_{i 
eq a} \sum_{j 
eq b} E(u_{ij}, u_{ab})
\]  

To evaluate this, we can examine each part separately, e.g.:

1. \( E(u_{ij}^2) = p \) \( \text{mn such terms} \)

2. \( E(u_{ij}, u_{ib}) = E(u_{ij}) E(u_{ib}) = p^2 \) \( \text{m(m-1)n(n-1) such terms} \)

3. \( E(u_{ij}, u_{ib}) = \Pr \left[ Y_j > Y_i, X_i > x \right] \ dF(x) \)

\[
= \sum_{x=0}^{\infty} \Pr \left[ Y_j > x, Y_b > x \right] \ dF(x)
\]

\[
= \sum_{x=0}^{\infty} \left[1 - G(x)\right]^2 \ dF(x)
\]

\[
= E \left[ \{1 - G(x)\}^2 \right] \ F
\]  \( \text{mn(n-1) such terms} \)

4. by similar argument \( E(u_{ij}, u_{aj}) = \sum_{y=0}^{\infty} y^2 (y) \ dG(y) \)

\[
= E \left[y^2(y) \right] \ G
\]  \( \text{mn(m-1) such terms} \)
Thus:

\[ V(U) = mn + mn(m-1)(n-1) p^2 + mn(n-1) E \left[ (1-G)^2 \bigg| F \right] + mn(m-1) E \left[ F^2 \bigg| G \right] - m^2 n^2 p^2 \]

**Exercise:** If \( H \) is true, verify that \( E(F^2) = E(1-G)^2 = \frac{1}{3} \)

and then \( \text{Var}(U) = \frac{mn}{12} (m + n + 1) \)

Mann-Whitney in their further work on Wilcoxon's test proved that

\[ \frac{U - \frac{mn}{2}}{\sqrt{\frac{mn}{12} (m+n+1)}} \]

is \( \text{AN}(0,1) \). This they found by discovering a recursion formula to get all the moments of the distribution of \( U \), then observing that the limits of the moments, as \( m \to \infty, n \to \infty \) were the moments of the normal distribution.

\( U \) -- may be used for one-sided or two-sided tests.

-- the most complete tables are given by Hodges + Fix, *Annals of Math. Stat.*, 1955.

**Problem 64:**

Take 10 observations of (a) \( X \) which is \( \text{N}(0,1) \) and

(b) \( Y \) which is \( \text{N}(1,1) \)

Apply each of the five tests to the data to obtain tests for

\( H_0: F = G \) against \( H_1: F \neq G \).

**Problem 65:**

Let \( a_1, a_2, \ldots, a_{k-1} \) be fixed points.

Let \( X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n \) be independent observations from \( F(x), G(y) \).

Define: \( z_i = F_m(a_i) - G_n(a_i) \)

\[ Z = \sum_{i=1}^{k-1} z_i \]

(a) Find \( E(Z) \), \( \text{Var}(Z) \) in general and for the case \( F = G \).

denote: \( F(a_i) = f_i \quad G(a_i) = g_i \)
(b) Find $E(Z)$, $Var(Z)$ for the case when

\[ F(u) = u \quad G(u) = 0 \quad 0 \leq u < b \quad 0 \leq b \leq \frac{1}{2} \]

\[ = 1 + \frac{u - \frac{1}{2} + b}{1 - 2b} \quad b \leq u \leq 1 - b \]

\[ = 1 \quad 1 - b < u \leq 1 \]

In this problem let $a_i = \frac{1}{k}$

(c) Assuming $Z$ is AN, is it consistent for alternatives of the form of $G$ in (b).

\[ \text{Kolmogorov-Smirnov Test (Test No. 4):} \]

Our basic sample $(X X X X Y Y X Y Y)$ could be expressed graphically in two ways:

(a) \[ D_{mn} = .8 \]

(b) \[ Y \]

\[ \text{i.e., the sample could be plotted as a two-dimensional random walk reaching the point (5,5)--reject if the walk strays beyond a line parallel to the 45^\circ \text{ line}.} \]

Asymptotic distributions of $D_{mn}$, if $H$ is true:

\[ \lim_{m,n \to \infty} \Pr \left[ \sqrt{\frac{mn}{m+n}} D_{mn} \leq z \right] = L(z) = 2 \sum_{i=1}^{\infty} (-1)^{i+1} e^{-2i^2 z^2} \]

$L(z)$ has been tabled in the Annals of Math. Stat., 1948. Also, $D_{mn}$ have the same limiting distribution as $D_n^+$ (one-sample statistic) with the normalization $\sqrt{\frac{mn}{m+n}}$ (see p. 127)
Test: Reject $H$ if $D_{mn} > d$.

Van der Waerden's $X$-test (Test No. 5):

$$X = \sum_{i=1}^{m} \psi \left( \frac{R_i^X}{N+1} \right) \quad \text{where} \quad \psi'(u) = \phi^{-1}(u)$$

$$\phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} \, dt$$

$X$ is $\chi^2(N, \frac{mn}{N-1}, Q)$ where $Q = \frac{1}{N} \sum_{i=1}^{N} \psi^2 \left( \frac{R_i}{N+1} \right)$

$N = m + n$

Example: XXXXXXYYYYY

$$R_i^X = 1 \quad 2 \quad 3 \quad 4 \quad 8$$

$$R_i = 1 \quad 2 \quad 3 \quad 4 \quad 8$$

$$\frac{N}{N+1} = 11 \quad 11 \quad 11 \quad 11 \quad 11$$

$= .09, .18, .27, .36, .73$

using normal deviate tables

$$\psi \left( \frac{R_i}{N+1} \right) = .09, .18, .27, .36, .73$$

$$= -1.34, -.91, -.60, -.35, +.60$$

$$X = \sum_{i} \psi \left( \frac{R_i^X}{N+1} \right) = -2.60$$

For determining the variance of $X$, tables of $Q$ have been given by Van der Waerden in an appendix to his text.

Theorem 32 (Pitman's Theorem on A.R.E.):

Assume: $T_n^1, T_n^2$ are A. N. Statistics

$\Omega^1$ is a subset of $\Omega$ indexed by $\rho$ such that when $H$ is true $\rho = 0$.

Let the sequence of $\rho$'s tends to $\theta$, i.e., $\rho_1 \rho_2 \ldots \rho_n \to 0$. 

\textbf{[-End-]}
Test: \[ T \rightarrow \text{reject } H \text{ if } T_n > t_{n-\alpha} \]
\[ T \rightarrow \text{reject } H \text{ if } T^* > t^*_{n-\alpha} \]

Assumptions:

1. \( \frac{d}{d\varrho} \frac{E_{\varrho} (T_n)}{\sigma_0} > 0 \)

2. \( \lim_{\varrho \to 0} \frac{d}{d\varrho} \frac{E_{\varrho} (T_n)}{\sigma_0} = c \quad \text{if } \varrho_n = \frac{k}{\sqrt{n}} \)

3. \( \lim_{n \to \infty} \frac{d}{d\varrho} \frac{E_{\varrho} (T_n)}{\sigma_0} = 1 \)

4. \( \lim_{n \to \infty} \frac{\sigma_n (T_n)}{\sigma_0 (T_n)} = 1 \)

Theorem: Under these regularity conditions, the limiting power of the \( T_n \) test for alternatives \( \varrho_n = \frac{k}{\sqrt{n}} \) as \( n \to \infty \) is \( 1 - \Phi (z_{1-\alpha} - k\alpha) \).

The A.R.E. of \( T_n^* \) to \( T^* \) is

\[
= (\frac{c}{c^*})^2 = \lim_{n \to \infty} \left[ \frac{d}{d\varrho} \frac{E_{\varrho} (T_n)}{\sigma_0} \bigg| \varrho = 0 \right]^2 \frac{\sigma_n^2 (T^*_n)}{\sigma_0^2 (T_n)}
\]
Proof:

\[ \Pr \left[ \frac{T_n - E_0(T_n)}{\sigma_0} > \frac{t_n - E_0(T_n)}{\sigma_0} \right] = \alpha \]

For large \( n \), since \( T \) is asymptotically normal, \( t_n \sim N(0, \sigma_0^2) \)

\[ \lim_{n \to \infty} \Pr \left[ \text{Reject } H \mid \rho_n \right] = \lim_{n \to \infty} \Pr \left[ \frac{T_n - E_0(T_n)}{\rho_n} > \frac{z \sigma_0}{\sigma_0} + \frac{E_0(T_n) - E_0(T_n)}{\rho_n} \right] \]

\[ = \lim_{n \to \infty} \Pr \left[ z > \frac{\sigma_0}{\rho_n} + \frac{E_0(T_n) - E_0(T_n)}{\rho_n} \sigma_0 \right] \]

\[ \lim_{n \to \infty} z \left( \frac{\sigma_0}{\rho_n} + \frac{E_0(T_n) - E_0(T_n)}{\rho_n} \sigma_0 \right) = z - kc \]

from assumption 4, as \( n \to \infty, \frac{\sigma_0}{\rho_n} \to 1 \)

using a Taylor series expansion:

\[ E_{\rho_n}(T_n) = E_0(T_n) + \rho_n \left( \frac{d}{d\rho} E_\rho(T_n) \right) \rho_n \]

Putting \( \rho_n = \frac{k}{\sqrt{n}} \)

\[ \lim_{n \to \infty} \left[ \frac{E_0(T_n) - E_{\rho_n}(T_n)}{\sigma_0} \right] = \lim_{n \to \infty} \left[ - \frac{k}{\sqrt{n}} \frac{d}{d\rho} E_\rho(T_n) \right] \rho = \rho_n \]

\[ = - kc \]

Therefore:

\[ \lim_{n \to \infty} \Pr \left[ \text{Reject } H \mid \rho_n \right] = 1 - \phi (z - kc) \]

By a similar argument, the limiting power of the \( T^* \) test is

\[ 1 - \phi (z - k^* c^*) \]

where \( c^* = \lim_{n \to \infty} \frac{d}{d\rho} E_{\rho_n}^{(T^*)} \bigg|_{\rho = 0} \sqrt{n} \sigma_0 \).
We want to determine sequences \( \{ n_1 \} \), \( \{ n_1^* \} \) such that \( 1 - \emptyset (Z_{1-x} - kc) =\)
\( 1 - \emptyset (Z_{1-x} - k^*c^*) \) which means that \( kc = k^*c^* \). Also, for the two sequences to be the same \( \rho_n = \rho_n^* \) or \( \frac{k}{\sqrt{n_1}} = \frac{k^*}{\sqrt{n_1^*}} \).

Thus we can determine the equality of the following ratios:

\[
\frac{n_1^*}{n_1} = \left( \frac{k^*}{k} \right)^2 = \left( \frac{c^*}{c} \right)^2 = \lim_{n \to \infty} \left( \frac{n}{n_0} \frac{c^2}{c^2_0} \right) \left( \frac{d}{dp} \frac{Ep (T_n^*)}{|\rho=0^*} \right)^2
\]

The A.R.E. of \( T \) to \( T^* \) is given by any of these ratios, or as stated in the theorem:

\[
A.R.E. = \left( \frac{c^*}{c} \right)^2 = \lim_{n \to \infty} \left( \frac{c^2}{c^2_0} \right) \left( \frac{d}{dp} \frac{Ep (T_n^*)}{|\rho=0^*} \right)^2
\]

Example: Obtaining the A.R.E. for the Wilcoxon test versus the normal mean test, i.e.

\[ U \text{ versus } Z = \frac{\bar{Y} - \bar{X}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \] which is asymptotically equivalent to the two-sample t-test.

\( X \) and \( Y \) have d.f. \( F(x) \) under \( H_0 \).
\( X \) has d.f. \( F(x) \), \( Y \) has d.f. \( F(x - \mu) \) under \( H_1 \).
In both cases the variance is \( \sigma^2 \).

For Wilcoxon's test:

\[
p = \Pr [\bar{Y} > \bar{X}] = \int_0^\infty [1 - G(x)] dF(x) = \int_{-\infty}^\infty [1 - F(x - \mu)] f(x) \, dx
\]

\[
\frac{dp}{d\mu} \bigg|_{\mu=0} = \int_{-\infty}^\infty f(x - \mu) f(x) \, dx = \int_{-\infty}^\infty f^2(x) \, dx
\]
\[-144-\]

\[ E(U) = mn \quad \frac{d}{d\mu} E(u) \bigg|_{\mu=0} = mn \int_{-\infty}^{\infty} f^2(x) \, dx \]

\[ \text{Var} (U) = \frac{mn(m + n + 1)}{12} \]

\[ E(Z) = \frac{\int x \, dF(x - \mu) - \int x \, dF(x)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{\mu}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \]

\[ \frac{dE(Z)}{d\mu} = \frac{1}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} = \frac{1}{\sigma \sqrt{\frac{m+n}{mn}}} \quad \text{Var} (Z) = 1 \]

Therefore, the A.R.E. of U to Z is

\[ \text{A.R.E.} = \lim_{n \to \infty} \left( \frac{1}{\frac{mn(m + n + 1)}{12}} \right)^2 \left( \frac{\int_{-\infty}^{\infty} f^2(x) \, dx}{\frac{\int_{-\infty}^{\infty} f^2(x) \, dx}{\sigma \sqrt{\frac{m+n}{mn}}}^2} \right) \]

which reduces to:

\[ \text{A.R.E. (U to Z)} = 12 \sigma^2 \left[ \int_{-\infty}^{\infty} f^2(x) \, dx \right]^2 \]

Thus we can compare U and Z for any f(x) whatsoever.

For instance, if \( f(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-x^2/2\sigma^2} \) then \( \int f^2(x) \, dx = \frac{1}{2\pi\sigma^2} \int e^{-x^2/2\sigma^2} \, dx \)

using the transformation \( \frac{x}{\sigma} = \frac{y}{\sqrt{2}} \)

\[ \int f^2(x) \, dx = \frac{1}{\sqrt{2\pi\sigma}} \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} e^{-y^2/2} \, dy \]

\[ = \frac{1}{2\sigma\sqrt{\pi}} \]

and \( 12 \sigma^2 \left[ \int_{-\infty}^{\infty} f^2(x) \, dx \right]^2 = 12 \sigma^2 \left( \frac{1}{2\sigma\sqrt{\pi}} \right)^2 \)

\[ = \frac{12 \sigma^2}{4\pi \sigma^2} = \frac{3}{\pi} = .955 \)
All of which says that the \( \text{A.R.E.} \) of the Wilcoxon test to the Normal (two-sample) test, if the underlying populations are normal, and we are testing "slippage of the mean" is \( \frac{3}{n} \).

**Problem 66:**

(a) Evaluate the \( \text{A.R.E.} \) of \( U \) and \( Z \) when

\[
(1) \ f(x) = 1 \quad 0 \leq x \leq 1, \ i.e. \\
F(x) = x \quad 0 \leq x \leq 1 \\
F(x - \mu) = x - \mu \quad \mu < x < 1 + \mu \\
(2) \ f(x) = e^{-x} \quad 0 \leq x < \infty
\]

(b) Find an \( f(x) \) such that the \( \text{A.R.E.} \) of \( U \) to \( Z \) is \( + \infty \).

**Remark:** It has been shown that the \( \text{A.R.E.} \) of \( U \) to \( Z \) in this case, i.e. testing slippage, is always \( \geq 0.864 \). - Hodges and Lehmen, Annals of Math. Stat., 1955.

<table>
<thead>
<tr>
<th>Test</th>
<th>( \text{A.R.E.} ) of test to ( Z ) (testing for slippage)</th>
<th>Consistency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Median</td>
<td>2/n</td>
<td>yes, if the median of ( F_0 \neq ) median of ( F_1 )</td>
</tr>
<tr>
<td>2. Runs</td>
<td>0</td>
<td>consistent for all ( F_0 = F_1 )</td>
</tr>
<tr>
<td>3. U</td>
<td>3/n</td>
<td>yes, if the median of ( F_0 \neq ) median of ( F_1 )</td>
</tr>
<tr>
<td>4. K-S</td>
<td>???</td>
<td>consistent for all ( F_0 = F_1 )</td>
</tr>
<tr>
<td>5. X</td>
<td>1</td>
<td>yes, if the median of ( F_0 \neq ) median of ( F_1 )</td>
</tr>
</tbody>
</table>

Robustness of a test (as propounded by Box) refers to the behavior of a test when the various assumptions made for the validity of the test are not fulfilled.

Type 1 Error - \( \Pr [ \text{reject } H \text{ when true under the assumptions} ] \)

Power - \( \Pr [ \text{reject } H \text{ when false under the assumptions} ] \)

A test is said to be robust if \( \Pr [ \text{reject } H \text{ when true if assumptions are not fulfilled} ] \) remains close to \( \alpha \) regardless of the assumptions.

**Note:** the \( Z \)-test, or two-tailed t-test, is robust.

The proponents of distribution-free statistics argue that the disadvantage of the \( Z \)-test is that the power may slip if the assumptions (of normalcy, etc) are not satisfied.
V. k-Sample Tests:

1. Median Tests
2. \( x^2 \)-Test with arbitrary groupings
3. Kruskal-Wallis Rank Order Test
4. Kolmogorov-Smirnov Type Tests

samples:

\[
\begin{align*}
1. & \quad X_{11} \quad X_{12} \quad \cdots \quad X_{1n_1} \\
2. & \quad X_{21} \quad X_{22} \quad \cdots \quad X_{2n_2} \\
\vdots \\
k. & \quad X_{k1} \quad X_{k2} \quad \cdots \quad X_{kn_k}
\end{align*}
\]

\[N = \sum_{j=1}^{k} n_j\]

1. Median Test is made by setting up a 2xk table:

<table>
<thead>
<tr>
<th>sample</th>
<th>1</th>
<th>2</th>
<th>\cdots</th>
<th>i</th>
<th>\cdots</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above</td>
<td></td>
<td></td>
<td></td>
<td>( m_i )</td>
<td>\cdots</td>
<td>N/2</td>
</tr>
<tr>
<td>Below</td>
<td></td>
<td></td>
<td></td>
<td>( n_i - m_i )</td>
<td>N/2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( n_1 )</td>
<td>( n_2 )</td>
<td>\cdots</td>
<td>( n_i )</td>
<td>\cdots</td>
<td>( n_k )</td>
</tr>
</tbody>
</table>

where \( m_i \) = the number of observations in sample \( i \) above the median of the combined sample.

Use a \( x^2 \)-test of the null hypothesis with the expected values = \( n_i/2 \) when \( N \)
is even, and \( x^2 = \sum_{i=1}^{k} \frac{2(n_i - n_i/2)^2}{n_i/2} \) with \( k-1 \) d.f. under the null hypothesis

that the \( k \) samples all came from the same distribution.

2. \( x^2 \)-Test:

Being given or arbitrarily choosing groups \( A_j \) (\( a_j \leq X \leq a_{j+1} \)) define \( n_{ij} \) = the number of \( X \)'s in sample \( i \) that fall in \( A_j \).

Under the null hypothesis \( \Pr[X \text{ falls in } A_j] = p_j \) independently of \( i \).

This can be tested by \( x^2 \) in the usual manner -- as an \( r \times k \) test of homogeneity, where \( x^2 \) has \((r-1)(k-1)\) d.f.
3. Kruskal-Wallis Test:

\[ H = \frac{12}{N(N+1)} \sum_{i=1}^{k} n_i \left( \overline{R_i} - \frac{N+1}{2} \right)^2 = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \]

where \( R_i \) = sum of the ranks of sample \( i \) taken within the combined sample
\( \overline{R_i} = R_i/n_i \)

\( H \) is asymptotically distributed as \( \chi^2 \) with \( k-1 \) d.f.

ref: March 1959 JASA for small sample approximations.

Problem 67: What does \( H \) reduce to when \( k = 2 \)?
Prove your answer.

4. K-S Type Tests:

Would involve drawing a step-function for each sample on the same graph. Unfortunately nothing is now known about the distributions.


Consistency:

The Median and \( H \) are consistent against all alternatives if at least one of the sub-group medians differs from the others.

A.R.E.:

A.R.E. for slippage alternatives:

- median test against the standard ANOVA test \( 2/\pi \)
- \( H \) test against the standard ANOVA test \( 3/\pi \)

where the underlying distribution is normal.

If the underlying distribution is rectangular, then the A.R.E.'s become:

- median against ANOVA \( 1/3 \)
- \( H \) against ANOVA \( 1 \)
CHAPTER VI
TESTING OF HYPOTHESES -- PARAMETRIC THEORY -- POWER

Refs: Gramer, ch. 35
Kendall, vol. 2, chs. 26-27
Fraser, "Nonparametric Methods in Statistics", ch. 5

1. Generalities

\[ X_1, X_2, \ldots, X_n \text{ have } d_x F(x, \theta) \]

-- usually the X's are independent with density \( f(x, \theta) \), the density having a specified parametric form with one or more unknown parameters.

For the parameter space \( \bigcap \mathcal{H}_0: \theta \in \omega_0 \quad \mathcal{H}_1: \theta \in \omega_1 \)

Recall that \( \phi(x) \) is a test function of size \( \alpha \) such that

\[ \phi(x) = 1 \quad \text{reject } H \text{ with probability 1} \]
\[ = k \quad \text{reject } H \text{ with probability } k \]
\[ = 0 \quad \text{do not reject } H \text{ (accept } H) \]

where \( E \left[ \phi \mid \theta \in \omega_0 \right] \leq \alpha \)

Power function: \( \beta_{\phi}(\theta) = E \left[ \phi \mid \theta \right] \)

Ref\textsuperscript{s} Defs. 33-38 in chapter 5.

Def\textsuperscript{l}\textsubscript{l}: \( \phi^* \) is a uniformly most powerful (u.m.p.) test of size \( \alpha \), if \( \phi \) being an other size \( \alpha \) test

\[ \beta_{\phi^*}(\theta) \geq \beta_{\phi}(\theta) \text{ for all } \theta \in \omega_1 \]

2. Probability Ratio Tests

Neyman-Pearson Theorem:

X is a continuous random variable with density \( f(x, \theta) \).

\[ H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1 \]

(simple hypothesis) (simple alternative)

Assume: \( f(x, \theta_0) > 0; f(x, \theta_1) > 0 \) for the same set S.

Assume: \( f(x, \theta_1) \) is a continuous random variable.
Theorem 33 (Neyman-Pearson):

The most powerful test of $H_0$ against $H_1$ is given by $\phi^*(x)$ defined as follows:

$$\phi^*(x) = \begin{cases} 1 & \text{when } f(x, \Theta_1) > k f(x, \Theta_0) \text{ or } f_1 > k f_0 \\ 0 & \text{elsewhere} \end{cases}$$

where $k$ can be chosen so that $\phi^*$ is of size $\alpha$.

If the test $\phi^*$ is independent of $\Theta_1$ for $\Theta_1 \in \Omega_1$ then $\phi^*$ is the u.m.p. test of $H_0$ against $H_1$: $\Theta_1 \in \Omega_1$.

Remark: This is what has been called the probability ratio test, since Neyman-Pearson originally expressed the theorem that

$$\phi^* = 1 \quad \text{if } \frac{f_1}{f_0} > k$$

Proof:

1. To show there is a required $k$ that makes $\phi^*$ of size $\alpha$.

define: $\alpha(k) = \Pr \left[ \frac{f_1(x)}{f_0(x)} > k \mid X \text{ has density } f_0 \right]$

$\alpha(0) = 1 \quad \alpha(\infty) = 0$

since $\frac{f_1}{f_0}$ is a continuous random variable, $1 - \alpha(k)$, which is the d.f. of this random variable, is a continuous function and is monotone non-decreasing, hence for some $k'$ we must have $\alpha(k') = \alpha$. 
2. To show that \( \phi^* \) is u.m.p.

Let \( \phi \) be any other test of size \( \alpha \).

We want to show that:

\[
\int \phi^*(x) f(x, \theta_1) \, dx \geq \int \phi(x) f(x, \theta_1) \, dx
\]

Consider:

\[
\left[ \phi^*(x) - \phi(x) \right] \left[ f(x, \theta_1) - k \cdot f(x, \theta_0) \right] \, dx \geq 0
\]

That this integral is \( \geq 0 \) follows since

when \( f_1 - k f_0 > 0 \), then \( \phi^* = 1 \), and \( \phi^* - \phi > 0 \)

when \( f = k f \leq 0 \), then \( \phi^* = 0 \), and \( \phi^* - \phi \leq 0 \)

Expanding this integral we get:

\[
\int \phi^* f_1 \, dx = \int \phi f_1 \, dx - k \left[ \int \phi^* f_0 \, dx - \int \phi f_0 \, dx \right] \geq 0
\]

but \( \int \phi^* f_0 \, dx = \int \phi f_0 \, dx = \alpha = \text{the size condition} \)

Therefore:

\[
\int \phi^* f_1 \, dx \geq \int \phi f_1 \, dx
\]

and \( \phi^* \) is u.m.p.

**Example 18** \( X_1, X_2, \ldots, X_n \) are \( \text{N}(\mu, \sigma^2) \) \( \sigma^2 \) known

\[
H_0 : \mu = 0 \quad H_1 : \mu = \mu_1 > 0
\]

\[
q.e.d.
\]

\[
f_1 = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma^2}} \quad f_0 = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{\sum x_i^2}{2\sigma^2}}
\]

Reject \( H_0 \) when \( \frac{f_1}{f_0} > k \)

\[
\frac{\sum x_i - 2\mu_1 \sum x_i + n\mu_1^2}{2\sigma^2} = \frac{\sum x_i^2}{2\sigma^2} > k
\]

or \( \frac{2\mu_1 \sum x_i - n\mu_1^2}{2\sigma^2} \geq \ln k \)
or \[ \sum x_i > \frac{(\ln k) 2\sigma^2 + n\mu_1^2}{2\mu_1} \]

i.e., if \( \bar{x} > k \)

To satisfy the size condition, \( K = \frac{z_{1-\alpha}}{\sigma/\sqrt{n}} \) (which is independent of \( \mu_1 \))

If \( H_0 \) is true, \( \bar{x} \) is \( N(0, \frac{\sigma^2}{n}) \)

\[ \Pr \left[ \frac{\bar{x}}{\sigma/\sqrt{n}} > z_{1-\alpha} \right] = \alpha \]

Remark: If the most powerful test of \( H_0 \) against \( H_1 \), \( \theta = \theta_1 \) is independent of \( \theta_1 \) for some family \( \omega_1 \), then the probability ratio test, \( \phi^* \), is u.m.p. for \( H_0: \theta = \theta_0 \) against \( H_1: \theta = \theta_1 \).

Therefore, \( \bar{x} > z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \) is u.m.p. for \( H_0 \) against \( H_1: \mu > 0 \).

For \( H_0: \mu = 0 \) against \( H_1: \mu < 0 \) the u.m.p. test is \( \bar{x} < z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \)

Example 2: \( x_1, x_2, \ldots, x_n \) are NID \((\mu, \sigma^2)\)

\( H_0: \mu = 0 \quad H_1: \mu \neq 0 \)

There can be no u.m.p. test for this problem since the u.m.p. tests for \( \mu < 0 \) and \( \mu > 0 \) differ.

\[
\begin{array}{ccc}
\text{u.m.p. here:} & \bar{x} < K & \text{u.m.p. here:} \bar{x} > K \\
\mu_1 & 0 & \mu_1
\end{array}
\]

Problem 68: \( x_1, x_2, \ldots, x_n \) are independently distributed with density

\[ f(x) = \theta e^{-\theta x} \quad x > 0 \quad \theta > 0 \]

\( H_0: \theta = \theta_0 \quad H_1: \theta = \theta_1 > \theta_0 \)

(1) Find the u.m.p. test explicitly (i.e., find the distribution of the test statistic).

(2) Write down the power function in terms of a familiar tabulation.
Remark: \( \frac{f_1(x)}{f_0(x)} \) is required to be a continuous random variable.

If this ratio is not a continuous random variable, then \( \alpha(k) = \Pr \left[ \frac{f_1}{f_0} > k \right] \)

may be discontinuous, so that \( \alpha(k) = \alpha \) has no solution. (e.g. small binomial situations where one can't get \( \alpha = .05 \) exactly.)

\( \alpha^* \) in this case is defined to be:

\[
\begin{align*}
\alpha^* = 1 & \quad \text{when} \quad \frac{f_1(x)}{f_0(x)} > k \\
\alpha^* = 0 & \quad \text{when} \quad \frac{f_1(x)}{f_0(x)} < k \\
\alpha^* = \varepsilon & \quad \text{when} \quad \frac{f_1(x)}{f_0(x)} = k
\end{align*}
\]

where \( \varepsilon \) is chosen so that the size of the test comes out as \( \alpha \)

\[
\varepsilon = \frac{\alpha - \alpha(k+0)}{\alpha(k) - \alpha(k+0)}
\]

Problem 69: For \( f_1 \) let \( S_1 \) be the set where \( f_1 > 0 \).

For \( f_0 \) let \( S_0 \) be the set where \( f_0 > 0 \).

1. If \( S_0 \) and \( S_1 \) are not coincident then there may be no test of size \( \alpha \).

2. If there is no test of size \( \alpha \) given by the Neyman-Pearson Lemma

(Theorem 33) there is a test of size \( \alpha \) with power = 1.

Hypotheses noted thus far have been simple hypotheses. If \( \Theta \) has more than one point, the hypothesis is called a composite hypothesis, e.g.:

\[ X_1, X_2, \ldots, X_n \text{ are NID (} \mu, \sigma^2 \text{)} \]
\[ H_0: \mu \leq 0 \quad H_1: \mu > 0 \]

\[ \beta(\theta) = \Pr[\text{reject } H_0] \]

for \( H_0: \mu = 0 \) against

\[ H_1: \mu > 0 \]

Test: \( X > z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \)

---

**Extension of the u.m.p. test to composite hypotheses by means of a "most unfavorable distribution of \( \theta \)."**

\[ H_0: \theta \in \omega \quad H_1: \theta = \theta_1 \]

Let \( \lambda(\theta) \) be a distribution of \( \theta \) over \( \omega \).

Then \( X \) has a density \( f(x, \theta) \).

\[ h_\lambda(x) = \int_\omega f(x, \theta) d\lambda(\theta) = \text{density of } X \text{ under } H_0 \text{ plus the additional information.} \]

Let \( \tilde{H}_0 \) be: \( X \) has density \( h_\lambda(x) \)

\[ i.e., \text{ the density of } X \text{ is } f(x, \theta_1) \]

Let \( \phi^\lambda \) be the most powerful test of \( \tilde{H}_0 \) against \( H_1 \).

**Theorem 31:** If \( \phi^\lambda \) is of size \( \alpha \) for the original test, it is m.p. for this test.

**Proof:** Let \( \phi \) be any other test of \( H_0 \).

\[ \phi(x) f(x, \theta) \, dx \leq \alpha \text{ for all } \theta \in \omega \]

Then

\[ \alpha \geq \int_\omega \left[ \int_x \phi(x) f(x, \theta) \, dx \right] d\lambda(\theta) \]

\[ = \int_x \phi(x) \left\{ \int_\omega f(x, \theta) d\lambda(\theta) \right\} \, dx \]

\[ = \int_x \phi(x) h_\lambda(\theta) \, dx \]

Thus \( \phi \) is of required size \( \alpha \) for \( \tilde{H}_0 \).

and:

\[ \int \phi^\lambda(x) f(x, \theta) \, dx \geq \int \phi(x) f(x, \theta) \, dx \quad \text{q.e.d.} \]
Example: $X_1, X_2, \ldots, X_n$ are $N(\mu, \sigma^2)$, $\sigma^2$ known

$H_0$: $\mu \leq 0$ 

$H_1$: $\mu > 0$

$\phi^*$ reject if $\bar{X} > z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ which was found to be u.m.p., for

$H_0'$: $\mu = 0$ against $H_1$: $\mu > 0$. We know that Pr [rejecting $H$ with $\phi^*|\mu]\leq \alpha$ for $\mu \leq 0$.

Put $\Lambda(\Theta) = 0 \quad \Theta < 0$

$= 1 \quad \Theta \geq 0$

(i.e., concentrate the distribution at 0 which is the worst spot from the standpoint of testing or distinguishing)

This makes our composite distribution:

$$h_\lambda(x) = \int f(x, \Theta) d\lambda(\Theta) = f(x, 0)$$

so that our problem is back to $H_0'$ and we have for $H_0'$ a u.m.p. test which is also a test of the original $H_0$.

One way to get an optimum test in the absence of a u.m.p. test is to restrict the class of tests to be considered and look for probability ratio test for restricted class. Such restrictions are:

1. Unbiasedness

$\phi$ is an unbiased test if $\beta_\phi(\Theta) \geq \alpha$ for all $\Theta \in \Omega_1$

2. Similarity

$\phi$ is a similar test if $E_\theta(\phi) = \alpha$ for all $\Theta \in \Omega_0$

3. Invariance

$X$ has density $f(x, \Theta)$

$G$ = a family of transformations of the sample space of $X$ onto itself.

$\Theta \in \Omega_0$ 

$g(x) = cx$ change of scale 

$= x + d$ translation of axes

Let $g(X)$ have density $f\left[ x, g(\Theta) \right]$, $g(\Theta)$ a $\bigcup$

$g$ is a transformation of the parameter space induced by the transformation of the sample space.
If $X$ is $N(\mu, \sigma^2)$ and $cX$ is $N(c\mu, c^2\sigma^2)$,
\[ g(\mu, \sigma^2) = c\mu, c^2\sigma^2 \]

if $X$ is $N(\mu, \sigma^2)$ and $X + d$ is $N(\mu + d, \sigma^2)$,
\[ g(\mu, \sigma^2) = \mu + d, \sigma^2 \]

If we have $H_0 \subseteq \phi \subseteq \omega$ and $H_1 \subseteq \phi \subseteq \bigcap - \omega$ then a transformation induced by $g$ leaves the problem invariant if
\[ g(\phi) \subseteq \omega \text{ when } \phi \subseteq \omega \]
\[ g(\phi) \subseteq \bigcap - \omega \text{ when } \phi \subseteq \bigcap - \omega \]

Example: $X$ is $N(\mu, \sigma^2)$

\[ H_0 \subseteq \mu = 0 \]
\[ g(X) = cX \quad c > 0 \]
\[ g(\mu, \sigma^2) = c\mu, c^2\sigma^2 \]
\[ \omega \text{ is the } \sigma^2 \text{ axis} \]

under $H_0 \ g(X)$ is $N(0, c^2\sigma^2)$

$H_1 \ g(X)$ is $N(\mu_1, c^2\sigma^2)$ which doesn't change $\omega$

Def. 4.4: A test, $\phi(x)$, is invariant under a transformation $g$ (for which the corresponding $g$ leaves the problem invariant) if $\phi(g(x)) = \phi(x)$

Example: $t = \frac{X}{\sqrt{n}} = \frac{1}{n} \sum \frac{X_i}{\sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}}^{1/2}$ (under $g = cX$)
\[ \frac{1}{n} \sum cX_i \quad \frac{\sum (cX_i - \bar{x})^2}{(n-1) n}^{1/2} \]
\[ \frac{1}{n} \sum \frac{X_i}{\sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}}^{1/2} = t \]

Def. 4.5: If among all unbiased (or similar or invariant) tests there is a $\phi^*$ which is u.m.p. then $\phi^*$ is u.m.p.u. (or u.m.p.s. or u.m.p.i.).
Uniformly Most Powerful Unbiased (u.m.p.u.) Tests:

Single parameter \( \theta \)

\( H_0: \theta = \theta_0 \) which is inside an open interval of the parameter space.

"f" is differentiable with respect to \( \theta \).

Remark: If \( \phi \) is unbiased, then

\[
\frac{\partial \beta_\phi (\theta)}{\partial \theta} \bigg|_{\theta = \theta_0} = 0
\]

Proof:

\( \beta_\phi (\theta) \leq \alpha \)

\( \beta_\phi (\theta) > \alpha \) for \( \theta \neq \theta_0 \)

Hence \( \beta_\phi (\theta) \) has a minimum at \( \theta = \theta_0 \)

\( \beta_\phi (\theta) = \int \phi (x, \theta) \ dx \)

\( \beta_\phi (\theta) \) is differentiable with respect to \( \theta \)

and hence

\[
\frac{\partial \beta_\phi (\theta)}{\partial \theta} \bigg|_{\theta = \theta_0} = 0
\]

Note: For unbiasedness, alternatives must be two-sided (otherwise the power curve has no minimum).

Assuming that \( \int f(x, \theta) \ dx \) is differentiable under the integral sign, and with

\( H_0: \theta = \theta_0 \)

\( H_1: \theta = \theta_1 \) (two-sided) we can get:

Theorem 35: If there exist \( k_1, k_2 \) so that

\[
\phi_u^* = 1 \text{ when } f(x, \theta_1) > k_1 f(x, \theta_0) + k_2 \frac{\partial f}{\partial \theta} \bigg|_{\theta = \theta_0} = 0 \text{ elsewhere}
\]

is of size \( \alpha \) for \( H_0 \) and unbiased, then it is m.p. for alternatives \( \theta_1 \). If the test does not depend on \( \theta_1 \) it is u.m.p.u. for \( H_1: \theta \in \Omega \).

Proof: (ref. Cramer p. 532)

Let \( \phi \) be any other unbiased test.

Then

\[
\int \phi_u^* f_0 \ dx = \alpha = \int \phi f_0 \ dx - \text{size conditions}
\]

\[
\int \phi_u^* \frac{\partial f}{\partial \theta} \bigg|_{\theta = \theta_0} \ dx = 0 = \int \phi \frac{\partial f}{\partial \theta} \bigg|_{\theta = \theta_0} \ dx - \text{unbiasedness conditions}
\]
We want to show that:

\[ \int \phi_{\ast} f \, dx \geq \int \phi f \, dx \]

\[ \int (\phi^* - \phi) f \, dx = \int (\phi^* - \phi) \left[ f - k_1 f - k_2 \frac{df}{d\phi} \right] \bigg|_{\phi = \phi_0} \geq 0 \]

This integrand must always be \( \geq 0 \)

since if \( f_1 - k_1 f_0 - k_2 \frac{df}{d\phi} \bigg|_{\phi = \phi_0} \geq 0 \) \( \phi^* \geq \phi \)

\[ \leq 0 \quad 0 = \phi^* \leq \phi \]

Thus the desired relationship always holds.

Comment: If you are trying to find a bounded function, \( a \leq \phi \leq b \) which maximizes \( \int \phi f \, dx \) subject to side conditions; \( \int f_1 \, dx = c_1, i = 1, 2, \ldots, n \), then this maximum will be given by choosing:

\[ \phi = b \quad \text{where} \quad f > \sum k_i f_i(x) \quad \text{where} \quad k_i \text{ are chosen to satisfy the n side conditions} \]

\[ = a \quad \text{otherwise} \]

Example: \( X_1, X_2, \ldots, X_n \) are \( \text{NID}(\mu, \sigma^2) \) \( \sigma^2 \) known \( H_0 \): \( \mu = 0 \) \( H_1 \): \( \mu \neq 0 \)

Consider a particular alternative \( \mu_1 \):

\[ f(x, \mu_1) = \frac{1}{(\sqrt{2\pi} \sigma)^n} e^{-\frac{\sum (x_i - \mu_1)^2}{2\sigma^2}} \]

\[ \frac{df}{d\mu_1} = \frac{1}{(\sqrt{2\pi} \sigma)^n} \left( -\frac{\sum (x_i - \mu_1)^2}{2\sigma^2} \right) \left( \frac{\sum (x_i - \mu)}{\sigma^2} \right) \]
If we can find proper $k$'s, the u.m.p.u. test is:

Reject $H$ if

\[
\frac{1}{\sqrt{2\pi \sigma^2}} \sum (x_i - \mu)^2 \geq k_1 \frac{1}{\sqrt{2\pi \sigma^2}} \sum x_i^2 + k_2 \frac{1}{\sqrt{2\pi \sigma^2}} \sum x_i^2 \geq 0
\]

(derivative evaluated at $\mu = 0$)

\[
\frac{\mu}{\sigma^2} \sum x_i - \frac{n \mu^2}{2\sigma^2} \geq k_1 + \frac{k_2}{\sigma^2} \sum x_i
\]

If we set $y_1$ = the left hand side of this inequality, $y_2$ = the right hand side, and restrict ourselves to the cases where $\mu^2 \geq 0$, then we could get the following type of graphs:

\[y_2, y_2', y_2'', y_2'''\] are possible $y_2$ lines

The test says to reject $H$ if $\bar{x} < a$ or $\bar{x} > b$ where $a$ may be $-\infty$ and $b$ may be $+\infty$

\[y_2\] gives a two-tailed test (finite $a$, $b$) -- $y_2'$, $y_2''$, $y_2'''$ all give one-tailed tests (only one intersection with $y_1$)
Requiring that the test be of size \( \alpha \) specifies that:

\[
\Pr \left[ a < \bar{x} < b \right] = 1 - \alpha
\]

i.e.,

\[
\frac{\sqrt{n}}{2\pi \sigma} \int_a^b e^{-\frac{n\bar{x}^2}{2\sigma^2}} \, d\bar{x} = 1 - \alpha
\]

Also, the unbiasedness condition requires that:

\[
\frac{\partial}{\partial \mu_1} \left[ \frac{\sqrt{n}}{2\pi \sigma} \int_a^\infty e^{-\frac{n(\bar{x} - \mu_1)^2}{2\sigma^2}} \, d\bar{x} + \frac{\sqrt{n}}{2\pi \sigma} \int_b^\infty e^{-\frac{n(\bar{x} - \mu_1)^2}{2\sigma^2}} \, d\bar{x} \right]_{\mu_1=0} = 0
\]

or

\[
\frac{\partial}{\partial \mu_1} \left[ \frac{\sqrt{n}}{2\pi \sigma} \int_a^b e^{-\frac{n(\bar{x} - \mu_1)^2}{2\sigma^2}} \, d\bar{x} \right]_{\mu_1=0} = 0
\]

or

\[
-\frac{\sqrt{n}}{2\pi \sigma} \int_a^b e^{-\frac{n\bar{x}^2}{2\sigma^2}} \cdot \frac{n\bar{x}}{\sigma^2} \, d\bar{x} = 0
\]

[2]

The function under the integral in [2] is an odd function, so that the integral is zero only if \( a = -b \) (i.e., if \( a, b \) are symmetrical).

[1] thus becomes:

\[
\frac{\sqrt{n}}{2\pi \sigma} \int_a^{-a} e^{-\frac{n\bar{x}^2}{2\sigma^2}} \, d\bar{x} = 1 - \alpha
\]

so that we can determine \( a \) as:

\[
a = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}
\]
Thus the test is:

\[ \text{Reject } H \text{ if } \left| \bar{x} \right| > z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \]

and since it is independent of \( \mu \) it is \( u.m.p. \).

**Problem 70:** \( X_1, X_2, \ldots, X_n \) are \( \text{NID}(\mu, \sigma^2) \) \( \mu \) known

\[ H_0: \sigma = \sigma_0 \quad \quad H_1: \sigma \neq \sigma_0 \]

Find the \( u.m.p. \) test for \( H_0 \).

Note: \( \sum (X_i - \mu)^2 \) is sufficient for \( \sigma^2 \).

**Theorem 36:** If \( T \) is sufficient for \( \theta \) then given any test \( \phi(x) \) there exists a test \( \psi(T) \) with the same power function. Hence in looking for optimum tests, only functions of \( T \) need be considered.

**Proof:** define \( \psi(T) = E \left[ \phi(x) \left| T \right. \right] \) which is independent of \( \theta \) by definition

Thus:

\[ E \left[ \psi(T) \right] = E_{\tilde{F}_X} \left[ \phi(x) \left| T \right. \right] = E \left[ \phi(x) \right] \quad \text{Q.E.D.} \]
Invariant Tests

Refs: Lehmann, "Notes on Testing Hypotheses", ch. 4
Fraser, "Non Parametric Methods in Statistics", ch. 2

We have: observations: $X$ parameter: $\theta$ density: $f(x, \theta)$

$H_0$: $\theta \in \Omega_0$

$H_1$: $\theta \in \Omega_1$

Transformations: $x' = g(x)$ -- $x'$ has density $f \left[ x', g(\theta) \right]$  

i.e., there is an induced transformation on the parameter space:

$\tilde{\theta} = \tilde{g}(\theta)$

$\tilde{g}$ the group of transformations that leave the problem invariant, i.e.

$\tilde{g}(\theta) \in \Omega_0$ if $\theta \in \Omega_0$

$\tilde{g}(\theta) \in \Omega_1$ if $\theta \in \Omega_1$

Examples: $X_1, X_2, \ldots, X_n$ are NID ($\mu, \sigma^2$)

$x' = cx + d$ 

$x'$ is NID ($c\mu + d, c^2\sigma^2$)

$g(x) = cx + d$ 

$\tilde{g}(\mu, \sigma^2) = (c\mu + d, c^2\sigma^2)$

$H_0$: $\mu = 0$

$H_1$: $\mu > 0$

If we set $d = 0$ and $c > 0$, then the problem is invariant, and we have:

$x' = g(x) = cx$

$\tilde{g}(\mu, \sigma^2) = (c\mu, c^2\sigma^2)$

Sufficient statistics for $\mu, \sigma^2$ are $\bar{X}$ and $S = \sum (X_i - \bar{X})^2$

Under the transformation:

$t = \frac{\bar{X}}{\sqrt{S}}$ is invariant (the inclusion of constants does not affect invariance).

(discussion to be continued after def. 46 and theorem 37.)

Def. 46: $m(x)$ is a maximal invariant function under a group of transformations if

1) $m \left[ g(x) \right] = m(x)$

2) $m(x_1) = m(x_2) \implies$ there is a $g$ such that $g(x_2) = x_1$ of vice versa
Theorem 37: The necessary and sufficient condition that the test function \( \phi (x) \) be invariant under \( G \) is that it depend only on \( m(x) \).

Proof: 1) \( \phi (x) = \Psi [m(x)] \)
\[
\phi [g(x)] = \Psi \{m[g(x)]\} = \Psi [m(x)] = \phi (x)
\]

2) Suppose that \( \phi (x) \) is invariant. We have to show that \( \phi (x) \) is a function of \( m(x) \).

Given \( \phi [g(x)] = \phi (x) \)

show \( m(x_1) = m(x_2) \implies \phi (x_1) = \phi (x_2) \)

\( m(x_1) = m(x_2) \implies \text{for some } g, \text{ call it } g', g'(x_2) = x_1 \)

thus \( \phi (x_1) = \phi [g'(x_2)] = \phi (x_2) \) which is what we set out to prove.

Returning to the "Student" problem:

It remains to show that \( t = \frac{\bar{X}}{\sqrt{S}} \) is maximal invariant.

Setting \( t = \frac{\bar{X}}{\sqrt{S}} \); \( t' = \frac{\bar{X}'}{\sqrt{S'}} \) we have to show that given \( t = t' \), we can find \( g \) such that \( \bar{X}', S' = g(\bar{X}, S) \).

Consider \( \frac{\bar{X}'}{\bar{X}} \) and call this ratio "a".

\[
\frac{\bar{X}}{\sqrt{S}} = \frac{\bar{X}'}{\sqrt{S'}} \implies \frac{\bar{X}'}{\bar{X}} = \frac{\sqrt{S'}}{\sqrt{S}} = a \quad \text{or } S' = a^2 S
\]

But this is just one of the members of the original family of transformations so \( t \) is maximal invariant.

Hence the problem is reduced to finding the u.m.p. test based on \( t \).

In summary: \( X_1, X_2, \ldots, X_n \) are NID (\( \mu, \sigma^2 \))

\( H_0: \mu = 0 \quad H_1: \mu > 0 \)
1) Reduce by using sufficient statistics: \( \bar{x}, s \).

2) Impose the invariance conditions under the transformation \( x' = cx, c > 0 \). This reduces the problem to tests based on \( t \).

3) Find the distribution of \( t \) under \( H_0 \) and \( H_1 \) and apply Theorem 33 (Probability Ratio Test).

---

**Distribution of Non-Central \( t \) (t-distribution under \( H \))**

**Refs:**
Neyman + Tokarska, *JASA* 1936, pp. 318-326 (tables for the one-sided case)
Welch + Johnson, *Biometrika*, 1946, pp. 382-389
Resnikoff + Lieberman, "Tables of the Non-Central t-distribution", Stanford University Press, 1957

\( \tau \), the non-central \( t \)-variable, is defined by:

\[
\tau = \tau(\delta, f) = \frac{z + \delta}{\sqrt{w/f}}
\]

where:
- \( z \) is \( \text{N}(0, 1) \)
- \( w \) is \( \chi^2 \) with \( f \) d.f.
- \( \delta \) is a constant \( > 0 \)

The usual \( t \)-variable is

\[
t = \tau(0, f) = \frac{\sqrt{n} \bar{x}}{\sqrt{\frac{s^2}{n-1}}} = \frac{\sqrt{n} \bar{x}}{\sigma} \frac{1}{\sqrt{\frac{s^2}{\sigma^2(n-1)}}}
\]

when \( H_0 \) is true

If \( H_1 \) is true, \( \mu = \mu_1 \) and \( \frac{\sqrt{n}(\bar{x} - \mu_1)}{\sigma} \) is \( \text{N}(0, 1) \)

\[
t = \frac{\sqrt{n}(\bar{x} - \mu_1)}{\sigma} + \sqrt{n} \frac{\mu_1}{\sigma}\frac{1}{\sqrt{\frac{s^2}{\sigma^2(n-1)}}}
\]
The joint density of $z$ and $w$ is:

$$f(z, w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \frac{1}{2^{f/2}} e^{-\frac{w}{2}} \left(\frac{f}{2}\right)^{-1/2} e^{-\frac{w}{2}}$$

$$-\infty < z < \infty \quad w \geq 0$$

To get the density of $\tau$ (the non-central $t$) we make the transformation:

$$\tau = \frac{z + \delta}{\sqrt{\frac{w}{f}}} \quad u = w \quad |J| = \frac{\sqrt{u}}{\sqrt{f}}$$

Thus:

$$f(\tau) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{f/2}} e^{-\frac{\tau^2}{2}} \int_0^\infty e^{-\frac{1}{\sqrt{f}} \left(\frac{\sqrt{u} \tau}{\sqrt{f}} - \delta\right)^2} \frac{f-1}{2} e^{-\frac{u}{2}} du$$

To get the distribution of the usual $t$, put $\delta = 0$.

The integral in $f(\tau)$ becomes

$$\int_0^\infty u \frac{f+1}{2} - 1 - \frac{u}{2} \left(\frac{\tau^2}{f} + 1\right) du$$

which can be readily evaluated recalling that

$$\int_0^\infty e^{-ax} x^{b-1} dx = \frac{\Gamma(b)}{a^b}$$

Thus,

$$f(t) = \frac{1}{\sqrt{\pi f}} \left(\frac{f+1}{2}\right) \left(\frac{f}{2}\right)^{1/2} \left(1 + \frac{t^2}{f}\right)^{-1/2}$$

Ref: Cramer, p. 238
The u.m.p.i. test is to reject $H_0$ if:

$$R(\tau) = \frac{f(\tau, \delta)}{f(\tau, 0)} = c \left[ \int_0^\infty e^{-\frac{1}{2} \left( \frac{u - \delta}{\sqrt{s}} \right)^2} - \frac{u}{2} \left( \frac{u + 1}{2} - 1 \right) \left( 1 + \tau \frac{f+1}{2} \right)^2 > k \right]$$

Note: This ratio is a monotone increasing function of $\tau$. (The simplest proof of the required monotonicity is given by Kruskal, Annals of Math Stat, 1954, pp. 162-3.)

Since $t > t_0$ when the above ratio $> k$, the probability ratio test thus reduces to:

Reject $H_0$ when $t > t_0$

Final result is that the m.p.i. test for $H_1: \mu = \mu_1 > \mu_0$ is:

Reject $H_0$ when $t = \frac{\sqrt{n} \bar{X}}{\sqrt{s}} > t_{1-\alpha} (n-1)$

This is independent of $\mu_1$ and hence is u.m.p.i. for $H_0$ against $H_1$.

Example: $X_1, X_2, \ldots, X_m$ are NID ($\mu_1, \sigma^2$)

$Y_1, Y_2, \ldots, Y_n$ are NID ($\mu_2, \sigma^2$)

$\sigma^2$ known

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

Sufficient statistics for the three parameters are:

$$\bar{x}, \bar{y}, s_p = \sqrt{\frac{\sum (X_i - \bar{x})^2 + \sum (Y_i - \bar{y})^2}{m+n-2}}$$
Consider: \( X' = aX + b \)
\( Y' = cY + d \)

\( X' \) is \( N(a \mu_1 + b, a^2 \sigma^2) \)
\( Y' \) is \( N(c \mu_2 + d, c^2 \sigma^2) \)

Invariance requires that:

\[ a \mu_1 + b = c \mu_2 + d \]
when \( \mu_1 = \mu_2 \) for all \( \mu \)

\[ a \mu_1 + b > c \mu_2 + d \]
when \( \mu_1 > \mu_2 \)

The first line requires that: \( b = d; \quad a = c \)

The second line adds the requirement that: \( a = c > 0 \)

Therefore: \( X' = aX + b, \ Y' = aY + b, \ a > 0 \) leaves the problem invariant.

To be proven:

\[ t = \frac{\bar{X} - \bar{Y}}{s_p} \]

is a maximal invariant statistic.

\[ t = t' \implies \text{there exists an } a, b \text{ such that } X' = aX + b, \ Y' = aY + b \]

\[ t = \frac{\bar{X} - \bar{Y}}{s_p} = \frac{\bar{X}' - \bar{Y}'}{s_p'} = t' \]

define:

\[ \frac{s_p'}{s_p} = c > 0 \]

\[ \frac{\bar{X} - \bar{Y}}{s_p} = \frac{\bar{X}' - \bar{Y}'}{c s_p} \]

\[ c(\bar{X} - \bar{Y}) = \bar{X}' - \bar{Y}' \]

now let \( \bar{X}' - c\bar{X} = d \)

\[ c(\bar{X} - \bar{Y}) = c\bar{X} + d = \bar{Y}' \]

\[ -c\bar{X} = d - \bar{Y}' \]

or:

\[ \bar{Y}' = c\bar{X} + d \]

\[ \bar{X}' = c\bar{X} + d \]

so that the same transformation has been applied to \( \bar{X} \) and \( \bar{Y} \).

The u.m.p. test invariant under the family of transformations:

\( X' = aX + b, \ Y' = aY + b, \ a > 0 \)

is:

Reject \( H \) if

\[ t = \frac{\bar{X} - \bar{Y}}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}} > t_{1-\alpha} (m+n-2) \]
Under the alternative \( \mu_1 - \mu_2 = d \), thus

\[
\frac{(\bar{X} - \bar{Y} - d)}{\frac{s}{\sqrt{\frac{1}{m} + \frac{1}{n}}}} + \frac{d}{c}
\]

has a non-central t distribution with parameters

\[
\left[ \frac{d}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}}, \ m+n-2 \right]
\]

Exercise: \( \frac{d}{\sigma} = 0.8 \), \( \alpha = 0.05 \)

Find \( m, n \) required to have the power \( \geq 0.90 \) (put \( m = n \)).

Try to find \( m, n \) also by normal approximation.

Answer: \( d_{df} = 2(n-1) \), non-centrality parameter \( \rho = \frac{d \sqrt{n}}{\sigma \sqrt{2}} = \frac{0.8 \sqrt{n}}{\sqrt{2}} \)

\[
\begin{array}{cc}
30 & 3.1 \\
28 & 2.99 \\
27 & 2.94 \\
25 & 2.8 \\
\end{array}
\]

For the power to exceed 0.90, from the Neyman-Tokarska tables we need:

\( \rho = 2.99 \) at \( d_{df} = 30 \)

\( \rho = 2.93 \) at \( d_{df} = \infty \)

Thus, by rough interpolation, the minimum size required is \( n = m = 28 \).

From the normal approximation: \( n = 26.7 \) or 27

Problem 71: \( X_1, X_2, \ldots, X_m \) are NID \( (\mu_1, \sigma_1^2) \)

\( Y_1, Y_2, \ldots, Y_n \) are NID \( (\mu_2, \sigma_2^2) \)

\( H_0: \sigma_1^2 = \sigma_2^2 \)

\( H_1: \sigma_1^2 > \sigma_2^2 \)

a) Find the group of transformations leaving \( H_0 \) invariant.

b) Find sufficient statistics for the parameters.

c) Find the maximum invariant function.

d) Find the u.m.p.i. test.
e) Find the u.m.p.u.i. test (i.e., set down conditions to get the rejection region as in problem 70) for: \( H_0: \sigma_1^2 = \sigma_2^2 \) against \( H_1: \sigma_1^2 \neq \sigma_2^2 \).

f) Show that the usual test which is to reject if \( F > F_{1-\alpha} (m-1, n-1) \) where:

\[
\begin{align*}
\bar{s}_x^2 &= \frac{\sum (x_i - \bar{x})^2}{n-1} \\
\bar{s}_y^2 &= \frac{\sum (y_i - \bar{y})^2}{n-1} \\
F &= \max \left( \frac{\bar{s}_x^2}{\bar{s}_y^2}, \frac{\bar{s}_y^2}{\bar{s}_x^2} \right)
\end{align*}
\]

is a test of size \( 2\alpha \) with "equal tail probabilities".

g) Plot the power of the test of (f) for \( m = n = 10 \) with \( 2\alpha = 0.05 \).

(Include the points \( \sigma_1^2/\sigma_2^2 = 0.5, 0.8, 1.25, 1.5, 2.0 \).

Tests for Variances:

\( X_1, X_2, \ldots, X_n \) are NID \((\mu, \sigma^2)\)

\( H_0: \sigma^2 = \sigma_0^2 \)

\( H_1: \sigma^2 > \sigma_0^2 \)

\( H_1': \sigma^2 \neq \sigma_0^2 \)

1) \( \mu \) known \( \sum (X_i - \mu)^2 \) is sufficient for \( \sigma^2 \).

Reject \( H_0 \) if \( \sum (X_i - \mu)^2 > K \) -- u.m.p.

Reject \( H_0 \) if \( \sum (X_i - \mu)^2 < K_1 \) or \( > K_2 \) -- u.m.p.

ref: problem 70

2) \( \mu \) unknown \( \bar{x}, \frac{\sum (X_i - \bar{x})^2}{n-1} \) are sufficient for \( \mu, \sigma^2 \)

Problem is invariant under translation, i.e.,

\( X' = X + a \quad \bar{x}' = \bar{x} + a \quad (s'^2) = s^2 \)

To find a maximum invariant function

\( f(\bar{x}', s'^2) = f(\bar{x} + a, s^2) = f(\bar{x}, s^2) \) must hold for all \( a \).

This says that \( f(\bar{x}, s^2) \) is independent of \( \bar{x} \). Thus invariant functions are functions of \( s^2 \) only.

Hence, since \( \frac{(n-1)s^2}{\sigma^2} \) is \( \chi^2 \) with \( n-1 \) d.f. when \( H_0 \) is true, the problem is exactly that of (1) with the d.f. reduced by 1 (i.e., \( n-1 \) vice \( n \)).
Summary of Normal Tests:

\[ X_1, X_2, \ldots, X_m \text{ are NID } (\mu_x, \sigma^2_x) \]
\[ Y_1, Y_2, \ldots, Y_n \text{ are NID } (\mu_y, \sigma^2_y) \]
\[ X_i's, Y_i's \text{ are independent} \]

<table>
<thead>
<tr>
<th>Alternative Hypothesis</th>
<th>Test: Reject H if</th>
<th>Classification (for H against the given alternative)</th>
<th>Invariant under transformations of the form</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. [ H_0: \mu_x = 0 ] [ (\sigma^2_x \text{ known}) ]</td>
<td>[ \bar{x} &gt; z_{1-\alpha} \frac{\sigma_x}{\sqrt{m}} ]</td>
<td>u.m.p.</td>
<td>[ X' = cX \quad c &gt; 0 ]</td>
</tr>
<tr>
<td>[ H_1: \mu_x &gt; 0 ]</td>
<td>[</td>
<td>\bar{x}</td>
<td>&gt; z_{1-\alpha} \frac{\sigma_x}{\sqrt{m}} ]</td>
</tr>
<tr>
<td>[ H_1: \mu_x \neq 0 ]</td>
<td>[ t &gt; t_{1-\alpha} ]</td>
<td>u.m.p.i.</td>
<td></td>
</tr>
<tr>
<td>2. [ H_0: \mu_x = \mu_0 ] [ (\sigma^2_x \text{ unknown}) ]</td>
<td>[ \sum (X_i - \mu)^2 &gt; x^2_{1-\alpha}(m) ]</td>
<td>u.m.p.</td>
<td></td>
</tr>
<tr>
<td>[ H_1: \mu_x &gt; \mu_0 ]</td>
<td>[ \sum (X_i - \mu)^2 &lt; K_1 \text{ or } &gt; K_2 ] (for equations for ( K_1, K_2 ), see problem 70)</td>
<td>u.m.p.u.</td>
<td></td>
</tr>
<tr>
<td>[ H_1: \mu_x \neq \mu_0 ]</td>
<td>[</td>
<td>t</td>
<td>&gt; t_{1-\alpha} ]</td>
</tr>
<tr>
<td>3. [ H_0: \sigma^2_x = \sigma^2_0 ] [ (\mu \text{ known}) ]</td>
<td>[ \sum (X_i - \bar{x})^2 &gt; x^2_{1-\alpha}(m-1) ]</td>
<td>u.m.p.i.</td>
<td>[ X' = X + a ]</td>
</tr>
<tr>
<td>[ H_1: \sigma^2_x &gt; \sigma^2_0 ]</td>
<td>[ \sum (X_i - \bar{x})^2 &lt; K_1 \text{ or } &gt; K_2 ] (for ( K_1, K_2 ) see problem 70 - use ( m-1 ) d.f.)</td>
<td>u.m.p.u.i.</td>
<td></td>
</tr>
</tbody>
</table>
5. \( H_0: \mu_x = \mu_y \) (\( \sigma_x^2, \sigma_y^2 \) known)

\[
\begin{align*}
H_1: \mu_x > \mu_y & \quad \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}} > z_{1-\alpha} \\
& \quad \text{u.m.p.i.} \\
& \quad X' = X + a \\
& \quad Y' = Y + a
\end{align*}
\]

\[
\begin{align*}
H_1: \mu_x \neq \mu_y & \quad \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma_x^2}{m} + \frac{\sigma_y^2}{n}}} > z_{1-\alpha/2} \\
& \quad \text{u.m.p.u.i.} \\
& \quad X' = X + a \\
& \quad Y' = Y + a
\end{align*}
\]

6. \( H_0: \mu_x = \mu_y \) (\( \sigma_x^2, \sigma_y^2 \) unknown)

\[
\begin{align*}
a. \quad \sigma_x^2 = \sigma_y^2 & \quad X, Y, s^2_p = \frac{\sum (X_i - \bar{X})^2 + \sum (Y_i - \bar{Y})^2}{m + n - 2} \quad \text{are sufficient for} \quad \mu_x, \mu_y, \sigma^2 \\
& \quad \text{u.m.p.i.} \\
& \quad X' = aX + b \\
& \quad Y' = aY + b
\end{align*}
\]

\[
\begin{align*}
H_1: \mu_x > \mu_y & \quad \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}} > t_{1-\alpha(m+n-2)} \\
& \quad \text{u.m.p.i.} \\
& \quad X' = aX + b \\
& \quad Y' = aY + b
\end{align*}
\]

\[
\begin{align*}
H_1: \mu_x \neq \mu_y & \quad \frac{|\bar{x} - \bar{y}|}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}} > t_{\frac{1-\alpha}{2}(m+n-2)} \\
& \quad \text{u.m.p.u.i.} \\
& \quad X' = aX + b \\
& \quad Y' = aY + b
\end{align*}
\]

b. \( \sigma_x^2 \neq \sigma_y^2 \)

\[
X, Y, s^2_x, s^2_y \quad \text{are sufficient statistics for} \quad \mu_x, \mu_y, \sigma_x^2, \sigma_y^2.
\]

This is the classical Fisher-Behrens Problem—no exact test is known. The approximations thus far have tried to keep the size of the test under control, and very little attention has been paid to the power. The approximation used is:

\[
t' = \frac{\bar{x} - \bar{y}}{s^2 \sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \\
\text{t' is approximately distributed as} \ t \ \text{with modified d.f. (i.e., modifications by: Smith-Satterthwaite, Cochran-Cox, and Dixon-Massey)}
\]

Ref: Anderson and Bancroft, p. 80.

Tables for \( t' \) in the one-sided case (\( \alpha = 0.05, 0.01 \)) have been given by Aspen in Biometrika, 1949.
If \( m = n \)

\[
t' = \frac{-171-}{\frac{X - Y}{\sqrt{\frac{s_x^2 + s_y^2}{n}}}}
\]

\[
t = \frac{X - Y}{\left(\frac{(n-1)s_x^2 + (n-1)s_y^2}{2(n-1)}\right)^{\frac{1}{2}} \left(\frac{1}{n} + \frac{1}{n}\right)^{\frac{1}{2}}} = t'
\]

Test proposed, when \( m = n \), is to reject \( H \) if \( t > t_{1-\alpha}(n-1) \) for the one-sided case.

Empirical results:

based on 1000 samples, \( c_1^2 = 1, c_2^2 = 4 \):

\[
m = n = 15 \quad \text{significance level} \quad 5\% \quad 1\%
\]

actual rejections \( 4.9\% \quad 1.1\%
\]

rejections based on 28 d.f.

\[
m = n = 5 \quad \text{significance level} \quad 5\% \quad 1\%
\]

actual rejections \( 6.4\% \quad 1.3\%
\]

rejections based on 8 d.f.

In re power in the Fisher-Behrens problem:

Ref: Gronow; Biometrika, 1951, pp. 252-256

He gives a note on the power of the U and the t tests for the 2 sample problems with unequal variances.

With \( n_1 \neq n_2, c_1^2 \neq c_2^2 \), the U test stays fairly close to \( \alpha \) in size, whereas the t-test jumps wildly and a comparison of the power becomes very difficult.
7. \( H_0: \sigma^2_x = \sigma^2_y \) (\( \mu_x, \mu_y \) known)

- Test: Reject \( H \) if

\[
\frac{\sum (x_i - \mu_x)^2 / m}{\sum (y_i - \mu_y)^2 / n} > F_{1-\alpha}(m, n) \quad u.m.p.i.
\]

\[
Y' = (aX + b)Y + c
\]

\( H_1: \sigma^2_x > \sigma^2_y \)

\[
\frac{\sum (x_i - \mu_x)^2 / m}{\sum (y_i - \mu_y)^2 / n} \quad < F_1 \text{ or } > F_2 \quad u.m.p.u.i.
\]

\[
Y' = (aX + b)Y + c
\]

Where \( F_1, F_2 \) are chosen to satisfy the unbiasedness conditions as in problem 71.

Since \( F_1, F_2 \) depend on complicated equations, we usually use the test:

\[
\text{Reject } H_0 \text{ if } \max \left( \frac{s_x^2}{s_y^2}, \frac{s_y^2}{s_x^2} \right) > F_{1-\alpha}(m, n \text{ or } n, m)
\]

\[ \text{d.f., depending on which term is in the numerator.} \]

8. \( H_0: \sigma^2_x = \sigma^2_y \) (\( \mu_x, \mu_y \) unknown)

- Maximal invariant statistic is

\[
\frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}
\]

- Test: Reject \( H \) if

\[
\frac{\sum (x_i - \bar{x})^2 / m-1}{\sum (y_i - \bar{y})^2 / n-1} > F_{1-\alpha}(m-1, n-1) \quad u.m.p.i.
\]

\[
Y' = (aX + b)Y + c
\]

\( H_1: \sigma^2_x > \sigma^2_y \)

\[
\frac{\sum (x_i - \bar{x})^2 / m-1}{\sum (y_i - \bar{y})^2 / n-1} \quad < F_1 \text{ or } > F_2 \quad u.m.p.u.i.
\]

\[
Y' = (aX + b)Y + c
\]

Same comments on the determination of \( F_1 \) and \( F_2 \) apply as in No. 7, and the same alternative approach is usually taken (with the appropriate modification of d.f.)
3. **Maximum Likelihood Ratio Tests**

\[ H_0 : \theta \in \omega_0 \quad \text{versus} \quad H_1 : \theta \in \Omega - \omega_0 \]

Observations: \( x_1, x_2, \ldots, x_n \)

\[ \lambda = \frac{\max_{\theta \in \omega_0} f(x; \theta)}{\max_{\theta \in \Omega} f(x; \theta)} \]

Maximum Likelihood Ratio Test is to reject \( H_0 \) if \( \lambda < \lambda_0 \) where \( \lambda_0 \) is chosen to satisfy the size conditions.

For small samples in general this procedure may not give reasonable results. For large samples, as with the m.l.e., results are fairly good.

**Remark:** Under suitable regularity conditions \(-2 \ln \lambda \) is asymptotically distributed as \( \chi^2 \). The degrees of freedom depend upon the number of parameters specified by the hypothesis, i.e., if \( \theta \) has \( m \) components in \( \Omega \) and \( k \) component in \( \Omega_0 \) then the d.f. in the asymptotic distribution of \(-2 \ln \lambda \) (under \( H_0 \)) are \((m - k)\).


Wald has proven that the m.l.r. test is asymptotically most powerful or asymptotically most powerful unbiased test.

**Ref:** A. Wald: *Annals of Math Stat*, 1941
Transactions of American Math Society, 1943
(both papers in his collected papers)

**Example of m.l.r. tests:**

\( x_1, x_2, \ldots, x_n \) are \( N(\mu, \sigma^2) \) \( \sigma^2 \) unknown

\( H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0 \)

\[ f(x; \mu, \sigma^2) = \frac{1}{(\sqrt{2\pi} \sigma)^n} e^{-\frac{\sum (x_i - \bar{x})^2}{2\sigma^2}} \]
When $H_0$ is true, $\frac{\sum X_i^2}{n}$ is the m.l.e. of $\sigma^2$.

And in general, $\bar{x}$, $\frac{\sum (x_i - \bar{x})^2}{n}$ are the m.l.e. of $\mu, \sigma^2$.

$$\lambda = \frac{\frac{1}{2n} \left( \frac{\sum X_i^2}{n} \right)^{n/2}}{\frac{1}{2n} \left( \frac{\sum (x_i - \bar{x})^2}{n} \right)^{n/2}}$$

$$\lambda = \left[ \frac{\sum (x_i - \bar{x})^2 \sum X_i^2}{\sum X_i^2 \sum (x_i - \bar{x})^2} \right]^{n/2}$$

$$\frac{2^n}{n} = \frac{\sum (x_i - \bar{x})^2}{\sum X_i^2} = \frac{(n-1) s^2}{(n-1) s^2 + n \bar{x}^2}$$

$$\lambda = 1 + \left( \frac{n}{n-1} \right) \frac{s^2}{\bar{x}^2}$$

$$\sqrt{n-1} \left( \lambda - \frac{2}{n} - 1 \right)^{1/2} = \sqrt{n} \left( \frac{\bar{x}}{s} \right) = t$$

$$h(\lambda) = \sqrt{n-1} \left( \lambda - \frac{2}{n} - 1 \right)^{1/2}$$ can be used if it is a monotone function of $\lambda$. 
\[ h^*(\lambda) = \sqrt{\frac{1}{n-1}} \left( \frac{1}{2} \right) (\lambda^{-2/n} - 1)^{-1/2} \lambda^{(2/n) - 1} \left( \frac{2}{n} \right) < 0 \]

Therefore \( h(\lambda) \) is a monotone decreasing function of \( \lambda \).

Reject \( H_0 \) when:

\[
\begin{align*}
&h(\lambda) > h_0 \\
&|t| > t_0
\end{align*}
\]

\[ \lambda < \lambda_0 \]

**Problem 72:** Find the \( m.r.r. \) test for \( H_0^\prime: \mu = 0 \) against \( H_1^\prime: \mu > 0 \).

Does it reduce to the \( u.m.p.i. \) test?

**Problem 73:** \( X_1, X_2, \ldots, X_n \) are \( \text{NID}(\mu, \sigma^2) \) \( \mu, \sigma^2 \) unknown

\[
H_0^\prime: \frac{\sigma}{\mu} = c_0 \quad \text{(specified)} \quad H_1^\prime: \frac{\sigma}{\mu} = c_1 > c_0
\]

Find a \( u.m.p.i. \) test for \( H_0 \) against \( H_1 \).

Perform the test where \( \bar{X} = 5, s = 12, c_0 = 1, n = 20 \)

---

**4. General Linear Hypothesis:**

**Assumptions:** \( X_i \) are \( \text{NID}(\mu_i, \sigma^2) \) \( i = 1, 2, \ldots, N \)

\[
\mu_i = \sum_{j=1}^{p} a_{ij} \beta_j \quad \beta_1, \ldots, \beta_p \text{ are unknown parameters}
\]

\[ p \leq N \]

\[ \text{rank } A = (a_{ij}) \text{ is } p \]

\[
H_0^\prime: \sum_{j=1}^{p} b_{kj} \beta_j = 0 \quad k = 1, 2, \ldots, s \quad s \leq p
\]

i.e., \( s \) linearly independent equations in the parameters.

**Alternative formulation:**

We may solve for \( \beta_1, \beta_2, \ldots, \beta_p \) in terms of \( p \) of the \( \mu_i \)'s and then get as the assumptions:

\[
\sum_{i=1}^{N} \lambda_{li} \mu_i = 0 \quad k = 1, 2, \ldots, N-p
\]

\( H_0 \) can be written as an additional set of \( s \) equations in the \( \mu_i \)'s:

\[
H_0': \sum_{i=1}^{N} \rho_{ki} \mu_i = 0 \quad k = 1, 2, \ldots, s
\]
Example:

\[ X_{ij} \text{ are } N(\mu_{ij}, \sigma^2) \]

\[ \mu_{ij} = \mu + \alpha_i + \beta_j \]

\[ \sum \alpha_i = \sum \beta_j = 0 \quad \text{for } i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, b \]

ab observations

Parameters are \( \mu, \alpha_1, \ldots, \alpha_a, \beta_1, \ldots, \beta_b \)

\[ a + b - 1 \] parameters in all

\[ H_0: \alpha_1 = 0 \]
\[ \alpha_2 = 0 \]
\[ \ldots \]
\[ \alpha_{a-1} = 0 \]

are the \( a-1 \) equations in the parameters

Lemma: If \( \sum_{j=1}^{n} a_{ij}y_j \) \( (i = 1, 2, \ldots, m) \) are \( m \) linearly independent equations in \( n \) unknowns \( (m \leq n) \) then there exists an equivalent set of equations with matrix \( C \) which is orthogonal, i.e.

\[ \sum_{j=1}^{n} c_{ij}^2 = 1 \quad \text{for } i = 1, 2, \ldots, m \]

\[ \sum_{j=1}^{n} c_{ij}c_{lj} = 0 \quad i \neq l \]

Proof: Ref: Mann, Analysis and Design of Experiments.

Given \( m \) equations in \( n \) unknowns:

\[ a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n = 0 \]

\[ a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n = 0 \]

\[ \cdots \quad \cdots \quad \cdots \]

\[ a_{m1}y_1 + a_{m2}y_2 + \cdots + a_{mn}y_n = 0 \]
The first row in the equivalent set is determined by setting

\[ c_{1j} = \frac{a_{1j}}{n} \sum_{\ell=1}^{n} a_{\ell j}^2, \quad \sum_{j=1}^{n} c_{1j}^2 = 1 \]

To get the second row:

\[ c_{2j} = a_{2j} - \lambda c_{1j} \]

\[ \sum_{j} c_{1j} c_{2j}' = \sum_{j} c_{1j} a_{2j} - \lambda \sum_{j} c_{1j}^2 \]

This = 0 (as required) if \( \lambda = \sum_{j} c_{1j} a_{2j} \)

Thus, we set \( c_{2j} = \frac{c_{2j}'}{n} \sum_{\ell=1}^{n} (c_{2\ell})^2 \) which will hold if the denominator ≠ 0 -- but by virtue of the independence of the original equations the equality can not hold.

For the third row:

\[ c_{3j} = a_{3j} - \lambda_1 c_{1j} - \lambda_2 c_{2j} \]

\[ \sum_{j} c_{1j} c_{3j}' = \sum_{j} a_{3j} c_{1j} - \lambda_1(1) - \lambda_2(0) \]

This = 0 if \( \lambda_1 = \sum_{j} a_{3j} c_{1j} \)

Similarly \( \lambda_2 = \sum_{j} a_{3j} c_{2j} \)

Finally we set:

\[ c_{3j} = \frac{c_{3j}'}{n} \sum_{\ell=1}^{n} (c_{3\ell})^2 \]

The completion of the proof follows readily by induction.
In the alternative formulation of the general linear hypothesis, we had the following \( N-p+s ) ( \leq N ) \) linearly independent equations:

\[
\sum_{i=1}^{N} \lambda_{i1} \mu_i = 0 \quad \ell = 1, 2, \ldots, N-p
\]

\[
\sum_{i=1}^{N} \rho_{ki} \mu_i = 0 \quad k = 1, 2, \ldots, s
\]

From the lemma we may assume these form the first \( N-p+s \) rows of an orthogonal matrix. These can be extended to form a complete orthogonal matrix \((N \times N)\) — this is a well-known matrix algebra lemma, proof is in Cramer or Mann. If we call the complete orthogonal matrix \( \Lambda \), we have

\[
\Lambda = \begin{pmatrix}
\lambda_{11} & \cdots & \lambda_{1N} \\
\vdots & \ddots & \vdots \\
\lambda_{N-p, 1} & \cdots & \lambda_{N-p, N}
\end{pmatrix}
\]

\( \lambda \)'s, \( \rho \)'s orthogonalized \( \tau \)'s added to complete the matrix.

\[
Y = \Lambda X
\]

\[
E(Y_{\ell}) = \sum_{i=1}^{N} \lambda_{i1} \mu_i = 0
\quad \ell = 1, 2, \ldots, N-p
\]

by the assumptions

\[
E(Y_{N-p+k}) = \sum_{i=1}^{N} \rho_{ki} \mu_i = 0
\quad k = 1, 2, \ldots, s
\]

if \( H_0 \) is true

\( Y \)'s are independent normal variables with means as shown and variances \( \sigma^2 \) (an orthogonal transformation changes NID variables into other NID variables with the same variances).
This is the canonical form of the general linear hypothesis.

\[ Y_i \text{ are NID} \left( \mu_i, \sigma^2 \right) \text{ where } \mu_i = 0 \quad i = 1, 2, \ldots, N-p \]

\[ H_0: \mu_i = 0 \quad i = N-p+1, \ldots, N-p+s \]

\[ H_1: \text{one or more of these } \mu_i \text{'s } \neq 0 \]

M.L.R. Test for the canonical form:

\[ f(Y) = \frac{1}{(\sqrt{2\pi} \sigma)^N} e^{-\frac{1}{2\sigma^2} \left[ \sum_{i=1}^{N-p} Y_i^2 + \sum_{i=N-p+1}^{N-p+s} (Y_i - \mu_i)^2 + \sum_{i=N-p+s+1}^{N} (Y_i - \mu_i)^2 \right]} \]

m.l.e. of the \( \mu \)'s in \( \Omega \) are obtained by setting \( \mu_i = Y_i (i=N-p+1, \ldots, N) \)

and \( \hat{\sigma}^2 = \frac{\sum_{i=1}^{N-p} Y_i^2}{N} \)

m.l.e. of the \( \mu \)'s in \( \omega \) are found by setting \( \mu_i = Y_i (i=N-p+s+1, \ldots, N) \)

and \( \hat{\sigma}^2 = \frac{\sum_{i=1}^{N-p+s} Y_i^2}{N} \)

\[ \lambda = \frac{\left( \frac{1}{\hat{\sigma}^2} \right)^N e^{-N/2}}{\left( \frac{1}{\sigma^2} \right)^N e^{-N/2}} = \left( \frac{\hat{\sigma}^2}{\sigma^2} \right)^{N/2} \]

\[ = \left( \frac{\sum_{i=1}^{N-p} Y_i^2}{\sum_{i=1}^{N-p+s} Y_i^2} \right)^{N/2} = \left( \frac{Q_a}{Q_r} \right)^{N/2} \]

\[ a = \text{absolute minimum (nothing specified about the hypothesis)} \]

\[ r = \text{relative minimum} \]

\[ Q_r = Q_a + \sum_{i=N-p+1}^{N-p+s} Y_i^2 \]

\[ \chi^2 = \frac{Q_a}{Q_r} \]
\[ \lambda = \frac{Q_r}{Q_a} = 1 + \frac{Q_r - Q_a}{Q_a} \]

\[ \frac{-2}{\lambda - 1} = \frac{Q_r - Q_a}{Q_a} \]

We can use any monotone function of \( \lambda \) for test purposes, therefore the test is:

Reject \( H_0 \) if \( \frac{Q_r - Q_a}{Q_a} > K \)

which is equivalent to rejecting \( H_0 \) if \( \lambda < \lambda_0 \).

\[ Q_r = Q_a = \sum_{i=N-p+1}^{N-p+s} Y_i^2 \text{ is } \chi^2(s) \]

\[ Q_a = \sum_{i=1}^{N-p} Y_i^2 \text{ is } \chi^2(N-p) \]

Therefore \( \frac{(Q_r - Q_a)/s}{Q_a/(N-p)} \) has the \( F \)-distribution with parameters \((s, N-p)\).

If \( H_1 \) is true, then \( E(Y_i) = d_i \quad i = N-p+1, \ldots, N-p+s \)

\[ \sum_{i=N-p+1}^{N-p+s} d_i^2 > 0 \]

Recall that \( \sum_{i=N-p+1}^{N-p+s} Y_i^2 \) is the sum of squares of variables with non-zero means, and hence has a non-central \( \chi^2 \)-distribution with parameters \((s, \sum_{i=1}^{s} d_i^2)\).

The ratio of a non-central \( \chi^2 \) to a central \( \chi^2 \) is a non-central \( F \), thus \( \frac{(Q_r - Q_a)/s}{Q_a/(N-p)} \)

under \( H_1 \) has a non-central \( F \)-distribution with parameters \((s, N-p, \sum_{i=1}^{s} d_i^2)\).

Thus the problem is solved in terms of the \( Y \)'s (the orthogonalized \( X \)'s).

The original problem was:

\( X \)'s are NID \((\mu_i, \sigma^2)\)

1) \( \sum_{i=1}^{N} \lambda_{\ell i} \mu_i = 0 \quad \ell = 1, 2, \ldots, N-p \)

2) \( \sum_{i=1}^{n} \rho_{ki} \mu_i = 0 \quad k = 1, 2, \ldots, s \)
M.L.R. test:

Define:

\[ Q'_a = \min \sum (X_i - \mu_i)^2 \quad \text{under restrictions (1)} \]

\[ Q'_r = \min \sum (X_i - \mu_i)^2 \quad \text{under restrictions (1) and (2)} \]

\[ \hat{\sigma} \text{ (in } \Omega \text{)} = (Q'_a)^{1/2} \]

\[ \hat{\sigma} \text{ (in } \omega \text{)} = (Q'_r)^{1/2} \]

\[ \lambda = \frac{\left(\frac{(Q'_a)^{N/2}}{Q'_a} e^{-N/2}\right)}{\left(\frac{(Q'_r)^{N/2}}{Q'_r} e^{-N/2}\right)} \]

As in the other case, this is a monotone decreasing function of \( \frac{Q'_r - Q'_a}{Q'_a} \)

Recall that under an orthogonal transformation, sums of squares are preserved.

Thus:

- \( Q'_a \) is carried into \( Q_a \)
- \( Q'_r \) is carried into \( Q_r \)

Hence:

\[ \frac{Q'_r - Q'_a}{Q'_a} = \frac{Q_r - Q_a}{Q_a} \]

and has the same distributions under \( H_0 \) (F-distribution) and under \( H_1 \) (non-central F).

**Problem 74:**

\[ X_{ijk} \text{ is } N(\mu_{ij}, \sigma^2) \quad i = 1, 2, \ldots, a \]

\[ j = 1, 2, \ldots, b \]

\[ \mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij} \]

\[ \sum_i \alpha_i = \sum_j \beta_j = 0 \]

\[ \sum_i (\alpha \beta)_{ij} = \sum_j (\alpha \beta)_{ij} = 0 \]

\[ H_0^* : \text{ all } \alpha_i = 0 \quad H_0^* : \text{ all } \alpha_i = 0; \text{ all } (\alpha \beta)_{ij} = 0 \]

Find the usual F-test for:

1) \( H_0 \)

2) \( H_0^* \)
Power of the ANOVA test:

Power of the ANOVA test has been tabled by Tang (Statistical Research Memoirs, Vol. 2 for \( \alpha = 0.05, \alpha = 0.01 \) for various values of:

\[
\phi = \left[ \frac{2\lambda}{s+1} \right]^{1/2}, \quad \lambda = \frac{\sum d_i^2}{2\sigma^2}
\]

Tables in: Mann
Kempthorne

\[
2\sigma^2 \lambda = \sum_{k=1}^{s} d_i^2 = \sum_{k} \left[ \sum_{i=1}^{N} \rho_{ki} \mu_i \right]^2
\]

\[
Y_{N-p+k} = \sum_{i=1}^{N} \rho_{ki} x_i
\]

\[
Q_r - Q_a = \sum_{k=1}^{s} \sum_{N-p+k}^{2} = \sum_{k=1}^{s} \left( \sum_{i=1}^{N} \rho_{ki} x_i \right)^2
\]

hence \( 2\sigma^2 \lambda \) is \( Q_r - Q_a \) with \( x_i \) replaced by \( E(x_i) = \mu_i \)

Example:

\( X_{ijk} \) is \( N(\mu_{ij}, \sigma^2) \)

\[
\mu_{ij} = \mu + \alpha_i + \beta_j + (\varphi)_{ij}
\]

or \( X_{ijk} = \mu + \alpha_i + \beta_j + (\varphi)_{ij} + e_{ijk} \)

\[
\sum_{i} \alpha_i = 0, \quad \sum_{j} \beta_j = 0, \quad \sum_{i} (\varphi)_{ij} = 0, \quad \sum_{j} (\varphi)_{ij} = 0
\]

\( H_0: \alpha_i = 0 \)

\( H_1: \alpha_i \) not all zero

\( Q_r - Q_a = \sum \sum \sum (\bar{X}_{i.} - \bar{X}_{..})^2 \)

under \( H_1 \):

\( E(\bar{X}_{i.} - \bar{X}_{..}) = \mu + \alpha_i + 0 + 0 \)

\( E(\bar{X}_{..}) = \mu \)

\( E(\bar{X}_{i.} - \bar{X}_{..}) = \alpha_i \)
hence \(2\sigma^2 \lambda = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} \chi_{i,j,k}^2 = nb \sum_{i=1}^{a} \chi_i^2\)

thus \(\phi = \left(\frac{2nb \sum \chi_i^2}{2\sigma^2 a}\right)^{1/2} = \left(\frac{nb \sum \chi_i^2}{\sigma^2 a}\right)^{1/2}\)

**Exercise:** If \(a = 3, b = 5\) how large should \(n\) be so that we can detect

\(\chi_1 = -0.25 \sigma\); \(\chi_2 = -0.25 \sigma\); \(\chi_3 = 0.50 \sigma\)

with probability \(0.75 (\alpha = 0.05)\).

Try \(n\), calculate \(\phi\), enter the tables and find the power. After successive trials \(n\) will be obtained to give the required power. (d.f. for the numerator \(= 2\); for the denominator \(= ab(n-1) = 15(n-1)\). Verify that \(n = 13\).

**Randomized blocks:**

<table>
<thead>
<tr>
<th>(X_{ij})</th>
<th>(\mu_{ij}) are (N(\mu_{ij}, \sigma^2))</th>
<th>(i = 1, 2, \ldots, a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu_{ij})</td>
<td>(\mu + \chi_i + b_j)</td>
<td>(j = 1, 2, \ldots, b)</td>
</tr>
</tbody>
</table>

\(b_j\) are \(N(0, \sigma_b^2)\) (this is the additional assumption of randomized blocks)

or \(X_{ij} = \mu + \chi_i + b_j + \varepsilon_{ij}\) where \(\varepsilon_{ij} \sim N(0, \sigma^2)\)

\(b_j \sim N(0, \sigma_b^2)\)

and they are independent.

\(H_0: \) all \(\chi_i = 0\)

\(H_1: \) some \(\chi_i \neq 0\)

Given \(b_j, \bar{X}_i\), are \(N(\mu + \chi_i + \frac{\sum b_j}{b}; \sigma^2 / b)\)

If \(H_0\) is true, \(\sqrt{b} \bar{X}_i\) is \(N(\mu, \sigma^2)\), and

\[\sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij} - \bar{X}_{..})^2 = b \sum_{i=1}^{a} (\bar{X}_{i..} - \bar{X}_{..})^2\]

has a \(\chi^2\)-distribution with \(a-1\) d.f. This conditional (on the \(b_j\)) distribution does not involve the \(b_j\), therefore the unconditional distribution of

\[\sum_{i=1}^{a} \sum_{j=1}^{b} (X_{ij} - \bar{X}_{..})^2\]

is \(\chi^2(a-1)\).
In general, given $b_j$,

$$
\sum \sum (X_{ij} - \bar{X}_{i.} - \bar{X}_{..} + \bar{X}.)^2 \text{ is } \chi^2 (a-1)(b-1).
$$

Again we have the same result for the unconditional distribution.

Hence a test of $H_0$ is based on the statistic:

$$
\frac{\sum \sum (\bar{X}_{i.} - \bar{X}_{..})^2 / (a-1)}{\sum \sum (X_{ij} - \bar{X}_{i.} - \bar{X}_{..} + \bar{X}.)^2 / (a-1)(b-1)}
$$

which has the F-distribution with $(a-1), (a-1)(b-1)$ d.f.

If $H_0$ is false, $\sqrt{b_i} \bar{X}_{i.}$ is $N(\mu_i' + \sqrt{b_i} \alpha_i, \sigma^2)$. $\sum \sum (\bar{X}_{i.} - \bar{X}_{..})^2$ has a non-central $\chi^2$-distribution with parameter $b_i \sum \alpha_i^2$, d.f. $a-1$.

The Power is also calculated as if the $b_i$s were fixed parameters.

**Random model:** (Hierarchical Classification or Nested Sampling)

$$
X_{ij} | \mu_{ij} \text{ is } N(\mu_{ij}, \sigma^2) \quad i = 1, 2, \ldots, a \quad j = 1, 2, \ldots, n
$$

$$
\mu_{ij} = \mu + a_i
$$

where the $a_i$ are $N(0, \sigma_a^2)$

or $X_{ij} = \mu + a_i + \epsilon_{ij}$

**ANOVA Table**

<table>
<thead>
<tr>
<th></th>
<th>d.f.</th>
<th>$S_2 S_0$</th>
<th>$E(\text{MS})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$'s</td>
<td>$a-1$</td>
<td>$\sum \sum (\bar{X}<em>{i.} - \bar{X}</em>{..})^2$</td>
<td>$\sigma^2 + n \sigma_a^2$</td>
</tr>
<tr>
<td>error</td>
<td>$a(n-1)$</td>
<td>$\sum \sum (X_{ij} - \bar{X}_{i.})^2$</td>
<td>$\sigma^2$</td>
</tr>
</tbody>
</table>

Given the $a_i$, $X_{ij}$ is $N(\mu + a_i, \sigma^2)$

$$
\sum \sum (\bar{X}_{ij} - \bar{X}_{i.})^2 / \sigma^2 \text{ has a } \chi^2 \text{-distribution with } a(n-1) \text{ d.f.}
$$
Since this is conditional on the $\xi_i$, but independent of them, it is also the unconditional distribution.

$\xi_i$ is $N(\mu, \frac{\sigma^2}{n} + \sigma_a^2)$

--- unconditional distribution.

e.g., in terms of moment generating functions

$$E_{a_i} \left[ \frac{\xi_i t}{a_i} \right] = e^{\mu t + \frac{t^2}{2} \left[ \frac{\sigma^2}{n} + \sigma_a^2 \right]}$$

$$n \sum_{i=1}^{a} (\bar{x}_i - \bar{x}_o)^2$$

is $\chi^2$ with a - 1 d.f.

$$\frac{\sum \sum (\bar{x}_{ij} - \bar{x}_o)^2}{\sum \sum (\bar{x}_{ij} - \bar{x}_i)^2} \cdot \frac{\sigma^2}{\sigma^2 + n \sigma_a^2}$$

has an $F$-distribution with $\left[ (a-1), a(n-1) \right]$ d.f.

**Problem 75:** $X_{ijk}$ is $N(\mu_{ij}, \sigma^2)$

$i = 1, 2, \ldots, a$

$j = 1, 2, \ldots, b$

$k = 1, 2, \ldots, n$

a) $\mu_{ij} = \mu + \xi_i + b_{ij}$

where $b_{ij}$ are $N(0, \sigma_b^2)$

b) $\mu_{ij} = \mu + a_i + b_{ij}$

where $b_{ij}$ are $N(0, \sigma_b^2)$

$a_i$ are $N(0, \sigma_a^2)$

and the $a$'s and $b$'s are independent.

For $n = 2$, $b = 2$, $a = 5$, $\alpha = 0.05$, plot the power of the tests:

1) in (a) against

$$\frac{\sum \xi_i^2}{\sigma^2 + 2\sigma_a^2}$$

2) in (b) against

$$\frac{\sigma_a^2}{\sigma^2 + 2\sigma_b^2}$$
Multiple Comparisons:

Tests of all contrasts with the general linear hypothesis model following ANOVA.

Tukey's Procedure:

observations: $X_1, X_2, \ldots, X_n$ are NID(0, $\sigma^2$)

Let $R = \max \left\{ \frac{X_i}{\hat{\sigma}_i} \right\} - \min \left\{ \frac{X_i}{\hat{\sigma}_i} \right\}$

for $1 \leq i \leq n$

$s_e^2$ is an estimate of $\sigma^2$ which is independent of $X_1, X_2, \ldots, X_n$ with $f$ d.f.

(i.e., $f s_e^2/\sigma^2$ has a $\chi^2$-distribution with $f$ d.f.)

The distribution of $R/\sigma$ can be found in a general form, and in particular when the $X$'s are normal (since the $X_i/\sigma$ are N(0, 1) -- it is free of any parameters.

The distribution of $R^t = \frac{R}{s_e} = \frac{R/\sigma}{s_e/\sigma} = \text{Studentized Range}$ has been found by numerical integration and percentage points have been tabulated by Hartley and Pearson in Biometrika Tables, Table 29.

Consider the one-way ANOVA situation:

$X_{ij}$ are NID($\mu_i$, $\sigma^2$) \hspace{1cm} i = 1, 2, \ldots, a

$X_{i.}$ are NID($\mu_i$, $\sigma^2$) \hspace{1cm} j = 1, 2, \ldots, n_i = n$

$$\text{MSE} = \frac{\sum \sum (X_{ij} - \bar{X}_{i.})^2}{a(n-1)}$$

is an independent estimate of $\sigma^2$ with a(n-1) d.f.

Consider $H_0$: $\mu_i = \mu_k$ or $\mu_i = \mu_k = 0$

Given $H_0$:

$$\text{Pr} \left[ \frac{\sqrt{n} \left( \frac{\bar{X}_{i.} - \bar{X}_{k.}}{\text{MSE}^{1/2}} \right)}{\sqrt{\sum \sum (X_{ij} - \bar{X}_{i.})^2}} > t \right] \leq \text{Pr} \left[ R^t > t \right] = \alpha$$

This $t$ is the statistic $Q(a, f)$ in Snedecor, section 10.6, p. 251 where $f = \text{d.f.} = a(n-1)$ in this case.

This inequality will hold for any possible pairwise comparison (i.e., any $i, k$).
Frame of reference is the total experiment, not the individual pairwise tests -- i.e., for the total experiment and making all possible pairwise tests, the probability of the Type I error is \( \leq \alpha \), where \( \alpha \) is the level chosen for the \( Q(a,f) \) statistic.

Hence we reject \( H_0: \mu_i = \mu_k = 0 \) for any \( i, k \) when

\[
\frac{\sqrt{n} (\bar{X}_i - \bar{X}_k)}{\sqrt{\text{MSE}}} > Q_{\alpha} (a,f)
\]

Scheffe Test for all contrasts (most general of all).

Ref: Scheffe, Biometrika, June 1952

\( X_{ij} \) are NID(\( \mu_i \), \( \sigma^2 \)) \( i = 1, 2, \ldots, a \)

\( j = 1, 2, \ldots, n_i \)

\( s_e^2 \) is an estimate of \( \sigma^2 \) which is independent of \( \bar{X}_i \). with \( f \) d.f.

\( H_0(c): \sum_{i=1}^{a} c_i \mu_i = 0 \quad \sum_{i=1}^{a} c_i \mu_i = 0 \)

Consider the totality of all such \( H_0(c) \).

Theorem 38: If the \( \mu_i \) are all equal, then

\[
Pr \left[ \frac{\left\{ \sum_{i=1}^{a} c_i (\bar{X}_i - \bar{X}_*) \right\}^2}{s_e^2 \sum_{i=1}^{a} c_i^2 / n_i} > t \right] > (a-1) \cdot C \left( \frac{a-1}{F^2} \right)^{t \left( \frac{a-1+1}{2} \right)} \left( 1 + \frac{a-1}{F} \right)^{\frac{a-1+f}{2}} \text{d}F
\]

where \( C = \left( \frac{a-1}{f} \right)^{\frac{a-1}{2}} \frac{\Gamma \left( \frac{a-1+f}{2} \right)}{\Gamma \left( \frac{a-1}{2} \right) \Gamma \left( \frac{f}{2} \right)} \)

hence, if we reject any such \( H_0(c) \) when

\[
\frac{\left\{ \sum_{i=1}^{a} c_i (\bar{X}_i - \bar{X}_*) \right\}^2}{s_e^2 \sum_{i=1}^{a} c_i^2 / n_i} > (a-1) F_{a-l, f(l-\alpha)}
\]

then the type I error \( \leq \alpha \).
Proof: Under $H_0$, $\sum c_i \bar{x}_i = 0$ and $\sum c_i \mu_i = 0$. For all $c_i$ with $\sum c_i = 0$;

$$\Pr\left\{ \frac{\left( \sum c_i \bar{x}_i \right)^2}{\sum c_i / n_i} > t \right\} \leq \Pr\left\{ \max_{c_1 \sum c_i = 0} \frac{\left( \sum \frac{\sqrt{n_i} c_i}{\sqrt{n_i}} (\bar{x}_i - \mu_i) \right)^2}{\sum c_i^2 / n_i} > t \right\}$$

Define:

$$y_i = \frac{\sqrt{n_i} (\bar{x}_i - \mu_i)}{\sigma}$$

$y_i$ are NID(0, 1)

$i = 1, 2, \ldots, a$

$$d_i = \frac{c_i / \sqrt{n_i}}{\sum c_i^2 / n_i} \quad \sum c_i = \sum \sqrt{n_i} d_i = 0$$

$$\sum d_i^2 = 1$$

Expression we wish to maximize is

$$\frac{\left( \sum d_i y_i \right)^2}{\sum c_i^2 / n_i}$$

with respect to $d_i$

$$\phi = \left( \sum d_i y_i \right)^2 = \lambda_1 \sum \sqrt{n_i} d_i - \lambda_2 \sum d_i^2$$

$$\frac{\partial \phi}{\partial d_j} = 2 \left[ \sum d_i y_i \right] y_j - \lambda_1 \sqrt{n_j} - 2 \lambda_2 d_j = 0 \quad j = 1, 2, \ldots, a$$

Multiply [1] by $\sqrt{n_j}$ and sum over $j$;

$$2 \left[ \sum \sqrt{n_j} y_j \right] \left[ \sum d_i \bar{y}_i \right] = \lambda_1 \text{ since } \sum \sqrt{n_i} d_i = 0$$

Multiply [1] by $d_j$ and sum over $j$;

$$2 \left[ \sum d_j y_j \right] \left[ \sum d_i \bar{y}_i \right] = 0 - 2 \lambda_2 (1) = 0 \text{ since: i) } \sum \sqrt{n_i} d_i = 0$$

$$\lambda_2 = \left[ \sum d_i \bar{y}_i \right]^2$$
put $\lambda_1, \lambda_2$ in $[1]$ to obtain the $j$th equation:

$$2y_j \left[ \sum d_i y_i \right] = 2 \frac{\sqrt{n_j} \left[ \sum n_i y_i \right] \left[ \sum d_i y_i \right]}{\sum n_j} - 2d_j \left[ \sum d_i y_i \right] = 0$$

$$y_j = \frac{\sqrt{n_j} \sum n_i y_i}{\sum n_i} = d_j \sum d_i y_i$$

multiply this by $y_j$ and sum over $j$:

$$\left[ \sum d_j y_j \right]^2 = \sum y_j^2 - \frac{\left[ \sum \sqrt{n_j} y_j \right]^2}{\sum n_j}$$

in the $n_j$ are equal, this $= \sum y_j^2 - \frac{\left[ \sum y_j \right]^2}{a} = \sum_{j=1}^{a} (Y_j - \bar{Y})$

Recall:

$$\Pr \left\{ \frac{\left[ \sum_{i=1}^{a} c_i \bar{x}_i \right]^2}{s_e^2 \sum n_i} > t \right\}$$

$$\max_{c_i} \left\{ \frac{\sum c_i \bar{x}_i^2}{s_e^2 \sum \frac{c_i}{n_i}} > t \right\}$$

$$= \Pr \left\{ \frac{\sum y_i^2 - \frac{\sum n_i y_i}{\sum n_i}^2}{s_e^2 \sigma^2} > t \right\}$$

$$y_i = \sqrt{n_i} (\bar{x}_i - \mu_i)$$

and are $\text{NID}(0, 1)$. 
We can make an orthogonal transformation of the form:

\[ z_1 = \frac{\sqrt{\sum_{i=1}^{a} n_i y_i}}{(\sum n_i)^{1/2}} \]

\[ z_2 = \ldots \]

\[ z_a = \ldots \]

Then

\[ \sum y_i^2 - \frac{(\sum \sqrt{n_i y_i})^2}{\sum n_i} \rightarrow \sum_{j=1}^{a} z_j^2 - z_1^2 = \sum_{j=2}^{a} z_j^2 \]

Hence

\[ \sum y_i^2 - \frac{(\sum \sqrt{n_i y_i})^2}{\sum n_i} \] has a \( \chi^2 \)-distribution with \((a-1)\) d.f.

Thus

\[ \Pr \left[ Z^2 \right] = \Pr \left[ \frac{\sum_{j=2}^{a} z_j^2 / a-1}{\frac{\sum_{j=2}^{a} z_j^2}{s_e^2 / \sigma^2} > t / a-1} \right] \]

but

\[ \frac{\sum z_j^2 / a-1}{s_e^2 / \sigma^2} \] has an \( F \)-distribution with \((a-1, f)\) d.f.

Therefore the type I error will be less than or equal to \( \alpha \) if we reject any \( H_0(c) \) if:

\[ \frac{\left( \sum_{i} c_i x_i \right)^2}{s_e^2 \sum c_i / n_i} > (a-1) F_{1-\alpha} (a-1, f) \]
5. \( \chi^2 \)-Tests

Single series of trials: Simple hypothesis

classes (or events): \( 1 \ 2 \ \ldots \ k \)
probabilities: \( p_1 \ p_2 \ \ldots \ p_k \ \sum p_j = 1 \)
observations: \( v_1 \ v_2 \ \ldots \ v_k \ \sum v_j = n \)

\( H_0: \ p_j = p_j^0 \)

If \( H_0 \) is true,

\[
X_j = \frac{v_j - np_j^0}{\sqrt{np_j^0}} \quad \text{is asymptotically normal subject to the restriction,}
\]

\[
\sum \sqrt{p_j^0} X_j = 0
\]

By an orthogonal transformation it can be shown that

\[
\sum_{j=1}^{k} X_j^2 = \sum_{j=1}^{k} \left( \frac{v_j - np_j^0}{np_j^0} \right)^2
\]

has an asymptotic \( \chi^2 \)-distribution with \( k-1 \) d.f. as \( n \to \infty \).

The power of the test \( \to 1 \) as \( n \to \infty \).

Power of a simple \( \chi^2 \)-test:

\( H_1: \ p_j = p_j' \quad p_j' - p_j^0 = \frac{c_j}{\sqrt{n}} \)

\[
X_j = \frac{v_j - np_j^0}{\sqrt{np_j^0}} = \frac{v_j - np_j^0}{\sqrt{np_j^0}} \sqrt{\frac{p_j'}{p_j^0}} + \sqrt{\frac{n(p_j' - p_j^0)}{np_j^0}}
\]

\[
= Y_j e_j + \frac{c_j}{\sqrt{p_j^0}}
\]

now \( e_j = \sqrt{\frac{p_j'}{p_j^0}} = \sqrt{1 + \frac{p_j' - p_j^0}{p_j^0}} \to 1 + \frac{c_j}{\sqrt{np_j^0}} \)

Under \( H_1 \) the \( Y_j \) are asymptotically normal.
With the same type of orthogonal transformation, we have $\sum x_j^2$ under $H_1$ is a sum of squares of $Y_j + \frac{\alpha_j}{\sqrt{p_j}}$, and hence has a non-central $\chi^2$-distribution with parameters

$$k-1, \sum_{j=1}^{k} \frac{\alpha_j^2}{p_j}.$$ 

The non-centrality parameter can also be written as

$$n \sum_{j=1}^{k} \frac{(p_j^1 - p_j^0)^2}{p_j^0}.$$ 

**Example:** $v_i$ = no. of families with $i$ boys in families of 2 children.

Assume the number of boys is binomially distributed with parameter $p$.

$H_0$: $p = 0.5$  
$H_1$: $p = 0.4$

<table>
<thead>
<tr>
<th>No. boys</th>
<th>$p_j^0$ under $H_0$</th>
<th>$p_j^1$ (under $H_1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.16</td>
</tr>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.48</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.36</td>
</tr>
</tbody>
</table>

$$\lambda = n \sum_{j=1}^{k} \frac{(p_j^1 - p_j^0)^2}{p_j^0} = n \left[ \frac{0.09^2}{0.25} + \frac{0.02^2}{0.50} + \frac{0.11^2}{0.25} \right]$$

$$\lambda = 0.816$$

For $n = 100$  
$\lambda = 8.16$

The power can be determined from the Tables of the Non-Central $\chi^2$ by E. Fix, University of California Publications in Statistics, Volume I, No. 2, pp. 15-19.

From the tables, for $\lambda = 8.16$, $k-1 = f = 2$, $\alpha = 0.05$

$$0.7 < \beta \text{ (power)} < 0.8$$

$\chi^2$-test may also be a one-sided test:

$H_0$: $p_j = p_j^0$  
$H_1$: $p_j > p_j^0$ for $j \leq k'$ ($k'$ exceed $p_j^0$)

$p_j < p_j^0$ for $j > k'$ ($k - k'$ fall below $p_j^0$)
\( \chi^2 \)-test is modified as follows:

for \( j \leq k' \):
  if \( v_j > E(v_j) = np_j \)
  calculate \( \frac{(v_j - np_j)^2}{np_j} \) as usual

  if \( v_j < np_j \)
  put the contribution = 0

for \( j > k' \):
  if \( v_j > np_j \)
  put the contribution = 0

  if \( v_j < np_j \)
  calculate \( \chi^2 \) as usual

\( \chi^2 \) is rejected if the sample value > \( \chi_{1-2\alpha(k-1)}^2 \)

Example:

<table>
<thead>
<tr>
<th></th>
<th>8(10)</th>
<th>12(10)</th>
<th>20</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8(10)</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>12(10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( H_0 \): \( p = 0.5 \)

\( \chi^2 = 0 \) therefore we fail to reject \( H_0 \)

Note: if \( v_{11} > 10 \) we would then calculate \( \chi^2 \).

Refs: (on the \( \chi^2 \)-test):
1954 Biometrics

\( \chi^2 \)-test of a composite hypothesis:

classes:

\[
\begin{array}{cccc}
1 & 2 & \ldots & k \\
\end{array}
\]

series of observations:

\[
\begin{array}{cccc}
v_{11} & v_{12} & \ldots & v_{1k} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
v_{s1} & v_{s2} & \ldots & v_{sk} \\
\end{array}
\]

\[
\sum_{j=1}^{k} v_{ij} = n_i
\]

probabilities:

\[
\begin{array}{cccc}
p_{11} & p_{12} & \ldots & p_{1k} \\
\vdots & \vdots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
p_{s1} & p_{s2} & \ldots & p_{sk} \\
\end{array}
\]

\[
p_{ij} = f(\theta)
\]
\[ x_{ij} = \frac{v_{ij} - n_{i}p_{ij}(\theta)}{\sqrt{n_{i}p_{ij}(\theta)}} \]

\[ \sum_{j} \sqrt{p_{ij}} x_{ij} = 0 \quad \text{for } i = 1, 2, \ldots, s \]

\[ H_{0}: \quad p_{ij} = p_{ij}^{0}(\theta) \]

**Theorem 39**: If \( \theta \) is estimated by m.l.e. or any asymptotically equivalent method then as \( n \to \infty \), under the regularity conditions as for estimation:

\[ \chi^{2} = \sum_{i} \sum_{j} \left( \frac{v_{ij} - n_{i}p_{ij}(\theta)}{n_{i}p_{ij}(\theta)} \right)^{2} \]

has a \( \chi^{2} \)-distribution with \( s(k-1)-t \) d.f. when \( H \) is true. (\( t = \text{no. of components in } \theta \))

If \( H_{1} \) is true: \( p_{ij} = p_{ij}^{1} = p_{ij}^{0} + \frac{c_{ij}}{\sqrt{n_{i}}} \) and \( \rho = \frac{n_{i}}{\sum n_{i}} \)

then \( \chi^{2} \) has a non-central \( \chi^{2} \)-distribution with d.f. as before, and non-centrality parameter \( \delta^{2} \left[ I - B(B^{t}B)^{-1} B^{t} \right] \delta \)

where \( \delta^{t} x_{ks} = \left( \frac{c_{ij} \sqrt{p_{ij}}}{\sqrt{p_{ij}}} \right) \) \( i = 1, 2, \ldots, s \)

\( j = 1, 2, \ldots, k \)

\( \ell = 1, 2, \ldots, t \)

\[ B_{ks \times t} = \left[ \sqrt{p_{ij}} \left( \frac{\partial p_{ij}}{\partial \theta_{\ell}} \right)_{\theta_{\ell} = \theta_{\ell}^{0}} \right] \]

**Ref:** Cramer: \( \chi^{2} \) section

**Mitra:** December 1958 Annals of Math Stat

Note on the general $\chi^2$ tests of composite hypotheses:

$$X_{ij} = \frac{v_{ij} - n_1 p_{ij}}{\sqrt{n_1 p_{ij}}}$$

are asymptotically normal subject to the linear restrictions that

$$\sum_j \sqrt{p_{ij}} x_{ij} = 0 \quad \text{for all } i = 1, 2, \ldots, s \quad [1]$$

For sufficiently large $n$ the m.l.e. reduces to solving the equation:

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial \ln L}{\partial \theta} \bigg|_{\theta = \theta_0} + \left( \frac{\partial^2 \ln L}{\partial \theta^2} \bigg|_{\theta = \theta_0} \right) (\theta - \theta_0) \quad [2]$$

Recall:

$$L = K \prod_i \prod_j \left[ p_{ij}(\theta) \right]^{v_{ij}}$$

$$\ln L = K' + \sum_i \sum_j v_{ij} \ln p_{ij}(\theta)$$

$\ln L$ is linear in the $v_{ij}$, as are $\frac{\partial^h \ln L}{\partial \theta^h}$, $h = 1, 2$

Hence the estimation of $\Theta$ (single component) means one additional linear restriction on the $v_{ij}$ or on the $X_{ij}$ (since the two are linearly related).

If this restriction [2] is linearly independent of the $s$ restrictions in [1], then we can transform the $s+1$ restrictions to $s+1$ orthogonal linear restrictions and then by the usual extension get an orthogonal transformation that takes

$$\sum_i \sum_j X_{ij}^2$$

into

$$\sum_i \sum_j Y_{ij}^2$$

with $s(k-1) - 1 = sk -(s+1)$ terms and hence we get the result of the general $\chi^2$ theorem (theorem 39).

Example: Power of a $\chi^2$ test of homogeneity, $2 \times 2$ table.

<table>
<thead>
<tr>
<th>Probabilities:</th>
<th>$P_{11}$</th>
<th>$P_{12} = 1 - P_{11}$</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{21}$</td>
<td>$P_{22} = 1 - P_{21}$</td>
<td>1</td>
</tr>
</tbody>
</table>

$H_0$: $P_{11} = P_{21} = \theta$

$P_{12} = P_{22} = 1 - \theta$

$H_1$: $P_{11} = \theta + c/\sqrt{n}$

$P_{12} = 1 - \theta - c/\sqrt{n}$

$P_{21} = \theta - c/\sqrt{n}$

$P_{22} = 1 - \theta + c/\sqrt{n}$
\[
\lambda = \delta' \left[ I - B(B'B)^{-1}B' \right] \delta
\]

Take \( \rho_1 = \rho_2 = \frac{1}{2} \) (sampling fraction)

\[
B' = \left( \begin{array}{ccc}
\frac{1}{\sqrt{2} \Theta} & -1 & \frac{1}{\sqrt{2} \Theta} \\
-1 & \frac{1}{\sqrt{2(1-\Theta)}} & \frac{1}{\sqrt{2(1-\Theta)}} \\
\frac{1}{\sqrt{2(1-\Theta)}} & \frac{1}{\sqrt{2(1-\Theta)}} & \frac{1}{\sqrt{2(1-\Theta)}} \\
\end{array} \right)
\]

\[
B'B = \frac{1}{\Theta} + \frac{1}{1-\Theta} = \frac{1}{\Theta(1-\Theta)}
\]

\[
(B'B)^{-1} = \Theta(1-\Theta)
\]

\[
B(B'B)^{-1}B' = \left( \begin{array}{ccc}
\frac{1-\Theta}{2} & -\Theta(1-\Theta) & \frac{1-\Theta}{2} \\
-\Theta(1-\Theta) & \Theta & -\Theta(1-\Theta) \\
\frac{1-\Theta}{2} & -\Theta(1-\Theta) & \frac{1-\Theta}{2} \\
\end{array} \right)
\]

\[
\delta' = \left( \begin{array}{ccc}
c & -\Theta & c \\
\frac{\sqrt{2}}{\Theta} & \frac{\sqrt{2(1-\Theta)}}{\Theta} & \frac{\sqrt{2}}{\Theta} \\
\frac{\sqrt{2}}{\Theta} & \frac{\sqrt{2(1-\Theta)}}{\Theta} & \frac{\sqrt{2}}{\Theta} \\
\end{array} \right)
\]

\[
\delta' I \delta = \frac{c^2}{\Theta} + \frac{c^2}{1-\Theta} = \frac{c^2}{\Theta(1-\Theta)}
\]

\[
\delta' \left[ B(B'B)^{-1}B' \right] \delta = 0 \quad \text{(patterns of + and - signs in } \delta, B \text{ are opposite and sufficient to cancel all terms.)}
\]

Hence the non-centrality parameter, \( \lambda = \frac{c^2}{\Theta(1-\Theta)} \)

Which can be expressed in terms of the \( p \)'s as:

\[
\Theta = \frac{p_{11} + p_{21}}{2} \quad 1 - \Theta = \frac{p_{12} + p_{22}}{2}
\]

\[
p_{11} - p_{21} = \frac{2c}{\sqrt{n}} \quad c^2 = \frac{n}{4} (p_{11} - p_{21})^2
\]
\[ \lambda = \frac{n(p_{11} - p_{21})^2}{(p_{11} + p_{21})(p_{12} + p_{22})} \]

In the December 1958 Annals of Math Stat Mitra gives the following formula for \( \lambda \) in the 2 x k contingency table:

\[
\begin{align*}
\text{probabilities} & \quad p_{11} & \quad p_{12} & \quad \ldots & \quad p_{1k} \\
& \quad p_{21} & \quad p_{22} & \quad \ldots & \quad p_{2k}
\end{align*}
\]

\[ H_0: \quad p_{1j} = p_{2j} = \theta_j \quad \sum \theta_j = 1 \]
\[ H_1: \quad p_{1j} = \theta + c_{1j}/\sqrt{n} \quad \sum c_{1j} = 0 \]
\[ p_{2j} = \theta + c_{2j}/\sqrt{n} \quad \sum c_{2j} = 0 \]

\[ \lambda = \rho_1 \rho_2 \sum_{j=1}^{k} \frac{(c_{1j} - c_{2j})^2}{\theta_j} \]

For the above example: \( \rho_1 = \rho_2 = \frac{1}{2} \quad c_{1j} = c \quad c_{2j} = -c \)

\[ \lambda = \frac{1}{4} \left( \frac{(2c)^2}{\theta} + \frac{(2c)^2}{1-\theta} \right) = \frac{c^2}{\theta(1-\theta)} \]

6. Other Approaches to Testing Hypotheses and other problems:

1. Most Stringent Tests:

Let the family of tests of size \( \alpha \) be \( \Phi(\alpha) \).

Let \( \bar{\beta}(\theta) = \sup_{\phi \in \Phi(\alpha)} \beta_{\phi}(\theta) \)
\( \emptyset (x) \) is most stringent if:

i) it is of size \( \prec \)

ii) for any other size \( \prec \) test \( \emptyset' \)

\[
\sup_\emptyset \left[ \beta_\emptyset (x) - \beta_{\emptyset'} (x) \right] \leq \sup_\emptyset \left[ \beta_{\emptyset} (x) - \beta_{\emptyset'} (x) \right]
\]

2. Minimax Tests: Decision Theory Approach

\( L \left[ D(x), \emptyset \right] \) = the loss when \( D(x) \) is the decision made and \( \emptyset \) is the true value of the parameter.

Example:
\( X \) is \( N(\mu, 1) \)
\( H_0: \mu = 0 \)
\( D(x) = 1 \)
\( D(x) = 0 \) means accept \( H_0: \mu = 0 \)

\( L \left[ D(x), \mu \right] = c \mu^2 \) when \( D(x) = 0 \)
\( L \left[ D(x), \mu \right] = c/\mu \) when \( D(x) = 1 \)

Having set up a loss function, we then have a risk function defined as:

\[
r_D(\emptyset) = E(L) = \int_{-\infty}^{\infty} L \left[ D(x), \emptyset \right] f(x, \emptyset) \, dx
\]

It is frequently impossible to minimize this universally with respect to \( \emptyset \), thus:

\( D(x) \) is a minimax decision rule if it minimizes (with respect to all possible decision rules).

\[
\sup_\emptyset r(\emptyset)
\]

or expressed another way we determine:

\[
\inf_D \sup_\emptyset r_D (\emptyset)
\]

ref: Blackwell and Girshick, Theory of Games, Wiley

3. Admissible Decision Rules:

\( D(x) \) is admissible if there is no \( D'(x) \) such that

\[
r_{D'}(\emptyset) \leq r_D(\emptyset) \quad \text{for all } \emptyset
\]

\[
r_{D'}(\emptyset) < r_D(\emptyset) \quad \text{for some } \emptyset
\]

i.e., you cannot improve on \( D(x) \) uniformly.
A. Confidence Regions:

\( X \) is a random variable with d.f. \( f(x, \theta) \)
\( \theta \in \Omega \)

\( B(\omega) = \) the totality of subsets of \( \Omega \)

Let \( A(\bar{x}) \) be a function from \( X \) (sample space) to \( B(\omega) \)

Def. 4.7: If \( \Pr[A(\bar{x}) \supseteq \Theta] > 1 - \alpha \) then \( A(\bar{x}) \) is a confidence region with confidence coefficient \( 1 - \alpha \).

Theorem 4.0: If a non-randomized test of size \( \alpha \) exists for \( H_0: \theta = \theta_0 \) for every \( \theta_0 \), then there exists a confidence region for \( \theta \) of size \( 1 - \alpha \).

Proof: Let \( R(\bar{x}, \theta_0) \) be the set of \( \bar{x} \) for which \( H_0: \theta = \theta_0 \) is rejected, e.g.,

\[
\int_{R(\bar{x}, \theta_0)} f(\bar{x}, \theta_0) \, d\bar{x} \leq \alpha
\]

This \( R(\bar{x}, \theta_0) \) is defined and satisfies [1] for every \( \theta_0 \).

Define: \( A(\bar{x}) = \bigcup_{\theta} X \notin R(\bar{x}, \theta) \)

\[
\Pr[A(\bar{x}) \supseteq \Theta_0] = \Pr[\bar{x} \notin R(\theta)] = 1 - \Pr[\bar{x} \in R(\theta)] \geq 1 - \alpha \text{ q.e.d.}
\]

Def. 4.8: A confidence region, \( A(\bar{x}) \), is uniformly most powerful if \( \Pr[A(\bar{x}) \supseteq \theta \mid \theta' \text{ is true}] \) is minimized for all \( \theta' \neq \theta \).

Note: Kendall uses this same definition with "uniformly most powerful" replaced with "uniformly most selective." Neyman replaces "u.m.p." with "shortest."
Def. 12: A confidence interval is "shortest" if the length of the interval is minimized uniformly in $\theta$.

Confidence Interval for the ratio of two means:

$X_i, Y_i$ are bivariate normal $N(\mu, \sigma^2 X, \sigma^2 Y, \rho)$ for $i = 1, 2, \ldots, n$

where $\rho$ = correlation coefficient and may = 0.

Problem: find a confidence interval for $\kappa = \frac{E(Y)}{E(X)}$

$Z_i = \frac{X_i - Y_i}{\sigma X \sigma Y}$ is $N(0, \kappa^2 \sigma^2 X - 2\kappa \sigma X \sigma Y + \sigma^2 Y)$

$\sum (Z_i - \bar{Z})^2$ and $\bar{Z}$ are independent and distributed as $x^2$ and normal respectively.

$$\frac{\sum (Z_i - \bar{Z})^2}{n-1} = \frac{\sum (X_i - Y_i)^2 - n(\bar{X} - \bar{Y})^2}{n-1}$$

$$= \frac{\kappa^2 \sum (X_i - \bar{X})^2 - 2\kappa \sum (X_i - \bar{X})(Y_i - \bar{Y}) + \sum (Y_i - \bar{Y})^2}{n-1}$$

$$= \frac{\kappa^2 s^2_x - 2\kappa s_{xy} + s^2_y}{n-1}$$

$$\bar{Z} = \frac{\bar{X} - \bar{Y}}{\sqrt{n}}$$

therefore:

$$\frac{\sqrt{n} (\bar{X} - \bar{Y})}{(\kappa^2 s^2_x - 2\kappa s_{xy} + s^2_y)^{1/2}}$$

has a t-distribution with $n-1$ d.f.

$$Pr \left[ \frac{\sqrt{n} (\bar{X} - \bar{Y})}{(\kappa^2 s^2_x - 2\kappa s_{xy} + s^2_y)^{1/2}} \leq t \right] = 1 - \varepsilon$$

We can solve this and get confidence limits for $\kappa$. This yields the following quadratic equation in $\kappa$:

$$Q(\kappa) = \kappa^2 (n\bar{X}^2 - t^2 s^2_x) - 2\kappa (n\bar{X}\bar{Y} - t^2 s_{xy}) + (n\bar{Y}^2 - t^2 s^2_y) = 0$$
Among the various possible solutions of this quadratic equation we can have the following situations:

A. If $n\overline{x}^2 - t^2 s_x^2 > 0$:

It can be shown that 2 roots always exist and this the confidence interval is:

$$\alpha_1 < \bar{x} < \alpha_2$$

For the above inequality to hold, $\overline{x}$ must be significantly away from the origin; i.e., we must reject $H_0: \mu = 0$ on the basis of $X_1, X_2, \ldots, X_n$ at the $\varepsilon$ level. e.g., the following condition must hold:

$$\frac{\sqrt{n} |\overline{x}|}{s_x} > t_1 - \frac{\varepsilon}{2}$$

Note: One could, if desired, change $t$ (and thus $\varepsilon$) to insure that this inequality always held and thus that a "real" confidence interval exists.

B. If $n\overline{x}^2 - t^2 s_x^2 < 0$

$\alpha_1, \alpha_2$ may be real or complex

If the roots are real, then the "confidence interval" is $(-\infty, \alpha_1), (\alpha_2, +\infty)$, i.e., we accept "in the tails."

C. If

$$4(n\overline{xy} - t^2 s_{xy})^2 < 4(n\overline{x}^2 - t^2 s_x^2)(n\overline{y}^2 - t^2 s_y^2),$$

i.e., $b^2 < 4ac$, then the roots are complex and we have a confidence interval $(-\infty, +\infty)$.

[Thus we can say that $-\infty < \bar{x} < +\infty$ with confidence $1 - \varepsilon$.]

D. If we get

$$Q(\alpha) \leq 0 \iff \frac{\sqrt{n} |\overline{x} - \overline{y}|}{(\overline{x}^2 - 2\overline{xy} + \overline{y}^2)s_{xy}^{1/2}} \leq t$$

thus we always accept $H_0$ since the test statistic never exceeds $t$.

Refs: Fieller, Journal of the Royal Statistical Society, 1940
Paulson, Annals of Math Stat, 1942
Bennett, Sankhya, 1957
## LIST OF THEOREMS

<table>
<thead>
<tr>
<th>Theorem</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Properties of distribution functions</td>
<td>5</td>
</tr>
<tr>
<td>2. 1-1 Correspondence of distribution function and probability measure</td>
<td>6</td>
</tr>
<tr>
<td>3. Discontinuities of a distribution function</td>
<td>6</td>
</tr>
<tr>
<td>4. Multivariate distribution function properties</td>
<td>7</td>
</tr>
<tr>
<td>5. Addition law for expectations of random variables</td>
<td>22</td>
</tr>
<tr>
<td>6. Multiplication law for expectations of random variables</td>
<td>22</td>
</tr>
<tr>
<td>7. Conditions for unique determination of a distribution function by moments (Second Limit Theorem)</td>
<td>23</td>
</tr>
<tr>
<td>8. Tchebycheff's Inequality</td>
<td>24</td>
</tr>
<tr>
<td>9. Bernoulli's Theorem (Law of large numbers for binomomial random variables)</td>
<td>25</td>
</tr>
<tr>
<td>10. Characteristic function of a linear function of a random variable</td>
<td>28</td>
</tr>
<tr>
<td>11. Characteristic function of a sum of independent random variables</td>
<td>29</td>
</tr>
<tr>
<td>12. Inversion theorem for characteristic functions</td>
<td>30</td>
</tr>
<tr>
<td>13. Continuity theorem for distribution functions (First Limit Theorem)</td>
<td>33</td>
</tr>
<tr>
<td>14. Central Limit Theorem (independently and identically distributed random variables)</td>
<td>33</td>
</tr>
<tr>
<td>15. Liapounoff Theorem (Central limit theorem for non-identically distributed random variables)</td>
<td>35</td>
</tr>
<tr>
<td>16. (Weak) Law of Large Numbers (LLN)</td>
<td>41</td>
</tr>
<tr>
<td>17. Khintchine's Theorem (Weak LLN assuming only ( E(x) &lt; \infty ))</td>
<td>41</td>
</tr>
<tr>
<td>18. A Convergence Theorem (Cramer)</td>
<td>44</td>
</tr>
<tr>
<td>19. Convergence in probability of a continuous function</td>
<td>46</td>
</tr>
<tr>
<td>20. Strong Law of Large Numbers (SLLN)</td>
<td>51</td>
</tr>
<tr>
<td>21. Gauss-Markov Theorem (b.l.u.e.)</td>
<td>61</td>
</tr>
<tr>
<td>22. Asymptotic properties of m.l.e. (scalar case)</td>
<td>65</td>
</tr>
<tr>
<td>23. Asymptotic properties of m.l.e. (multiparameter case)</td>
<td>70</td>
</tr>
<tr>
<td>Theorem</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>24. Information Theorem (Cramer-Rao) (scalar case)</td>
<td>76</td>
</tr>
<tr>
<td>25. Information Theorem (Cramer-Rao) (multiparameter case)</td>
<td>81</td>
</tr>
<tr>
<td>26. Blackwell's Theorem for improving an unbiased estimate</td>
<td>88</td>
</tr>
<tr>
<td>27. Complete sufficient statistics and m.v.u.e.</td>
<td>92</td>
</tr>
<tr>
<td>28. Complete sufficient statistics in the non-parametric case</td>
<td>96</td>
</tr>
<tr>
<td>29. Bayes, constant risk, and minimax estimators</td>
<td>109</td>
</tr>
<tr>
<td>30. Density of a sample quantile</td>
<td>116</td>
</tr>
<tr>
<td>31. Convergence of a sample distribution function</td>
<td>124</td>
</tr>
<tr>
<td>32. Pitman's Theorem on Asymptotic Relative Efficiency</td>
<td>140</td>
</tr>
<tr>
<td>33. Neyman-Pearson Theorem (Probability ratio test)</td>
<td>149</td>
</tr>
<tr>
<td>34. Extension of Probability Ratio Test for composite hypotheses</td>
<td>153</td>
</tr>
<tr>
<td>35. Sufficient conditions for u.m.p.u. test</td>
<td>156</td>
</tr>
<tr>
<td>36. Reduction of test functions to those depending on sufficient statistics</td>
<td>160</td>
</tr>
<tr>
<td>37. Conditions for invariance of a test function</td>
<td>162</td>
</tr>
<tr>
<td>38. Scheffe's Test of all linear contrasts</td>
<td>187</td>
</tr>
<tr>
<td>39. Asymptotic power of the $\chi^2$ test (Mitra)</td>
<td>194</td>
</tr>
<tr>
<td>40. Correspondence of confidence regions and tests</td>
<td>199</td>
</tr>
</tbody>
</table>
# LIST OF DEFINITIONS

<table>
<thead>
<tr>
<th>Definition</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Notation of sets</td>
<td>1</td>
</tr>
<tr>
<td>2. Families of sets</td>
<td>2</td>
</tr>
<tr>
<td>3. Borel Sets</td>
<td>3</td>
</tr>
<tr>
<td>4. Additivity of sets</td>
<td>3</td>
</tr>
<tr>
<td>5. Probability measure</td>
<td>3</td>
</tr>
<tr>
<td>6. Point function</td>
<td>4</td>
</tr>
<tr>
<td>7. Distribution function</td>
<td>6</td>
</tr>
<tr>
<td>8. Density function</td>
<td>8</td>
</tr>
<tr>
<td>9. Random variable</td>
<td>10</td>
</tr>
<tr>
<td>10. Conditional probability</td>
<td>10</td>
</tr>
<tr>
<td>11. Joint distribution function</td>
<td>11</td>
</tr>
<tr>
<td>12. Discrete random variable</td>
<td>16</td>
</tr>
<tr>
<td>13. Continuous random variable</td>
<td>17</td>
</tr>
<tr>
<td>14. Expectation of g(X)</td>
<td>21</td>
</tr>
<tr>
<td>15. Moments</td>
<td>22</td>
</tr>
<tr>
<td>16. Characteristic function</td>
<td>25</td>
</tr>
<tr>
<td>17. Cumulant generating function</td>
<td>28</td>
</tr>
<tr>
<td>18. Convergence in probability</td>
<td>41</td>
</tr>
<tr>
<td>19. Convergence almost everywhere</td>
<td>49</td>
</tr>
<tr>
<td>20. Risk function</td>
<td>55</td>
</tr>
<tr>
<td>21. Minimum variance unbiased estimates (m.v.u.e.)</td>
<td>56</td>
</tr>
<tr>
<td>22. Best linear unbiased estimates (b.l.u.e.)</td>
<td>57</td>
</tr>
<tr>
<td>23. Bayes estimates</td>
<td>58</td>
</tr>
<tr>
<td>24. Minimax estimates</td>
<td>60</td>
</tr>
<tr>
<td>Definition</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>25. Constant risk estimates</td>
<td>60</td>
</tr>
<tr>
<td>26. Consistent estimates</td>
<td>60</td>
</tr>
<tr>
<td>27. Best asymptotically normal estimates (b.a.n.e.)</td>
<td>60</td>
</tr>
<tr>
<td>28. Least Squares estimates</td>
<td>61</td>
</tr>
<tr>
<td>29. Likelihood function</td>
<td>64</td>
</tr>
<tr>
<td>30. Maximum likelihood estimates (m.l.e.)</td>
<td>64</td>
</tr>
<tr>
<td>31. Sufficient statistics</td>
<td>86</td>
</tr>
<tr>
<td>32. Complete sufficient statistics</td>
<td>92</td>
</tr>
<tr>
<td>33. Test</td>
<td>113</td>
</tr>
<tr>
<td>34. Test of size &lt;</td>
<td>113</td>
</tr>
<tr>
<td>35. Power of a test</td>
<td>113</td>
</tr>
<tr>
<td>36. Consistent sequence of tests</td>
<td>113</td>
</tr>
<tr>
<td>37. Index of tests</td>
<td>113</td>
</tr>
<tr>
<td>38. Asymptotic relative efficiency (A.R.E.)</td>
<td>114</td>
</tr>
<tr>
<td>39. Quantile</td>
<td>115</td>
</tr>
<tr>
<td>40. Sample quantile</td>
<td>116</td>
</tr>
<tr>
<td>41. Uniformly most powerful tests</td>
<td>148</td>
</tr>
<tr>
<td>42. Unbiased tests</td>
<td>154</td>
</tr>
<tr>
<td>43. Similar tests</td>
<td>154</td>
</tr>
<tr>
<td>44. Invariant tests</td>
<td>155</td>
</tr>
<tr>
<td>45. Uniformly most powerful unbiased tests (u.m.p.u.)</td>
<td>155</td>
</tr>
<tr>
<td>46. Maximal invariant function</td>
<td>161</td>
</tr>
<tr>
<td>47. Confidence region</td>
<td>199</td>
</tr>
<tr>
<td>48. U.M.P.o confidence region</td>
<td>199</td>
</tr>
<tr>
<td>49. &quot;Shortest&quot; confidence interval</td>
<td>200</td>
</tr>
<tr>
<td>Problem</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1. Pr(Sum of sets) ≤ Σ Pr(sets)</td>
<td>3</td>
</tr>
<tr>
<td>2. Prove Pr(( \lim_{n \to \infty} S_n )) = ( \lim_{n \to \infty} ) Pr(S_n)</td>
<td>4</td>
</tr>
<tr>
<td>3. Cumulative probability at a point</td>
<td>6</td>
</tr>
<tr>
<td>4. Joint and marginal distribution functions</td>
<td>9</td>
</tr>
<tr>
<td>5. Transformations involving uniform distributions</td>
<td>13</td>
</tr>
<tr>
<td>6. Distribution of the sum of probabilities (combination of probabilities)</td>
<td>19</td>
</tr>
<tr>
<td>7. Prove that the density function of the negative binomial sums to 1</td>
<td>21</td>
</tr>
<tr>
<td>8. Find the mean of a censored normal distribution</td>
<td>21</td>
</tr>
<tr>
<td>9. Uniqueness of moments (uniform)</td>
<td>24</td>
</tr>
<tr>
<td>10. Prove the normal density function integrates to 1</td>
<td>28</td>
</tr>
<tr>
<td>11. Derive the characteristic function of N(0, 1)</td>
<td>28</td>
</tr>
<tr>
<td>12. Density of the product of n geometric distributions</td>
<td>32</td>
</tr>
<tr>
<td>13. Factorial m.g.f. for the binomial distribution</td>
<td>32</td>
</tr>
<tr>
<td>14. Inversion of the characteristic function (Cauchy)</td>
<td>33</td>
</tr>
<tr>
<td>15. Limiting distribution of the Poisson</td>
<td>35</td>
</tr>
<tr>
<td>16. Use of Mellin transforms</td>
<td>36</td>
</tr>
<tr>
<td>17. Transformations on NID(0, ( \sigma^2 ))</td>
<td>38</td>
</tr>
<tr>
<td>18. Limiting distribution of the multinomial distribution</td>
<td>38</td>
</tr>
<tr>
<td>19. Limiting distribution of the negative binomial</td>
<td>41</td>
</tr>
<tr>
<td>20. Show E [g(X)] = ( E_{Y</td>
<td>X} [g(X)</td>
</tr>
<tr>
<td>21. Distribution of the length of first run in a series of binomial trials</td>
<td>44</td>
</tr>
<tr>
<td>22. Distribution of the minimum observation from a uniform (0, 1) population</td>
<td>44</td>
</tr>
<tr>
<td>23. Prove theorem 19</td>
<td>46</td>
</tr>
<tr>
<td>24. Asymptotic distribution of ((X - \lambda)^2/X) where X is Poisson ( \lambda )</td>
<td>48</td>
</tr>
<tr>
<td>25. Example of convergence almost everywhere</td>
<td>50</td>
</tr>
<tr>
<td>Problem</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>26. Inversion of $\phi(t) = \cos ta$</td>
<td>50</td>
</tr>
<tr>
<td>27. Asymptotic distribution of the product of two independent sample means</td>
<td>50</td>
</tr>
<tr>
<td>28. Show that $\bar{x}$ is a b.l.u.e of $\mu$</td>
<td>57</td>
</tr>
<tr>
<td>29. Example of minimum risk estimation</td>
<td>57</td>
</tr>
<tr>
<td>30. Determination of risk function for Binomial</td>
<td>59</td>
</tr>
<tr>
<td>31. Find a constant risk estimate for Binomial $p$</td>
<td>60</td>
</tr>
<tr>
<td>32. Proof of consistency</td>
<td>60</td>
</tr>
<tr>
<td>33. B.l.u.e. of parameters in simple linear regression</td>
<td>63</td>
</tr>
<tr>
<td>34. Find the m.l.e. of &quot;a&quot; where $f(x) = a^2 e^{-a^2 x}$</td>
<td>69</td>
</tr>
<tr>
<td>35. M.l.e. of Poisson $\lambda$ and its asymptotic distribution</td>
<td>69</td>
</tr>
<tr>
<td>36. M.l.e. of parameter &quot;a&quot; of $R(0, a)$</td>
<td>70</td>
</tr>
<tr>
<td>37. M.l.e. of the parameter of the Laplace distribution</td>
<td>74</td>
</tr>
<tr>
<td>38. M.l.e. of $\lambda, \mu$ where $X$ is Poisson $\lambda$, $Y$ is Poisson $\mu \lambda$</td>
<td>74</td>
</tr>
<tr>
<td>39. Application of the multiparameter Cramer-Rao lower bound</td>
<td>75</td>
</tr>
<tr>
<td>40. C-R lower bound for unbiased estimates of the Poisson parameter $\lambda$</td>
<td>78</td>
</tr>
<tr>
<td>41. C-R lower bound for estimates of the geometric parameter</td>
<td>80</td>
</tr>
<tr>
<td>42. Application of the C-R lower bound to the negative binomial</td>
<td>84</td>
</tr>
<tr>
<td>43. Derivation of sufficient statistics</td>
<td>87</td>
</tr>
<tr>
<td>44. Sufficient statistics and unbiased estimators for the negative binomial</td>
<td>89</td>
</tr>
<tr>
<td>45. Variance of sufficient statistics for the Poisson parameter $\lambda$</td>
<td>91</td>
</tr>
<tr>
<td>46. M.v.u.e. of a quantile</td>
<td>95</td>
</tr>
<tr>
<td>47. M.v.u.e. of $Pr(x \leq a)$</td>
<td>98</td>
</tr>
<tr>
<td>48. Minimum and constant risk estimates of $\sigma^2$ (normal)</td>
<td>98</td>
</tr>
<tr>
<td>49. M.l.e. of the multinomial parameters ($p_{ij} = \rho_i \tau_j$)</td>
<td>101</td>
</tr>
<tr>
<td>50. M.l.e. and minimum modified $\chi^2$ estimates of the multinomial parameters ($p_{ij} = \alpha_i + \beta_j$)</td>
<td>107</td>
</tr>
</tbody>
</table>
- 208 -

Problem

51. M.l.e. and minimum modified $\chi^2$ estimates of $p_1 = e^{-\lambda}$ ............................................... 108
52. Comparison of variance of minimax and m.l. estimators for Binomial $p$ .............................................. 110
53. Example of Wolfowitz minimum distance estimate of $\mu$ ................................................................................. 112
54. A.R.E. of median test to mean test (normal) ................................................................................................. 114
55. Power of the normal approximation for testing Binomial $p$ ................................................................. 115
56. A.R.E. of median test to mean test (general) ................................................................................................. 117
57. Distribution of intervals between $R(0, 1)$ variables ..................................................................................... 120
58. Example of the $U_1$ test (Wilkinson Combination Procedure) ............................................................... 123
59. Distribution of the range .................................................................................................................................. 124, 6
60. Power of the $D_n^u$ test (Kolmogorov statistic) ............................................................................................ 128
61. Expected value of $\omega_n^2$ (Von Mises - Smirnov statistic) ........................................................................ 129
62. Wilcoxon's $U$ test .......................................................................................................................................... 134
63. Distribution and properties of the number of $Y_i$ exceeding max $X_i$ (two-sample test) ......................... 134
64. Application of two sample tests ...................................................................................................................... 138
65. Moments and properties of a particular non-parametric two-sample test statistic .............................................. 138
66. A.R.E. of Wilcoxon's $U$ test to the normal test ............................................................................................ 145
67. Equivalence of the Kruskal-Wallis $H$ test and Wilcoxon's $U$ test for $k=2$ .................................................. 147
68. U.m.p. test for negative exponential parameter ............................................................................................. 151
69. Example of non-existence of tests of size $\alpha$ .............................................................................................. 152
70. U.m.p.u. test for $\sigma$ (normal) ...................................................................................................................... 160
71. Example of application of Maximal Invariant Function ............................................................................... 167
72. Maximum likelihood ratio test for $\mu$ (normal) .............................................................................................. 175
73. U.m.p.i. test for the coefficient of variation (normal) ..................................................................................... 175
74. Derivation of $F$-test for the general linear hypothesis (two-way ANOVA) ................................................. 181
75. Power of the two-way ANOVA $F$-test ............................................................................................................ 185