1. *Let $X_1, \ldots, X_n$ be a random sample from a population with density $f(x|\theta) = exp(-(x-\theta))$, $x > \theta$, where $-\infty < \theta < \infty$ is unknown. Consider testing at level α*

$$H_0 : \theta \leq 0 \text{ versus } H_1 : \theta > 0.$$

   (a) *Show that the conditions required for the existence of UMP test are satisfied here.*
   (b) *Derive the UMP test of level α.*
   (c) *Find the minimal sufficient statistic for θ.*

Solution: The joint pdf of $X_1, \ldots, X_n$ is

$$f(\mathbf{x}|\theta) = exp(-\sum_{i=1}^{n}(x_i - \theta))I(\min_i x_i > \theta).$$

Using the Factorization Theorem, we get that $X_{(1)} = \min_i X_i$ is the sufficient statistic for θ.

(a) The distribution of $X_{(1)}$ is

$$g(y|\theta) = exp(-n(y-\theta))I(y > \theta).$$

The family of pdfs $\{g(y|\theta) : \theta \in (-\infty, \infty)\}$ has a monotone likelihood ratio (MLR), as for every $\theta_2 > \theta_1$, the ratio $\frac{g(y|\theta_2)}{g(y|\theta_1)}$ is a monotone function of y on $\{y : g(y|\theta_1) > 0 \text{ or } g(y|\theta_2) > 0\}$. This can be seen as

$$\begin{aligned} \frac{g(y|\theta_2)}{g(y|\theta_1)} = exp(n(\theta_2 - \theta_1))\frac{I(y>\theta_2)}{I(y>\theta_1)} &= exp(n(\theta_2 - \theta_1)), \text{ for } y > \theta_2 \\ &= 0 \text{ for } \theta_1 < y \leq \theta_2. \end{aligned}$$

Using the theorem due to Karlin and Rubin, for any y_0, the test that rejects H_0 if and only if $X_{(1)} > y_0$ is a UMP level α test, where $\alpha = P_{H_0}(X_{(1)} > y_0)$.

(b) Choosing

$$y_0 = -log(\alpha)/n,$$

we get $P_{H_0}(X_{(1)} > y_0) = \alpha$. For testing H_0 against H_1, the UMP test of level α, is

$$\begin{aligned} \phi(\mathbf{x}) &= 1, \text{ for } \min_i x_i > -log(\alpha)/n \\ &= 0 \text{ otherwise.} \end{aligned}$$

(c) Let $\mathbf{x}$ and $\mathbf{y}$ be two sample points. Then, the ratio of the densities

$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)} = \frac{exp(-\sum_{i=1}^{n}(x_i - \theta))I(\min_i x_i > \theta)}{exp(-\sum_{i=1}^{n}(y_i - \theta))I(\min_i y_i > \theta)},$$

is independent of θ if and only if $\min_i x_i = \min_i y_i$. Thus, $X_{(1)} = \min_i X_i$ is the minimal sufficient statistic for θ.

$\square$

2. *Suppose $X_1, \ldots, X_n$ is a random sample from $N(\mu, \sigma^2)$ where both μ and σ^2 are unknown.*

   (a) *Derive the generalized likelihood ratio level α test for testing $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 \neq 1$.*

   (b) *Is this also the UMP level α test? Justify?.*

Solution: The likelihood function is defined as

$$L(\mu, \sigma^2|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}).$$

(a) The entire parameter space is $\Theta = \{\mu, \sigma^2 : -\infty < \mu < \infty, \sigma^2 \geq 0\}$. The parameter space under H_0 is $\Theta_0 = \{\mu, \sigma^2 : -\infty < \mu < \infty, \sigma^2 = 1\}$. The LRT statistic is

$$\lambda(\mathbf{x}) = \frac{\underset{\Theta_0}{sup} L(\mu, 1|\mathbf{x})}{\underset{\Theta}{sup} L(\mu, \sigma^2|\mathbf{x})}.$$

The MLE for μ over the parameter space $\{\mu : -\infty < \mu < \infty\}$ is $\bar{x}_n = \sum_{i=1}^n x_i/n$. The MLE of μ and σ^2 over Θ, respectively, are $\bar{x}_n$ and $s_n^2 = \left(\sum_{i=1}^n (x_i - \bar{x}_n)^2\right)/n$. Then,

$$\lambda(\mathbf{x}) = \frac{exp(-\frac{\sum_{i=1}^n (x_i - \bar{x}_n)^2}{2})}{\frac{1}{(s_n^2)^{n/2}} exp(-\frac{n}{2})} = (s_n^2)^{n/2} exp(-\frac{s^2 n}{2} + \frac{n}{2}).$$

The generalized likelihood ratio test rejects H_0 for small values of $\lambda(\mathbf{x})$. The critical region can be written as $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c\}$.
Under H_0, $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ follows a $\chi^2(n-1)$ distribution, where $\chi^2(n-1)$ denotes the χ^2 distribution with $(n-1)$ degrees. The critical region for generalized likelihood ratio test that rejects H_0 is

$$(ns_n^2)^{n/2} exp(-\frac{(ns_n^2)}{2}) \leq c_\alpha exp(-\frac{n}{2}) n^{n/2} = c'_\alpha$$

where c'_α is chosen to give a size α test.
The above critical region for generalized likelihood ratio test of size α that rejects H_0 can be written as

$$ns_n^2 < c'_{1,\alpha}, \text{ and } ns_n^2 > c'_{2,\alpha},$$

where $c'_{1,\alpha}$ and $c'_{2,\alpha}$ are chosen such that

$$P_{H_0}(\sum_{i=1}^n (X_i - \bar{X}_n)^2 < c'_{1,\alpha}, \sum_{i=1}^n (X_i - \bar{X}_n)^2 > c'_{2,\alpha}) = \alpha.$$

For an equal tail test,

$$c'_{1,\alpha} = \chi^2(n-1)(\frac{\alpha}{2}), \text{ and } c'_{2,\alpha} = \chi^2(n-1)(1 - \frac{\alpha}{2}),$$

where $\chi^2(n-1)(\frac{\alpha}{2})$ is the $\alpha \times 100^{th}$ percentile of the $\chi^2(n-1)$ distribution.

(b) The generalized likelihood ratio level α test for testing $H_0 : \sigma^2 = 1$ versus $H_1 : \sigma^2 \neq 1$ derived in (a) is not a UMP test. Let $\phi(\mathbf{x})$ denote the test derived in (a).

For testing H_0 against $H_1' : \sigma^2 > 1$, a test of the form

$$\begin{aligned} \phi_1(\mathbf{x}) &= 1, \text{ for } \textstyle\sum_{i=1}^n (x_i - \bar{x}_n)^2 > \chi^2(n-1)(1-\alpha), \\ &= 0 \text{ otherwise,} \end{aligned}$$

is UMP test of size α.

The power of test ϕ_1 cannot exceed the power of test ϕ for testing H_0 against $H_1'' : \sigma^2 < 1$. Similarly, the power of test ϕ cannot exceed the power of ϕ_1 test for testing H_0 against $H_1' : \sigma^2 > 1$. Therefore, $\phi(\mathbf{x})$ cannot be the UMP level α test for testing H_0 against H_1.

$\square$

3. *Let X denote the number of independent Bernoulli(θ) trials before the first success occurs.*

   (a) What is the probability mass function of X?

   (b) Find the Fisher Information $I_1(\theta)$ contained in X.

   Let $X_1, \dots, X_n$ be a random sample from the distribution of X with $0 < \theta < 1$ unknown.

   (c) Find an estimator $T_n = T_n(X_1, \dots, X_n)$ such that

$$\sqrt{n}(T_n - \theta) \to N\Big(0, \frac{1}{I_1(\theta)}\Big).$$

   (d) Is it true that any estimator as in (c) above is a consistent estimator of θ? Why?

Solution:

(a) The probability mass function for X is

$$p(x|\theta) = (1-\theta)^x \theta, x = 0, 1, \dots.$$

(b)

$$logp(x|\theta) = xlog(1-\theta) + log\theta.$$

The Fisher's information number is

$$E_\theta\Big[\Big(\frac{\partial logp(X|\theta)}{\partial \theta}\Big)^2\Big] = E_\theta\Big[\Big(\frac{-X}{1-\theta} + \frac{1}{\theta}\Big)^2\Big] = \frac{1}{(1-\theta)^2} E_\theta\Big[\Big(-X + \frac{1-\theta}{\theta}\Big)^2\Big] = \frac{1}{\theta^2(1-\theta)}.$$

(c) Let $logL(\theta|\mathbf{x})$ denote the log likelihood function. Then,

$$logL(\theta|\mathbf{x}) = \sum_{i=1}^n x_i(1-\theta) + nlog\theta.$$

The partial derivative, with respect to θ is

$$\frac{\partial logL(\theta|\mathbf{x})}{\partial \theta} = -\frac{\sum_{i=1}^n x_i}{(1-\theta)} + \frac{n}{\theta}.$$

Setting the partial derivative to 0 and solving the equation yields the following unique solution

$$\hat{\theta}_n = \frac{1}{\bar{x}_n + 1},$$

where $\bar{x}_n = \sum_{i=1}^{n} x_i/n$. Also,

$$\frac{d^2 logL(\theta|\mathbf{x})}{d(\theta)^2}\Big|_{\theta=\hat{\theta}_n} < 0.$$

The MLE for θ is $\hat{\theta}_n$.

As the distribution function of X_1 satisfies the regularity conditions, the MLE $\hat{\theta}_n$ is asymptotically normally distributed, *i.e.*

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N\Big(0, \frac{1}{I_1(\theta)}\Big), \text{ as } n \to \infty.$$

One can choose the estimator T_n as $\hat{\theta}_n$.

(d) Yes, any estimator as in (c) above will be a consistent estimator. Let Z follow a standard normal distribution. As

$$T_n - \theta = \sqrt{n}(T_n - \theta)I_1(\theta)\Big[\frac{1}{I_1(\theta)\sqrt{n}}\Big] \to Z \lim_{n\to\infty}\Big[\frac{1}{I_1(\theta)\sqrt{n}}\Big] = 0,$$

$T_n - \theta$ converges in distribution to 0, as $n \to \infty$. Hence, $T_n - \theta$ converges in probability to 0, as $n \to \infty$.

□

4. *In an ecological study 5 independent attempts were made to photographically capture (or to camera trap) a particular tiger. The fourth attempt provided the only success. The success probability, θ, is known as the detection probability. Assume that the prior distribution on θ is $Beta(0.2, 1)$.*

   (a) *Derive the posterior distribution of θ given the data.*

   (b) *Find the highest posterior density estimate of θ.*

   (c) *Find the posterior mean and posterior standard deviation of θ.*

   (d) *Consider testing $H_0 : \theta \leq 0.25$ versus $H_1 : \theta > 0.25$. Explain the Bayesian approach for this.*

Solution: Let Y be the number of successes of camera trapping the particular tiger in n attempts. Then,

$$f(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}, y = 0, \ldots, n.$$

The prior distribution of θ is

$$g(\theta) = \frac{\Gamma(1.2)}{\Gamma(0.2)\Gamma(1)}\theta^{-0.8}, 0 < \theta < 1.$$

(a) The joint distribution of Y and θ is

$$p_1(y, \theta) = f(y|\theta)g(\theta).$$

The marginal pdf of Y is

$$p(y) = \int_0^1 p_1(y,\theta)d\theta = \binom{n}{y} \frac{\Gamma(1.2)}{\Gamma(0.2)\Gamma(1)} \frac{\Gamma(y+0.2)\Gamma(n-y+1)}{\Gamma(n+1.2)}.$$

The posterior distribution of θ given data $(y=1, n=5)$, is

$$g_1(\theta|y) = \frac{p(y,\theta)}{p(y)} = \frac{\Gamma(6.2)}{\Gamma(1.2)\Gamma(5)} \theta^{0.2}(1-\theta)^4, 0<\theta<1,$$

i.e. the posterior distribution of θ given data $(y=1, n=5)$, is $Beta(1.2, 5)$.

(b) The posterior density $g(\theta|y)$ is unimodal, with the mode

$$\frac{0.2}{4.2} = \frac{1}{21}.$$

The highest posterior density estimate of θ is $1/21 = 0.04762$.

(c) Let X follow a $Beta(\alpha,\beta)$ distribution. Then, the mean

$$E(X) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^\alpha (1-x)^{\beta-1}dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\alpha)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta}.$$

Similarly, it can be shown that

$$E(X^2) = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.$$

The variance is,

$$E(X^2) - (E(X))^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Putting $\alpha = 1.2$ and $\beta = 5$, the posterior mean and posterior standard deviation of θ are 0.19354 and 0.14724, respectively.

(d) The given problem of hypothesis testing can be viewed as a special decision problem. Given the observation y, the decision $d(y) \in \mathcal{A}$, where $\mathcal{A} = \{a_1, a_2\}$ is the action set, with a_1 corresponding to acceptance of H_0 and a_2 corresponding to rejection of H_0. The decision function d divides the space of values of Y into set C and its complement C^c, such that if $y \in C$ we take the decision $d(y) = a_1$ (accept H_0), and if $y \in C^c$, we take decision $d(y) = a_2$ (reject H_0). In Bayesian approach, we choose the decision function d that minimizes the risk $R(g,d) = E(E_\theta(L(\theta,d)))$, for a loss function L defined on $(0,1)\times\mathcal{A}$. One can choose the Loss function L as $(a(\theta), b(\theta) > 0)$,

$$\begin{aligned} L(\theta,a_1) &= 0, \text{ for } \theta \le 0.25 \\ &= a(\theta), \text{ for } \theta > 0.25, \\ L(\theta,a_2) &= 0, \text{ for } \theta > 0.25 \\ &= b(\theta), \text{ for } \theta \le 0.25. \end{aligned}$$

We choose d that minimizes $R(g,d)$. As $R(g,d) = E(E_\theta(L(\theta,d)))$, it suffices to minimize $E(L(\theta,d(y))|Y=y)$. Then,

$$\begin{aligned} E(L(\theta,d(y))|Y=y) &= \textstyle\int_{\theta>0.25} a(\theta)g_1(\theta|y)d\theta \text{ for } d(y)=a_1, \\ &= \textstyle\int_{\theta\le 0.25} b(\theta)g_1(\theta|y)d\theta \text{ for } d(y)=a_2. \end{aligned}$$

It follows that we reject H_0, *i.e.* take action $d(y) = a_2$ if

$$\int_{\theta\le 0.25} b(\theta)g_1(\theta|y)d\theta \le \int_{\theta>0.25} a(\theta)g_1(\theta|y)d\theta.$$

For $a(\theta) \equiv 1$, for $\theta > 0.25$ and $b(\theta) \equiv 1$, for $\theta \leq 0.25$, the above can be written as

$$\int_0^{0.25} g(\theta|y)d\theta \leq \frac{1}{2}.$$

The above inequality implies that H_0 is rejected in favor of H_1 if and only if the median of the posterior distribution $Beta(1.2, 5)$ is greater than 0.25.

□

5. (a) *Let S and T be two statistics such that S has finite variance. Show that*

$$Var(S) = Var(E(S|T)) + E(Var(S|T)).$$

(b) *Suppose $(X_1, X_2, \ldots, X_n)$ has probability distribution $P_\theta, \theta \in \Theta$. Let $\delta(X_1, X_2, \ldots, X_n)$ be an estimator of θ with finite variance. Suppose that T is sufficient for θ, and let $\delta^\star(t) = E(\delta(X_1, X_2, \ldots, X_n)|T = t)$, be the conditional expectation of $\delta(X_1, X_2, \ldots, X_n)$ given $T = t$. Then, arguing as in (a), and without applying Jensen's inequality, prove that*

$$E(\delta^\star(T) - \theta)^2 \leq E(\delta(X_1, X_2, \ldots, X_n) - \theta)^2,$$

with strict inequality unless $\delta = \delta^\star$ (i.e., δ is already a function of T).

Solution:

(a)

$$\begin{aligned} Var(S) = E(S^2) - [E(S)]^2 &= E(E(S^2|T)) - E([E(S|T)]^2) + E([E(S|T)]^2) - [E(S)]^2 \\ &= E(Var(S|T)) + E([E(S|T)]^2) - [E(E(S|T))]^2 \\ &= E(Var(S|T)) + Var([E(S|T)]). \end{aligned}$$

(b) $E(\delta(X_1, X_2, \ldots, X_n)) = \theta$. In (a), put $S = \delta(X_1, X_2, \ldots, X_n)$. Then,

$$\begin{aligned} E(\delta(X_1, X_2, \ldots, X_n) - \theta)^2 &= Var(E(\delta(X_1, X_2, \ldots, X_n)|T)) + E(Var(\delta(X_1, X_2, \ldots, X_n)|T)) \\ &\geq Var(E(\delta(X_1, X_2, \ldots, X_n)|T)), \qquad (1) \end{aligned}$$

because $E(Var(\delta(X_1, X_2, \ldots, X_n)|T)) \geq 0$. As

$$Var(E(\delta(X_1, X_2, \ldots, X_n)|T)) = Var(\delta^\star(T)) = E(\delta^\star(T) - \theta)^2$$

we get

$$E(\delta(X_1, X_2, \ldots, X_n) - \theta)^2 \geq E(\delta^\star(T) - \theta)^2.$$

The equality exists if $E(Var(\delta(X_1, X_2, \ldots, X_n)|T)) = 0$, i.e.

$$\begin{aligned} &E(E[(\delta(X_1, X_2, \ldots, X_n) - \delta^\star(T))^2|T]) = 0 \\ &\implies \delta(X_1, X_2, \ldots, X_n) - E(\delta(X_1, X_2, \ldots, X_n)|T) = 0, \end{aligned}$$

i.e. δ is a function of T.

□