2 Testing Statistical Hypotheses

The primary reference for this Chapter is Testing Statistical Hypotheses (3rd edition) by Lehmann and Romano. We will abbreviate this as TSH. We will basically cover Chapters 3 - 6 of TSH.

General problem:

We have a (parametric) model $p(x|\theta) \equiv p_\theta(x)$, and we wish to test $H_0 : \theta \in \mathbf{\Theta}_0$ versus $H_1 : \theta \in \mathbf{\Theta}_1, \mathbf{\Theta} = \mathbf{\Theta}_0 \cup \mathbf{\Theta}_1$.
We will denote the test procedure by $\phi(\boldsymbol{X})$ = critical function (decision rule). $\phi(\boldsymbol{X})$ gives us the form of the rejection region for rejecting (not rejecting) H_0, i.e. $\phi(x)$ denotes the decision to reject H_0 when $\boldsymbol{X} = x$ is observed. Here, $\boldsymbol{X} = (X_1, \cdots, X_n)$ denotes a random sample of size n.

We first consider testing a simple null versus a simple alternative, and thus $\mathbf{\Theta}_0 = \{0\} \equiv \{\theta_0\}$ and $\mathbf{\Theta}_1 = \{1\} \equiv \{\theta_1\}$.

$$\begin{aligned} \text{Let } \alpha &= \text{significance level or size} \\ &\equiv E_0[\phi(\boldsymbol{x})] = \int \phi(\boldsymbol{x}) p_{\theta_0}(x)\, d\mu \\ &= \int_{\boldsymbol{\mathcal{X}}} \phi(\boldsymbol{x}) p_{\theta_0}(x)\, dx \quad (\mu = \text{Lebesgue measure}) \end{aligned}$$

$$\begin{aligned} \beta = \text{power} &= E_1[\phi(\boldsymbol{x})] = \int \phi(\boldsymbol{x}) p_{\theta_1}(x)\, d\mu \\ &= \int \phi(\boldsymbol{x}) p_{\theta_1}(x)\, dx \end{aligned}$$

$\alpha = P(\text{type I error}) = P(\text{rejecting } H_0 \text{ when } H_0 \text{ is true})$
$1 - \beta = P(\text{type II error}) = P(\text{accepting } H_0 \text{ when } H_0 \text{ is false})$

The basic approach in hypothesis testing is to select $\phi(\boldsymbol{x})$ so as to maximize the power $\beta = \int \phi(\boldsymbol{x}) p_\theta(x)\, dx, \theta \in \boldsymbol{\Theta}_1$, subject to the condition

$$E_\theta[\phi(\boldsymbol{x})] \leq \alpha \quad \text{for all} \quad \theta \in \Theta_0.$$

We are now led to the following Theorem:

Theorem 2.1 (Neyman-Pearson lemma)

Let P_0 and P_1 have densities p_0 and p_1 with respect to some dominating measure μ ($\mu = P_0 + P_1$ always works). Let $0 \leq \alpha \leq 1$. Then
(i) There exists a constant k and a critical function ϕ of the form

$$\phi(x) = \begin{cases} 1 & \text{if} \quad p_1(x) > k\, p_0(x) \\ \gamma & \text{if} \quad p_1(x) = k\, p_0(x) \\ 0 & \text{if} \quad p_1(x) < k\, p_0(x) \end{cases} \tag{2.1}$$

such that

$$E_0[\phi(x)] = \alpha. \tag{2.2}$$

(ii) The test of (2.1) and (2.2) is a most powerful α level test of P_0 versus P_1.
(iii) If ϕ^* is most powerful test of size α, then for a.e. μ,

$$\phi^*(x) = \begin{cases} 1 & p_1(x) > kp_0(x) \\ 0 & p_1(x) < kp_0(x). \end{cases}$$

Corollary 2.1

If $0 < \alpha < 1$ and β is the power of the most powerful level α test, then $\alpha < \beta$ unless $P_0 = P_1$.

Proof of Theorem 2.1:

Let $0 < \alpha < 1$.

(i) Now

$$P_0(p_1(x) > cp_0(x)) = P_0\left(Y \equiv \frac{p_1(x)}{p_0(x)} > c\right) = 1 - F_Y(c).$$

Let $k \equiv \inf\{c : 1 - F_Y(c) < \alpha\}$, and if $P_0(Y = k) > 0$, let

$$\gamma \equiv \frac{\alpha - P_0(Y > k)}{P_0(Y = k)}.$$

Thus with

$$\phi(x) = \begin{cases} 1 & \text{if} \quad p_1(x) > k\, p_0(x) \\ \gamma & \text{if} \quad p_1(x) = k\, p_0(x) \\ 0 & \text{if} \quad p_1(x) < k\, p_0(x) \end{cases} \tag{2.3}$$

we have

$$E_0[\phi(x)] = P_0(Y > k) + \gamma P_0(Y = k) = \alpha.$$

(ii) Let ϕ^* be another test with $E_0[\phi^*] \le \alpha$.

$$\begin{aligned}\text{Now} \quad & \int_{\mathcal{X}} (\phi - \phi^*)(p_1 - kp_0)\, d\mu \\ = \ & \int_{[\phi - \phi^* > 0] \cup [\phi - \phi^* < 0]} (\phi - \phi^*)(p_1 - kp_0)\, d\mu \ge 0,\end{aligned}$$

and this implies that

$$\begin{aligned}\beta_\phi - \beta_{\phi^*} &= \int_{\mathcal{X}} (\phi - \phi^*) p_1\, d\mu \\ &\ge k \int_{\mathcal{X}} (\phi - \phi^*) p_0\, d\mu \\ &= k(\alpha - E_0\phi^*) \ge 0.\end{aligned}$$

Thus ϕ is most powerful.

(iii) Let ϕ^* be another most powerful test of size α. Define

$$A = \{x : \phi(x) \neq \phi^*(x), p_1(x) \neq kp_0(x)\}.$$

Then, $[\phi(x) - \phi^*(x)][p_1(x) - kp_0(x)] > 0$ when $x \in A$ and $= 0$ when $x \in A^c$, and

$$\int [\phi(x) - \phi^*(x)][p_1(x) - kp_0(x)]d\mu = 0,$$

since both $\phi(x)$ and ϕ^* are most powerful tests of size α. Therefore, $\mu(A) = 0$.
Note that P_0 and P_1 in Theorem 2.1 need not be two parametric models nested within a class of models indexed by θ.

For example, $P_0 = N(0,1), P_1 = t_q$, or $P_0 = N(0,1), P_1 = N(1,1)$, so that in this case $\theta_0 = 0$ and $\theta_1 = 1$, where $P_\theta = N(\theta, 1)$.
For $P_0 = N(0,1), P_1 = t_q$, P_0 and P_1 are not embedded within a parametric class P_θ, and in this case $\mathbf{\Theta}_0 = \{0\}$ and $\mathbf{\Theta}_1 = \{1\}$ do not correspond to "parameter spaces" in the usual sense.

Proof of Corollary 2.1:

$\tilde{\phi} = \alpha$ has power α, so $\beta \geq \alpha$. If $\beta = \alpha$, then $\tilde{\phi} \equiv \alpha$ is in fact most powerful, and hence (iii) shows that $\phi(x) = \alpha$ satisfies (i); that is, $p_1(x) = kp_0(x)$ a.e. μ. Thus $k = 1$ and $p_1 = p_0$.

Remark 2.1

1. The way to read $\phi(x)$ in (2.3) is that
   a) We reject H_0 with probability 1 if $p_1 > kp_0$, or equivalently if $\frac{p_1}{p_0} > k,\ p_0 \neq 0$.
   b) We reject H_0 with probability $\gamma \equiv \gamma(x)$ if $p_1 = kp_0$, or equivalently if $\frac{p_1}{p_0} = k,\ p_0 \neq 0$.
   c) We reject H_0 with probability 0 if $p_1 < kp_0$, or equivalently if $\frac{p_1}{p_0} < k,\ p_0 \neq 0$.
2. Suppose the desired level of the test is α. Then if there exists a k such that $P_0(p_1 > kp_0) = \alpha$, then use that k and set $\gamma = 0$. In this case, the test is a nonrandomized test and reduces to

$$\phi(x) = \begin{cases} 1 & p_1 > k\,p_0 \\ 0 & p_1 < k\,p_0 \end{cases} .$$

If no such k exists, then pick k so that
$P_0(p_1 > kp_0) < \alpha \leq P_0(p_1 \geq kp_0)$,
and let

$$\gamma(x) = \frac{\alpha - P_0(p_1 > kp_0)}{P_0(p_1 = kp_0)}. \tag{2.4}$$

If X is continuous, γ is typically equal to 0. γ may not be 0 in some continuous X situations (see Example 2.3). When X is discrete, γ is typically not equal to 0.

3 $p_1 > kp_0$ for rejecting H_0 is the same as $\frac{p_0}{p_1} < k$. Thus we reject H_0 with probability 1 if $\frac{p_1}{p_0} > k$ (or $\frac{p_0}{p_1} < k$).

Example 2.1

Suppose $X_1, \cdots, X_n$ is a random sample from a Poisson(θ) distribution. We wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$, $\theta_1 > \theta_0 > 0$. By the Neyman-Pearson (NP) lemma, we can find the most powerful α level test as

$$
\begin{aligned}
\frac{p_1}{p_0} &= \frac{\theta_1^{\sum_{i=1}^{n} X_i} e^{-n\theta_1}}{\theta_0^{\sum_{i=1}^{n} X_i} e^{-n\theta_0}} > k \\
&\Leftrightarrow \left(\frac{\theta_1}{\theta_0}\right)^{\sum_{i=1}^{n} X_i} > k^* \\
&\Leftrightarrow \sum_{i=1}^{n} X_i > k^{**},
\end{aligned}
$$

or just $\sum_{i=1}^{n} X_i > k$. Thus

$$
\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} X_i > k \\ \gamma & \text{if} \quad \sum_{i=1}^{n} X_i = k \\ 0 & \text{if} \quad \sum_{i=1}^{n} X_i < k \end{cases} .
$$

k must be chosen so that

$$\begin{aligned}\alpha &= E_0[\phi(\boldsymbol{x})] \\ &= P_{\theta_0}\left(\sum_{i=1}^{n} X_i > k\right) + \gamma P_{\theta_0}\left(\sum_{i=1}^{n} X_i = k\right) \\ &= 1 - \sum_{t=0}^{k} \frac{(n\theta_0)^t e^{-n\theta_0}}{t!} + \gamma\left\{\frac{(n\theta_0)^k e^{-n\theta_0}}{k!}\right\}\end{aligned}$$

Note that $\sum_{i=1}^{n} X_i \sim \text{Poisson}(n\theta)$.
Now suppose $\theta_0 = .1, \theta_1 = .2, n = 10$, and $\alpha = .05$.
Let us find k in this situation.

$\phi(\boldsymbol{x})$	$\sum_{i=1}^{n} X_i > 0$	$\sum_{i=1}^{n} X_i > 1$	$\sum_{i=1}^{n} X_i > 2$	$\sum_{i=1}^{n} X_i > 3$
α	.6321	.2642	.0803	.0190

Since none of these are equal to .05 exactly, we must randomize, and thus the randomized test is of the form

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} X_i > 3 \\ \gamma & \text{if} \quad \sum_{i=1}^{n} X_i = 3 \\ 0 & \text{if} \quad \sum_{i=1}^{n} X_i < 3 \end{cases} .$$

$$\begin{aligned} .05 &= P_{\theta_0}\left(\sum_{i=1}^{n} X_i > 3\right) + \gamma P_{\theta_0}\left(\sum_{i=1}^{n} X_i = 3\right) \\ &= .0190 + \gamma(.0613) \\ \Rightarrow \gamma &= .506. \end{aligned}$$

So our randomized test becomes

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} X_i > 3 \\ .506 & \text{if} \quad \sum_{i=1}^{n} X_i = 3 \\ 0 & \text{if} \quad \sum_{i=1}^{n} X_i < 3 \end{cases} .$$

Note that

$$\begin{aligned} P_{\theta_0}\left(\sum_{i=1}^{n} X_i = 3\right) &= P_{\theta_0}\left(\sum_{i=1}^{n} X_i > 2\right) - P_{\theta_0}\left(\sum_{i=1}^{n} X_i > 3\right) \\ &= .0803 - .0190 \\ &= .0613. \end{aligned}$$

Example 2.2

Suppose $X_1, \cdots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where μ is known and σ^2 is unknown. We wish to test $H_0 : \sigma^2 = \sigma_0^2$ versus $H_1 : \sigma^2 = \sigma_1^2$, where $\sigma_1^2 > \sigma_0^2 > 0$.

The NP test of level α takes the form

$$
\begin{aligned}
& \frac{p_1}{p_0} > k \\
\Leftrightarrow \quad & \frac{(2\pi)^{-\frac{n}{2}} \sigma_1^{-n} e^{-\frac{1}{2\sigma_1^2} \sum_{i=1}^{n} (X_i - \mu)^2}}{(2\pi)^{-\frac{n}{2}} \sigma_0^{-n} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \mu)^2}} > k \\
\Leftrightarrow \quad & \sum_{i=1}^{n} (X_i - \mu)^2 > k \quad .
\end{aligned}
$$

Thus

$$
\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} (X_i - \mu)^2 > k \\ 0 & \text{if} \quad \sum_{i=1}^{n} (X_i - \mu)^2 < k \quad . \end{cases}
$$

To make the test of level α, we need to solve

$$
\alpha = P_0 \left(\sum_{i=1}^{n} (X_i - \mu)^2 > k \right)
$$

Now $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma_0^2} \overset{H_0}{\sim} \chi_n^2$, so that

$$\begin{aligned}
\alpha &= P_0\left(\sum_{i=1}^{n}(X_i-\mu)^2 > k\right) \\
&= P(\sigma_0^2\chi_n^2 > k) = P\left(\chi_n^2 > \frac{k}{\sigma_0^2}\right) \\
&= 1 - F_{\chi_n^2}\left(\frac{k}{\sigma_0^2}\right), \text{ so that } k = \sigma_0^2\left(F_{\chi_n^2}^{-1}(1-\alpha)\right).
\end{aligned}$$

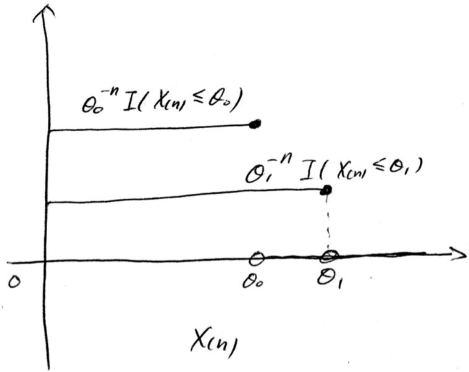
$$\begin{aligned}
\beta &= E_1[\phi(\boldsymbol{x})] \\
&= P_1\left(\sum_{i=1}^{n}(X_i-\mu)^2 > k\right) \\
&= P\left(\chi_n^2 > \frac{k}{\sigma_1^2}\right) = 1 - F_{\chi_n^2}\left(\frac{k}{\sigma_1^2}\right).
\end{aligned}$$

Example 2.3

Suppose $X_1, \cdots, X_n$ is a random sample from a $U(0, \theta)$ distribution ($U =$ uniform distribution). We wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1,\ \theta_1 > \theta_0 > 0$. Let us find the most powerful α level test.

$$
\begin{aligned}
p_\theta(\boldsymbol{x}) &= \prod_{i=1}^{n} p_\theta(x_i) \\
&= \prod_{i=1}^{n} \left(\frac{1}{\theta}\right) I(0 \le X_i \le \theta) \\
&= \theta^{-n} I(X_{(n)} \le \theta). \\
p_0(\boldsymbol{x}) &= \theta_0^{-n} I(X_{(n)} \le \theta_0). \\
p_1(\boldsymbol{x}) &= \theta_1^{-n} I(X_{(n)} \le \theta_1).
\end{aligned}
$$

$$
\begin{aligned}
& \frac{p_1(\boldsymbol{x})}{p_0(\boldsymbol{x})} > k \\
\Leftrightarrow\quad & \frac{\theta_1^{-n} I(X_{(n)} \le \theta_1)}{\theta_0^{-n} I(X_{(n)} \le \theta_0)} > k, \quad 0 < \theta_0 < \theta_1 < \infty.
\end{aligned}
$$


$\theta_0^{-n} I(X_{(n)} \le \theta_0)$
$\theta_1^{-n} I(X_{(n)} \le \theta_1)$
0
θ_0
θ_1
$X_{(n)}$

When $X_{(n)} > \theta_0$, we never accept H_0 regardless of k, so we want $\phi(\boldsymbol{x}) = 1$ for this case. When $X_{(n)} \leq \theta_0$, we need randomization. So, the α level NP test is given by

$$\phi(\boldsymbol{x}) = \begin{cases} \gamma & \text{if} \quad X_{(n)} \leq \theta_0 \\ 1 & \text{if} \quad X_{(n)} > \theta_0 \end{cases},$$

where γ is chosen so that $\alpha = E_{\theta_0}[\phi(\boldsymbol{x})]$.

$$\begin{aligned} \alpha &= \gamma P_{\theta_0}(X_{(n)} \leq \theta_0) + P_{\theta_0}(X_{(n)} > \theta_0) \\ &= \gamma(1) + 0 = \gamma, \end{aligned}$$

and thus $\alpha = \gamma$, and therefore our NP α level test is given by

$$\phi(\boldsymbol{x}) = \begin{cases} \alpha & \text{if} \quad X_{(n)} \leq \theta_0 \\ 1 & \text{if} \quad X_{(n)} > \theta_0 \end{cases}.$$

Note that the density of $X_{(n)}$ is given by

$$p_{X_{(n)}}(x) = \begin{cases} n\theta^{-n}x^{n-1} & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}.$$

Composite Null and Alternative Hypotheses

When testing composite null and alternative hypotheses, it is not clear what a "best" or "most powerful" test means, since several values of θ can be entertained in H_0 as well as in H_1. We need new concepts in this situation.

Definition 2.1

If the family of densities $\{p_\theta : \theta \in [\theta_0, \theta_1] \subset \Re\}$ is such that $\frac{p_{\theta'}(x)}{p_\theta(x)}$ is nondecreasing in $T(x)$ for each $\theta < \theta'$, then the family is said to have a monotone likelihood ratio (MLR) .

Definition 2.2

A test ϕ is of size α if

$$\sup_{\theta \in \mathbf{\Theta}_0} E_\theta[\phi(x)] = \alpha.$$

Let $C_\alpha \equiv \{\phi : \phi \text{ is of size } \alpha\}$. A test ϕ_0 is uniformly most powerful of size α (UMP of size α) if it has size α and $E_\theta[\phi_0(x)] \geq E_\theta[\phi(x)]$ for all $\theta \in \mathbf{\Theta}_1$ and all $\phi \in C_\alpha$.

Definition 2.3

The power function of a test ϕ is defined as

$$\beta(\theta) = E_\theta[\phi(x)].$$

Thus, UMP tests for composite nulls and alternatives have maximum power against any test of no greater size for any $\theta \in \boldsymbol{\Theta}_1$. This is therefore a generalization of the NP lemma.
We are led to the following theorem:

Theorem 2.2

Suppose that X has a density p_θ with MLR in $T(x)$.

(i) Then there exists a UMP level α test of $H_0 : \theta \leq \theta_0$ versus

$H_1 : \theta > \theta_0$ which is of the form

$$\phi(x) = \begin{cases} 1 & \text{if} \quad T(x) > k \\ \gamma & \text{if} \quad T(x) = k \\ 0 & \text{if} \quad T(x) < k \end{cases} \tag{2.5}$$

with $E_{\theta_0}[\phi(x)] = \alpha$.

(ii) $\beta(\theta) \equiv E_\theta[\phi(x)]$ is increasing in θ for $\beta < 1$.

(iii) For all θ', this same test is the UMP level $\alpha' \equiv \beta(\theta')$ test of $H_0' : \theta \leq \theta'$ versus $H_1' : \theta > \theta'$.

(iv) For all $\theta \leq \theta_0$, the test of (i) minimizes $\beta(\theta)$ among tests satisfying $\alpha = E_{\theta_0}[\phi]$.

Proof:

(i) and (ii) The most powerful α level test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ is the ϕ in (2.5), by the NP lemma, which guarantees the existence of k and γ.

Thus, ϕ is UMP of $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. According to the NP lemma (ii), this same test is most powerful of θ' versus θ''; thus (ii) follows from the NP Corollary (Corollary 2.1). Thus ϕ is also level α in the smaller class of tests of H_0 versus H_1, and hence is UMP there also. Note that with

$$C_\alpha = \{\phi : \sup_{\theta \le \theta_0} E_\theta[\phi] = \alpha\}$$

and

$$C_\alpha^{\theta_0} = \{\phi : E_{\theta_0}[\phi] \le \alpha\}, \quad C_\alpha \subset C_\alpha^{\theta_0}.$$

(iii) The same argument works.

(iv) To minimize power, just apply the NP lemma with the inequalities reversed.

Example 2.4

We present an example demonstrating the concept of a UMP test. Suppose $X_1, \cdots, X_n$ are iid $N(\theta, 1)$ and we wish to test

$$\begin{array}{l} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{array} \quad .$$

Let us find the UMP α level test. We seek a test $\phi^*(\boldsymbol{x})$ that is uniformly best out of the class of tests $\phi(\boldsymbol{x})$ for which

$$\sup_{\theta \in \boldsymbol{\Theta}_0} E_\theta[\phi(x)] = \alpha.$$

To solve this problem, we solve an easier but related problem. Let us find the best test $\phi^*(\boldsymbol{x})$ for $H_0' : \theta = \theta_0$ versus $H_1' : \theta = \theta_1, \theta_1 > \theta_0$. This brings us back to the NP framework. By the NP lemma, it is easily seen that for H_0' versus H_1', that

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \overline{X} > k \\ 0 & \text{if} \quad \overline{X} < k \end{cases} \quad ,$$

where k is chosen so that

$$\begin{aligned} \alpha &= E_{\theta_0}[\phi^*(\boldsymbol{x})] = P_{\theta_0}(\overline{X} > k) \\ &= 1 - \Phi(\sqrt{n}(k - \theta_0)), \\ \Rightarrow k &= \frac{1}{\sqrt{n}}\Phi^{-1}(1-\alpha) + \theta_0. \end{aligned}$$

From this calculation, we see that k depends <u>only</u> on α and θ_0 and <u>not</u> on θ_1. Thus any value of θ_1 under the alternative would lead to the same k. Thus $\phi^*(\boldsymbol{x})$ is UMP of level α for $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$. Thus $\phi^*(\boldsymbol{x})$ is UMP out of the class of all test for which

$$E_{\theta_0}[\phi(\boldsymbol{x})] \leq \alpha. \tag{2.6}$$

Now to see that $\phi^*(\boldsymbol{x})$ is UMP for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ involves a simple observation. Note that the power function of $\phi^*(\boldsymbol{x})$ is given by

$$\beta(\theta) = P_\theta(\overline{X} > k) = 1 - \Phi(\sqrt{n}(k - \theta)). \tag{2.7}$$

(2.7) is an increasing function of θ with value α at $\theta = \theta_0$. Thus $\beta(\theta)$ satisfies

$$E_\theta[\phi(x)] \leq \alpha \quad \text{for} \quad \theta \leq \theta_0. \tag{2.8}$$

Now since $\phi^*(\boldsymbol{x})$ is UMP out of the class of all tests satisfying (2.6), it is also UMP out of the class of all tests satisfying (2.8) since the class in (2.6) is a subclass of the class in (2.8). Thus

$$\phi^*(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \overline{X} > k \\ 0 & \text{if} \quad \overline{X} < k \end{cases}$$

is UMP of level α for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

All families with the MLR property have UMP tests derived using the same ideas as in Example 2.4, and the form of the UMP test is the same as the NP test of the simple versus simple case.

Example 2.5

Suppose $X_1, \cdots, X_n$ are iid Poisson(θ). Let us show that this family of distributions has the MLR property. Let $\theta_2 > \theta_1 > 0$.

$$\begin{aligned} \frac{p_{\theta_2}(\boldsymbol{x})}{p_{\theta_1}(\boldsymbol{x})} &= \left(\frac{\theta_2}{\theta_1}\right)^{\sum\limits_{i=1}^{n} X_i} e^{-n(\theta_2-\theta_1)} \\ &= \left(\frac{\theta_2}{\theta_1}\right)^{T(\boldsymbol{x})} e^{-n(\theta_2-\theta_1)}, \end{aligned}$$

where $T(\boldsymbol{x}) = \sum\limits_{i=1}^{n} X_i$. Since $\frac{\theta_2}{\theta_1} > 1$, we see that $\frac{p_{\theta_2}(\boldsymbol{x})}{p_{\theta_1}(\boldsymbol{x})}$ is a nondecreasing function of $T(\boldsymbol{x}) = \sum\limits_{i=1}^{n} X_i$. Thus, the Poisson distribution has the MLR property in $T(\boldsymbol{x}) = \sum\limits_{i=1}^{n} X_i$. Thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \overline{X} > k \\ \gamma & \text{if} \quad \overline{X} = k \\ 0 & \text{if} \quad \overline{X} < k \end{cases}$$

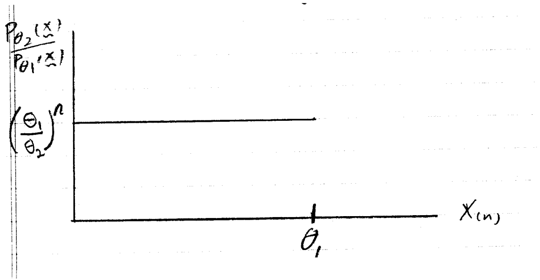
is UMP for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

Example 2.6

Suppose $X_1, \cdots, X_n$ are iid $U(0, \theta)$. Show that this family of distributions has the MLR property in $X_{(n)}$. Let $\theta_2 > \theta_1 > 0$.

$$\begin{aligned} \frac{p_{\theta_2}(\boldsymbol{x})}{p_{\theta_1}(\boldsymbol{x})} &= \left(\frac{\theta_1}{\theta_2}\right)^n \frac{I(X_{(n)} \leq \theta_2)}{I(X_{(n)} \leq \theta_1)} \\ &= \begin{cases} \left(\frac{\theta_1}{\theta_2}\right)^n & 0 \leq X_{(n)} \leq \theta_1 \\ \infty & X_{(n)} \geq \theta_1 \end{cases}. \end{aligned}$$

Thus we see that $\frac{p_{\theta_2}(\boldsymbol{x})}{p_{\theta_1}(\boldsymbol{x})}$ is a nondecreasing function of $X_{(n)}$.


$\frac{P_{\theta_2}(\underset{\sim}{x})}{P_{\theta_1}(\underset{\sim}{x})}$
$\left(\frac{\theta_1}{\theta_2}\right)^n$
θ_1
$X_{(n)}$

Thus, we see that the $U(0,\theta)$ distribution has the MLR property in $T(\boldsymbol{x}) = X_{(n)}$.

Thus $\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad X_{(n)} > k \\ 0 & \text{if} \quad X_{(n)} < k \end{cases}$

is UMP for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

Example 2.7 (Hypergeometric)

Suppose that we sample without replacement n items from a population of N items of which $\theta = D$ are defective. Let $X =$ number of defective items in the sample. Then

$$P_D(X = x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}},$$

for $x = \max\{0, n - N + D\}, 1, 2, \cdots, \min\{n, D\}$.

Since

$$\frac{p_{D+1}(x)}{p_D(x)} = \left(\frac{D+1}{N-D}\right)\left(\frac{N-D-n+x}{D+1-x}\right)$$

is increasing in x, there is MLR in $T(x) = x$. Thus, the UMP test of $H_0 : D \leq D_0$ versus $H_1 : D > D_0$ rejects if X is "too big", i.e., $\phi(x) = I(X > k) + \gamma I(X = k)$, and k is chosen to make the test of level α, i.e., $P_{D_0}(X > k) + \gamma P_{D_0}(X = k) = \alpha$.

For Examples 2.5 and 2.6, the k needed to make the test of level α is:

Example 2.5 (continued)

$$\begin{aligned}\alpha &= P_{\theta_0}(\overline{X} > k) + \gamma P_{\theta_0}(\overline{X} = k) \\ &= 1 - \sum_{t=0}^{nk} \frac{(n\theta_0)^t e^{-n\theta_0}}{t!} + \gamma \frac{(n\theta_0)^{nk} e^{-n\theta_0}}{(nk)!}.\end{aligned}$$

Example 2.6 (continued)

$$\begin{aligned}\alpha &= P_{\theta_0}(X_{(n)} > k) \\ &= \int_k^{\theta_0} n\theta_0^{-n} x^{n-1}\, dx \\ &= 1 - \theta_0^{-n} k^n, \\ \Rightarrow k &= \theta_0 (1-\alpha)^{1/n}.\end{aligned}$$

Example 2.8

Suppose that $p_\theta(x)$ is the density of a 1 parameter exponential family,

$$p_\theta(x) = c(\theta)\exp\{Q(\theta)T(x)\}h(x),$$

where $Q(\theta)$ is increasing in θ. Then

$$\phi(x) = \begin{cases} 1 & \text{if} \quad T(x) > k \\ \gamma & \text{if} \quad T(x) = k \\ 0 & \text{if} \quad T(x) < k \end{cases}$$

with $E_{\theta_0}[\phi(x)] = \alpha$ is UMP of level α for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. (See TSH 67-68 for other examples).

Example 2.9

The Binomial, Poisson, $U(0, \theta)$, hypergeometric, gamma, beta, negative binomial, geometric, normal, and exponential distributions all have the MLR property.

Example 2.10

The noncentral t, χ^2, and F distributions have the MLR property. See TSH pages 224, 307.

Example 2.11

The Cauchy family $p_\theta(x) = \pi^{-1}(1 + (x - \theta)^2)^{-1}$ does not have the MLR.

Theorem 2.2^{*} **(Dual of Theorem 2.2)**

(i) For testing $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$

$$\phi(x) = \begin{cases} 1 & \text{if} \quad T(x) < k \\ \gamma & \text{if} \quad T(x) = k \\ 0 & \text{if} \quad T(x) > k \end{cases}$$

is UMP of level α, with $E_{\theta_0}[\phi(x)] = \alpha$.

(ii) $\beta(\theta) \equiv E_\theta[\phi(x)]$ is decreasing in θ for $\beta < 1$.

(iii) For all θ', this same test is the UMP level $\alpha' \equiv \beta(\theta')$ test of $H_0' : \theta \geq \theta'$ versus $H_1' : \theta < \theta'$.

(iv) For all $\theta \geq \theta_0$, the test of (i) minimizes $\beta(\theta)$ among tests satisfying $\alpha = E_{\theta_0}[\phi(x)]$.

Example 2.8 (continued)

If $Q(\theta)$ is decreasing in θ, then a UMP α level test for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$ is given by

$$\phi(x) = \begin{cases} 1 & \text{if} \quad T(x) < k \\ \gamma & \text{if} \quad T(x) = k \\ 0 & \text{if} \quad T(x) > k \end{cases}$$

where $\alpha = E_{\theta_0}[\phi(x)]$.

Example 2.12

Suppose $X_1, \cdots, X_n$ are iid Gamma(θ, β), where β is known. We wish to test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

Let us find a UMP α level test.

$$\begin{aligned} p_\theta(\boldsymbol{x}) &= \frac{\beta^{n\theta}}{(\Gamma(\theta))^n}\left(\prod_{i=1}^n X_i^{\theta-1}\right)\left(\exp\left\{-\beta\sum_{i=1}^n X_i\right\}\right) \\ &= \exp\left\{\theta\sum_{i=1}^n \log X_i - \sum_{i=1}^n \log X_i - \beta\sum_{i=1}^n X_i + n\theta\log\beta - n\log(\Gamma(\theta))\right\} \end{aligned}$$

Let $T(\boldsymbol{x}) = \sum_{i=1}^n \log X_i$, $Q(\theta) = \theta$, $c(\theta) = \exp\{n\theta\log\beta - n\log(\Gamma(\theta))\}$ and $h(\boldsymbol{x}) = \exp\left\{-\sum_{i=1}^n \log X_i - \beta\sum_{i=1}^n X_i\right\}$.

Thus, this is a 1 parameter exponential family. Since $\theta > 0$, $Q(\theta) = \theta$ is an increasing function of θ, and thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^n \log X_i > k \\ 0 & \text{if} \quad \sum_{i=1}^n \log X_i < k \end{cases}$$

is UMP of level α for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

$$\alpha = E_{\theta_0}[\phi(\boldsymbol{x})] = P_{\theta_0}\left(\sum_{i=1}^{n} \log X_i > k\right) \quad .$$

Suppose we wanted to test $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$. Then, using Theorem 2.2*, the UMP α level test is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} \log X_i < k \\ 0 & \text{if} \quad \sum_{i=1}^{n} \log X_i > k \end{cases} \quad ,$$

where $\alpha = P_{\theta_0}(\sum_{i=1}^{n} \log X_i < k)$.

Example 2.3 (continued)

$X_1, \cdots, X_n$ are iid. $U(0, \theta)$. The α level NP test for testing $H_0 : \theta = \theta_0$

versus $H_1 : \theta = \theta_1$ is

$$\phi^*(\boldsymbol{x}) = \begin{cases} \alpha & \text{if} \quad X_{(n)} \leq \theta_0 \\ 1 & \text{if} \quad X_{(n)} > \theta_0 \end{cases} . \tag{2.9}$$

What is the UMP α level test for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$? We showed in Example 2.6 that the UMP α level test is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad X_{(n)} > k \\ 0 & \text{if} \quad X_{(n)} < k \end{cases} , \tag{2.10}$$

where $k = \theta_0(1-\alpha)^{1/n}$.
(2.9) and (2.10) are <u>not</u> the same. The test in (2.9) is also UMP. Thus, the UMP test is not unique here.

However,

(1) (2.9) and (2.10) are equivalent in the NP sense, that is, they are both level α and have the same power at each $\theta = \theta_1,\ \theta_1 > \theta_0$.

(2) (2.10) is based on the sufficient statistic $X_{(n)}$ (based on MLR) and (2.9) was based on $X_1, \cdots, X_n$. Thus looking over $X_1, \cdots, X_n$ led us to randomize (i.e., (2.9)), while looking at $X_{(n)}$ alone led to the non-randomized test in (2.10).

(3) Note that for the test in (2.10), $E_\theta[\phi(x)] \to 0$ as $\theta \to 0$, whereas for (2.9), $E_\theta[\phi(x)]$ remains constant over $0 < \theta \leq \theta_0$. Thus (2.10) exhibits more desirable properties than (2.9). Thus, basing tests on sufficient statistics leads to desirable properties of the test.

Example 2.13

Suppose $X_1, \cdots, X_n$ are iid from

$$p_\theta(x) = \begin{cases} \frac{\theta}{x^2} & x \geq \theta\,,\ \theta > 0 \\ 0 & \text{otherwise} \end{cases} .$$

We wish to test

$$\begin{aligned} H_0 &: \theta \leq \theta_0 \\ H_1 &: \theta > \theta_0 \end{aligned} .$$

We wish to find a UMP α level test of this hypothesis. This family of distributions is <u>not</u> in the exponential family, so we then examine if this family of distributions has the MLR property.

Let $\theta_2 > \theta_1 > 0$.

$$\begin{aligned} \frac{p_{\theta_2}(\boldsymbol{x})}{p_{\theta_1}(\boldsymbol{x})} &= \left(\frac{\theta_2}{\theta_1}\right)^n \frac{I(\theta_2 \le X_{(1)})}{I(\theta_1 \le X_{(1)})} \\ &= \begin{cases} 0 & \theta_1 \le X_{(1)} < \theta_2 \\ \left(\frac{\theta_2}{\theta_1}\right)^n & \theta_2 \le X_{(1)} \end{cases}. \end{aligned}$$

Since $\frac{\theta_2}{\theta_1} > 1$, $p_\theta(\boldsymbol{x})$ has the MLR property in $T(\boldsymbol{x}) = X_{(1)}$. Thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad X_{(1)} > k \\ 0 & \text{if} \quad X_{(1)} < k \end{cases}$$

is UMP of level α, where

$$\begin{aligned} \alpha &= P_{\theta_0}(X_{(1)} > k) \\ &= 1 - P_{\theta_0}(X_{(1)} \le k) = 1 - \int_{\theta_0}^{k} \left(\frac{n\theta_0}{x^2}\right) \left(\frac{\theta_0}{x}\right)^{n-1} dx \\ &= 1 + \left(\frac{\theta_0}{x}\right)^n \Big|_{\theta_0}^{k} = \left(\frac{\theta_0}{k}\right)^n, \end{aligned}$$

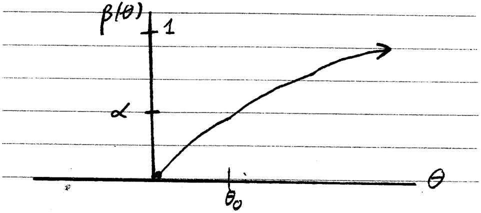
and thus $k = \theta_0 \alpha^{-1/n}$.

Note that $p_{X_{(1)}}(x) = \begin{cases} \left(\frac{n\theta}{x^2}\right)\left(\frac{\theta}{x}\right)^{n-1} & \theta \le x \\ 0 & \text{otherwise} \end{cases}$.

The power function of this test is

$$\begin{aligned}
\beta(\theta) &= P_\theta(X_{(1)} > k) \\
&= 1 - \int_\theta^k \left(\frac{n\theta}{x^2}\right)\left(\frac{\theta}{x}\right)^{n-1} dx \\
&= 1 + \left(\frac{\theta}{x}\right)^n \Big|_\theta^k \\
&= \left(\frac{\theta}{k}\right)^n \\
&= \alpha \left(\frac{\theta}{\theta_0}\right)^n.
\end{aligned}$$

We see that $\beta(\theta_0) = \alpha$.


$\beta(\theta)$
1
α
θ_0
θ

Consistency of NP Tests

Let P and Q be probability measures, and suppose that p and q are their densities with respect to a common $\sigma-$finite measure μ on $(\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{A}})$. Recall that the Hellinger distance $\mathcal{H}(P, Q)$ between P and Q is given by

$$\begin{aligned}\mathcal{H}^2(P, Q) &= \frac{1}{2}\int(\sqrt{p} - \sqrt{q})^2 \, d\mu \\ &= 1 - \int \sqrt{pq}\, d\mu \\ &= 1 - \rho(P, Q),\end{aligned}$$

where

$$\rho(P,Q) = \int \sqrt{pq}\, d\mu$$

is the affinity between P and Q. We have $0 \leq \rho(P,Q) \leq 1$.

Proposition 2.1

$\mathcal{H}(P,Q) = 0$ iff $p = q$ a.e. μ iff $\rho(P,Q) = 1$.
(Furthermore, $\rho(P,Q) = 0$ iff $\sqrt{p} \perp \sqrt{q}$ in the Hilbert space $\mathcal{L}_2(\mu)$).

Proposition 2.2

Let $X_1, \cdots, X_n$ be iid P (or Q) with joint densities

$$\begin{aligned} p_n(\boldsymbol{x}) &\equiv p(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i) \\ \text{or} \quad q_n(\boldsymbol{x}) &\equiv q(\boldsymbol{x}) = \prod_{i=1}^{n} q(x_i). \end{aligned}$$

Then

$$\rho(P_n, Q_n) = [\rho(P, Q)]^n \to 0$$

unless $p = q$ a.e. μ.
Note here that

$$\begin{aligned}
\rho(P_n, Q_n) &= \int (p_n q_n)^{1/2}\ d\mu \\
&= \int \left[\prod_{i=1}^{n} p(x_i) q(x_i)\right]^{1/2}\ d\mu \\
&= \prod_{i=1}^{n} \int (p(x_i) q(x_i))^{1/2}\ d\mu \\
&= \prod_{i=1}^{n} \rho(P, Q) \\
&= [\rho(P, Q)]^n .
\end{aligned}$$

Theorem 2.3 (Size and power of NP-type tests)

For testing p versus q, the test

$$\phi_n(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad q_n(\boldsymbol{x}) > k_n\, p_n(\boldsymbol{x}) \\ 0 & \text{if} \quad q_n(\boldsymbol{x}) < k_n\, p_n(\boldsymbol{x}) \end{cases}$$

with $0 < a_1 \leq k_n \leq a_2 < \infty$ for all $n \geq 1$ is size and power consistent: both probabilities of error, $\alpha \equiv \alpha_n \to 0$ and $1 - \beta_n \to 0$ as $n \to \infty$.

Proof:

For the Type I error probabilities,

$$
\begin{aligned}
E_P[\phi_n(\boldsymbol{x})] &= \int \phi_n(\boldsymbol{x}) p_n(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \\
&= \int \phi_n(\boldsymbol{x}) p_n^{1/2}(\boldsymbol{x}) p_n^{1/2}(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \\
&\leq k_n^{-1/2} \int \phi_n(\boldsymbol{x}) p_n^{1/2}(\boldsymbol{x}) q_n^{1/2}(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \\
&\leq k_n^{-1/2} \rho(P_n, Q_n) \\
&= k_n^{-1/2} [\rho(P, Q)]^n \to 0 \quad \text{as} \quad n \to \infty.
\end{aligned}
$$

The argument for Type II error is similar:

$$
\begin{aligned}
E_Q(1 - \phi_n(\boldsymbol{x})) &= \int (1 - \phi_n(\boldsymbol{x})) q_n(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \\
&= \int (1 - \phi_n(\boldsymbol{x})) q_n^{1/2}(\boldsymbol{x}) q_n^{1/2}(\boldsymbol{x})\, d\mu(\boldsymbol{x}) \\
&\leq k_n^{1/2} \int (1 - \phi_n(\boldsymbol{x})) p_n^{1/2}(\boldsymbol{x}) q_n^{1/2}(\boldsymbol{x})\, d\mu(\boldsymbol{x})
\end{aligned}
$$

$$\begin{aligned} &\leq k_n^{1/2}\rho(P_n, Q_n) \\ &= k_n^{1/2}[\rho(P, Q)]^n \to 0 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Unbiased Tests

In this section, we will discuss

1. Unbiased tests.
2. Application to 1 parameter exponential families.
3. Uniformly Most Powerful Unbiased (UMPU) tests for families with nuisance parameters via conditioning.

Consider testing

$$H_0 : \theta \in \boldsymbol{\Theta}_0 \quad \text{versus} \quad H_1 : \theta \in \boldsymbol{\Theta}_1$$

where $X \sim P_\theta$, for some $\theta \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$.
Let ϕ denote the critical function.

Definition 2.4

ϕ is unbiased if $\beta_\phi(\theta) \geq \alpha$ for all $\theta \in \boldsymbol{\Theta}_1$ and $\beta_\phi(\theta) \leq \alpha$ for all $\theta \in \boldsymbol{\Theta}_0$.

Remark: If ϕ is UMP, then ϕ is unbiased.

Definition 2.5

A uniformly most powerful unbiased (UMPU) level α test is a test ϕ_0 for which

$$E_\theta[\phi_0] \geq E_\theta[\phi] \quad \text{for all} \quad \theta \in \boldsymbol{\Theta}_1$$

for all unbiased level α tests ϕ.

Application to One-parameter Exponential Families:

Consider

$$p_\theta(\boldsymbol{x}) = c(\theta)\exp(\theta T(\boldsymbol{x}))h(\boldsymbol{x})$$

with respect to a σ-finite measure μ on $\Re^n$.
We wish to test:

(1) $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$,

(2) $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ versus $H_1 : \theta_1 < \theta < \theta_2$,

(3) $H_0 : \theta_1 \leq \theta \leq \theta_2$ versus $H_1 : \theta < \theta_1$ or $\theta > \theta_2$,

(4) $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

Theorem 2.4

(1) The test ϕ_1 with $E_{\theta_0}[\phi(T)] = \alpha$ given by

$$\phi_1(T(\boldsymbol{x})) = \begin{cases} 1 & \text{if} \quad T(\boldsymbol{x}) > k \\ \gamma & \text{if} \quad T(\boldsymbol{x}) = k \\ 0 & \text{if} \quad T(\boldsymbol{x}) < k \end{cases}$$

is UMP of level α for H_0 versus H_1 in (1).

(2) The test ϕ_2 with $E_{\theta_i}[\phi_2(T)] = \alpha, \quad i = 1,2$, given by

$$\phi_2(T(\boldsymbol{x})) = \begin{cases} 1 & \text{if} \quad k_1 < T(\boldsymbol{x}) < k_2 \\ \gamma_i & \text{if} \quad T(\boldsymbol{x}) = k_i, i = 1,2 \\ 0 & \text{if} \quad \text{else} \end{cases}$$

is UMP of level α for H_0 versus H_1 in (2).

(3) The test ϕ_3 with $E_{\theta_i}[\phi_3(T)] = \alpha, \quad i = 1,2$, given by

$$\phi_3(T(\boldsymbol{x})) = \begin{cases} 1 & \text{if} \quad T(\boldsymbol{x}) < k_1 \text{ or } T(\boldsymbol{x}) > k_2 \\ \gamma_i & \text{if} \quad T(\boldsymbol{x}) = k_i, i = 1,2 \\ 0 & \text{if} \quad \text{else} \end{cases}$$

is UMPU of level α for H_0 versus H_1 in (3). A UMP test for (3) does not exist.

(4) The test ϕ_4 with $E_{\theta_0}[\phi_4(T)] = \alpha$ and $E_{\theta_0}[T\phi_4(T)] = \alpha E_{\theta_0}[T]$, given by

$$\phi_4(T(\boldsymbol{x})) = \begin{cases} 1 & \text{if} \quad T(\boldsymbol{x}) < k_1 \text{ or } T(\boldsymbol{x}) > k_2 \\ \gamma_i & \text{if} \quad T(\boldsymbol{x}) = k_i, i = 1, 2 \\ 0 & \text{if} \quad \text{else} \end{cases}$$

is UMPU for H_0 versus H_1 in (4). Furthermore, if T is symmetrically distributed about a under θ_0, then $E_{\theta_0}[\phi_4(T)] = \alpha$, $k_2 = 2a - k_1$ and $\gamma_1 = \gamma_2$ determine the constants. A UMP test for (4) does not exist.

Example 2.14

Suppose $X_1, \cdots, X_n$ are Bernoulli(θ), and we wish to test

$$\begin{array}{l} H_0 : \theta_1 \leq \theta \leq \theta_2 \\ H_1 : \theta < \theta_1 \text{ or } \theta > \theta_2 \end{array} \quad .$$

Let us find the UMPU test of level α. Here, $T(\boldsymbol{x}) = \sum_{i=1}^{n} X_i$ for the Bernoulli distribution, so

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} X_i < k_1 \text{ or } \sum_{i=1}^{n} X_i > k_2 \\ \gamma_1 & \text{if } \sum_{i=1}^{n} X_i = k_1 \\ \gamma_2 & \text{if } \sum_{i=1}^{n} X_i = k_2 \\ 0 & \text{if } k_1 < \sum_{i=1}^{n} X_i < k_2 \end{cases} ,$$

where $\alpha = E_{\theta_1}[\phi(\boldsymbol{x})] = E_{\theta_2}[\phi(\boldsymbol{x})]$, and thus

$$\begin{aligned}\alpha &= P(T(\boldsymbol{x}) < k_1 \text{ or } T(\boldsymbol{x}) > k_2) + \gamma_1 P(T(\boldsymbol{x}) = k_1) \\ &\quad + \gamma_2 P(T(\boldsymbol{x}) = k_2) \\ \Rightarrow 1 - \alpha &= \sum_{t=k_1}^{k_2} \binom{n}{t} \theta_1^t (1-\theta_1)^{n-t} - \sum_{i=1}^{2} \gamma_i \binom{n}{k_i} \theta_1^{k_i} (1-\theta_1)^{n-k_i}.\end{aligned}$$

Also,

$$1 - \alpha = \sum_{t=k_1}^{k_2} \binom{n}{t} \theta_2^t (1-\theta_2)^{n-t} - \sum_{i=1}^{2} \gamma_i \binom{n}{k_i} \theta_2^{k_i} (1-\theta_2)^{n-k_i}.$$

Given α, we solve these two equations for k_1, k_2, γ_1 and γ_2.

Example 2.15

Suppose $X_1, \cdots, X_n$ is a random sample from $N(0, \sigma^2)$. Let us find the UMPU test of $H_0 : \sigma^2 = \sigma_0^2$ versus $H_0 : \sigma^2 \neq \sigma_0^2$.

This family of distributions is in the 1 parameter exponential family with $T(\boldsymbol{x}) = \sum_{i=1}^{n} X_i^2$. Thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} X_i^2 < k_1 \text{ or } \sum_{i=1}^{n} X_i^2 > k_2 \\ \gamma_1 & \text{if} \quad \sum_{i=1}^{n} X_i^2 = k_1 \\ \gamma_2 & \text{if} \quad \sum_{i=1}^{n} X_i^2 = k_2 \\ 0 & \text{if} \quad k_1 < \sum_{i=1}^{n} X_i^2 < k_2 \end{cases}$$

is UMPU of level α, where

$$\alpha = E_0[\phi(\boldsymbol{x})] \tag{2.11}$$

$$\alpha E_0[T(\boldsymbol{x})] = E_0[\phi(\boldsymbol{x}) T(\boldsymbol{x})]. \tag{2.12}$$

Since

$$P_{\sigma^2}\left(\sum_{i=1}^{n} X_i^2 = k_1\right) = P_{\sigma^2}\left(\sum_{i=1}^{n} X_i^2 = k_2\right) = 0,$$

we pick $\gamma_1 = \gamma_2 = 0$. Thus, for (2.11), we have

$$\begin{aligned}
\alpha &= P_{\sigma_0^2}\left(\sum_{i=1}^{n} X_i^2 < k_1 \text{ or } \sum_{i=1}^{n} X_i^2 > k_2\right) \\
&= 1 - P_{\sigma_0^2}\left(k_1 \le \sum_{i=1}^{n} X_i^2 \le k_2\right) \\
&= 1 - P\left(\frac{k_1}{\sigma_0^2} \le \chi_n^2 \le \frac{k_2}{\sigma_0^2}\right),
\end{aligned}$$

since under H_0, $\frac{\sum_{i=1}^{n} X_i^2}{\sigma_0^2} \sim \chi_n^2$.
Thus

$$1 - \alpha = F_{\chi_n^2}\left(\frac{k_2}{\sigma_0^2}\right) - F_{\chi_n^2}\left(\frac{k_1}{\sigma_0^2}\right). \tag{2.13}$$

From (2.12), we have

$$\alpha E_{\sigma_0^2}\left[\sum_{i=1}^{n} X_i^2\right] = E_{\sigma_0^2}\left[\sum_{i=1}^{n} X_i^2 \phi(\boldsymbol{x})\right] \tag{2.14}$$

The left side of (2.14) equals $\alpha n \sigma_0^2$. The right side of (2.14) is

$$\begin{aligned}
E_{\sigma_0^2}\left[\sum_{i=1}^{n} X_i^2 \phi(\boldsymbol{x})\right] &= E_{\sigma_0^2}[T(\boldsymbol{x})\phi(\boldsymbol{x})] \\
&= \int T(\boldsymbol{x})\phi(\boldsymbol{x}) p_T \, dt \quad \text{where} \quad p_T = \text{density of} \quad T \\
&= \int_0^\infty t\, \phi(t) p(t)\, dt \\
&= \int_0^{k_1} t\, p(t)\, dt + \int_{k_2}^\infty t\, p(t)\, dt.
\end{aligned}$$

But

$$\int_0^\infty t\, p(t)\, dt = E_{\sigma_0^2}\left[\sum_{i=1}^{n} X_i^2\right] = E\left[\sigma_0^2 \chi_n^2\right] = n\sigma_0^2.$$

Thus, we have

$$\begin{aligned} \alpha n\sigma_0^2 &= n\sigma_0^2 - \int_{k_1}^{k_2} t\, p(t)\, dt \\ \Rightarrow \int_{k_1}^{k_2} t\, p(t)\, dt &= n\sigma_0^2(1-\alpha). \end{aligned}$$

Now since $p(t)$ =density of $\sigma_0^2\chi_n^2 = \frac{1}{\sigma_0^2} p_{\chi_n^2}\left(\frac{t}{\sigma_0^2}\right)$, we have

$$\begin{aligned} \int_{k_1}^{k_2} \frac{t}{\sigma_0^2 2^{n/2}\Gamma\left(\frac{n}{2}\right)} \left(\frac{t}{\sigma_0^2}\right)^{\frac{n}{2}-1} e^{-\left(\frac{t}{2\sigma_0^2}\right)} dt &= n\sigma_0^2(1-\alpha) \\ \Rightarrow \int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} \frac{y^{n/2} e^{-y/2}}{2^{n/2}\Gamma(\frac{n}{2})}\, dy &= n(1-\alpha). \end{aligned}$$

We can integrate by parts n times to get the left-hand side equal to

$$\left(\frac{k_1}{\sigma_0^2}\right)^{n/2} e^{-\left(\frac{k_1}{2\sigma_0^2}\right)} - \left(\frac{k_2}{\sigma_0^2}\right)^{n/2} e^{-\left(\frac{k_2}{2\sigma_0^2}\right)}, \qquad (2.15)$$

or we can use the recursion $tp(t) = n\,p^*(t)$,

where $p^*(t) = \frac{1}{\sigma_0^2} p_{\chi^2_{n+2}}\left(\frac{t}{\sigma_0^2}\right)$, $p(t) = \frac{1}{\sigma_0^2} p_{\chi^2_n}\left(\frac{t}{\sigma_0^2}\right)$, and

$$\int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} np^*(t)\,dt = n\left[F^*\left(\frac{k_2}{\sigma_0^2}\right) - F^*\left(\frac{k_1}{\sigma_0^2}\right)\right],$$

where F^* is the cdf of p^*. Also, note that

$$\begin{aligned}
&\int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} \frac{y^{n/2}e^{-y/2}}{2^{n/2}\Gamma(\frac{n}{2})}\,dy \\
= \quad & n\int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} \frac{y^{n/2+1-1}e^{-y/2}}{2^{n/2+1}\Gamma(\frac{n}{2}+1)}\,dy \\
= \quad & n\int_{\frac{k_1}{\sigma_0^2}}^{\frac{k_2}{\sigma_0^2}} p_{\chi^2_{n+2}}(y)\,dy \\
= \quad & n\left(F_{\chi^2_{n+2}}\left(\frac{k_2}{\sigma_0^2}\right) - F_{\chi^2_{n+2}}\left(\frac{k_1}{\sigma_0^2}\right)\right).
\end{aligned}$$

Thus our UMPU α level test is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} X_i^2 < k_1 \text{ or } \sum_{i=1}^{n} X_i^2 > k_2 \\ 0 & \text{if} \quad k_1 < \sum_{i=1}^{n} X_i^2 < k_2 \end{cases},$$

where k_1 and k_2 are determined by (2.13), (2.14) and (2.15).

UMPU Tests For Exponential Families With Nuisance Parameters

Consider the exponential family $P = \{p_{\theta,\xi}\}$ given by

$$p_{\theta,\xi} = h(x)c(\theta,\xi)\exp\left[\theta u(\boldsymbol{x}) + \sum_{i=1}^{k}\xi_i T_i(\boldsymbol{x})\right] \tag{2.16}$$

with respect to a $\sigma-$finite measure μ on some subset of $\Re^n$, where $\boldsymbol{\Theta}$ is convex, has dimension $k+1$, $\boldsymbol{\Theta} = \{(\theta, \xi_1, \cdots, \xi_k)\}$, and contains interior points $\theta_i, i = 0, 1, 2$, and $\boldsymbol{x} = (X_1, \cdots, X_n)$.

We wish to test:

(1) $H_0 : \theta \leq \theta_0$
$H_1 : \theta > \theta_0$,

(2) $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$
$H_1 : \theta_1 < \theta < \theta_2$,

(3) $H_0 : \theta_1 \leq \theta \leq \theta_2$
$H_1 : \theta < \theta_1$ or $\theta > \theta_2$,

(4) $H_0 : \theta = \theta_0$
$H_1 : \theta \neq \theta_0$.

Theorem 2.5 (see TSH, Theorem 4.4.1, page 121)

The following are UMPU tests for (1) - (4) below.

(1) $$\phi_1(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad U > c(t) \\ \gamma(t) & \text{if} \quad U = c(t) \\ 0 & \text{if} \quad U < c(t) \end{cases}$$

where $E_{\theta_0}[\phi_1(U)|T = t] = \alpha$.

(2) $$\phi_2(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad c_1(t) < U < c_2(t) \\ \gamma_i & \text{if} \quad U = c_i(t) \\ 0 & \text{if} \quad \text{else} \end{cases}$$

where $E_{\theta_i}[\phi_2(U)|T = t] = \alpha, \quad i = 1, 2.$

(3) $$\phi_3(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad U < c_1(t) \text{ or } U > c_2(t) \\ \gamma_i & \text{if} \quad U = c_i(t) \\ 0 & \text{if} \quad \text{else} \end{cases}$$

where $E_{\theta_i}(\phi_3(U)|T = t) = \alpha, \quad i = 1, 2.$

(4)
$$\phi_4(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad U < c_1(t) \text{ or } U > c_2(t) \\ \gamma_i & \text{if} \quad U = c_i(t) \\ 0 & \text{if} \quad \text{else} \end{cases} \qquad \text{(a)}$$

where $E_{\theta_0}[\phi_4(U)|T = t] = \alpha$ and (b)

$E_{\theta_0}[u\phi_4(U)|T = t] = \alpha E_{\theta_0}[U|T = t].$ (c)

Remark 2.2

(a) If $V \equiv h(U, T)$ is increasing in U for each fixed t and is independent of T on $\boldsymbol{\Theta}_B$, then

$$\phi_1(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad V > c \\ \gamma & \text{if} \quad V = c \\ 0 & \text{if} \quad V < c \end{cases}$$

is UMPU in hypothesis (1).

(b) If $V \equiv h(U, T) = a(t)U + b(t)$ with $a(t) > 0$, then the second constraint in (4) becomes

$$E_{\theta_0}\left[\frac{V - b(t)}{a(t)}\phi \middle| T = t\right] = \alpha E_{\theta_0}\left(\frac{V - b(t)}{a(t)}|T = t\right)$$

or $E_{\theta_0}[V\phi|T = t] = \alpha E_{\theta_0}[V|T = t]$, and if V is independent of T on the boundary, then the test for hypothesis (4) is unconditional.

(c) For hypotheses (2) and (3), if V is monotone in U and V is independent of T on the boundary, then the test is unconditional.

Example 2.16 (page 153, TSH)

Suppose $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$, both (μ, σ^2) are unknown, and we wish to test $H_0 : \sigma^2 \geq \sigma_0^2$ versus $H_0 : \sigma^2 < \sigma_0^2$.
By Theorem 2.5, there exists a UMPU test of the hypothesis $H_0 : \theta \geq \theta_0$ versus $H_0 : \theta < \theta_0$, for which $\theta_0 = -\frac{1}{2\sigma_0^2}$, which is equivalent to
$H_0 : \sigma^2 \geq \sigma_0^2$ versus $H_0 : \sigma^2 < \sigma_0^2$.

Note that

$$\begin{aligned} p(\boldsymbol{x}|\mu,\sigma^2) &= (2\pi)^{-n/2}\sigma^{-n}e^{-\frac{1}{2\sigma^2}\left[\sum\limits_{i=1}^{n}X_i^2-2\mu\sum\limits_{i=1}^{n}X_i+n\mu^2\right]} \\ &= (2\pi\sigma^2)^{-n/2}\exp\left\{\frac{-n\mu^2}{2\sigma^2}\right\}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}X_i^2+\frac{n\mu}{\sigma^2}\bar{X}\right) \\ &= c(\theta,\xi)\exp\left[\theta u(\boldsymbol{x})+\sum_{i=1}^{k}\xi_i T_i(\boldsymbol{x})\right], \end{aligned}$$

where $k=1$, $u(\boldsymbol{x})=\sum\limits_{i=1}^{n}X_i^2$, $\theta=-\frac{1}{2\sigma^2}$, $\xi_1\equiv\xi=\frac{n\mu}{\sigma^2}$,
$T_1(\boldsymbol{x})\equiv T(\boldsymbol{x})=\bar{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$, and $c(\theta,\xi)=(2\pi\sigma^2)^{-n/2}\exp\left\{\frac{-n\mu^2}{2\sigma^2}\right\}$.
The rejection region of this test can be obtained from Theorem 2.5, part (1), with the inequality reversed because the hypothesis is now $\theta\geq\theta_0$. In the present case, this becomes

$$\sum_{i=1}^{n}X_i^2\leq C_0(\bar{X}),$$

here $P_{\sigma_0^2}\left(\sum_{i=1}^{n} X_i^2 \le C_0(\bar{X}) \middle| \bar{X}\right) = \alpha$.
If this is written as

$$\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \le C_0'(\bar{X}),$$

it follows from the independence of $\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$ and $\bar{X}$ that $C_0'(\bar{X})$ does <u>not</u> depend on $\bar{X}$. The test therefore rejects when $\sum_{i=1}^{n}(X_i - \bar{X})^2 \le C_0'$, or equivalently when

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} \le C_0, \tag{2.17}$$

with C_0 determined by $P_{\sigma_0^2}\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma_0^2} \le C_0\right) = \alpha$.

Since $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma_0^2} \sim \chi^2_{n-1}$, C_0 is determined as

$$\int_0^{C_0} p_{\chi^2_{n-1}}(y)\,dy = \alpha$$

$$\Rightarrow F_{\chi^2_{n-1}}(C_0) = \alpha \Rightarrow C_0 = F^{-1}_{\chi^2_{n-1}}(\alpha).$$

This same result can be obtained via Remark 2.2, part (a).
To show an unconditional UMPU via Remark 2.2, part (a), a statistic $V = h(U,T)$ is required, where V is independent of $T = \sum_{i=1}^{n} X_i$ on $\boldsymbol{\Theta}_B$ (i.e., at $\theta = \theta_0$ and all μ). Take

$$V = \sum_{i=1}^{n}(X_i - \bar{X})^2 = U - nT^2 \equiv h(U,T).$$

V is independent of $\bar{X}$ for all μ and σ^2. Since $h(u,t)$ is an increasing function of u for each t, it follows that the UMPU test has a rejection region of the form $V \leq C_0'$.

This derivation also shows that the UMPU rejection region for $H_0 : \sigma^2 \le \sigma_1^2$ or $\sigma^2 \ge \sigma_2^2$ is

$$C_1 < \sum_{i=1}^{n} (X_i - \bar{X})^2 < C_2,$$

where the C_i's are given by

$$\int_{\frac{C_1}{\sigma_1^2}}^{\frac{C_2}{\sigma_1^2}} p_{\chi^2_{n-1}}(y)\,dy = \int_{\frac{C_1}{\sigma_2^2}}^{\frac{C_2}{\sigma_2^2}} p_{\chi^2_{n-1}}(y)\,dy = \alpha.$$

Since $h(u,t)$ is linear in u, it is further seen that the UMPU test of $H_0 : \sigma^2 = \sigma_0^2$ has the acceptance region

$$C_1' < \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma_0^2} < C_2',$$

where C_1' and C_2' are determined by

$$\int_{C_1'}^{C_2'} p_{\chi^2_{n-1}}(y)\,dy = \frac{1}{n-1} \int_{C_1'}^{C_2'} y\, p_{\chi^2_{n-1}}(y)\,dy = 1 - \alpha.$$

Theorem 2.5 (and Remark 2.2) shows for this and other hypotheses considered, that the UMPU test depends only on V.
The power function of the above tests can be obtained explicitly in terms of the χ^2 distribution. In the case of the one-sided test corresponding to (2.17), for example, the power function of σ^2 is given by

$$\begin{aligned}
\beta(\sigma^2) &= P_{\sigma^2}\left\{ \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \leq \frac{C_0\sigma_0^2}{\sigma^2} \right\} \\
&= \int_0^{C_0\sigma_0^2/\sigma^2} p_{\chi^2_{n-1}}(y)\, dy \\
&= F_{\chi^2_{n-1}}\left(\frac{C_0\sigma_0^2}{\sigma^2}\right).
\end{aligned}$$

This same method can be applied to the problems of testing the hypothesis $H_0 : \mu \leq \mu_0$ against $H_1 : \mu > \mu_0$, and $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. There is no loss of generality in assuming $\mu_0 = 0$, since we can transform to $X_i - \mu_0$. In this case, we can identify

$$\theta = \frac{n\mu}{\sigma^2}, \quad \xi = -\frac{1}{2\sigma^2}, \quad u(\boldsymbol{x}) = \bar{X}, \quad T(\boldsymbol{x}) = \sum_{i=1}^{n} X_i^2.$$

Theorem 2.5 then shows that UMPU test exists for $H_0 : \theta \leq 0$ or $H_0 : \theta = 0$, which are equivalent to $H_0 : \mu \leq 0$ and $H_0 : \mu = 0$, repectively.

Since

$$V = \frac{\bar{X}}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}} = \frac{U}{\sqrt{T - nU^2}}$$

is independent of $T = \sum_{i=1}^{n} X_i^2$ when $\mu = 0$, it follows from Theorem 2.5 and Remark 2.2 that the UMPU rejection region for $H_0 : \mu \leq 0$ is $V \geq C_0'$ or equivalently

$$t(x) \geq C_0,$$

where

$$t(x) = \frac{\sqrt{n}\bar{X}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}}. \tag{2.18}$$

In order to apply the theorem to $H_0 : \mu = 0$, let $W = \frac{\bar{X}}{\sqrt{\sum_{i=1}^{n} X_i^2}}$. This is also independent of $\sum_{i=1}^{n} X_i^2$ when $\mu = 0$, and in addition, is linear in $U = \bar{X}$.

The distribution of W is symmetric about 0 when $\mu = 0$ and conditions (a), (b), (c), of Theorem 2.5 with W in place of V are therefore satisfied for the rejection region $|W| \geq C'$ with $P_{\mu=0}(|W| \geq C') = \alpha$. Since

$$t(x) = \frac{\sqrt{(n-1)n}\, W(x)}{\sqrt{1 - nW^2(x)}},$$

the absolute value of $t(x)$ is an increasing function of $|W(x)|$, and the rejection region is equivalent to

$$|t(x)| \geq C.$$

We know from previous courses that, under H_0, $t(x) \sim t_{n-1}$, and C_0 and C of the one-sided and two-sided tests are determined by

$$\int_{C_0}^{\infty} p_{t_{n-1}}(y)\,dy = \alpha \quad \text{and} \quad \int_{C}^{\infty} p_{t_{n-1}}(y)\,dy = \frac{\alpha}{2}.$$

When $\mu \neq 0$, $t(x) \sim$ noncentral t, with non-centrality parameter $\delta^2 = \frac{n\mu^2}{\sigma^2}$, denoted $t(x) \sim t\left(n-1, \delta^2\right)$. We obtain a central t when $\mu = 0$.
When $\mu_0 \neq 0$, (2.18) becomes

$$t(x) = \frac{\sqrt{n}(\bar{X} - \mu_0)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}} \quad \overset{H_0}{\sim} \quad t_{n-1}$$

$$\overset{H_1}{\sim} \quad t\left(n-1, \frac{n(\mu-\mu_0)^2}{\sigma^2}\right).$$

Example 2.17 (comparing two Poisson populations ,page 125, TSH)

Suppose $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$, and X and Y are independent. Their joint distribution is given by

$$P(X = x, Y = y) \equiv p(x, y) = \frac{e^{-(\lambda+\mu)}}{x!\,y!} \exp\left[y \log\left(\frac{\mu}{\lambda}\right) + (x + y)\log(\lambda)\right].$$

By Theorem 2.5, there exist UMPU tests of the four hypotheses (1) - (4) of Theorem 2.5 concerning the parameter $\theta = \log(\frac{\mu}{\lambda})$, or equivalently concerning the ratio $\rho = \frac{\mu}{\lambda}$. This includes in particular the hypotheses $H_0 : \mu \leq \lambda$ against $H_1 : \mu > \lambda$, and $H_0 : \mu = \lambda$ against $H_1 : \mu \neq \lambda$.

Comparing the joint distribution of (X, Y) with the multiparameter exponential family form in (2.16), we have $U = Y$, $T = X + Y$, and by Theorem 2.5, the tests are performed conditionally on the integer points of the line segment $X + Y = t$ in the positive quadrant of the (x, y) plane.

The conditional distribution of $Y|X+Y=t$ is

$$P(Y=y|X+Y=t)=\binom{t}{y}\left(\frac{\mu}{\lambda+\mu}\right)^{y}\left(\frac{\lambda}{\lambda+\mu}\right)^{t-y},\quad y=0,1,\cdots,t,$$

which is a binomial distribution corresponding to t trials and probability $p=\frac{\mu}{\lambda+\mu}$ of success. The original hypotheses therefore reduce to the corresponding ones about the parameter p of a binomial distribution.

The hypothesis $H_0:\mu\le a\lambda$, for example, becomes $H_0:p\le\frac{a}{a+1}$, which is rejected when Y is too large.

Example 2.18 (comparing two Binomial populations,page 126, TSH)

The comparison of two binomial populations is quite similar to the comparison of two Poisson populations. Let $X\sim\text{Binomial}(m,p_1)$, $Y\sim\text{Binomial}(n,p_2)$. X and Y are independent, and have joint distribution

$$
\begin{aligned}
P(X = x, Y = y) &= \binom{m}{x} p_1^x (1 - p_1)^{m-x} \binom{n}{y} p_2^y (1 - p_2)^{n-y} \\
&= \binom{m}{x}\binom{n}{y}(1 - p_1)^m (1 - p_2)^n \\
&\quad \times \exp\left[y \left(\log\left(\frac{p_2}{1 - p_2}\right) - \log\left(\frac{p_1}{1 - p_1}\right)\right)\right. \\
&\quad \left. + (x + y)\log\left(\frac{p_1}{1 - p_1}\right)\right].
\end{aligned}
$$

The four hypotheses of Theorem 2.5, (1), (2), (3), (4), can be tested concerning the parameter

$$
\theta = \log\left[\frac{\frac{p_2}{1-p_2}}{\frac{p_1}{1-p_1}}\right]
$$

or equivalently concerning the <u>odds ratio</u>

$$
\rho = \left(\frac{\frac{p_2}{1-p_2}}{\frac{p_1}{1-p_1}}\right).
$$

This includes, in particular, the problems of testing $H_0 : p_2 \leq p_1$ against $H_1 : p_2 > p_1$ and $H_0 : p_2 = p_1$ against $H_1 : p_2 \neq p_1$. As in the Poisson case, $U = Y$ and $T = X + Y$, and the test is carried out in terms of the conditional distribution of Y on the line segment $X + Y = t$. This distribution is given by

$$P(Y = y|X + Y = t) = C_t(\rho)\binom{m}{t-y}\binom{n}{y}\rho^y, \quad y = 0, 1, 2, \cdots, t$$

where

$$C_t(\rho) = \frac{1}{\sum_{j=0}^{t} \binom{m}{t-j}\binom{n}{j}\rho^j}.$$

In the particular case of the hypotheses (1) and (4), the boundary value θ_0 of (1) and (4) is 0, and the corresponding value of ρ is $\rho_0 = 1$. The conditional distribution then reduces to

$$P(Y = y|X + Y = t) = \frac{\binom{m}{t-y}\binom{n}{y}}{\binom{m+n}{t}},$$

which is a hypergeometric distribution.

Likelihood Ratio Tests (LRT)

UMP and UMPU tests may not always exist. For example, in the presence of nuisance parameters, UMPU tests are quite hard to construct, especially when the dimension of the nuisance parameter is high. In k-parameter exponential families, for example, if we want to test one parameter treating all others as nuisance, UMPU tests typically do not exist.

Example 2.19

Suppose $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$, and we wish to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$, where σ^2 is a nuisance (unknown) parameter. We have seen how to construct a UMPU test in this situation. A more systematic and "automated" procedure for constructing tests when (i) UMP or UMPU tests do not exist or (ii) UMPU tests are too hard to construct, is a procedure called the likelihood ratio test.

Suppose we are interested in testing $H_0 : \theta \in \mathbf{\Theta}_0$ versus $H_1 : \theta \in \mathbf{\Theta}_1$, where $\mathbf{\Theta} \subset \Re^k$, so that $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$ are subsets of $\Re^k$. Suppose $X_1, \cdots, X_n$ are iid from $p(x|\theta)$.

The likelihood ratio test is defined by

$$\Lambda = \frac{\sup_{\theta \in \mathbf{\Theta}_0} p(\boldsymbol{x}|\theta)}{\sup_{\theta \in \mathbf{\Theta}_0 \cup \mathbf{\Theta}_1} p(\boldsymbol{x}|\theta)}, \tag{2.19}$$

where $p(\boldsymbol{x}|\theta) = \prod_{i=1}^{n} p(x_i|\theta)$.

We reject H_0 if Λ is "too small". Thus, the likelihood ratio test of level α is given by

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \Lambda < k \\ \gamma & \text{if} \quad \Lambda = k \\ 0 & \text{if} \quad \Lambda > k, \end{cases}$$

where k is chosen so that $\alpha = \sup_{\theta \in \boldsymbol{\Theta}_0} E[\phi(\boldsymbol{x})]$. In some situations, a k may not exist that satisfies $\alpha = \sup_{\theta \in \boldsymbol{\Theta}_0} E[\phi(\boldsymbol{x})]$.

Interpretation of Λ

The "best" θ in H_0 is compared to the "best" θ overall, that is for all $\theta \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1$. We thus compare the best explanation of the data in H_0 with the best overall explanation of the data. The likelihood ratio test is thus a very sensible procedure and often corresponds to UMP, UMPU, or nearly UMPU tests for many hypothesis testing problems. That is, for a great many hypothesis testing problems, the likelihood ratio test is often equivalent to a UMP or UMPU test when a UMP or UMPU test exist.

The likelihood ratio test is well suited for dealing with many nuisance parameters and for situations in which a UMP or UMPU test does not exist.

Example 2.20

Suppose $X_1, \cdots, X_n$ are iid Poisson(θ). Suppose we wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$

We know from previous results that a UMPU test exists.

Let us construct the likelihood ratio test (LRT).

$$\sup_{\theta\in\mathbf{\Theta}_0} p(\boldsymbol{x}|\theta) = p(\boldsymbol{x}|\theta_0) = \frac{\theta_0^{\sum\limits_{i=1}^{n} X_i} e^{-n\theta_0}}{\prod\limits_{i=1}^{n} X_i!}, \quad \mathbf{\Theta}_0 = \{\theta_0\}$$

$$= \frac{\theta_0^{n\bar{X}} e^{-n\theta_0}}{\prod\limits_{i=1}^{n} X_i!},$$

where $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$. For the denominator of Λ, we have

$$\sup_{\theta\in\mathbf{\Theta}_0\cup\mathbf{\Theta}_1} p(\boldsymbol{x}|\theta) = \sup_{\theta\in\mathbf{\Theta}} p(\boldsymbol{x}|\theta),$$

where $\mathbf{\Theta}_0 \cup \mathbf{\Theta}_1 = \{\theta : \theta > 0\}$. Maximizing $p(\boldsymbol{x}|\theta)$ w.r.t. θ is the same as finding the MLE of θ, which is $\bar{X}$. Thus $\hat{\theta} = \bar{X}$. Therefore,

$$\sup_{\theta\in\mathbf{\Theta}} p(\boldsymbol{x}|\theta) = p(\boldsymbol{x}|\hat{\theta}) = \frac{\bar{X}^{n\bar{X}} e^{-n\bar{X}}}{\prod\limits_{i=1}^{n} X_i!}.$$

Thus

$$\Lambda = \frac{\theta_0^{n\bar{X}} e^{-n\theta_0}}{\bar{X}^{n\bar{X}} e^{-n\bar{X}}}.$$

We reject H_0 with probability 1 if $\Lambda < k$, that is,

$$\Leftrightarrow (\frac{\theta_0}{\bar{X}})^{\bar{X}} e^{\bar{X}} < k.$$

Thus, the likelihood ratio test of level α is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad (\frac{\theta_0}{\bar{X}})^{\bar{X}} e^{\bar{X}} < k \\ \gamma & \text{if} \quad (\frac{\theta_0}{\bar{X}})^{\bar{X}} e^{\bar{X}} = k \\ 0 & \text{if} \quad (\frac{\theta_0}{\bar{X}})^{\bar{X}} e^{\bar{X}} > k \end{cases},$$

where $\alpha = E_{\theta_0}[\phi(\boldsymbol{x})]$. To find k explicitly from the equation $\alpha = E_{\theta_0}[\phi(\boldsymbol{x})]$, we would need to derive the distribution of $(\frac{\theta_0}{\bar{X}})^{\bar{X}} e^{\bar{X}}$, which has no closed form. Is the LRT equivalent to a UMP or UMPU test here?

Example 2.21

Suppose $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$ where both μ and σ are unknown. We wish to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

The LRT is given by

$$\begin{aligned} \Lambda &= \frac{\sup\limits_{(\mu,\sigma^2)\in\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\mu,\sigma^2)}{\sup\limits_{(\mu,\sigma^2)\in\boldsymbol{\Theta}_0\cup\boldsymbol{\Theta}_1} p(\boldsymbol{x}|\mu,\sigma^2)}. \\ \boldsymbol{\Theta}_0 &= \{(\mu,\sigma^2) : \mu = \mu_0, \sigma^2 > 0\}. \\ \boldsymbol{\Theta} &= \{(\mu,\sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\}. \end{aligned}$$

<u>Numerator of Λ</u>:
Under H_0, $\mu = \mu_0$ and σ is unrestricted.

$$\sup_{(\mu,\sigma^2)\in\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\mu,\sigma^2) = \sup_{\sigma^2>0,\mu=\mu_0} p(\boldsymbol{x}|\mu,\sigma^2) = \sup_{\sigma^2>0} p(\boldsymbol{x}|\mu_0,\sigma^2).$$

$$p(\boldsymbol{x}|\mu_0,\sigma^2) = (2\pi)^{-n/2}\sigma^{-n}e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(X_i-\mu_0)^2}.$$

$$\log p(\boldsymbol{x}|\mu_0,\sigma^2) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu_0)^2.$$

Maximizing $\log p(\boldsymbol{x}|\mu_0,\sigma^2)$ w.r.t. σ, we get

$$\hat{\sigma}^2 \equiv \hat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)^2.$$

Thus

$$\begin{aligned}\sup_{(\mu,\sigma^2)\in\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\mu,\sigma^2) &= (2\pi)^{-n/2}\hat{\sigma}_0^{-n}e^{-\frac{1}{2\hat{\sigma}_0}\sum\limits_{i=1}^{n}(X_i-\mu_0)^2} \\ &= (2\pi)^{-n/2}\hat{\sigma}_0^{-n}e^{-n/2}.\end{aligned}$$

Denominator of Λ:
In the denominator, both (μ, σ^2) are unrestricted.

$$p(\boldsymbol{x}|\mu, \sigma^2) = (2\pi)^{-n/2}\sigma^{-n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}.$$

Maximizing with respect to (μ, σ^2), we get

$$\hat{\mu} \equiv \hat{\mu}_1 = \bar{X}, \quad \hat{\sigma}^2 \equiv \hat{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Thus

$$\begin{aligned} \sup_{\theta\in\boldsymbol{\Theta}} p(\boldsymbol{x}|\mu, \sigma^2) &= p(\boldsymbol{x}|\hat{\mu}_1, \hat{\sigma}_1^2) \\ &= (2\pi)^{-n/2}\hat{\sigma}_1^{-n}e^{-n/2}, \end{aligned}$$

and thus, we reject H_0 if

$$\Lambda = \frac{(2\pi)^{-n/2}\hat{\sigma}_0^{-n}e^{-n/2}}{(2\pi)^{-n/2}\hat{\sigma}_1^{-n}e^{-n/2}} < k$$

$$\Leftrightarrow \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_0^2}\right) < k$$

$$\Leftrightarrow \frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} > k.$$

Thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2} > k \\ 0 & \text{if} \quad \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2} < k \end{cases}.$$

To compute k, we need to know the distribution of $\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\sum_{i=1}^{n}(X_i-\bar{X})^2}$.

To facilitate this, we use the decomposition

$$\sum_{i=1}^{n}(X_i-\mu_0)^2=\sum_{i=1}^{n}(X_i-\bar{X})^2+n(\bar{X}-\mu_0)^2$$

$$\begin{aligned}\Rightarrow \frac{\sum\limits_{i=1}^{n}(X_i-\mu_0)^2}{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2} &= \frac{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2+n(\bar{X}-\mu_0)^2}{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}\\ &= 1+\frac{n(\bar{X}-\mu_0)^2}{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2},\end{aligned}$$

and

$$\begin{aligned}1+\frac{n(\bar{X}-\mu_0)^2}{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2} &> k\\ \Leftrightarrow \frac{n(\bar{X}-\mu_0)^2}{\sum\limits_{i=1}^{n}(X_i-\bar{X})^2} &> k_1\end{aligned}$$

$$\Leftrightarrow \frac{\sqrt{n}|\bar{X}-\mu_0|}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \quad > \quad k_2$$

$$\Leftrightarrow \frac{\sqrt{n}|\bar{X}-\mu_0|}{\sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}} \quad > \quad k_3$$

$$\Leftrightarrow \frac{\sqrt{n}|\bar{X}-\mu_0|}{S} \quad > \quad k, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

Thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \frac{\sqrt{n}|\bar{X}-\mu_0|}{S} > k \\ 0 & \text{if} \quad \frac{\sqrt{n}|\bar{X}-\mu_0|}{S} < k, \end{cases} \tag{2.20}$$

which is precisely the one-sample t-test derived earlier when discussing UMPU tests. Now

$$\frac{\sqrt{n}\,(\bar{X}-\mu_0)}{S} \quad \overset{H_0}{\sim} \quad t_{n-1}$$

and thus our level α test is (2.20), with $k = (1-\frac{\alpha}{2}) \times 100$ percentile of the t_{n-1} distribution, that is $k = t_{n-1,1-\frac{\alpha}{2}}$.

For the one-sided hypotheses $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$,

the LRT is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} > k \\ 0 & \text{if} \quad \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} < k \end{cases} \quad , \quad k = t_{n-1,\alpha}.$$

For $H_0 : \mu \geq \mu_0$ versus $H_1 : \mu < \mu_0$,

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} < k \\ 0 & \text{if} \quad \frac{\sqrt{n}(\bar{X}-\mu_0)}{S} > k \end{cases}, \quad k = -t_{n-1,\alpha}.$$

When do LRT's correspond to UMP or UMPU tests?

Theorem 2.6

Suppose X has a distribution in the 1 parameter exponential family,

$$p_\theta(\boldsymbol{x}) = c(\theta)\exp\{Q(\theta)T(\boldsymbol{x})\}h(\boldsymbol{x}),$$

where $Q(\theta)$ is a strictly increasing function of θ.

(i) For testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, there exists a LRT whose rejection region is the same as the UMP test given by Theorem 2.4, Part (1).

(ii) For testing $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ versus $H_1 : \theta_1 < \theta < \theta_2$, there exists a LRT whose rejection region is the same as the UMP test given by Theorem 2.4, Part (2).

(iii) For testing the hypotheses (3) or (4) of Theorem 2.4, there exists a LRT whose rejection region is the same as the UMPU test given by Theorem 2.4, Part (3) and (4).

Proof:

(i) Let $\hat{\theta}$ be the MLE of θ. Note that $p(\boldsymbol{x}|\theta) \equiv p_\theta(\boldsymbol{x})$ is increasing when $\theta \leq \hat{\theta}$ and decreasing when $\theta > \hat{\theta}$. Thus

$$\Lambda = \begin{cases} 1 & \text{if} \quad \hat{\theta} \leq \theta_0 \\ \frac{p(\boldsymbol{x}|\theta_0)}{p(\boldsymbol{x}|\hat{\theta})} & \text{if} \quad \hat{\theta} > \theta_0 \end{cases} .$$

Then $\Lambda < k$ is the same as $\hat{\theta} > \theta_0$ and $\frac{p(\boldsymbol{x}|\theta_0)}{p(\boldsymbol{x}|\hat{\theta})} < k$.

It can also be shown that $\log p(\boldsymbol{x}|\hat{\theta}) - \log p(\boldsymbol{x}|\theta_0)$ is strictly increasing in $T(\boldsymbol{x})$ when $\hat{\theta} > \theta_0$ and strictly decreasing in $T(\boldsymbol{x})$ when $\hat{\theta} < \theta_0$ (exercise).
Hence, for any $c \in \Re^1$, $\hat{\theta} > \theta_0$ and $\frac{p(\boldsymbol{x}|\theta_0)}{p(\boldsymbol{x}|\hat{\theta})} < k$ is equivalent to $T(\boldsymbol{x}) > c$ for some $0 < k < 1$.

(ii) The proof is similar to that in (i). Note that

$$\Lambda = \begin{cases} 1 & \text{if} \quad \hat{\theta} < \theta_1 \ \text{ or } \ \hat{\theta} > \theta_2 \\ \frac{\max\{p(\boldsymbol{x}|\theta_1), p(\boldsymbol{x}|\theta_2)\}}{p(\boldsymbol{x}|\hat{\theta})} & \text{if} \quad \theta_1 \leq \hat{\theta} \leq \theta_2 \end{cases} .$$

Hence $\Lambda < k$ is equivalent to $c_1 < T(\boldsymbol{x}) < c_2$.

(iii) The proof of (iii) is left as an exercise.

Theorem 2.6 can be applied to problems concerning one parameter exponential families such as the binomial, Poisson, negative binomial, and normal (with one parameter known) families. The following example shows that the same result holds in a situation where Theorem 2.6 is not applicable.

Example 2.22

Suppose $X_1, \cdots, X_n$ are iid $U(0, \theta)$ and we wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. It can be shown that the UMP size α test for this hypothesis (see TSH, problem 3.2) has rejection region $X_{(n)} > \theta_0$ or $X_{(n)} < \theta_0 \alpha^{1/n}$. We now show that the LRT is the same as the UMP test for this example.
Note that

$$p(\boldsymbol{x}|\theta) = \theta^{-n} I(X_{(n)} \leq \theta),$$

and therefore

$$\begin{aligned}\Lambda &= \frac{\sup_{\theta\in\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\theta)}{\sup_{\theta\in\boldsymbol{\Theta}} p(\boldsymbol{x}|\theta)} = \frac{\theta_0^{-n} I(X_{(n)} \le \theta_0)}{X_{(n)}^{-n}} \\ &= \left(\frac{X_{(n)}}{\theta_0}\right)^n I(X_{(n)} \le \theta_0).\end{aligned}$$

Thus

$$\Lambda = \begin{cases} \left(\frac{X_{(n)}}{\theta_0}\right)^n & \text{if} \quad X_{(n)} \le \theta_0 \\ 0 & \text{if} \quad X_{(n)} > \theta_0 \end{cases},$$

and $\Lambda < k$ is equivalent to

$$X_{(n)} > \theta_0 \ \text{ or } \ \frac{X_{(n)}}{\theta_0} < k^{1/n}.$$

Taking $k = \alpha$ ensures that the LRT has size α.

Example 2.23

Consider the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}),$$

where $(\boldsymbol{\beta}, \sigma^2)$ are unknown, $\boldsymbol{\beta}$ is $p \times 1$, $\boldsymbol{X}$ is $n \times p$ of rank p, and $\boldsymbol{y}$ is an $n \times 1$ vector of responses. Here $\boldsymbol{X}$ is an $n \times p$ matrix of fixed covariates. The linear model above implies $\boldsymbol{y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I})$. Suppose we wish to test
$H_0 : E(\boldsymbol{y}) \in C(\boldsymbol{X}_0)$ versus $H_1 : E(\boldsymbol{y}) \in C(\boldsymbol{X}) \cap C(\boldsymbol{X}_0)^c$,
where $C(\boldsymbol{X})$ denotes the column space of $\boldsymbol{X}$, and $C(\boldsymbol{X}_0) \subset C(\boldsymbol{X})$. The model under H_0 is thus the reduced model, which is nested in H_1.

The model under H_0 is $\boldsymbol{y} = \boldsymbol{X}_0\gamma_0 + \boldsymbol{\epsilon}$, $\boldsymbol{X}_0$ is $n \times q$, γ_0 is $q \times 1$, and the likelihood function under H_0 is given by

$$p(\boldsymbol{y}|\gamma_0, \sigma^2) = (2\pi)^{-n/2}\sigma^{-n}e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{\gamma}_0)'(\boldsymbol{y}-\boldsymbol{X}_0\boldsymbol{\gamma}_0)}.$$

To get the numerator of the LRT, we need to maximize $p(\boldsymbol{y}|\gamma_0, \sigma^2)$ w.r.t. $(\boldsymbol{\gamma}_0, \sigma^2)$. From the theory of linear models, we know that

$$\hat{\gamma}_0 = (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{y} \quad (\boldsymbol{X}_0 \text{ is rank } q)$$

$$\hat{\sigma}^2 \equiv \hat{\sigma}_0^2 = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{M}_0)\boldsymbol{y}}{\boldsymbol{n}} \equiv \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n},$$

where $\boldsymbol{M}_0 = \boldsymbol{X}_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'$, $\hat{\boldsymbol{y}} = \boldsymbol{X}_0\hat{\gamma}_0$, and $\hat{Y}_i$ is the *ith* component of $\hat{\boldsymbol{y}}$.

To compute the denominator of the LRT, we compute $\sup\limits_{(\boldsymbol{\beta},\sigma^2)\in\boldsymbol{\Theta}} p(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2)$.

The likelihood function for the "full" model is

$$p(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2) = (2\pi)^{-n/2}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})\right\},$$

for which

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \\
\hat{\sigma}^2 &= \frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{n}, \\
\boldsymbol{M} &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.
\end{aligned}$$

Thus

$$\begin{aligned}
\Lambda &= \frac{\sup\limits_{(\boldsymbol{\gamma}_0,\sigma^2)} p(\boldsymbol{y}|\boldsymbol{\gamma}_0,\sigma^2)}{\sup\limits_{(\boldsymbol{\beta},\sigma^2)} p(\boldsymbol{y}|\boldsymbol{\beta},\sigma^2)} \\
&= \frac{\hat{\sigma}_0^{-n}\exp\left\{-\frac{1}{2\hat{\sigma}_0^2}(\boldsymbol{y}-\boldsymbol{X}_0\hat{\boldsymbol{\gamma}}_0)'(\boldsymbol{y}-\boldsymbol{X}_0\hat{\boldsymbol{\gamma}}_0)\right\}}{\hat{\sigma}^{-n}\exp\left\{-\frac{1}{2\hat{\sigma}^2}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})\right\}}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}}\right)^{n/2} \frac{\exp\left\{\frac{-1}{2\hat{\sigma}_0^2}(\boldsymbol{y}-\boldsymbol{M}_0\boldsymbol{y})'(\boldsymbol{y}-\boldsymbol{M}_0\boldsymbol{y})\right\}}{\exp\left\{\frac{-1}{2\hat{\sigma}^2}(\boldsymbol{y}-\boldsymbol{M}\boldsymbol{y})'(\boldsymbol{y}-\boldsymbol{M}\boldsymbol{y})\right\}} \\
&= \left(\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}}\right)^{n/2} \frac{\exp\left\{\frac{-n}{2\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}}(\boldsymbol{y}-\boldsymbol{M}_0\boldsymbol{y})'(\boldsymbol{y}-\boldsymbol{M}_0\boldsymbol{y})\right\}}{\exp\left\{\frac{n}{2\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}(\boldsymbol{y}-\boldsymbol{M}\boldsymbol{y})'(\boldsymbol{y}-\boldsymbol{M}\boldsymbol{y})\right\}} \\
&= \left(\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}}\right)^{n/2} \frac{\exp\{-n/2\}}{\exp\{-n/2\}} \\
&= \left(\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}}\right)^{n/2}.
\end{aligned}
$$

Thus, we reject H_0 if

$$
\left(\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}}\right)^{n/2} \leq k.
$$

Taking both sides to the $2/n$ power, this is equivalent to rejecting H_0 when

$$\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y}} \le k_1. \tag{2.21}$$

Now write $\boldsymbol{I}-\boldsymbol{M}_0 = (\boldsymbol{I}-\boldsymbol{M}) + (\boldsymbol{M}-\boldsymbol{M}_0)$ so that

$$\begin{aligned} \boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{y} &= \boldsymbol{y}'[(\boldsymbol{I}-\boldsymbol{M}) + (\boldsymbol{M}-\boldsymbol{M}_0)]\boldsymbol{y} \\ &= \boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{y}. \end{aligned}$$

Thus (2.21) becomes

$$\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{y}} \le k_1. \tag{2.22}$$

Now (2.22) is equivalent to

$$\frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y} + \boldsymbol{y}'(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}} \ge k_2, \tag{2.23}$$

where $k_2 = 1/k_1$. We can write (2.23) as

$$1 + \frac{\boldsymbol{y}'(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}} \ge k_2,$$

which is equivalent to

$$\frac{\boldsymbol{y}'(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}} \geq k_3,$$

where $k_3 = k_2 - 1$. Finally, the expression above is equivalent to

$$\frac{\boldsymbol{y}'(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{y}/r(\boldsymbol{M}-\boldsymbol{M}_0)}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}/r(\boldsymbol{I}-\boldsymbol{M})} \geq k_4 \equiv k. \tag{2.24}$$

We see that the likelihood ratio test is equivalent to the F test, since the statistic in (2.24) is precisely the F statistic which is the UMPU test. In (2.24), $r(\boldsymbol{M}-\boldsymbol{M}_0) =$ rank of $(\boldsymbol{M}-\boldsymbol{M}_0) = p - q$, and $r(\boldsymbol{I}-\boldsymbol{M}) = n - p$. Also $k = F(p-q, n-p, 1-\alpha)$, which makes (2.24) of level α.

The power function of the test is given by

$$\beta(\boldsymbol{\theta}) = P(F(p-q, n-p, \delta) \geq F(p-q, n-p, 1-\alpha)),$$

$$\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2), \quad \delta = \text{ non-centrality parameter}$$

$$\delta = \frac{\|(\boldsymbol{M}-\boldsymbol{M}_0)\boldsymbol{X}\boldsymbol{\beta}\|^2}{2\sigma^2} = \frac{\|(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{X}\boldsymbol{\beta}\|^2}{2\sigma^2} = \frac{\boldsymbol{\beta}'\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{M}_0)\boldsymbol{X}\boldsymbol{\beta}}{2\sigma^2}.$$

Under H_0, $\boldsymbol{\beta} = \boldsymbol{\gamma}_0$ and $\boldsymbol{X} = \boldsymbol{X}_0$, so that $\delta = 0$, since $\boldsymbol{M}_0\boldsymbol{X}_0 = \boldsymbol{X}_0$.

LRT for two-sample problems

Now suppose we have two independent samples, $X_1, \cdots, X_n$ from $p(\boldsymbol{x}|\boldsymbol{\theta}^{(1)})$, and $Y_1, \cdots, Y_m$ from $p(\boldsymbol{y}|\boldsymbol{\theta}^{(2)})$, where $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \cdots, \theta_s^{(1)})$, and $\boldsymbol{\theta}^{(2)} = (\theta_1^{(2)}, \cdots, \theta_s^{(2)})$. We often wish to test whether some function of $\boldsymbol{\theta}^{(1)}$ is equal to some function of $\boldsymbol{\theta}^{(2)}$.

Example 2.24

Suppose $X_1, \cdots, X_n$ are iid Exponential(θ_1) and $Y_1, \cdots, Y_m$ are iid Exponential(θ_2), and we wish to test $H_0 : \theta_1 = \theta_2$ versus $H_1 : \theta_1 \neq \theta_2$. Now

$$\begin{aligned} \boldsymbol{\Theta}_0 &= \{(\theta_1, \theta_2) : \theta_1 = \theta_2, \theta_1 > 0, \theta_2 > 0\}, \\ \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 &= \{(\theta_1, \theta_2) : \theta_1 > 0, \theta_2 > 0\}. \end{aligned}$$

Here, both θ_1 and θ_2 are unknown. Let us call the common value of θ_1

and θ_2 under H_0 θ, and thus $H_0 : \theta_1 = \theta_2 = \theta$, θ unknown. Now

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{y}|\theta_1, \theta_2) &= p(\boldsymbol{x}|\theta_1)p(\boldsymbol{y}|\theta_2) \\ &= \left(\theta_1^n e^{-\theta_1 \sum\limits_{i=1}^{n} X_i}\right) \left(\theta_2^m e^{-\theta_2 \sum\limits_{j=1}^{m} Y_j}\right). \end{aligned}$$

Under H_0, $\theta_1 = \theta_2 = \theta$, and therefore

$$p(\boldsymbol{x}, \boldsymbol{y}|\theta) = \theta^{n+m} e^{-\theta\left(\sum\limits_{i=1}^{n} X_i + \sum\limits_{j=1}^{m} Y_j\right)}.$$

$$\begin{aligned} \log p(\boldsymbol{x}, \boldsymbol{y}|\theta) &= (n+m)\log(\theta) - \theta\left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j\right) \\ &= (n+m)\log(\theta) - \theta\left(n\bar{X} + m\bar{Y}\right). \end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial\theta}\log p(\boldsymbol{x},\boldsymbol{y}|\theta) &= 0\\ \Rightarrow \hat{\theta} &= \frac{n+m}{n\bar{X}+m\bar{Y}},\end{aligned}$$

and so

$$\begin{aligned}&\sup_{(\theta_1,\theta_2)\in\boldsymbol{\Theta}_0} p(\boldsymbol{x},\boldsymbol{y}|\theta_1,\theta_2)\\ &= \left(\frac{n+m}{n\bar{X}+m\bar{Y}}\right)^{n+m} e^{-\frac{n+m}{n\bar{X}+m\bar{Y}}}(n\bar{X}+m\bar{Y})\\ &= \left(\frac{n+m}{n\bar{X}+m\bar{Y}}\right)^{n+m} e^{-(n+m)}.\end{aligned}$$

Under H_1, we have

$$\sup_{(\theta_1,\theta_2)\in\boldsymbol{\Theta}} p(\boldsymbol{x},\boldsymbol{y}|\theta_1,\theta_2) = \sup_{\theta_1>0,\theta_2>0} p(\boldsymbol{x},\boldsymbol{y}|\theta_1,\theta_2).$$

$$p(\boldsymbol{x}, \boldsymbol{y}|\theta_1, \theta_2) = \left(\theta_1^n e^{-\theta_1 \sum\limits_{i=1}^{n} X_i}\right) \left(\theta_2^m e^{-\theta_2 \sum\limits_{j=1}^{m} Y_j}\right).$$

$\frac{\partial}{\partial \theta_j} \log p(\boldsymbol{x}, \boldsymbol{y}|\theta_1, \theta_2) = 0, \quad j = 1, 2$, leads to $\hat{\theta}_1 = \frac{1}{\bar{X}}$ and $\hat{\theta}_2 = \frac{1}{\bar{Y}}$, and thus

$$\begin{aligned} \sup_{(\theta_1,\theta_2)\in\boldsymbol{\Theta}} p(\boldsymbol{x}, \boldsymbol{y}|\theta_1, \theta_2) &= \left(\bar{X}^{-n} e^{-\bar{X}^{-1} n\bar{X}}\right) \left(\bar{Y}^{-m} e^{-\bar{Y}^{-1} m\bar{Y}}\right) \\ &= \bar{X}^{-n} \bar{Y}^{-m} e^{-(n+m)}. \end{aligned}$$

Thus the LRT rejects H_0 if

$$\begin{aligned} \Lambda &= \frac{\left(\frac{n+m}{n\bar{X}+m\bar{Y}}\right)^{n+m} e^{-(n+m)}}{\bar{X}^{-n} \bar{Y}^{-m} e^{-(n+m)}} < k \\ &\Leftrightarrow \frac{\bar{X}^n \bar{Y}^m}{(n\bar{X} + m\bar{Y})^{n+m}} < k. \end{aligned}$$

Thus,

$$\phi(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} 1 & \text{if} \quad \frac{\bar{X}^n\bar{Y}^m}{(n\bar{X}+m\bar{Y})^{n+m}} < k \\ 0 & \text{if} \quad \frac{\bar{X}^n\bar{Y}^m}{(n\bar{X}+m\bar{Y})^{n+m}} > k \end{cases},$$

where k is chosen to make the test level α.
Likelihood ratio tests can also be applied for the following one sample or two sample problems.

one-sample case

Suppose $X_1, \cdots, X_n$ are iid with joint density $p(\boldsymbol{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_s)$. Suppose we wish to test

$$\begin{array}{l} H_0 : \theta_1 = \theta_{10}, \theta_2 = \theta_{20}, \cdots, \theta_s = \theta_{s0} \\ H_1 : \theta_1 = \theta_{11}, \theta_2 = \theta_{21}, \cdots, \theta_s = \theta_{s1} \end{array},$$

where $\theta_{10}, \cdots, \theta_{s0}, \theta_{11}, \cdots, \theta_{s1}$, are specified values. This is just a simple versus simple situation and the LRT is just the NP test in this case.

Subsets of the above hypotheses can also be considered, such as
$H_0 : \theta_1 = \theta_{10}, \theta_2 = \theta_{20}, \theta_3 = \theta_4 = \cdots = \theta_s = \theta,$
θ unknown, and so forth. Any time we test simple versus simple, the NP test=LRT.

Example 2.25

Suppose $X_1, \cdots, X_n$ are iid from $\mathrm{N}(\mu, \sigma^2)$, where (μ, σ^2) are both unknown. We wish to test

$$H_0 : \mu = \mu_0, \sigma^2 = \sigma_0^2 \quad \text{versus} \quad H_1 : \mu = \mu_1, \sigma^2 = \sigma_1^2,$$

where $\mu_1 > \mu_0, \sigma_1^2 < \sigma_0^2$. We can use the NP lemma to test this hypothesis, which is a special case of the LRT. We reject H_0 if

$$\Lambda = \frac{\sup\limits_{\mu=\mu_0, \sigma^2=\sigma_0^2} p(\boldsymbol{x}|\mu, \sigma^2)}{\sup\limits_{(\mu_0,\sigma_0^2)\cup(\mu_1,\sigma_1^2)} p(\boldsymbol{x}|\mu, \sigma^2)}$$

$$= \quad \frac{(2\pi)^{-n/2}\sigma_0^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i-\mu_0)^2}}{(2\pi)^{-n/2}\sigma_1^n e^{-\frac{1}{2\sigma_1^2}\sum_{i=1}^n (X_i-\mu_1)^2}} < k$$

$$\Rightarrow \quad \sigma_1^2 \sum_{i=1}^n (X_i - \mu_0)^2 - \sigma_0^2 \sum_{i=1}^n (X_i - \mu_1)^2 > k.$$

Thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \sigma_1^2 \sum_{i=1}^n (X_i - \mu_0)^2 - \sigma_0^2 \sum_{i=1}^n (X_i - \mu_1)^2 > k \\ 0 & \text{if} \quad \sigma_1^2 \sum_{i=1}^n (X_i - \mu_0)^2 - \sigma_0^2 \sum_{i=1}^n (X_i - \mu_1)^2 < k \end{cases} .$$

$\phi(\boldsymbol{x})$ =LRT=NP since we were testing simple versus simple.
This same idea extends to two-sample problems.

Example 2.26

Suppose $X_1, \cdots, X_n$ are iid from Exponential(θ_1) and $Y_1, \cdots, Y_m$ are iid

Exponential(θ_2). Suppose we wish to test

$$H_0 : \theta_1 = 1, \theta_2 = 2 \quad \text{versus} \quad H_1 : \theta_1 = 2, \theta_2 = 3.$$

$$p(\boldsymbol{x}, \boldsymbol{y} | \theta_1, \theta_2) = \theta_1^n \theta_2^m e^{-\theta_1 \sum\limits_{i=1}^{n} X_i} e^{-\theta_2 \sum\limits_{j=1}^{m} Y_j}.$$

We reject H_0 when

$$\begin{aligned}
\Lambda &= \frac{p(\boldsymbol{x}, \boldsymbol{y} | \theta_1 = 1, \theta_2 = 2)}{p(\boldsymbol{x}, \boldsymbol{y} | \theta_1 = 2, \theta_2 = 3)} \\
&= \frac{1^n e^{-\sum\limits_{i=1}^{n} X_i} 2^m e^{-2\sum\limits_{j=1}^{m} Y_j}}{\left(2^n e^{-2\sum\limits_{i=1}^{n} X_i}\right)\left(3^m e^{-3\sum\limits_{j=1}^{m} Y_j}\right)} < k \\
&\Leftrightarrow \quad e^{\sum\limits_{i=1}^{n} X_i + \sum\limits_{j=1}^{m} Y_j} < k \\
&\Rightarrow \quad \sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j < k.
\end{aligned}$$

Thus

$$\phi(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 1 & \text{if} \quad \sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j < k \\ 0 & \text{if} \quad \sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j > k \end{cases} .$$

To make this an α level test, we need small

$$\alpha = P_0 \left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j < k \right).$$

$$\text{Now} \quad \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta_1), \quad \sum_{j=1}^{m} Y_j \sim \text{Gamma}(m, \theta_2).$$

Under H_0, $\left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j \right) \sim \text{Gamma}(n, 1) + \text{Gamma}(m, 2)$.

If $\theta_1 = \theta_2 = \theta$, say then $\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j \sim \text{Gamma}(n + m, \theta)$. So if we test

$H_0 : \theta_1 = \theta_2 = 1$ versus $H_1 : \theta_1 = \theta_2 = 2$, then

$$\begin{aligned}\alpha &= P_0\left(\sum_{i=1}^{n} X_i + \sum_{j=1}^{m} Y_j < k\right) \\ &= F(k), \quad F = \text{cdf of Gamma}(n+m, 1),\end{aligned}$$

so that $k = F^{-1}(\alpha)$.

Example 2.27

Suppose $X_1, \cdots, X_n$ are iid, where

$$p(x|\lambda, \theta) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x-\theta}{\lambda}} & x \geq \theta, \lambda > 0 \\ 0 & \text{otherwise} \end{cases},$$

where λ and θ are both unknown. Suppose we wish to test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.

We wish to derive the LRT for this hypothesis. We have
$\boldsymbol{\Theta} = \{(\lambda, \theta) : \lambda > 0, -\infty < \theta < \infty\}$,
$\boldsymbol{\Theta}_0 = \{(\lambda, \theta) : \theta \leq \theta_0, \lambda > 0\}$,
$\boldsymbol{\Theta}_1 = \{(\lambda, \theta) : \theta > \theta_0, \lambda > 0\}$.

$$\begin{aligned} p(\boldsymbol{x}|\lambda, \theta) &= \prod_{i=1}^{n} \frac{1}{\lambda} e^{-\frac{(X_i - \theta)}{\lambda}} I\left(X_i \geq \theta\right) \\ &= \lambda^{-n} e^{-\frac{1}{\lambda}\sum_{i=1}^{n}(X_i - \theta)} I\left(X_{(1)} \geq \theta\right). \end{aligned}$$

Under H_0,

$$\sup_{\theta \in \boldsymbol{\Theta}_0} p(\boldsymbol{x}|\lambda, \theta) = \sup_{\lambda > 0} p(\boldsymbol{x}|\lambda, \theta_0).$$

$$p(\boldsymbol{x}|\lambda, \theta_0) = \lambda^{-n} e^{-\frac{1}{\lambda}\sum_{i=1}^{n}(X_i - \theta_0)} I\left(X_{(1)} \geq \theta_0\right),$$

$$\log p(\boldsymbol{x}|\lambda, \theta_0) = -n \log \lambda - \frac{1}{\lambda} \sum_{i=1}^{n} (X_i - \theta_0).$$

$$\frac{\partial}{\partial \lambda} \log p(\boldsymbol{x}|\lambda, \theta_0) = -\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum_{i=1}^{n} (X_i - \theta_0) = 0$$

$$\Rightarrow \quad n\lambda = \sum_{i=1}^{n} (X_i - \theta_0)$$

$$\Rightarrow \quad \hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \theta_0).$$

So

$$\sup_{\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\lambda, \theta_0) = \left(\frac{\sum\limits_{i=1}^{n} (X_i - \theta_0)}{n} \right)^{-n} e^{-n} I\left(X_{(1)} \geq \theta_0\right).$$

Under $H_0 \cup H_1$,

$$\sup_{\boldsymbol{\Theta}} p(\boldsymbol{x}|\lambda, \theta) = \left\{ \sup_{\boldsymbol{\Theta}} \lambda^{-n} e^{-\frac{1}{\lambda} \sum\limits_{i=1}^{n} (X_i - \theta)} I\left(X_{(1)} \geq \theta\right) \right\}.$$

It follows that the MLE of θ is $X_{(1)}$, and thus $\hat{\theta} = X_{(1)}$, and the MLE of

λ is $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n}(X_i - X_{(1)})$. Therefore,

$$\sup_{\boldsymbol{\Theta}} p(\boldsymbol{x}|\lambda,\theta) = \left(\frac{\sum_{i=1}^{n}\left(X_i - X_{(1)}\right)}{n}\right)^{-n} e^{-n}.$$

Thus

$$\begin{aligned}
\Lambda &= \frac{\left(\frac{\sum_{i=1}^{n}(X_i-\theta_0)}{n}\right)^{-n} e^{-n} I\left(X_{(1)} \geq \theta_0\right)}{\left(\frac{\sum_{i=1}^{n}\left(X_i-X_{(1)}\right)}{n}\right)^{-n} e^{-n}} < k \\
&\Leftrightarrow \frac{\sum_{i=1}^{n}\left(X_i - X_{(1)}\right)}{\sum_{i=1}^{n}(X_i - \theta_0)} I\left(X_{(1)} \geq \theta_0\right) < k,
\end{aligned}$$

and thus

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \frac{\sum_{i=1}^{n}(X_i - X_{(1)})}{\sum_{i=1}^{n}(X_i - \theta_0)} I\left(X_{(1)} \geq \theta_0\right) < k \\ 0 & \frac{\sum_{i=1}^{n}\left(X_i - X_{(1)}\right)}{\sum_{i=1}^{n}(X_i - \theta_0)} I\left(X_{(1)} \geq \theta_0\right) > k \end{cases}$$

To make this an α level test, we need

$$\begin{aligned} \alpha &= E_{\theta_0}[\phi(\boldsymbol{x})] \\ &= P_{\theta_0}\left(X_{(1)} \geq \theta_0 \text{ and } \frac{\sum_{i=1}^{n}(X_i - X_{(1)})}{\sum_{i=1}^{n}(X_i - \theta_0)} < k \right). \end{aligned}$$

Suppose that λ was known in this example, and we wish to test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.
Under H_0,

$$\sup_{\theta \in \boldsymbol{\Theta}_0} p(\boldsymbol{x}|\lambda, \theta) = \lambda^{-n} e^{-\frac{1}{\lambda} \sum_{i=1}^{n} (X_i - \theta_0)} I\left(X_{(1)} \geq \theta_0\right).$$

Under $H_0 \cup H_1$, $\hat{\theta} = X_{(1)}$, and

$$\sup_{\theta \in \boldsymbol{\Theta}} p(\boldsymbol{x}|\lambda, \theta) = \lambda^{-n} e^{-\frac{1}{\lambda} \sum_{i=1}^{n} \left(X_i - X_{(1)}\right)}.$$

Thus

$$\begin{aligned} \Lambda &= \frac{\lambda^{-n} e^{-\frac{1}{\lambda} \sum_{i=1}^{n} (X_i - \theta_0)}}{\lambda^{-n} e^{-\frac{1}{\lambda} \sum_{i=1}^{n} \left(X_i - X_{(1)}\right)}} I\left(X_{(1)} \geq \theta_0\right) < k \\ &\Rightarrow e^{-\frac{n}{\lambda}(\theta_0 - X_{(1)})} I\left(X_{(1)} \geq \theta_0\right) < k. \end{aligned}$$

Now

$$\begin{aligned} & e^{-\frac{n}{\lambda}\left(\theta_0 - X_{(1)}\right)} < k \\ \Leftrightarrow \quad & X_{(1)} > \theta_0 - \frac{\lambda}{n}\log(k). \end{aligned}$$

Thus

$$\begin{aligned} \phi(\boldsymbol{x}) &= \begin{cases} 1 & X_{(1)} \geq \theta_0 \text{ and } X_{(1)} > \theta_0 - \frac{\lambda}{n}\log(k) \\ 0 & \text{otherwise} \end{cases} \\ \Leftrightarrow \phi(\boldsymbol{x}) &= \begin{cases} 1 & X_{(1)} \geq \max(\theta_0, k_1) \\ 0 & \text{otherwise,} \end{cases} \\ k_1 &= \theta_0 - \frac{\lambda}{n}\log(k), \text{ and so} \\ \phi(\boldsymbol{x}) &= \begin{cases} 1 & X_{(1)} \geq \max(\theta_0, k) \\ 0 & X_{(1)} < \max(\theta_0, k). \end{cases} \end{aligned}$$

To make this test of level α, we need

$$\alpha = P_{\theta_0}\left(X_{(1)} \geq \max(\theta_0, k)\right).$$

Let

$$\begin{aligned} m &= \max(\theta_0, k). \\ \alpha &= P_{\theta_0}\left(X_{(1)} \geq m\right) \\ &= \int_m^{\infty} p_{X(1)}(x)\, dx \\ &= \int_m^{\infty} \frac{n}{\lambda} e^{-\frac{n}{\lambda}(x-\theta_0)}\, dx \\ &= -e^{-\frac{n}{\lambda}(x-\theta_0)}\Big|_m^{\infty} = e^{-\frac{n}{\lambda}(m-\theta_0)}. \end{aligned}$$

Thus

$$\alpha = e^{-\frac{n}{\lambda}(m-\theta_0)} \Rightarrow m = \theta_0 - \frac{\lambda}{n}\log(\alpha)$$

$$
\begin{aligned}
\Rightarrow \quad & \max(\theta_0, k) = \theta_0 - \frac{\lambda}{n}\log(\alpha) \\
\Rightarrow \quad & k = \theta_0 - \frac{\lambda}{n}\log(\alpha).
\end{aligned}
$$

$\max(\theta_0, k) = k$ since $\theta_0 - \frac{\lambda}{n}\log(\alpha) > \theta_0$.
Thus, the LRT is

$$
\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad X_{(1)} \geq \theta_0 - \frac{\lambda}{n}\log(\alpha) \\ 0 & \text{if} \quad X_{(1)} < \theta_0 - \frac{\lambda}{n}\log(\alpha). \end{cases}
$$

Example 2.28

Suppose $X_1, \cdots, X_n$ are iid $\mathrm{U}(\lambda, \theta)$, where λ is known and θ is unknown.
Suppose we wish to test $H_0 : \theta \geq \theta_0$ versus $H_1 : \theta < \theta_0$.
Let us derive the LRT for this hypothesis.

$\boldsymbol{\Theta} = \{\theta : \lambda \leq \theta < \infty\}$,
$\boldsymbol{\Theta}_0 = \{\theta : \theta \geq \theta_0\}$,
$\boldsymbol{\Theta}_1 = \{\theta : \theta < \theta_0\}$.

$$p(\boldsymbol{x}|\theta) = (\theta - \lambda)^{-n} I\left(\lambda \le X_{(1)} \le X_{(n)} \le \theta\right),$$

$$\sup_{\theta \in \boldsymbol{\Theta}_0} p(\boldsymbol{x}|\theta) = (\theta_0 - \lambda)^{-n} I\left(\lambda \le X_{(1)} \le X_{(n)} \le \theta_0\right).$$

Under $H_0 \cup H_1$, $\hat{\theta} = X_{(n)}$, and so

$$\sup_{\boldsymbol{\Theta}} p(\boldsymbol{x}|\theta) = \left(X_{(n)} - \lambda\right)^{-n} I\left(\lambda \le X_{(1)} \le X_{(n)}\right).$$

Thus

$$\Lambda = \frac{(\theta_0 - \lambda)^{-n} I\left(\lambda \le X_{(1)} \le X_{(n)} \le \theta_0\right)}{\left(X_{(n)} - \lambda\right)^{-n} I\left(\lambda \le X_{(1)} \le X_{(n)}\right)} < k$$

$$\Rightarrow \phi(\boldsymbol{x}) = \begin{cases} 1 & \left(X_{(n)} - \lambda\right) I\left(\lambda \le X_{(1)} \le X_{(n)} \le \theta_0\right) < k \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow \phi(\boldsymbol{x}) = \begin{cases} 1 & X_{(n)} < \lambda + k \quad \text{and} \quad \lambda \le X_{(n)} \le \theta_0 \\ 0 & \text{otherwise} \end{cases}$$

$$
\begin{aligned}
\Rightarrow \phi(\boldsymbol{x}) &= \begin{cases} 1 & X_{(n)} < k \quad \text{and} \quad \lambda \le X_{(n)} \le \theta_0 \\ 0 & \text{otherwise} \end{cases} \\
\Rightarrow \phi(\boldsymbol{x}) &= \begin{cases} 1 & \lambda \le X_{(n)} \le \min(\theta_0, k) \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}
$$

To find k to make this of level α, let $m = \min(\theta_0, k)$,

$$
\begin{aligned}
\alpha &= P_{\theta_0}\left(\lambda \le X_{(n)} \le m\right) \\
&= \int_{\lambda}^{m} p_{X(n)}(x)\,dx \\
&= \int_{\lambda}^{m} \frac{n}{(\theta_0 - \lambda)^n}(x-\lambda)^{n-1}\,dx \\
&= \frac{(x-\lambda)^n}{(\theta_0-\lambda)^n}\bigg|_{\lambda}^{m} = \frac{(m-\lambda)^n}{(\theta_0-\lambda)^n} = \alpha \\
\Rightarrow m &= \alpha^{\frac{1}{n}}(\theta_0 - \lambda) + \lambda.
\end{aligned}
$$

Since $m = \min(\theta_0, k) = \alpha^{\frac{1}{n}}(\theta_0 - \lambda) + \lambda < \theta_0$,
we have, $k = \alpha^{\frac{1}{n}}(\theta_0 - \lambda) + \lambda$.
Thus, our α level test is

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \lambda \leq X_{(n)} \leq \lambda + \alpha^{\frac{1}{n}}(\theta_0 - \lambda) \\ 0 & \text{otherwise.} \end{cases}$$

Asymptotic Distribution of the Likelihood Ratio Statistic

As indicated in Theorem 2.6 and other examples, the LRT is often equivalent to a test based on a statistic whose distribution under H_0 can be used to determine the rejection region of the LRT with size α. When this technique fails, it is difficult or even impossible to find a LRT with size α, even if the cdf of Λ is continuous. The following theorem shows that in the iid case, we can obtain the asymptotic distribution (under H_0) of Λ so that an LRT having asymptotic significance level α can be obtained.

Theorem 2.7

Assume the "usual" regularity conditions for a likelihood function, [$p_\theta(x)$ is twice continuously differentiable, $I(\theta)$ exists and is finite].

Then,

$$\text{as } n \to \infty, \quad -2\log\Lambda \xrightarrow{H_0} \chi^2_r,$$

where $r = \text{dimension}(\boldsymbol{\Theta}) - \text{dimension}(\boldsymbol{\Theta}_0)$.

Example 2.29

Suppose $X_1, \cdots, X_n$ are iid Poisson(θ). We wish to test $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$.

As shown earlier,

$$\Lambda = \frac{\theta_0^{n\bar{X}} e^{-n\theta_0}}{\bar{X}^{n\bar{X}} e^{-n\bar{X}}}.$$

$$-2\log\Lambda = -2\left\{n\bar{X}(\log(\theta_0) - \log(\bar{X})) + n(\bar{X} - \theta_0)\right\}$$

$$-2\log\Lambda \xrightarrow{H_0} \chi_1^2.$$

Here, we have dimension$(\boldsymbol{\Theta}_0) = 0$, dimension$(\boldsymbol{\Theta}) = 1$.

Example 2.30

Suppose $X_1, \cdots, X_n$ are iid $\mathrm{N}(\mu, \sigma^2)$, where (μ, σ^2) are unknown parameters. We wish to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$.

$$\Lambda = \frac{\left[\sum_{i=1}^{n}(X_i - \mu_0)^2\right]^{-n/2}}{\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right]^{-n/2}}.$$

$$\begin{aligned} -2\log\Lambda &= n\left[\log\left(\sum_{i=1}^{n}(X_i - \mu_0)^2\right) - \log\left(\sum_{i=1}^{n}(X_i - \bar{X})^2\right)\right] \\ &\xrightarrow{H_0} \chi_1^2 \quad \text{as} \quad n \to \infty. \end{aligned}$$

Here, dimension$(\Theta_0) = 1$, dimension$(\Theta) = 2$, $r = 2 - 1 = 1$.

There are two other very popular tests used in biostatistics that are asymptotically equivalent to the LRT. These are the Wald test and the Score test.

Wald Test:

Suppose we wish to test: $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Wald (1943) introduced a test that rejects H_0 when the value of

$$W_n = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \boldsymbol{I}_n(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

is large, where $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_r)$, $\boldsymbol{I}_n(\boldsymbol{\theta})$ is the Fisher information based on $(X_1, \cdots, X_n)$, $\hat{\boldsymbol{\theta}}$ =MLE of $\boldsymbol{\theta}$.

Theorem 2.8

Assume the same regularity conditions as Theorem 2.7. Then as $n \to \infty$, $W_n \xrightarrow{H_0} \chi_r^2$, and therefore the test rejects H_0 at level α if $W_n > \chi^2_{r,1-\alpha}$.

Score Test:

Rao (1947) introduced a Score test, that rejects $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ when the value of

$$R_n = [\nabla l(\boldsymbol{\theta}_0)]' \left(\boldsymbol{I}_n(\boldsymbol{\theta}_0)\right)^{-1} [\nabla l(\boldsymbol{\theta}_0)]$$

is large, where $\nabla l(\boldsymbol{\theta}_0) = \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \log p_{\boldsymbol{\theta}}(x_i)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$, and $\boldsymbol{I}_n(\boldsymbol{\theta})$ is the Fisher information.

As $n \to \infty$, $R_n \xrightarrow{H_0} \chi_r^2$.

Proofs of these results are given in Shao, pages 432 - 435.

Example 2.31

Consider the logistic regression model

$$p(y_i|\boldsymbol{x}_i, \boldsymbol{\beta}) = \exp\left\{y_i\boldsymbol{x}_i'\boldsymbol{\beta} - \log(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}})\right\},$$

where y_i is a binary response, $p(y_i = 1|\boldsymbol{x}_i) = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}$, $\boldsymbol{x}_i$ is a $p \times 1$ vector of fixed covariates, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients. Let $\boldsymbol{y} = (y_1, \ldots, y_n)$. The likelihood function of $\boldsymbol{\beta}$ based on n observations is given by

$$p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}) = \exp\left\{\sum_{i=1}^{n}\left(y_i\boldsymbol{x}_i'\boldsymbol{\beta} - \log(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}})\right)\right\}.$$

Suppose we wish to test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. Here, the MLE of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, does not have a closed form and must be iteratively obtained by a Newton-Raphson algorithm.

The LRT is given by

$$\Lambda = \frac{p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}_0)}{p(\boldsymbol{y}|\boldsymbol{X}, \hat{\boldsymbol{\beta}})},$$

and thus

$$\begin{aligned} -2\log\Lambda &= -2\left[\log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}_0) - \log p(\boldsymbol{y}|\boldsymbol{X}, \hat{\boldsymbol{\beta}})\right] \\ &= 2\left[\log p(\boldsymbol{y}|\boldsymbol{X}, \hat{\boldsymbol{\beta}}) - \log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}_0)\right], \end{aligned}$$

where

$$\log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i \boldsymbol{x}_i' \boldsymbol{\beta} - \log(1 + e^{\boldsymbol{x}_i' \boldsymbol{\beta}}) \right),$$

and thus $\quad -2\log\Lambda \xrightarrow{H_0} \chi^2_p.$

<u>Note:</u>

In general regression problems

$$\begin{aligned}-2\log\Lambda &= 2\left[\text{log-likelihood (full model)}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(1)}}\right. \\ &\qquad \left.-\text{log-likelihood (reduced model)}\Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^{(0)}}\right],\end{aligned}$$

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{(0)} &= \text{estimate of } \boldsymbol{\beta} \text{ under } H_0, \\ \hat{\boldsymbol{\beta}}^{(1)} &= \text{estimate of } \boldsymbol{\beta} \text{ under } \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1.\end{aligned}$$

The Wald test is given by

$$W_n = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{I}_n(\hat{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0),$$

where

$$\begin{aligned}\boldsymbol{I}_n(\boldsymbol{\beta}) &= -E\left\{\frac{\partial^2 \log p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right\} \\ &= \boldsymbol{X}'\boldsymbol{V}\boldsymbol{X},\end{aligned}$$

where $\boldsymbol{X}$ is the $n \times p$ matrix of covariates, and $\boldsymbol{V}$ is an $n \times n$ diagonal matrix, with *ith* diagonal element $v_{ii} = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{(1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}})^2}$,

$$\boldsymbol{V} = \begin{pmatrix} \frac{e^{\boldsymbol{x}_1'\boldsymbol{\beta}}}{(1+e^{\boldsymbol{x}_1'\boldsymbol{\beta}})^2} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \frac{e^{\boldsymbol{x}_n'\boldsymbol{\beta}}}{(1+e^{\boldsymbol{x}_n'\boldsymbol{\beta}})^2} \end{pmatrix}_{n\times n}.$$

$$W_n \xrightarrow{H_0} \chi_p^2.$$

Finally, the score test is given by

$$R_n = [\nabla l(\boldsymbol{\beta}_0)]'\boldsymbol{I}_n(\boldsymbol{\beta}_0)^{-1}[\nabla l(\boldsymbol{\beta}_0)],$$

where

$$\begin{aligned} \nabla l(\boldsymbol{\beta}) &= \frac{\partial}{\partial \boldsymbol{\beta}} \log p(\boldsymbol{y}|\boldsymbol{X}, \boldsymbol{\beta}) \\ &= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^{n} \left\{ y_i \boldsymbol{x}_i' \boldsymbol{\beta} - \log(1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}) \right\} \\ &= \boldsymbol{X}'\boldsymbol{S}(\boldsymbol{\beta}), \end{aligned}$$

where

$$\boldsymbol{S}(\boldsymbol{\beta}) = \begin{pmatrix} y_1 - \mu_1 \\ \vdots \\ y_n - \mu_n \end{pmatrix}, \quad \mu_i = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}.$$

Thus

$$R_n = \left[\boldsymbol{S}'(\boldsymbol{\beta}_0)\boldsymbol{X}\right]\left[\boldsymbol{X}'\boldsymbol{V}(\boldsymbol{\beta}_0)\boldsymbol{X}\right]^{-1}\left[\boldsymbol{X}'\boldsymbol{S}(\boldsymbol{\beta}_0)\right],$$

and

$$R_n \xrightarrow{H_0} \chi^2_p \quad \text{as} \quad n \to \infty.$$

Note that

$$
\begin{aligned}
& \frac{\partial}{\partial \beta_j} \left\{ \sum_{i=1}^{n} \left(y_i \boldsymbol{x}_i' \boldsymbol{\beta} - \log(1 + e^{\boldsymbol{x}_i' \boldsymbol{\beta}}) \right) \right\} \\
= & \sum_{i=1}^{n} \left\{ y_i x_{ij} - \frac{e^{\boldsymbol{x}_i' \boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i' \boldsymbol{\beta}}} (x_{ij}) \right\} \\
= & \sum_{i=1}^{n} \left(y_i - \frac{e^{\boldsymbol{x}_i' \boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i' \boldsymbol{\beta}}} \right) x_{ij}
\end{aligned}
$$

$$
= \sum_{i=1}^{n} (y_i - \mu_i)\, x_{ij} \quad , \quad j = 1, \cdots, p.
$$

Thus, in matrix form,

$$
\frac{\partial}{\partial \boldsymbol{\beta}} \log p(\boldsymbol{y} | \boldsymbol{X}, \boldsymbol{\beta}) = \boldsymbol{X}' \boldsymbol{S}.
$$

Bayesian Hypothesis Testing

In classical hypothesis testing, a null hypothesis $H_0 : \theta \in \mathbf{\Theta}_0$ and an alternative hypothesis, $H_1 : \theta \in \mathbf{\Theta}_1$, are specified. A test procedure is evaluated in terms of the probabilities of Type I and Type II error. These probabilities of error represent the chance that a sample is observed for which the test procedure will result in the wrong hypothesis being accepted.

In Bayesian analysis, the task of deciding between H_0 and H_1 is conceptually more straightforward. One merely calculates the posterior probabilities $p(\mathbf{\Theta}_0|\boldsymbol{x})$ and $p(\mathbf{\Theta}_1|\boldsymbol{x})$ and decides between H_0 and H_1 accordingly. The conceptual advantage is that $p(\mathbf{\Theta}_0|\boldsymbol{x})$ and $p(\mathbf{\Theta}_1|\boldsymbol{x})$ are the actual probabilities of the hypotheses in light of the data and the prior distributions.

The quantity used to test hypotheses in the Bayesian framework is called the Bayes factor. The Bayes factor is the Bayesian analogue of the likelihood ratio test. Suppose we wish to test H_0 versus H_1, and let $\lambda_0 \equiv p(H_0)$, $\lambda_1 \equiv p(H_1)$ denote the prior probabilities of H_0 and H_1, respectively. Let $p(H_0|\boldsymbol{x})$ and $p(H_1|\boldsymbol{x})$ denote the posterior probabilities of the hypotheses. Then the Bayes factor in favor of H_0 is defined as the posterior to prior odds of H_0 divided by the posterior to prior odds of H_1. That is

$$B = \frac{\left\{\frac{p(H_0|\boldsymbol{x})}{p(H_0)}\right\}}{\left\{\frac{p(H_1|\boldsymbol{x})}{p(H_1)}\right\}} = \frac{\left\{\frac{p(H_0|\boldsymbol{x})}{p(H_1|\boldsymbol{x})}\right\}}{\left\{\frac{p(H_0)}{p(H_1)}\right\}}. \tag{2.25}$$

Now by Bayes theorem, we know that

$$p(H_0|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|H_0)p(H_0)}{p(\boldsymbol{x}|H_0)p(H_0) + p(\boldsymbol{x}|H_1)p(H_1)} \tag{2.26}$$

and a similar formula holds for $p(H_1|\boldsymbol{x})$.
Substituting (2.26) into (2.25) and the corresponding formula for $p(H_1|\boldsymbol{x})$, we can rewrite the Bayes factor in favor of H_0 as

$$B = \frac{p(\boldsymbol{x}|H_0)}{p(\boldsymbol{x}|H_1)}. \tag{2.27}$$

B in (2.27) has a very nice interpretation. Large value of B in (2.27) favor H_0. The following table was devised by Jeffreys:

$1 \le B \le 3$	weak evidence for H_0
$3 < B \le 12$	positive
$12 < B \le 150$	strong
$B > 150$	decisive

More formally, if $H_0 : \theta \in \mathbf{\Theta}_0$, $H_1 : \theta \in \mathbf{\Theta}_1$, $\mathbf{\Theta} = \mathbf{\Theta}_0 \cup \mathbf{\Theta}_1$, then the Bayes factor in favor of H_0 is defined as

$$
\begin{aligned}
B &= \frac{p(\boldsymbol{x}|H_0)}{p(\boldsymbol{x}|H_1)} \\
&\equiv \frac{\int_{\mathbf{\Theta}_0} p(\boldsymbol{x}|\theta, H_0)\lambda(\theta|H_0)\, d\theta}{\int_{\mathbf{\Theta}_1} p(\boldsymbol{x}|\theta, H_1)\lambda(\theta|H_1)\, d\theta},
\end{aligned}
$$

where $\lambda(\theta|H_i)$ is the prior distribution of θ under $H_i, i = 0, 1$. One immediate feature of the Bayes factor is that $p(\boldsymbol{x}|H_0)$ and $p(\boldsymbol{x}|H_1)$ are obtained by <u>integrating</u> over the parameter space, <u>not</u> maximizing over it. The Bayes factor has several advantages over the likelihood ratio test.

(1) We integrate over $\mathbf{\Theta}$, <u>not</u> maximize over it.

(2) B does <u>not</u> require nested models, whereas the LRT does.

(3) B has a nice interpretation of being a posterior to prior odds ratio.

(4) B reduces to the NP test in the simple versus simple case. That is, for $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$,

$$B = \frac{p(\boldsymbol{x}|\theta_0)}{p(\boldsymbol{x}|\theta_1)}.$$

(5) B can easily accommodate complicated composite hypotheses of any form.

$$B = \frac{\int_{\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\theta, H_0)\lambda(\theta|H_0)\, d\theta}{\int_{\boldsymbol{\Theta}_1} p(\boldsymbol{x}|\theta, H_1)\lambda(\theta|H_1)\, d\theta}.$$

Remarks:

(1) B is defined only when proper priors are used for θ. B is not well defined under improper priors.

(2) B may be sensitive to the choice of prior distribution and/or the choice of prior hyperparameters. For example, if $\theta \sim$N(μ_0, σ_0^2), B may be sensitive to the choice of σ_0^2 as $\sigma_0^2 \to \infty$.

Example 2.32

Suppose $X_1, \cdots, X_n$ are iid $\mathrm{N}(\theta, 1)$ and we wish to test $H_0 : \theta = 0$ versus $H_1 : \theta = 1$.
Then

$$\begin{aligned} B &= \frac{(2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right\}}{(2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} (X_i - 1)^2\right\}} \\ &= \exp\left\{\frac{n}{2} - \sum_{i=1}^{n} X_i\right\}. \end{aligned}$$

Suppose $n = 10$, $\sum_{i=1}^{n} X_i = 4.5$. Then

$$B = e^{\frac{10}{2} - 4.5} = e^{.5} = 1.65,$$

which is weak evidence in favor of $H_0 : \theta = 0$.

Example 2.33

Suppose $X_1, \cdots, X_n$ are iid $\mathrm{N}(\theta, 1)$ and we wish to test $H_0 : \theta = 0$ versus $H_1 : \theta \neq 0$. In this case,

$$B = \frac{p(\boldsymbol{x}|\theta = 0)}{\int_{\Theta_1} p(\boldsymbol{x}|\theta)\lambda(\theta)\, d\theta}.$$

Suppose, under H_1, we take $\lambda(\theta) = (2\pi)^{-1/2} e^{-\frac{1}{2}(\theta-1)^2} = \mathrm{N}(1, 1)$. Then

$$\begin{aligned} B &= \frac{(2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right\}}{\int_{-\infty}^{\infty} (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2\right\} (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(\theta - 1)^2\right\}\, d\theta} \\ &= \frac{(2\pi)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{n} X_i^2 - 2\theta\sum_{i=1}^{n} X_i + n\theta^2\right)\right\} \exp\left\{-\frac{1}{2}(\theta^2 - 2\theta + 1)\right\}\, d\theta} \end{aligned}$$

$$= \frac{(2\pi)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(1+\sum_{i=1}^{n} X_i^2\right)\right\} \exp\left\{\frac{\left(1+\sum_{i=1}^{n} X_i\right)^2}{2(n+1)}\right\} \exp\left\{-\frac{(n+1)}{2}\left(\theta - \frac{\left(1+\sum_{i=1}^{n} X_i\right)}{n+1}\right)^2\right\} d\theta}$$

$$= \frac{(2\pi)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} X_i^2\right\}}{(2\pi)^{\frac{1}{2}} (n+1)^{-1/2} \exp\left\{-\frac{1}{2}\left(1+\sum_{i=1}^{n} X_i^2\right)\right\} \exp\left\{\frac{(1+\sum_{i=1}^{n} X_i)^2}{2(n+1)}\right\} d\theta}$$

$$= (n+1)^{1/2} \exp\left\{\frac{1}{2}\right\} \exp\left\{-\frac{\left(1+\sum_{i=1}^{n} X_i\right)^2}{2(n+1)}\right\}.$$

If $n = 10$ and $\sum_{i=1}^{n} X_i = 5$, we get $B = 1.06$.

An alternative to the Bayes factor is the posterior probability of the hypothesis,

$$\begin{aligned} p(H_0|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|H_0)p(H_0)}{p(\boldsymbol{x}|H_0)p(H_0) + p(\boldsymbol{x}|H_1)p(H_1)} \\ &= 1 - p(H_1|\boldsymbol{x}). \end{aligned}$$

We can write the posterior probability in terms of the Bayes factor as

$$p(H_0|\boldsymbol{x}) = \frac{rB}{1 + rB},$$

where $r = \frac{p(H_0)}{p(H_1)}$.

Note that if $p(H_0) = p(H_1) = 1/2$, then $r = 1$ and $p(H_0|\boldsymbol{x}) = \frac{B}{1+B}$, so that B is a monotonic function of $p(H_0|\boldsymbol{x})$, and thus in this case, B is equivalent to the posterior probability. Now suppose we have k models that we want to compare, denoted by $m_1, \cdots, m_k$.

The posterior probability for model m_j is given by

$$p(m_j|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|m_j)p(m_j)}{\sum_{j=1}^{k} p(\boldsymbol{x}|m_j)p(m_j)},$$

where $p(m_j)$ =prior probability of model m_j, and

$$p(\boldsymbol{x}|m_j) = \int_{\boldsymbol{\Theta}_{m_j}} p(\boldsymbol{x}|\theta, m_j)\lambda(\theta|m_j)\, d\theta$$

is the marginal distribution of $\boldsymbol{x}$ under model m_j.

We can write

$$p(m_j|\boldsymbol{x}) = \frac{r_j B_{j1}}{\sum_{j=1}^{k} r_j B_{j1}},$$

where $r_j = \frac{p(m_j)}{p(m_1)}$, and B_{j1} is the Bayes factor in favor of m_j over m_1, $j = 1, \cdots, k$. Again, if the prior model probabilities are all equal, i.e., $p(m_1) = p(m_2) = \cdots = p(m_k) = \frac{1}{k}$, then the Bayes factor is a monotonic function of the posterior model probability and thus the two methods are equivalent in this case. This is clear since in this case $r_j = 1$ and

$$p(m_j|\boldsymbol{x}) = \frac{B_{j1}}{\sum_{j=1}^{k} B_{j1}}.$$

Note that $B_{11} = 1$.

The situation of comparing several models arises, for example, in variable subset selection problems in regression.

Let $\mathcal{M} = \{m_1, \cdots, m_k\}$ denote the collection of all models. $\mathcal{M}$ is often called the model space. In most model selection problems, such as variable subset selection, $\mathcal{M}$ is discrete and finite. In model selection problems, in addition to specifying prior distributions for all the parameters, we must also specify a discrete prior on the model space $\mathcal{M}$, that is, $p(m_j)$, where $\sum_{j=1}^{k} p(m_j) = 1$.

So a (fully) Bayesian approach to model selection (or hypothesis testing) is based on computing posterior model probabilities. Posterior model probabilities require

(1) Proper prior distributions for all of the parameters arising from the k possible models in the model space.
(2) A prior distribution on the model space $\mathcal{M}$.

A fully Bayesian approach to model selection is to compute posterior model probabilities for all possible models in the model space, and to select the model with the largest posterior probability.

There are also several **criterion-based** approaches to model selection that have a Bayesian motivation. A common criterion for model selection is called the BIC criterion (Schwarz, 1978, *Annals of Statistics*). BIC has a Bayesian motivation, and is given by

$$\text{BIC}_j = -2\log L(\hat{\theta}|m_j) + \log(n)p_j, \quad j = 1, \cdots, k,$$

where p_j =dimension of θ under model m_j and $\hat{\theta}$ is the MLE of θ (under model m).

There are many other Bayesian criteria for model selection. The advantage of criterion based methods for model selection is that:

1. One does not need to specify a prior on the model space.
2. Improper priors can be used to define the criteria.

On the other hand, posterior model probabilities require:

1. Proper prior distributions for all of the parameters arising from the k possible models in the model space. There are many priors to elicit (k proper priors).
2. A prior distribution on the model space $\mathcal{M}$.

Thus computing posterior model probabilities can be quite difficult in practice, especially when k is large. For example, in variable subset selection with 10 covariates, $k = 2^{10} = 1024$. Here, we would have to specify prior distributions for the regression coefficients for 1024 models, and we would need to specify 1024 probabilities for the model space.

When $k = 30$, we would need to specify $2^{30} \approx 10^9$ probabilities for the model space.

Criterion-based methods do not have such a requirement. Most criterion-based methods only require prior distributions on the parameters arising from the different models. BIC does not even require priors. However, criterion-based methods, even with a Bayesian motivation, are not fully Bayesian.

Example 2.34

Consider the linear regression model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i,$$

where the ϵ_i's are i.i.d. and $\epsilon_i \sim N(0, \sigma^2)$, $i = 1, \cdots, n$. We are interested in variable subset selection. In this problem, $\mathcal{M}$ consists of $k = 4$ models. In variable subset selection problems, we always include the intercept in every model by convention.

Thus we have

$$\begin{array}{ll} m_1 & \text{(intercept model)} \\ m_2 & (x_1) \\ m_3 & (x_2) \\ m_4 & (x_1, x_2) \end{array}$$

The notation (x_1) means the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i,$$

and so forth.
The regression coefficients have a <u>different</u> physical meaning from model to model, and thus to specify prior distributions, we have to think about the β_j's differently for the 4 models.

<u>Prior for the model space</u>
We have

$$\mathcal{M} = \{m_1, \cdots, m_4\}.$$

A uniform prior on $\mathcal{M}$ is $p(m_1) = \cdots = p(m_4) = \frac{1}{4}$.
Another prior might be
$p(m_1) = .1,\ p(m_2) = .3,\ p(m_3) = .4,\ p(m_4) = .2$, and so on. We need good methods to specify a prior for $\mathcal{M}$ because a uniform prior for $\mathcal{M}$ may not be desirable if k is large, say $k = 1024$.

Prior Distribution for the Regression Parameters

We have

m_1	intercept	(β_0)
m_2	(x_1)	(β_0, β_1)
m_3	(x_2)	(β_0, β_2)
m_4	(x_1, x_2)	$(\beta_0, \beta_1, \beta_2)$

$\tau = \frac{1}{\sigma^2}$ also needs a prior distribution. Thus we need to specify prior distributions for 9 parameters. Model selection for the linear model will be discussed in much more detail shortly.

Example 2.35

Suppose $X_1, \cdots, X_n$ are iid Bernoulli(θ), and we wish to test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$.
The Bayes factor for this hypothesis is

$$B = \frac{\int_{\boldsymbol{\Theta}_0} p(\boldsymbol{x}|\theta, H_0)\lambda(\theta|H_0)\, d\theta}{\int_{\boldsymbol{\Theta}_1} p(\boldsymbol{x}|\theta, H_1)\lambda(\theta|H_1)\, d\theta},$$

$$\begin{aligned} \boldsymbol{\Theta}_0 &= \{\theta : 0 < \theta \leq \theta_0\}, \\ \boldsymbol{\Theta}_1 &= \{\theta : \theta_0 < \theta < 1\}, \\ \boldsymbol{\Theta} &= \boldsymbol{\Theta}_0 \cup \boldsymbol{\Theta}_1 = \{\theta : 0 < \theta < 1\}. \end{aligned}$$

$\lambda(\theta|H_0)$ needs to have support on $\boldsymbol{\Theta}_0$, and $\lambda(\theta|H_1)$ needs to have support on $\boldsymbol{\Theta}_1$. That is

$$\int_{\boldsymbol{\Theta}_0} \lambda(\theta|H_0)\, d\theta = 1,$$

and

$$\int_{\boldsymbol{\Theta}_1} \lambda(\theta|H_1)\, d\theta = 1.$$

Suppose under H_0, $\theta \sim U(0, \theta_0)$, and under H_1, $\theta \sim U(\theta_0, 1)$, so that

$$\lambda(\theta|H_0) = \begin{cases} \frac{1}{\theta_0} & 0 < \theta < \theta_0 \\ 0 & \text{otherwise} \end{cases},$$

$$\lambda(\theta|H_1) = \begin{cases} \frac{1}{1-\theta_0} & \theta_0 < \theta < 1 \\ 0 & \text{otherwise} \end{cases}.$$

So

$$B = \frac{\int_0^{\theta_0} \theta^{\sum_{i=1}^{n} X_i} (1-\theta)^{n-\sum_{i=1}^{n} X_i} \left(\frac{1}{\theta_0}\right) d\theta}{\int_{\theta_0}^{1} \theta^{\sum_{i=1}^{n} X_i} (1-\theta)^{n-\sum_{i=1}^{n} X_i} \left(\frac{1}{1-\theta_0}\right) d\theta}. \tag{2.28}$$

The integrals in (2.28) are called incomplete beta integrals and have closed forms in only a few special cases.

Example 2.36

Suppose $X_1, \cdots, X_n$ are iid Exponential(θ), and we wish to test

$$\begin{aligned} H_0 &: \theta = \theta_0 \text{ or } \theta_1 \le \theta \le \theta_2 \\ H_1 &: \theta \ne \theta_0 \text{ and } (\theta < \theta_1 \text{ or } \theta > \theta_2). \end{aligned}$$

Such strange hypotheses are <u>not</u> easily handled in the frequentist setting, but pose no conceptual problems in the Bayesian setting.

$\boldsymbol{\Theta}_0 = \{\theta : \theta = \theta_0 \text{ or } \theta_1 \le \theta \le \theta_2\}$,

$\boldsymbol{\Theta}_1 = \{\theta : \theta \ne \theta_0 \text{ and } \theta < \theta_1 \text{ or } \theta > \theta_2\}$.

$\lambda(\theta|H_0)$ will consist of a point mass at $\theta = \theta_0$ and a continuous prior on $\theta_1 \le \theta \le \theta_2$. Thus,

$$\begin{aligned} & \int_{\boldsymbol{\Theta}_0} \lambda(\theta|H_0)\, d\theta \\ = \quad & P(\theta = \theta_0) + \int_{\theta_1}^{\theta_2} \lambda(\theta|H_0)\, d\theta = 1. \end{aligned}$$

Under H_1, $\lambda(\theta|H_1)$ can be continuous such that

$$\begin{aligned}\int_{\boldsymbol{\Theta}_1} \lambda(\theta|H_1)\,d\theta &= \int_0^{\theta_1} \lambda(\theta|H_1)\,d\theta + \int_{\theta_2}^{\infty} \lambda(\theta|H_1)\,d\theta \\ &= 1.\end{aligned}$$

Efficiency of Tests

The Power of Two Tests
The Power of the One-sample t-test:

Let $X_1, \cdots, X_n$ be i.i.d. (θ, σ^2). We wish to test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. The classical test of H_0 versus H_1 rejects H when $T_n \equiv \sqrt{n}(\bar{X} - \theta_0)/S > t_{n-1,\alpha}$.

(i) This test, asymptotically, has the correct level of significance (assuming $E[X^2] < \infty$ as we have) since

$$P_{\theta_0}(T_n > t_{n-1,\alpha}) \to P(N(0,1) > z_\alpha) = \alpha.$$

(ii) This test is consistent since, when a fixed $\theta > \theta_0$ is true

$$T_n = \frac{\sqrt{n}(\bar{X} - \theta)}{S} + \frac{\sqrt{n}(\theta - \theta_0)}{S}$$

$$\to_d N(0,1) + \infty,$$

and $t_{n-1,\alpha} \to z_\alpha$ so that $P_\theta(T_n > t_{n-1,\alpha}) \to 1$.

(iii) If $X_1, \cdots, X_n$ are i.i.d. $(\theta_n, \sigma^2) \equiv (\theta_0 + n^{-1/2}c_n, \sigma^2)$ where $c_n \to c$, then

$$T_n = \frac{\sqrt{n}(\bar{X}_n - \theta_n)}{S} + \frac{c_n}{S}$$

$$\to_d N(0,1) + \frac{c}{\sigma} = N\left(\frac{c}{\sigma}, 1\right).$$

Let $\beta_n^t(\theta)$ denote the power of the t-test based on $X_1, \cdots, X_n$ against the alternative θ. Then

$$\beta_n^t(\theta_n) = \beta_n^t(\theta_0 + n^{-1/2}c_n) \tag{2.29}$$
$$= P_{\theta_0 + c_n/\sqrt{n}}(T_n > t_{n-1,\alpha}) \to P(N(c/\sigma, 1) > z_\alpha).$$

The Power of the Sign-test:

Let $X_1, \cdots, X_n$ be i.i.d. with d.f. $F(x) = F_0(x - \theta)$, where F_0 has a unique median 0 (so that $F_0(0) = 1/2$). We wish to test $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Let $Y_i \equiv 1_{[\theta_0, \infty)}(X_i)$, $i = 1, \cdots, n$. The *sign test* of H_0 versus H_1 rejects H_0 when $S_n \equiv \sqrt{n}(\bar{Y} - 1/2)$ exceeds the upper α percentage point $s_{n,\alpha}$ of its distribution when θ_0 is true.

(i) When θ_0 is true, Y_i is Bernoulli$(1/2)$ so that $S_n \to_d N(0, 1/4)$. Since the exact distribution of $\sum_{i=1}^{n} Y_i$ is Binomial$(n, 1/2)$ for all df's F above, the test has exact level of significance α for all such F.

(ii) This test is consistent, when $\theta > \theta_0$

$$\begin{aligned} S_n &= \sqrt{n}(\bar{Y} - P_\theta(X > \theta_0)) + \sqrt{n}(P_\theta(X > \theta_0) - 1/2) \\ &\to_d N(0, p(1-p)) + \infty, \end{aligned}$$

with $p \equiv 1 - F_0(\theta_0 - \theta) > 1/2$ so that $P_\theta(S_n > s_{n,\alpha}) \to 1$.

(iii) If $X_1, \cdots, X_n$ are i.i.d. $F(x) = F_0(x - (\theta_0 + n^{-1/2} d_n))$ where $d_n \to d$ as $n \to \infty$ and where we now assume that F_0 has a strictly positive derivative f_0 at 0. Then, using $F_0(0) = 1/2$, we have

$$\begin{aligned} S_n &= \sqrt{n}(\bar{Y} - P_{\theta_0 + n^{-1/2} d_n}(X > \theta_0)) + \sqrt{n}(P_{\theta_0 + n^{-1/2} d_n}(X > \theta_0) - 1/2) \\ &= \frac{1}{\sqrt{n}} \left(\text{Binomial}(n, 1 - F_0(-d_n/\sqrt{n})) - n(1 - F_0(-d_n/\sqrt{n})) \right) \\ &\quad + \sqrt{n}(F_0(0) - F_0(-d_n/\sqrt{n})) \\ &\to_d N(0, 1/4) + d f_0(0) = N(d f_0(0), 1/4). \end{aligned}$$

Thus the power of the sign test $\beta_n^s(\theta)$ satisfies

$$\begin{aligned} \beta_n^s(\theta_0 + n^{-1/2} d_n) &\to P(N(d f_0(0), 1/4) > z_\alpha/2) \qquad (2.30) \\ &= P(N(2d f_0(0), 1) > z_\alpha). \end{aligned}$$

Problem 2.1

Use Liapunov's central limit theorem to show that

$$\frac{1}{\sqrt{n}}(\text{Binomial}(n, p_n) - np_n) \to_d N(0, p(1-p))$$

provided $p_n \to p$ as $n \to \infty$.

Pitman Efficiency

Definition 2.6

Pitman efficiency is defined to be the limiting ratio of the sample sizes that produces equal asymptotic power against the same sequence of alternatives.

Now equal asymptotic power β in (2.29) and (2.30) requires that

$$\frac{c}{\sigma} = 2d f_0(0). \tag{2.31}$$

If the t-test is based on N_t observations and the sign test based on N_s observations, then equal alternatives in (2.29) and (2.30) requires that

$$c_{N_t}/\sqrt{N_t} = d_{N_s}/\sqrt{N_s}. \tag{2.32}$$

Thus the Pitman efficiency $e_{s,t}$ of the sign test with respect to the t-test is just the limiting value of N_t/N_s subject to (2.31) and (2.32). Thus

$$\frac{N_t}{N_s} = \frac{c_{N_t}^2}{d_{N_s}^2} \to \left(\frac{c}{d}\right)^2 = 4\sigma^2 f_0^2(0) = e_{s,t}.$$

Problem 2.2

Evaluate $e_{s,t} = 4\sigma^2 f_0^2(0)$ in case:

(i) f_0 is Uniform$(-a, a)$.

(ii) f_0 is Normal$(0, a^2)$.

(iii) f_0 is Logistic$(0, a)$: $f_0(x) = a^{-1}e^{-x/a}/(1 + e^{-x/a})^2$ with variance$= \pi^2 a^2/3$.

(iv) f_0 is t with k degrees of freedom.

(v) f_0 is double-exponential(a); $f_0(x) = (2a)^{-1}\exp(-a|x|)$.

A General Calculation

We now consider the problem more generally. Suppose that $X_1, \cdots, X_N$ have a joint distribution P_θ where θ is a real-valued parameter. We wish to test $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$. Suppose that the T_1 test and the T_2 test are both consistent tests of H_0 versus H_1; and that the T_i test rejects H_0 if the statistic $T_{N,i}$ exceeds the upper α percent point of its distribution when $\theta = \theta_0$. Since both tests are consistent, it's useless to compare their limiting power against fixed alternatives; hence we will compare their power on a sequence of alternatives that approach θ_0 from above at the rate $1/\sqrt{N}$.

Suppose that for each $c > 0$ that the statistics $T_{N,i}$ satisfy

$$P_{\theta_0 + c_N/\sqrt{N}}(T_{N,i} \le x) \to P(N(c\mu_i, \sigma_i^2) \le \sigma_i x) \tag{2.33}$$

$$= P(N(c\mu_i/\sigma_i, 1) \le x)$$

for all x as $N \to \infty$ for any sequences of c_N's converging to c. Let the T_1-test (the T_2-test) use N_1 (use N_2) observations against the sequence of alternatives $c_{N_1}/\sqrt{N_1}$ (the sequence of alternatives $c_{N_2}/\sqrt{N_2}$), where

$c_{N_1} \to c_1$ (where $c_{N_2} \to c_2$). Equal asymptotic power requires

$$\frac{c_1 \mu_1}{\sigma_1} = \frac{c_2 \mu_2}{\sigma_2},$$

and equal alternative requires

$$\frac{c_{N_1}}{\sqrt{N_1}} = \frac{c_{N_2}}{\sqrt{N_2}};$$

solving these simultaneously leads to

$$\frac{N_2}{N_1} = \frac{c_{N_2}^2}{c_{N_1}^2} \to \frac{(\mu_1/\sigma_1)^2}{(\mu_2/\sigma_2)^2} = e_{1,2}. \tag{2.34}$$

Note that the efficiency $e_{1,2}$ is independent of the common level of significance α of the tests, of the particular value of the asymptotic power β, and of the particular sequences that converge to the values of c_1 and c_2 that are specified by the choice of β. Since so much is summarized in a single number, the procedure is bound to have some shortcomings; however it can be extremely useful and informative.

The quantity $\epsilon_i \equiv (\mu_i/\sigma_i)^2$ is called the *efficacy* of the T_i-test, and hence the efficiency $e_{1,2}$ is the ratio of the efficacies.

Problem 2.3

Define your idea of what the exact small sample efficiency $e_{s,t}(\alpha, \beta, n)$ of the sign test with respect to the t-test should be. Compute some values of it in case $X_1, \cdots, X_n$ are normal, and compare these values with the asymptotic value $e_{s,t} = 2/\pi = .6366\cdots$ that was obtained in problem 2.2.

Problem 2.4

Now redefine Pitman efficiency to be the ratio of the squared distances from the alternative to the hypothesized value θ_0 that produce equal asymptotic power as equal sample sizes approach infinity. Show that you get the same answer as before.

Now that if T_1 and T_2 are estimating the same thing (that is $\mu_1 = \mu_2$), then $e_{1,2}$ is just the ratio of the limiting variances.

Also note that the typical test of $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ is of the form: reject H_0 if

$$\frac{\sqrt{n}(T - E_{\theta_0}(T))}{\sqrt{\mathrm{Var}_{\theta_0}(\sqrt{n}T)}} > constant_\alpha \to z_\alpha .$$

Thus when $\theta_0 + c/\sqrt{n}$ is true, intuitively we have (letting $m(\theta) \equiv E_\theta(T)$ and $\sigma_0^2 \equiv \mathrm{Var}_{\theta_0}[\sqrt{n}T]$),

$$\begin{aligned}
\frac{\sqrt{n}(T - E_{\theta_0}(T))}{\sqrt{\mathrm{Var}_{\theta_0}(\sqrt{n}T)}} &= \frac{\sqrt{\mathrm{Var}_{\theta_0+c/\sqrt{n}}(T)}}{\sqrt{\mathrm{Var}_{\theta_0}(T)}} \frac{\sqrt{n}(T - E_{\theta_0+c/\sqrt{n}}(T))}{\sqrt{\mathrm{Var}_{\theta_0+c/\sqrt{n}}(\sqrt{n}T)}} \\
&\quad + \frac{\sqrt{n}(m(\theta_0 + c/\sqrt{n}) - m(\theta_0))}{\sqrt{\mathrm{Var}_{\theta_0}(\sqrt{n}T)}} \\
&\to_d \quad 1 \cdot N(0,1) + \frac{cm'(\theta_0)}{\sigma_0} = N(cm'(\theta_0)/\sigma_0, 1).
\end{aligned}$$

Thus we expect $(cm'(\theta_0)/\sigma_0)^2$ to be the efficacy.

Problem 2.5

Now consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ on the basis of a two-sided test based on either the T_1 or the T_2 statistic considered previously. Show that the same formula for Pitman efficiency is appropriate for the two-sided test also.

Problem 2.6

Again consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$; but suppose now that

$$T_{n,i} \to_d \chi^2_k(c^2\delta_i^2) \quad \text{as} \quad n \to \infty$$

under any sequence of alternatives $\theta_0 + c_n/\sqrt{n}$ having $c_n \to c > 0$ as $n \to \infty$. Here k is a fixed integer, and the limiting random variable has a noncentral chi-square distribution. Show that the Pitman efficiency criteria leads to $e_{1,2} = \delta_1^2/\delta_2^2$.

Some two-sample tests

The two-sample t-test

Let $X_1, \cdots, X_m$ be independent samples from F and $Y_1, \ldots, Y_n$ be independent sample from $G(y) = F(y - \theta)$ respectively. The classical test of $H_0 : \theta \le 0$ versus $H_1 : \theta > 0$ rejects H_0 if

$$T \equiv \sqrt{\frac{mn}{N}}(\bar{Y} - \bar{X}) \Big/ \left[\frac{m-1}{N-2}S_X^2 + \frac{n-1}{N-2}S_Y^2\right]^{1/2} > t_{m+n-2,\alpha}.$$

(This test can be shown to possess certain optimality properties when F is a normal distribution.) If F is any df having finite variance, then:

(i) When $\theta = 0$ we have $T \to_d N(0,1)$ provided $m \wedge n \to \infty$.

(ii) When $\theta > 0$ is true, then the test is consistent.

(iii) If $\lambda_N \equiv m/N \to \lambda \in (0,1)$ as $m \wedge n \to \infty$, then

$$P_{\theta + c_N/\sqrt{N}}(t > t_{m+n-2,\alpha}) \to P(N(c\sqrt{\lambda(1-\lambda)}/\sigma, 1) > z_\alpha).$$

Thus the efficacy of the two-sample t-test is $\epsilon_t = c^2\lambda(1-\lambda)/\sigma^2$.

Confidence Regions

The testing of UMP one-sided tests can be applied to the problem of obtaining a lower or upper bound for a real-valued parameter θ. This problem arises when θ is the toxicity of a drug or the probability of a undesirable event, for example. Here, we first consider the case of finding lower confidence bounds for θ. The case of finding upper confidence bounds is completely parallel, and thus will not be discussed. Let $\theta_L \equiv \theta_L(\boldsymbol{x})$ denote the lower confidence bound for θ. Since $\theta_L(\boldsymbol{x})$ will be a function of the observations, it cannot be required to fall below θ with certainty, but only with specified high probability. One selects a number $1-\alpha$, the confidence level, and restricts attention to bounds satisfying

$$P_\theta\left\{\theta_L(\boldsymbol{x}) \leq \theta\right\} \geq 1-\alpha \quad \text{for all} \quad \theta. \tag{2.35}$$

We call $\theta_L(\boldsymbol{x})$ a lower confidence bound for θ at confidence level $1-\alpha$. The infimum of the left-hand side of (2.35), which in practice will be equal to $1-\alpha$, is called the confidence coefficient of $\theta_L(\boldsymbol{x})$. Subject to (2.35), $\theta_L(\boldsymbol{x})$ should underestimate θ by as little as possible.

A function $\theta_L(\boldsymbol{x})$ for which

$$P_\theta\left\{\theta_L(\boldsymbol{x}) \leq \theta'\right\} = \text{minimum}$$

for all $\theta' < \theta$ subject to (2.35) is a uniformly most accurate lower confidence bound for θ at confidence level $1 - \alpha$.

A more general concept regarding the construction of confidence bounds is the notion of a confidence set. A family of subsets $S(\boldsymbol{x})$ of the parameter $\boldsymbol{\Theta}$ is said to constitute a family of confidence sets at confidence level $1 - \alpha$ if

$$P_\theta\left\{\theta \in S(\boldsymbol{x})\right\} \geq 1 - \alpha \quad \text{for all} \quad \theta \in \boldsymbol{\Theta},$$

that is, if the random set $S(\boldsymbol{x})$ covers the true parameter point with probability $\geq 1 - \alpha$. A lower confidence bound corresponds to the special case that $S(\boldsymbol{x})$ is one-sided interval

$$S(\boldsymbol{x}) = \left\{\theta : \theta_L(\boldsymbol{x}) \leq \theta < \infty\right\}.$$

Theorem 2.9

(i) For each $\theta_0 \in \mathbf{\Theta}$ let $A(\theta_0)$ be the acceptance region of a level α test for testing $H_0 : \theta = \theta_0$, and for each sample point X, let $S(\boldsymbol{x})$ denote the set of parameter values

$$S(\boldsymbol{x}) = \{\theta : X \in A(\theta), \theta \in \mathbf{\Theta}\}.$$

Then $S(\boldsymbol{x})$ is a family of confidence sets for θ at confidence level $1 - \alpha$.

(ii) If for all θ_0, $A(\theta_0)$ is UMP for testing H_0 at level α against the alternatives H_1, then for each θ_0 in $\mathbf{\Theta}_1$ (which does not cover the true θ), $S(\boldsymbol{x})$ minimizes the probability

$$P_\theta \{\theta_0 \in S(\boldsymbol{x})\} \quad \text{for all} \quad \theta_0 \in \mathbf{\Theta}_1$$

among all level $1 - \alpha$ families of confidence sets for θ.

Proof:

(i) By definition of $S(\boldsymbol{x})$,

$$\theta \in S(\boldsymbol{x}) \quad \text{iff} \quad X \in A(\theta), \tag{2.36}$$

and hence

$$P_\theta\left\{\theta \in S(\boldsymbol{x})\right\} = P_\theta\left\{X \in A(\theta)\right\} \geq 1 - \alpha.$$

(ii) If $S^*(\boldsymbol{x})$ is any other family of confidence sets at level $1-\alpha$, and if $A^*(\theta) = \{X : \theta \in S^*(\boldsymbol{x})\}$, then

$$P_\theta\left\{X \in A^*(\theta)\right\} = P_\theta\left\{\theta \in S^*(\boldsymbol{x})\right\} \geq 1 - \alpha,$$

so that $A^*(\theta_0)$ is the acceptance region of a level α test of H_0. It follows from the assumed property of $A(\theta_0)$ that for any $\theta \in \boldsymbol{\Theta}_1$

$$P_\theta\left\{X \in A^*(\theta_0)\right\} \geq P_\theta\left\{X \in A(\theta_0)\right\},$$

and hence that

$$P_\theta\left\{\theta_0 \in S^*(\boldsymbol{x})\right\} \geq P_\theta\left\{\theta_0 \in S(\boldsymbol{x})\right\},$$

as was to be proved.

The equivalence (2.36) shows that the structure of the confidence sets, $S(\boldsymbol{x})$, as the totality of parameter values θ for which the hypothesis H_0 is accepted when $\boldsymbol{x}$ is observed.

A confidence set can therefore be viewed as a combined statement regarding the tests of the various hypotheses H_0, which exhibits the values for which the hypothesis is accepted $\{\theta \in S(\boldsymbol{x})\}$ and those for which it is rejected $\{\theta \in S(\boldsymbol{x})^c\}$.

Theorem 2.10

Let the family of densities $p_\theta(x)$, $\theta \in \boldsymbol{\Theta}$, have the MLR property in $T(\boldsymbol{x})$, and suppose that the cdf $F_\theta(t)$ of $T = T(\boldsymbol{x})$ is a continuous function in each of the variables t and θ when the other is fixed.

(i) There exists a uniformly most accurate confidence bound $\theta_L(\boldsymbol{x})$ (or $\theta_U(\boldsymbol{x})$) for θ at each confidence level $1-\alpha$.

(ii) If $\boldsymbol{x} = (X_1, \cdots, X_n)$ and $t = T(\boldsymbol{x})$, and if the equation

$$F_\theta(t) = 1 - \alpha \tag{2.37}$$

has a solution $\theta = \hat{\theta}$ in $\boldsymbol{\Theta}$, then this solution is unique and $\theta_L(\boldsymbol{x}) = \hat{\theta}$.

Proof:

(i) There exists for each θ_0 a constant $C(\theta_0)$ such that

$$P_{\theta_0}\{T > C(\theta_0)\} = \alpha,$$

and by Theorem 2.2, $T > C(\theta_0)$ is a UMP level α rejection region for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$. By Corollary 2.1, the power of this test against any alternative $\theta_1 > \theta_0$ exceeds α, and hence $C(\theta_0) < C(\theta_1)$ so that the function C is strictly increasing; It is also continuous. Let $A(\theta_0)$ denote the acceptance region $T \leq C(\theta_0)$, and let $S(\boldsymbol{x})$ be defined by (2.36). It follows from the monotonicity of the function C that $S(\boldsymbol{x})$ consists of those values $\theta \in \boldsymbol{\Theta}$ which satisfy $\theta_L(\boldsymbol{x}) \leq \theta$, where $\theta_L(\boldsymbol{x}) = \inf\{\theta : T(\boldsymbol{x}) \leq C(\theta)\}$. By Theorem 2.9, the set $\{\theta : \theta_L(\boldsymbol{x}) \leq \theta\}$, restricted to possible values of the parameter, thus constitute a family of confidence sets at level $1 - \alpha$, which minimize $p_\theta\{\theta_L(\boldsymbol{x}) \leq \theta'\}$ for all $\theta > \theta'$. This shows $\theta_L(\boldsymbol{x})$ to be a uniformly most accurate confidence bound for θ.

(ii) It follows from Corollary 2.1 that $F_\theta(t)$ is a strictly decreasing function of θ at any point t for which $0 < F_\theta(t) < 1$, and hence that (2.37) can have at most one solution. Suppose now that t is the observed value of T and that the equation $F_\theta(t) = 1 - \alpha$ has the solution $\hat{\theta} \in \mathbf{\Theta}$. Then $F_{\hat{\theta}}(t) = 1 - \alpha$, and by definition of the function C, $C(\hat{\theta}) = t$. The inequality $t \leq C(\theta)$ is then equivalent to $C(\hat{\theta}) \leq C(\theta)$ and hence $\hat{\theta} \leq \theta$. It follows that $\theta_L(\boldsymbol{x}) = \hat{\theta}$, as was to be proved.

Under the same assumptions, the corresponding upper confidence bound with confidence coefficient $1 - \alpha$ is the solution $\theta_U(\boldsymbol{x})$ to the equation $P_\theta(T \geq t) = 1 - \alpha$ or equivalently $F_\theta(t) = \alpha$.

Finally, a confidence interval for θ at confidence level $1 - \alpha$ is defined as a set of random intervals with endpoints $(\theta_L(\boldsymbol{x}), \theta_U(\boldsymbol{x}))$ such that

$$P_\theta\{\theta_L(\boldsymbol{x}) \leq \theta \leq \theta_U(\boldsymbol{x})\} \geq 1 - \alpha \quad \text{for all} \quad \theta \in \mathbf{\Theta}.$$

Example 2.37

Suppose $X_1, \cdots, X_n$ are iid Exponential(θ). Let us construct an upper confidence bound for θ. The acceptance region of the most powerful test of $H_0 : \theta = \theta_0$ against $H_1 : \theta < \theta_0$ is $T(\boldsymbol{x}) \leq C$, where $T(\boldsymbol{x}) = \sum_{i=1}^{n} X_i$. Now $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta)$, and thus $2\theta \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$. Thus C is determined by $F_{\chi^2_{2n}}(C) = 1 - \alpha$, and therefore $C = F^{-1}_{\chi^2_{2n}}(1 - \alpha)$.
Therefore, the $1 - \alpha$ upper confidence bound for θ is given by

$$\theta \leq \frac{C}{2 \sum_{i=1}^{n} X_i}, \tag{2.38}$$

where $C = F^{-1}_{\chi^2_{2n}}(1 - \alpha)$. It follows from Theorem 2.9 that (2.38) is a uniformly most accurate upper confidence bound for θ.
Let us now construct a $1 - \alpha$ confidence interval for θ.

To find such an interval, we need to find numbers a, b such that

$$P\left(a < 2\theta \sum_{i=1}^{n} X_i < b\right) = 1 - \alpha.$$

Since $2\theta \sum_{i=1}^{n} X_i \sim \chi^2_{2n}$, it follows that a, b satisfy

$$\begin{aligned}
& P\left(a < \chi^2_{2n} < b\right) = 1 - \alpha \\
\Rightarrow \quad & F_{\chi^2_{2n}}(b) - F_{\chi^2_{2n}}(a) = 1 - \alpha \\
\Rightarrow \quad & a = \frac{\alpha}{2} \quad \text{percentile of} \quad \chi^2_{2n}, \\
& b = 1 - \frac{\alpha}{2} \quad \text{percentile of} \quad \chi^2_{2n}, \\
& a = F^{-1}_{\chi^2_{2n}}\left(\frac{\alpha}{2}\right), \quad b = F^{-1}_{\chi^2_{2n}}\left(1 - \frac{\alpha}{2}\right).
\end{aligned}$$

Thus the $(1 - \alpha) \times 100\%$ confidence interval for θ is

$$\frac{F^{-1}_{\chi^2_{2n}}\left(\frac{\alpha}{2}\right)}{2\sum_{i=1}^{n} X_i} < \theta < \frac{F^{-1}_{\chi^2_{2n}}\left(1 - \frac{\alpha}{2}\right)}{2\sum_{i=1}^{n} X_i}.$$

Remarks:

1. A function of the data and the parameters whose distribution does not depend on any parameters is called a pivotal quantity (or pivotal). We use pivotals to construct confidence bounds and confidence intervals.
2. When the statistic $T(\boldsymbol{x})$, which is used in the construction of a confidence bound (interval), has a continuous distribution with cdf $F_t(\theta)$, the quantity $F_T(T|\theta)$ is a pivotal since $F_T(T|\theta) \sim U(0,1)$, independent of θ. This is just a restatement of Theorem 2.10, (ii).

Example 2.38

Suppose $X_1, \cdots, X_n$ are iid $U(0,\theta)$. Let us find a $(1-\alpha)\times 100\%$ confidence interval for θ.

UMP tests for this family of distributions are based on $X_{(n)}$.

Let $T(\boldsymbol{x}) = X_{(n)}$.

$$F_{X_{(n)}}(t|\theta) = [F_X(t)]^n = \left(\frac{t}{\theta}\right)^n,$$

so it follows that

$$\left(\frac{T}{\theta}\right)^n \sim U(0,1), \quad T = X_{(n)}.$$

Thus

$$\begin{aligned}
& P\left(a < \left(\frac{T}{\theta}\right)^n < b\right) = 1-\alpha \\
\Rightarrow \quad & P\left(a < Y < b\right) = 1-\alpha, \quad Y \sim U(0,1), \\
\Rightarrow \quad & F_Y(b) - F_Y(a) = 1-\alpha, \\
& \text{so that } \; b = 1-\frac{\alpha}{2} \; \text{ and } \; a = \frac{\alpha}{2}.
\end{aligned}$$

Thus our confidence interval is

$$\frac{\alpha}{2} < \left(\frac{X_{(n)}}{\theta}\right)^n < 1-\frac{\alpha}{2}.$$

Solving for θ, we get

$$\frac{X_{(n)}}{\left(1-\frac{\alpha}{2}\right)^{1/n}} < \theta < \frac{X_{(n)}}{\left(\frac{\alpha}{2}\right)^{1/n}}.$$

Confidence Bounds and Intervals for Normal Population Parameters

Example 2.39

Suppose $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$, where (μ, σ^2) are both unknown.

(i) We wish to construct a $(1-\alpha) \times 100\%$ confidence interval for σ^2.

Let $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. We know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, so that $\frac{(n-1)S^2}{\sigma^2}$ is a pivotal. Thus

$$P\left(a < \frac{(n-1)S^2}{\sigma^2} < b\right) = 1-\alpha,$$

$$a = \frac{\alpha}{2} \quad \text{percentile of} \quad \chi^2_{n-1} = F^{-1}_{\chi^2_{n-1}}\left(\frac{\alpha}{2}\right),$$
$$b = 1 - \frac{\alpha}{2} \quad \text{percentile of} \quad \chi^2_{n-1} = F^{-1}_{\chi^2_{n-1}}\left(1 - \frac{\alpha}{2}\right).$$

Thus the confidence interval takes the form

$$\chi^2_{n-1,\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}},$$

and therefore,

$$\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}.$$

A $(1-\alpha)\times 100\%$ confidence interval for σ is

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}}.$$

(ii) Now suppose we wish to construct a $(1-\alpha)\times 100\%$ confidence interval for μ.

Now $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is a pivotal since $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$. So a $(1-\alpha)\times 100\%$

confidence interval for μ is specified as

$$P\left(a < \frac{\bar{X}-\mu}{S/\sqrt{n}} < b\right) = 1-\alpha,$$

$$a = t_{n-1,\frac{\alpha}{2}}, \quad b = t_{n-1,1-\frac{\alpha}{2}},$$

so that the interval becomes

$$t_{n-1,\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{n-1,1-\frac{\alpha}{2}}$$

$$\Rightarrow \bar{X} - \frac{S}{\sqrt{n}} t_{n-1,\frac{\alpha}{2}} < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{n-1,1-\frac{\alpha}{2}},$$

since the t distribution is symmetric about 0, $t_{n-1,\frac{\alpha}{2}} = -t_{n-1,1-\frac{\alpha}{2}}$.

(iii) Finally suppose we wanted to construct a joint confidence region for (μ, σ^2).

We know that

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{and} \quad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

are pivotals. Moreover, these pivotals are independent. So that

$$P\left\{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 < z_{1-\delta}^2, \quad \chi_\epsilon^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\epsilon}^2\right\}$$

$$\begin{aligned} &= P\left\{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 < z_{1-\delta}^2\right\} P\left\{\chi_\epsilon^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\epsilon}^2\right\} \\ &= (1-2\delta)(1-2\epsilon). \end{aligned}$$

Here

$$\begin{aligned} z_{1-\delta}^2 &= [1-\delta \quad \text{percentile of} \quad N(0,1)]^2 \\ \chi_\epsilon^2 &= \epsilon \quad \text{percentile of} \quad \chi_{n-1}^2. \end{aligned}$$

Thus if we want a $(1-\alpha)\times 100\%$ confidence region, we need to choose (δ, ϵ) so that

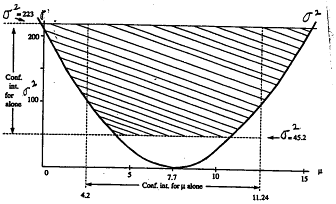
$$(1-2\delta)(1-2\epsilon) = 1-\alpha.$$

For example, if we wanted a 90% confidence region for (μ, σ^2), if we pick $\delta = \epsilon = .025$, we get $(1-2\delta)(1-2\epsilon) = .90$. Further suppose

$n = 14, \bar{X} = 7.714$, $S^2 = 85.9$. Then a 90% confidence region for (μ, σ^2) is given by

$$\left\{(\mu, \sigma^2) : \frac{(\mu - 7.714)^2}{\sigma^2/14} < (1.96)^2 \quad \text{and} \quad 5.01 < \frac{13(85.9)}{\sigma^2} < 24.7\right\}$$

This region is defined by the lines between $\sigma^2 = 45.2$ and $\sigma^2 = 222.9$, and inside the parabola $(\mu - 7.714)^2 < .2744\sigma^2$.



$\sigma^2 = 223$
σ^2
200
100
Conf. int. for σ^2 alone
$\sigma^2 = 45.2$
0
5
7.7
10
15
μ
Conf. int. for μ alone
4.2
11.24

So far, we have given only a formal development for constructing confidence intervals (bounds) for a single parameter θ. We have not given a formal development of constructing confidence intervals in the presence of nuisance parameters, although in Example 2.39, we constructed a confidence interval for μ with σ^2 treated as nuisance and vice-versa. We now want to give a more general treatment of construction of confidence intervals (sets, bounds) in the presence of nuisance parameters. When nuisance parameters $\boldsymbol{\xi}$ are present, the defining condition for a lower confidence bound $\theta_L(\boldsymbol{x})$ becomes

$$P_{\theta,\boldsymbol{\xi}}\left(\theta_L(\boldsymbol{x}) \leq \theta\right) \geq 1-\alpha \quad \text{for all} \quad (\theta, \boldsymbol{\xi}). \tag{2.39}$$

Similarly, confidence intervals for θ at confidence level $1-\alpha$ are defined as a set of random intervals with endpoints $(\theta_L(\boldsymbol{x}), \theta_U(\boldsymbol{x}))$ such that

$$P_{\theta,\boldsymbol{\xi}}\left(\theta_L(\boldsymbol{x}) \leq \theta \leq \theta_U(\boldsymbol{x})\right) \geq 1-\alpha \quad \text{for all} \quad (\theta, \boldsymbol{\xi}). \tag{2.40}$$

The infimum over $(\theta, \boldsymbol{\xi})$ of the left-hand side of (2.39) and (2.40) is the confidence coefficient associated with these statements. In the presence of nuisance parameters, we can still use the ideas established earlier in which we constructed confidence bounds (intervals) by finding the corresponding acceptance region of the UMP test. If we are testing $H_0 : \theta = \theta_0$ and $S(\boldsymbol{x}) = \{\theta : X \in A(\theta_0)\}$ then

$$\theta \in S(\boldsymbol{x}) \quad \text{iff} \quad X \in A(\theta),$$

and hence

$$P_{\theta,\boldsymbol{\xi}}\{\theta \in S(\boldsymbol{x})\} \geq 1 - \alpha \quad \text{for all} \quad (\theta, \boldsymbol{\xi}).$$

Example 2.40

Suppose $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$. The confidence interval for μ can be obtained from the acceptance region of the hypothesis $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, which is given by

$$\frac{|\sqrt{n}(\bar{X} - \mu_0)|}{S} \leq C, \tag{2.41}$$

where $C = t_{n-1,1-\frac{\alpha}{2}}$. Letting $\mu = \mu_0$ in (2.41), and solving for μ leads to

$$\bar{X} - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}.$$

(see TSH, page 163 for more discussion of this example.)

Unbiased Confidence Sets

Confidence sets can be viewed as a family of tests of the hypotheses $\theta \in H_0(\theta')$ against alternative $\theta \in H_1(\theta')$ for varying θ'. A confidence level of $1 - \alpha$ then simply expresses the fact that all the tests are to be at level α, and the condition therefore becomes

$$P_{\theta,\boldsymbol{\xi}}\{\theta' \in S(\boldsymbol{x})\} \geq 1 - \alpha \quad \text{for all} \quad \theta \in H_0(\theta') \quad \text{and all} \quad \boldsymbol{\xi}. \tag{2.42}$$

In the case that $H_0(\theta')$ is the hypothesis $\theta = \theta'$ and $S(\boldsymbol{x})$ is the interval $(\theta_L(\boldsymbol{x}), \theta_U(\boldsymbol{x}))$, this agrees with (2.40). In the one-sided case in which $H_0(\theta')$ is the hypothesis $\theta \leq \theta'$ and $S(\boldsymbol{x}) = \{\theta : \theta_L(\boldsymbol{x}) \leq \theta\}$, the condition reduces to $P_{\theta,\boldsymbol{\xi}}\{\theta_L(\boldsymbol{x}) \leq \theta'\} \geq 1 - \alpha$ for all $\theta' \geq \theta$, and this is seen to be equivalent to

$$P_{\theta,\boldsymbol{\xi}}(\theta_L(\boldsymbol{x}) \leq \theta) \geq 1 - \alpha \quad \text{for all} \quad (\theta, \boldsymbol{\xi}).$$

With this interpretation of confidence sets, the probabilities

$$P_{\theta,\boldsymbol{\xi}}\{\theta' \in S(\boldsymbol{x})\}, \quad \theta \in H_1(\theta'), \tag{2.43}$$

are the probabilities of false acceptance of $H_0(\theta')$ (Type II error). The smaller these probabilities are, the more desirable are the tests.

From the point of view of estimation, (2.43) is the probability of covering the wrong value θ'. Thus, (2.43) provides a measure of accuracy of the confidence sets.

In the presence of nuisance parameters, UMP tests usually do not exist and this implies nonexistence of confidence sets that are uniformly most accurate in the sense of minimizing

$$P_{\theta,\boldsymbol{\xi}}\{\theta' \in S(\boldsymbol{x})\}, \quad \theta \in H_1(\theta') \quad \text{for all} \quad \boldsymbol{\xi}.$$

This suggests restricting attention to confidence sets which in a suitable sense are unbiased. In analogy with the corresponding definition for tests, a family of confidence sets at confidence level $1-\alpha$ is said to be unbiased if

$$P_{\theta,\boldsymbol{\xi}}\{\theta' \in S(\boldsymbol{x})\} \leq 1-\alpha \tag{2.44}$$

for all θ' such that $\theta \in H_1(\theta')$ and for all $(\boldsymbol{\xi}, \theta)$, so that the probability of covering these false values does not exceed the confidence level.

In the two and one-sided cases, the condition (2.44) reduces to

$$P_{\theta,\boldsymbol{\xi}}(\theta_L(\boldsymbol{x}) \leq \theta' \leq \theta_U(\boldsymbol{x})) \leq 1-\alpha \quad \text{for all} \quad \theta' \neq \theta \quad \text{and all} \quad \boldsymbol{\xi},$$

and

$$P_{\theta,\boldsymbol{\xi}}\{\theta_L(\boldsymbol{x}) \leq \theta'\} \leq 1-\alpha \quad \text{for all} \quad \theta' < \theta \quad \text{and all} \quad \boldsymbol{\xi}.$$

With this definition of unbiasedness, unbiased families of tests lead to unbiased confidence sets and conversely. A family of confidence sets is uniformly most accurate unbiased at confidence level $1-\alpha$ if it minimizes the probabilities $P_{\theta,\boldsymbol{\xi}}\{\theta' \in S(\boldsymbol{x})\}$ for all θ' such that $\theta \in H_1(\theta')$ and for all $(\boldsymbol{\xi},\theta)$, subject to (2.42) and (2.44).

Example 2.41

Suppose $X_1,\cdots,X_n$ are iid $N(\mu,\sigma^2)$. The UMPU test of the hypothesis $H_0:\sigma^2=\sigma_0^2$ against $H_1:\sigma^2\neq\sigma_0^2$ is given by the acceptance region

$$C_1 \leq \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma_0^2} \leq C_2.$$

The most accurate unbiased confidence intervals for σ^2 are therefore

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{C_2} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{C_1},$$

where

$$C_2 = \chi^2_{n-1,1-\frac{\alpha}{2}}, \quad C_1 = \chi^2_{n-1,\frac{\alpha}{2}}.$$

Example 2.42

Suppose we have two independent normal populations $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$, where the X_j are iid $N(\mu_1, \sigma^2)$ and the Y_i are iid $N(\mu_2, \sigma^2)$. The UMPU acceptance region for the hypothesis $H_0 : \mu_1 = \mu_2$ versus

$H_1 : \mu_1 \neq \mu_2$ is

$$\frac{|\bar{X}-\bar{Y}-(\mu_1-\mu_2)|\Big/\sqrt{\frac{1}{m}+\frac{1}{n}}}{\sqrt{\frac{\sum_{j=1}^{m}(X_j-\bar{X})^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{m+n-2}}} \leq C,$$

where $C = t_{m+n-2,1-\frac{\alpha}{2}}$. Thus, the most accurate unbiased $(1-\alpha)\times 100\%$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X}-\bar{Y}-CS \leq \mu_1-\mu_2 \leq \bar{X}-\bar{Y}+CS,$$

where

$$\begin{aligned} S^2 &= \left(\frac{1}{m}+\frac{1}{n}\right)\left(\frac{\sum_{j=1}^{m}(X_j-\bar{X})^2+\sum_{i=1}^{n}(Y_i-\bar{Y})^2}{m+n-2}\right), \\ C &= t_{m+n-2,1-\frac{\alpha}{2}}. \end{aligned}$$

Example 2.43

Suppose $X_1, \cdots, X_m$ are iid $N(\mu_1, \sigma_1^2)$ and $Y_1, \cdots, Y_n$ are iid $N(\mu_2, \sigma_2^2)$, and we wish to construct a most accurate unbiased confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$.
The acceptance region for the UMPU test of $H_0 : \sigma_1^2 = \sigma_2^2$ versus $H_1 : \sigma_1^2 \neq \sigma_2^2$ is given by (level α)

$$C_1 \frac{\sum_{j=1}^{m}(X_j - \bar{X})^2/(m-1)}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq C_2 \frac{\sum_{j=1}^{m}(X_j - \bar{X})^2/(m-1)}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)},$$

where

$$\begin{aligned} C_2 &= F(n-1, m-1, 1-\frac{\alpha}{2}), \\ C_1 &= F(n-1, m-1, \frac{\alpha}{2}). \end{aligned}$$

Note that

$$\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/\sigma_2^2(n-1)}{\sum_{j=1}^{m}(X_j - \bar{X})^2/\sigma_1^2(m-1)} \sim F(n-1, m-1),$$

so that $\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/\sigma_2^2(n-1)}{\sum_{j=1}^{m}(X_j - \bar{X})^2/\sigma_1^2(m-1)}$ is a pivotal quantity. And the $(1-\alpha)\times 100\%$ confidence interval is

$$a \leq \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/\sigma_2^2(n-1)}{\sum_{j=1}^{m}(X_j - \bar{X})^2/\sigma_1^2(m-1)} \leq b$$

$$\Rightarrow a\,\frac{\sum_{j=1}^{m}(X_j - \bar{X})^2/(m-1)}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq b\,\frac{\sum_{j=1}^{m}(X_j - \bar{X})^2/(m-1)}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2/(n-1)}.$$

Example 2.44

Suppose $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Poisson}(\mu)$ and we wish to construct a $(1-\alpha) \times 100\%$ confidence interval for $\rho = \frac{\mu}{\lambda}$. Recall that $Y|X+Y=t \sim \text{Binomial}(p,t)$, where $p = \frac{\rho}{1+\rho}$.

The UMPU test $\phi(y,t)$ of the hypothesis $H_0 : \rho = \rho_0$ is defined for each t as the UMPU conditional test of the hypothesis $H_0 : p = \frac{\rho_0}{1+\rho_0}$. If

$$p_L(t) \leq p \leq p_U(t)$$

is the most accurate unbiased confidence interval for p given t, it follows that the most accurate unbiased confidence interval for $\rho = \frac{\mu}{\lambda}$ is

$$\frac{p_L(t)}{1-p_L(t)} \leq \frac{\mu}{\lambda} \leq \frac{p_U(t)}{1-p_U(t)}.$$

Example 2.45 (Confidence Intervals in Linear Regression)

Consider the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

where the ϵ_i are iid $N(0, \sigma^2)$, $i = 1, \cdots, n$.
Set $v_i = \frac{x_i - \bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$ and $\gamma + \delta v_i = \beta_0 + \beta_1 x_i$,

so that $\sum_{i=1}^{n} v_i = 0$ and $\sum_{i=1}^{n} v_i^2 = 1$.

Thus

$$\beta_0 = \gamma - \delta \frac{\bar{x}}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

and

$$\beta_1 = \frac{\delta}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}.$$

Now the joint density of $\boldsymbol{y} = (Y_1, \cdots, Y_n)$ is

$$p(\boldsymbol{y}|\gamma, \delta) = (2\pi)^{-n/2}\sigma^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\gamma-\delta v_i)^2}.$$

$p(\boldsymbol{y}|\gamma, \delta)$ is in the exponential family with

$$u = \sum_{i=1}^{n} v_i y_i, \quad T_1 = \sum_{i=1}^{n} y_i^2, \quad T_2 = \sum_{i=1}^{n} y_i,$$

$$\theta = \frac{\delta}{\sigma^2}, \quad \xi_1 = -\frac{1}{2\sigma^2}, \quad \xi_2 = \frac{\gamma}{\sigma^2}.$$

This representation implies the existence of a UMPU test of the hypothesis $H_0 : a\gamma + b\delta = c$, where a, b, c are given constants. To obtain the most accurate unbiased confidence interval for $\rho = a\gamma + b\delta$, we find the acceptance region of the UMPU test for $H_0 : \rho = \rho_0$. This acceptance region is given by

$$\frac{\left| b \sum_{i=1}^{n} v_i y_i + a\bar{y} - \rho_0 \right| \Big/ \sqrt{a^2/n + b^2}}{\sqrt{\left[\sum_{i=1}^{n} (y_i - \bar{y})^2 - \left(\sum_{i=1}^{n} v_i y_i \right)^2 \right] \Big/ (n-2)}} \leq C,$$

where $C = t_{n-2,1-\frac{\alpha}{2}}$. Thus a $(1-\alpha) \times 100\%$ confidence interval for $\rho = a\gamma + b\delta$ is given by

$$a\bar{y} + b\sum_{i=1}^{n} v_i y_i - C\,S \leq a\gamma + b\delta \leq a\bar{y} + b\sum_{i=1}^{n} v_i y_i + C\,S,$$

where

$$\begin{aligned} S^2 &= \frac{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2 - \left(\sum\limits_{i=1}^{n} v_i y_i\right)^2}{n-2}, \\ C &= t_{n-2,1-\frac{\alpha}{2}}. \end{aligned}$$

A confidence interval for the slope of the regression line is obtained by setting $a = 0$ and $b = \frac{1}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}}$.

It might also be of interest to construct a confidence interval for the regression line at a point x_0, that is, for $y_0 \equiv \beta_0 + \beta_1 x_0$. Since

$$\beta_0 + \beta_1 x_0 = \gamma + \frac{\delta(x_0 - \bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}},$$

the constants a and b are $a = 1$, $b = \frac{(x_0 - \bar{x})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$.

See TSH, pages 168 - 171 for more details on this example.

Example 2.46

Consider the linear model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$

where $\boldsymbol{\beta}$ is $p \times 1$ and $\boldsymbol{X}$ is $n \times p$ of rank p. A $(1 - \alpha) \times 100\%$ confidence region for $\boldsymbol{\beta}$ is

$$C(\boldsymbol{\beta}) = \left\{ \boldsymbol{\beta} : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq p\,\hat{\sigma}^2 F(p, n-p, 1-\alpha) \right\},$$

where

$$\begin{aligned}\hat{\sigma}^2 &= MSE = \frac{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{M})\boldsymbol{y}}{n-p},\\ \hat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},\\ p &= \operatorname{rank}(\boldsymbol{X}).\end{aligned}$$

$C(\boldsymbol{\beta})$ is a p dimensional ellipsoid with center $\hat{\boldsymbol{\beta}}$.

Note here that

$$\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{p\,\hat{\sigma}^2} \sim F(p, n-p),$$

so that $\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{p\,\hat{\sigma}^2}$ is a pivotal.

Now suppose $\boldsymbol{A}$ is a $q \times p$ matrix of constants of rank q. Then

$$\frac{(\boldsymbol{A}\boldsymbol{\beta}-\boldsymbol{A}\hat{\boldsymbol{\beta}})'(\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}')^{-1}(\boldsymbol{A}\boldsymbol{\beta}-\boldsymbol{A}\hat{\boldsymbol{\beta}})}{q\,\hat{\sigma}^2} \sim F(q, n-q),$$

so that this quantity is a pivotal.

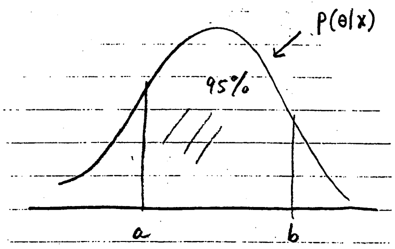
Thus a $(1-\alpha)\times 100\%$ confidence region for $\boldsymbol{A\beta}$ is

$$C(\boldsymbol{\beta}) = \left\{\boldsymbol{\beta} : (\boldsymbol{A\beta} - \boldsymbol{A}\hat{\boldsymbol{\beta}})'(\boldsymbol{A}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}')^{-1}(\boldsymbol{A\beta} - \boldsymbol{A}\hat{\boldsymbol{\beta}}) \leq q\,\hat{\sigma}^2 F(q, n-q, 1-\alpha)\right\}.$$

Highest Posterior Density Regions

The Bayesian "confidence interval" is called a highest posterior density (HPD) region (or a credible set). HPD regions and credible sets are *not* the same thing in general.

Suppose θ is a univariate parameter. A 95% HPD interval for θ is the interval such that 95% of the highest area of the posterior density is contained in this interval.


p(θ|x)
95%
a
b

A 95% HPD interval for θ is $\{\theta : a \le \theta \le b\}$.

If θ is multidimensional, they are called **HPD regions**. A formal definition is given as follows:

Definition 2.7

Let $p(\theta|\boldsymbol{x})$ denote the posterior density of θ. A region R in the parameter space $\boldsymbol{\Theta}$ is called an HPD region of content $1-\alpha$ if

(a) $P(\theta \in R|\boldsymbol{x}) = 1-\alpha$.

(b) For $\theta_1 \in R$ and $\theta_2 \notin R$, $p(\theta_1|\boldsymbol{x}) \ge p(\theta_2|\boldsymbol{x})$.

Definition 2.8

A credible set is a posterior region that is constructed by removing the upper $\frac{\alpha}{2}$ and lower $\frac{\alpha}{2}$ percentile of the posterior distribution. Such a set is called a $(1-\alpha)\times 100\%$ credible set. HPD region = credible set when the posterior distribution is symmetric (such as normal or t).

When the posterior distribution is skewed, it may be hard to construct HPD regions. Credible sets are easier to construct in this case. When the posterior distribution is symmetric, HPD regions are easy to construct. In many situations, credible set $\approx$ HPD region, especially for large n.

Some Properties of HPD Regions

1. It follows from the definition that for a given probability content $1-\alpha$, the HPD region has the smallest possible volume in the parameter space θ.
2. If we make the assumption that $p(\theta|\boldsymbol{x})$ is non-uniform over every region in the space of θ, then the HPD region of content $1-\alpha$ is unique. Further if θ_1 and θ_2 are two points such that $p(\theta_1|\boldsymbol{x}) = p(\theta_2|\boldsymbol{x})$, then these two points are simultaneously included (or excluded) by a $1-\alpha$ HPD region. The converse is also true.

That is, if $p(\theta_1|\boldsymbol{x}) \neq p(\theta_2|\boldsymbol{x})$, then there exists a $1-\alpha$ HPD region which includes one point but not the other.

Example 2.47

Suppose $X_1, \cdots, X_n$ are i.i.d. $N(\theta, 1)$, and take $\lambda(\theta) \propto 1$. Let $\boldsymbol{x} = (X_1, \cdots, X_n)$. We have established that

$$\theta | \boldsymbol{x} \sim N(\bar{\boldsymbol{x}}, \frac{1}{n}), \quad z | \boldsymbol{x} \sim N\left(\bar{\boldsymbol{x}}, 1 + \frac{1}{n}\right).$$

Note that the posterior mean of θ is the frequentist point estimate of θ, which is $\bar{\boldsymbol{x}}$. The posterior variance of θ is the frequentist variance of $\bar{\boldsymbol{x}}$, i.e., $\mathrm{Var}(\bar{\boldsymbol{x}}|\theta) = \frac{1}{n}$. Also, the predictive mean of z is the frequentist point estimate of a future value, and the predictive variance of z is the frequentist estimate. That is,

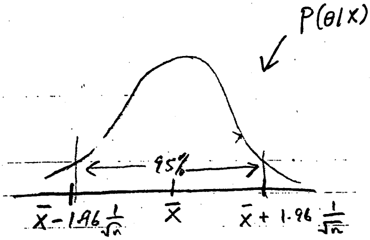
$$\mathrm{Var}(z - \bar{\boldsymbol{x}}|\theta) = 1 + \frac{1}{n}.$$

Note: The Bayesian predictive distribution of a future value z given the data $\boldsymbol{x}$ is

$$p(z|\boldsymbol{x}) = \int p(z|\theta) p(\theta|\boldsymbol{x})\, d\theta.$$

z can also be vector values. If $\boldsymbol{z}$ is $q \times 1$, then $p(\boldsymbol{z}|\boldsymbol{x})$ is a q dimensional predictive distribution. $p(z|\theta)$ above is the density of $z|\theta$. If $X_1, \cdots, X_n$ are iid $N(\theta, 1)$, then $Z|\theta \sim N(\theta, 1)$.

Now let us construct a 95% HPD interval for θ. We know that $\theta|\boldsymbol{x} \sim N(\bar{\boldsymbol{x}}, \frac{1}{n})$.


P(θ|X)
95%
$\bar{X} - 1.96 \frac{1}{\sqrt{n}}$
$\bar{X}$
$\bar{X} + 1.96 \frac{1}{\sqrt{n}}$

We have $P(\bar{\boldsymbol{x}} - 1.96\frac{1}{\sqrt{n}} \leq \theta \leq \bar{\boldsymbol{x}} + 1.96\frac{1}{\sqrt{n}} | \boldsymbol{x}) = .95$. Thus a 95% HPD interval for θ is

$$\left\{\theta : \bar{\boldsymbol{x}} - \frac{1.96}{\sqrt{n}} \leq \theta \leq \bar{\boldsymbol{x}} + \frac{1.96}{\sqrt{n}}\right\}.$$

Notice that this is the same interval as a frequentist confidence interval for θ, but is has a much different interpretation.

Suppose we want to construct a 95% **Highest Predictive Density** interval for z. This interval is given by

$$\left\{z : \bar{\boldsymbol{x}} - 1.96\sqrt{1 + \frac{1}{n}} \leq z \leq \bar{\boldsymbol{x}} + 1.96\sqrt{1 + \frac{1}{n}}\right\}.$$

This is the same interval as the frequentist predictive interval for z.

Example 2.48

Suppose $X_1, \cdots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, where (μ, σ^2) are both unknown. Let $\tau = \frac{1}{\sigma^2}$. Suppose we specify the joint prior

$$\lambda(\mu, \tau) \propto \tau^{-1}.$$

Now

$$p(\mu, \tau|\boldsymbol{x}) \ \propto \tau^{\frac{n}{2}-1} \ \exp\left\{-\frac{\tau}{2}\Sigma(x_i-\mu)^2\right\} \ .$$

Thus

$$p(\mu|\boldsymbol{x}) \propto \int_0^\infty \tau^{\frac{n}{2}-1} \ \exp\left\{-\frac{\tau}{2}\Sigma(x_i-\mu)^2\right\} \ d\tau$$

$$= \int_0^\infty \tau^{\frac{n}{2}-1} \ \exp\left\{-\frac{\tau}{2}\left[\Sigma x_i^2 - 2\mu\Sigma x_i + n\mu^2\right]\right\} \ d\tau$$

$$= \int_0^\infty \tau^{\frac{n}{2}-1} \exp\left\{\frac{n\tau}{2}\bar{x}^2\right\} \exp\left\{-\frac{n\tau}{2}\left[\frac{1}{n}\Sigma x_i^2\right]\right\} \exp\left\{-\frac{n\tau}{2}\left[\mu-\bar{x}\right]^2\right\} \ d\tau$$

$$= \int_0^\infty \tau^{\frac{n}{2}-1} \ \exp\left\{-\frac{\tau}{2}\left[\Sigma x_i^2 - n\bar{x}^2\right]\right\} \ \exp\left\{-\frac{n\tau}{2}\left[\mu-\bar{x}\right]^2\right\} \ d\tau$$

$$= \int_0^\infty \tau^{\frac{n}{2}-1} \ \exp\left\{-\frac{\tau}{2}\left[\Sigma(x_i-\bar{x})^2\right]\right\} \ \exp\left\{-\frac{n\tau}{2}\left[\mu-\bar{x}\right]^2\right\} \ d\tau.$$

Let $s^2 \ = \ \frac{1}{n-1}\Sigma(x_i-\bar{x})^2$. Now, we get

$$
\begin{aligned}
&= \int_0^{\infty} \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\left[(n-1)s^2 + n(\mu-\bar{x})^2\right]\right\} d\tau \\
&\propto \left[(n-1)s^2 + n(\mu-\bar{x})^2\right]^{-\frac{n}{2}} \\
&\propto \left[1 + \frac{n}{(n-1)s^2}(\mu-\bar{x})^2\right]^{-\frac{(n-1+1)}{2}} .
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mu|\boldsymbol{x} &\sim S_1(n-1, \bar{x}, \tfrac{s^2}{n}) \\
&= t(n-1, \bar{x}, \tfrac{s^2}{n})
\end{aligned}
$$

$$
\text{Hence} \qquad \left(\frac{\mu-\bar{x}}{\frac{s}{\sqrt{n}}}\middle|\boldsymbol{x}\right) \sim t_{n-1} .
$$

Therefore, a 95% HPD interval for μ is

$$
\left\{\mu :\ \bar{x} - t_{(n-1,.975)}\frac{s}{\sqrt{n}} \ \le\ \mu\ \le\ \bar{x} + t_{(n-1,.975)}\frac{s}{\sqrt{n}}\right\},
$$

where $t_{(n-1,.975)}$ corresponds to the 97.5^{th} percentile of the t distribution with $n-1$ degrees of freedom. This HPD interval corresponds to the

frequentist confidence interval for μ when σ^2 is unknown.
Let us construct a credible interval for $\tau = \frac{1}{\sigma^2}$.

$$\begin{aligned} p(\tau|\boldsymbol{x}) &\propto \int_{-\infty}^{\infty} \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\left[(n-1)s^2 + n(\mu-\bar{x})^2\right]\right\} d\mu \\ &= \tau^{\frac{n}{2}-1} \exp\left\{-\frac{\tau}{2}\left[(n-1)s^2\right]\right\} \int_0^{\infty} \exp\left\{-\frac{n\tau}{2}(\mu-\bar{x})^2\right\} d\mu \\ &\propto \tau^{\frac{n-1}{2}-1} \exp\left\{-\frac{\tau}{2}\left[(n-1)s^2\right]\right\}. \end{aligned}$$

Thus $\tau|\boldsymbol{x} \sim \text{Gamma}\left(\frac{n-1}{2}, \frac{(n-1)s^2}{2}\right)$.
Note that the above posterior distribution for τ implies

$$\boxed{(n-1)s^2\,\tau \;\sim\; \chi^2_{n-1}}.$$

Thus a 95% credible interval for τ is given by

$$\chi^2_{(n-1,.025)} \;\leq\; (n-1)s^2\tau \;\leq \chi^2_{(n-1,.975)},$$

and therefore,

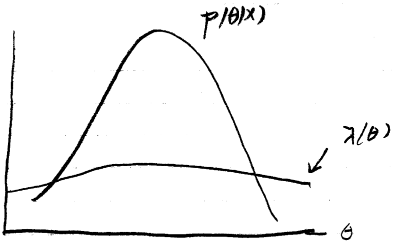
$$\left\{\tau : \frac{\chi^2_{(n-1,.025)}}{(n-1)s^2} \;\leq\; \tau \;\leq\; \frac{\chi^2_{(n-1,.975)}}{(n-1)s^2}\right\}.$$

Remarks on Prior Elicitation

There are essentially two classifications of prior distributions: noninformative priors and informative priors.

Noninformative Priors:

A prior is said to be noninformative if the likelihood dominates the prior, so that the likelihood (data) drives the inference. A noninformative prior can be proper or improper. An improper prior is necessarily noninformative.



Example of noninformative prior

Common choices of noninformative priors for location parameters are

(i) Uniform prior. That is, $\lambda(\boldsymbol{\theta}) \propto 1$.

(ii) Normal prior with large variance, for example, $\theta \sim N(\mu_0, \sigma_0^2)$, with σ_0^2 large.

(iii) Jeffreys prior:

$$\lambda_J(\boldsymbol{\theta}) \propto |I(\boldsymbol{\theta})|^{1/2},$$

where $I(\boldsymbol{\theta})$ = Fisher information for $\boldsymbol{\theta}$; where $\boldsymbol{\theta}$ may be vector-valued. If $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)$,

$$I(\boldsymbol{\theta}) = \left(-E\left[\frac{\partial^2 \log p(x|\boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right]\right)_{p\times p}.$$

Uniform priors are especially attractive as noninformative priors since

(i) they yield inferences that are very similar to maximum likelihood. For example, if $\lambda(\boldsymbol{\theta}) \propto 1$, then the MLE = posterior mode. If the posterior distribution is symmetric and $\lambda(\boldsymbol{\theta}) \propto 1$, then

$$\text{posterior mode} = \text{posterior mean} = \text{MLE}.$$

(ii) are easy to work with computationally.

Note that in most problems, a uniform prior is an improper prior.

Improper priors are attractive as noninformative priors, in general, since improper priors do not require the specification of prior hyperparameters. Jeffrey's prior is an automated way of constructing noninformative priors, which are usually improper. No hyperparameters specifications are required for Jeffrey's prior.

If $X \sim N(\theta, 1)$, then Jeffreys prior for θ is $\lambda_J(\theta) \propto 1$.

If $X \sim \text{Bernoulli}(\theta)$, then Jeffreys prior for θ is $\lambda_J(\theta) = \text{Beta}(\frac{1}{2}, \frac{1}{2})$.

If $X \sim \text{Poisson}(\theta)$, then Jeffreys prior for θ is $\lambda_J(\theta) \propto \theta^{-1/2}$.

For scale parameters, the uniform prior is typically not a good choice for a noninformative prior. For improper priors for scale parameters, one usually uses Jeffreys prior or some modification of it. Moreover, it is often more convenient to specify priors for the precision $= \frac{1}{\text{variance}}$.

If $X \sim N(\mu, \sigma^2)$, then letting $\tau = 1/\sigma^2$, the prior $\lambda(\tau) \propto \tau^{-1}$ is a good choice of noninformative prior for τ (it yields results similar to maximum likelihood). In general, "good" improper priors for scale parameters take the form $\lambda(\tau) \propto \tau^{-a}$ for $a > 0$.

One can also specify proper noninformative priors for scale parameters. The most popular choice is the gamma prior. If $\theta \sim \text{Gamma}(\delta_0, \gamma_0)$, then we obtain a noninformative prior by

(i) taking $\delta_0 = \gamma_0 =$small, such as .001, for example, that is

$$\theta \sim \text{Gamma}(.001, .001),$$

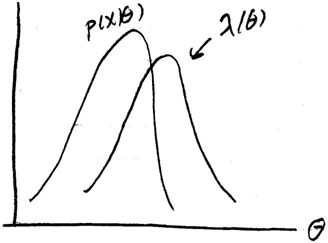
(ii) taking $\delta_0 \geq 1$, and $\gamma_0 \ll \delta_0$. For example,

$$\theta \sim \text{Gamma}(1, .001).$$

The notion of "noninformative" when proper priors are used is all relative to the likelihood. Thus, we only say that a proper prior is noninformative if the prior is dominated (or flat) relative to the likelihood. Of course, when specifying noninformative proper priors, prior hyperparameters must be specified. Improper priors, are by definition, "flat" relative to the likelihood.

Informative Priors

Priors not dominated by the likelihood are called informative priors. Informative priors have an impact on the posterior analysis for $\boldsymbol{\theta}$.


p(x|θ)
λ(θ)
θ

By definition informative priors must be proper. For location parameters, a $N(\mu_0, \sigma_0^2)$ prior is common where (μ_0, σ_0^2) are elicited based on expert opinion or historical data. For scale parameters, the Gamma(δ_0, γ_0) prior is a common choice. In general, conjugate priors may be good choices as informative priors.

When location and scale parameters are unknown in an inference problem and it is desired to specify a noninformative prior, it is common to take the parameters independent a priori. For example, suppose $X \sim N(\mu, \sigma^2)$, (μ, σ^2) are both unknown, and $\tau = 1/\sigma^2$.

A good choice for a joint noninformative prior for (μ, τ) is $\lambda(\mu, \tau) \propto \tau^{-1}$. (here μ has a uniform improper prior).

If we wish to specify an informative prior, or a noninformative proper prior, then a priori dependence between the location and scale parameters can be considered.

If

$$X \sim N(\mu, \sigma^2), \quad (\mu, \sigma^2) \text{ both unknown,}$$

then we can take

$$\mu | \tau \sim N(\mu_0, \tau^{-1}\sigma_0^2),$$

and

$$\tau \sim \text{Gamma}(\delta_0, \gamma_0),$$

where $(\mu_0, \sigma_0^2, \delta_0, \gamma_0)$ are specified hyperparameters. This prior specification is also a conjugate prior specification for (μ, τ).