Bayesian Framework

Prior distribution \( \pi(\theta) \)

Sampling distribution \( x|\theta \sim f(x|\theta) \)

Joint distribution \( \pi(\theta) f(x|\theta) \)

Marginal distribution \( m(x) = \int \pi(\theta) f(x|\theta) d\theta \)

Posterior distribution \( \pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} \)

Bayes Estimator is a posterior mean of \( \theta : E[\theta|x] \).

Bayesian Decision Theory

Loss Function \( L(\theta, \hat{\theta}) \) (e.g. \( (\theta - \hat{\theta})^2 \))

Risk Function is the average loss : \( R(\theta, \hat{\theta}) = E[L(\theta, \hat{\theta})|\theta] \).

For squared error loss \( L = (\theta - \hat{\theta})^2 \), the risk function is MSE

Bayes Risk is the average risk across all \( \theta : E[R(\theta, \hat{\theta})|\pi(\theta)] \).

Bayes Rule Estimator minimizes Bayes risk \( \iff \) minimizes posterior expected loss.
Asymptotics

Consistency Using law of large numbers, show variance and bias converge to zero, for any continuous mapping function \( \tau \)

Asymptotic Normality Using central limit theorem, Slutsky Theorem, and Delta Method

Asymptotic Relative Efficiency ARE(\( V_n, W_n \)) = \( \sigma^2_W/\sigma^2_V \).

Asymptotically Efficient ARE with CR-bound of unbiased estimator of \( \tau(\theta) \) is 1.

Asymptotic Efficiency of MLE Theorem 10.1.12 MLE is always asymptotically efficient under regularity condition.

Hypothesis Testing

Type I error \( \Pr(X \in R|\theta) \) when \( \theta \in \Omega_0 \)

Type II error \( 1 - \Pr(X \in R|\theta) \) when \( \theta \in \Omega_0^c \)

Power function \( \beta(\theta) = \Pr(X \in R|\theta) \)

\( \beta(\theta) \) represents Type I error under \( H_0 \), and power (=1-Type II error) under \( H_1 \).

Size \( \alpha \) test \( \sup_{\theta \in \Omega_0} \beta(\theta) = \alpha \)

Level \( \alpha \) test \( \sup_{\theta \in \Omega} \beta(\theta) \leq \alpha \)

LRT \( \lambda(x) = \frac{L(\hat{\theta}|x)}{L(\hat{\theta}_0|x)} \) rejects \( H_0 \) when \( \lambda(x) \leq c \)

\( \iff -2 \log \lambda(x) \geq -2 \log c = c^* \)

LRT based on sufficient statistics LRT based on full data and sufficient statistics are identical.

Asymptotic Tests and p-Values

Asymptotic Distribution of LRT For testing, \( H_0 : \theta = \theta_0 \) vs. \( H_1 : \theta = \theta_1 \),

\[ -2 \log \lambda(x) \xrightarrow{d} \chi^2_1 \] under regularity condition.

Wald Test If \( W_n \) is a consistent estimator of \( \theta \), and \( S^2_n \) is a consistent estimator of \( \text{Var}(W_n) \), then \( Z_n = (W_n - \theta_0)/S_n \) follows a standard normal distribution

- Two-sided test: \( |Z_n| > z_{\alpha/2} \)
- One-sided test: \( Z_n > z_{\alpha/2} \) or \( Z_n < -z_{\alpha/2} \)

p-Value A p-value \( 0 \leq p(x) \leq 1 \) is valid if, \( \Pr(p(X) \leq \alpha | \theta) \leq \alpha \) for every \( \theta \in \Omega_0 \) and \( 0 \leq \alpha \leq 1 \).

Constructing p-Value Theorem 8.3.27: If large \( W(X) \) value gives evidence that \( H_1 \) is true, \( p(x) = \sup_{\theta \in \Omega_0} \Pr(W(X) \geq W(x)|\theta) \) is a valid p-value

p-Value given sufficient statistics For a sufficient statistic \( S(X) \),

\( p(x) = \Pr(W(X) \geq W(x)|S(X) = S(x)) \) is also a valid p-value.
Interval Estimation

Coverage probability \( \Pr(\theta \in [L(X), U(X)]) \)

Coverage coefficient is 1 - \( \alpha \) if \( \inf_{\theta \in \Omega} \Pr(\theta \in [L(X), U(X)]) = 1 - \alpha \)

Confidence interval \([L(X), U(X)])\) is 1 - \( \alpha \) if
\[ \inf_{\theta \in \Omega} \Pr(\theta \in [L(X), U(X)]) = 1 - \alpha \]

Inverting a level \( \alpha \) test If \( A(\theta_0) \) is the acceptance region of a level \( \alpha \) test, then \( C(X) = \{ \theta : X \in A(\theta) \} \) is a 1 - \( \alpha \) confidence set (or interval).

Solution for (a)

For \( \theta_1 < \theta_2 \),
\[
\frac{f(x|\theta_2)}{f(x|\theta_1)} = \frac{e^{(x-\theta_2)}}{(1 + e^{x-\theta_2})^2} \frac{(1 + e^{x-\theta_1})^2}{e^{(x-\theta_1)}}
\]
\[
= e^{(\theta_1 - \theta_2)} \left( 1 + e^{x-\theta_1} \right)^2
\]

Let \( r(x) = (1 + e^{x-\theta_1})/(1 + e^{x-\theta_2}) \)
\[
r'(x) = \frac{e^{x-\theta_1} (1 + e^{x-\theta_2}) - (1 + e^{x-\theta_1}) e^{x-\theta_2}}{(1 + e^{x-\theta_2})^2}
\]
\[
= \frac{e^{x-\theta_1} - e^{x-\theta_2}}{(1 + e^{x-\theta_2})^2} > 0 \quad (\because x - \theta_1 > x - \theta_2)
\]

Therefore, the family of \( X \) has an MLR.

Solution for (b)

The UMP test rejects \( H_0 \) if and only if
\[
\frac{f(x|1)}{f(x|0)} = \frac{1 + e^x}{1 + e^{x-1}} > k
\]
\[
\frac{1 + e^x}{1 + e^{x-1}} > k^*
\]
\[
\frac{1 + e^x}{e + e^x} > k^{**}
\]
\[
X > x_0
\]

Because under \( H_0 \), \( F(x_0|\theta = 0) = \frac{e^x}{1 + e^x} \), the rejection region of UMP level \( \alpha \) test satisfies
\[
1 - F(x|\theta = 0) = \frac{1}{1 + e^{x_0}} = \alpha
\]
\[
x_0 \sim \log \left( \frac{1 - \alpha}{\alpha} \right)
\]
Solution for (c)

Because the family of $X$ has an MLR, UMP size $\alpha$ for testing $H_0 : \theta \leq 0$ vs. $H_1 : \theta > 0$ should be a form of

$$X > x_0$$

$$\Pr(X > x_0 | \theta = 0) = \alpha$$

Therefore, $x_0 = \log \left( \frac{1-\alpha}{\alpha} \right)$, which is identical to the test defined in (b).

Solution (a) - Consistency

1. Obtain $EX = 1/\theta$ (Derive yourself if not given)

$$EX = \int_0^\infty xf(x|\theta) dx = \int_0^\infty \theta x \exp(-\theta x) dx$$

$$= \left[-x \exp(-\theta x)\right]_0^\infty + \int_0^\infty \exp(-\theta x) dx$$

$$= 0 + \left[-\frac{1}{\theta} \exp(-\theta x)\right]_0^\infty = \frac{1}{\theta}$$

2. By LLN (Law of Large Number), $\bar{X} \xrightarrow{P} EX = 1/\theta$.

3. By Theorem of continuous map, $n/\sum_{i=1}^n X_i = 1/\bar{X} \xrightarrow{P} \theta$.

Practice Problem 2

Problem

Suppose $X_1, \ldots, X_n$ are iid random samples with pdf $f(x|\theta) = \theta \exp(-\theta x)$, where $x \geq 0, \theta > 0$

(a) Show that $\sum_{i=1}^n X_i / n$ is a consistent estimator for $\theta$.

(b) Show that $\sum_{i=1}^n X_i / n$ is asymptotically normal and derive its asymptotic distribution.

(c) Derive the Wald asymptotic size $\alpha$ test for $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$.

(d) Find an asymptotic $(1 - \alpha)$ confidence interval for $\theta$ by inverting the above test.

You may use the fact that $EX = 1/\theta$ and $\text{Var}(X) = 1/\theta^2$.

Solution (b) - Asymptotic Distribution

1. Obtain $\text{Var}(X) = 1/\theta^2$ (Derive if needed, omitted here).

2. Apply CLT (Central Limit Theorem),

$$\bar{X} \sim \mathcal{AN} \left( \frac{1}{\theta}, \frac{1}{\theta^2 n} \right)$$

3. Apply Delta method. Let $g(y) = 1/y$, then $g'(y) = -1/y^2$.

$$\frac{\sum X_i}{n} = 1/\bar{X} = g(\bar{X}) \sim \mathcal{AN} \left( g(1/\theta), \frac{[g'(1/\theta)]^2}{\theta^2 n} \right)$$

$$= \mathcal{AN} \left( \theta, \frac{\theta^2}{n} \right)$$

$$\Leftrightarrow \sqrt{n} \left( \frac{1}{\bar{X}} - \theta \right) = \mathcal{N} \left( 0, \theta^2 \right)$$
Solution (c) - Wald asymptotic size $\alpha$ test

1. Obtain a consistent estimator of $\theta$:

$$W(X) = \frac{\sum_{i=1}^{n} X_i}{n} \sim \mathcal{N}\left(\theta, \frac{\theta^2}{n}\right)$$

2. Obtain a constant estimator of $\text{Var}(W)$

$$\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \xrightarrow{P} \frac{\text{Var}(X)}{\theta^2} \quad \text{(CLT)}$$

$$\frac{n-1}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \xrightarrow{P} \theta^2 \quad \text{(Continuous Map Theorem)}.$$

$$S^2 = \frac{n-1}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \xrightarrow{P} \theta^2 \quad \text{(Slutsky’s Theorem)}.$$

Solution (d) - Asymptotic $1 - \alpha$ confidence interval

The acceptance region is

$$A = \left\{ x : \left| \frac{1}{\bar{X}} - \theta_0 \right| \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{X})^2}{n}} \leq z_{\alpha/2} \right\}$$

By inverting the acceptance region, the confidence interval is

$$C(X) = \left\{ \theta : \left| \frac{1}{\bar{X}} - \theta \right| \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}} \leq z_{\alpha/2} \right\}$$

which is equivalent to

$$C(X) = \left\{ \theta \in \left[ \frac{1}{\bar{X}} - \frac{z_{\alpha/2}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}}, \frac{1}{\bar{X}} + \frac{z_{\alpha/2}}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \right] \right\}$$

Practice Problem 3

**Problem**

The independent random variables $X_1, \cdots, X_n$ have the following pdf

$$f(x|\theta, \beta) = \frac{\beta x^{\beta-1}}{\theta^\beta} \quad 0 < x < \theta, \ \beta > 0$$

1. Find the MLEs of $\beta$ and $\theta$

2. When $\beta$ is a known constant $\beta_0$, construct a LRT testing $H_0 : \theta \geq \theta_0$ vs. $H_1 : \theta < \theta_0$.

3. When $\beta$ is a known constant $\beta_0$, find the upper confidence limit for $\theta$ with confidence coefficient $1 - \alpha$.  

Solution (c) - Wald Asymptotic size $\alpha$ test (cont’d)

3. Construct a two-sided asymptotic size $\alpha$ Wald test, whose rejection region is

$$|Z(X)| = \left| \frac{W(X) - \theta_0}{S/\sqrt{n}} \right| = \left| \frac{\sum_{i=1}^{n} X_i - \theta_0}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}} \right| = \left| \frac{1}{\bar{X}} - \theta_0 \right| \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2} \geq z_{\alpha/2}$$
(a) - MLE

\[ L(\theta, \beta | x) = \frac{\beta^n (\prod_{i=1}^{n} x_i)^{\beta-1}}{\hat{\theta}^{n\beta}} I(x(n) \leq \theta) \]

Because \( L \) is a decreasing function of \( \theta \) and positive only when \( \theta \geq x(n) \)

\[ \hat{\theta} = x(n) \]

\[ \ell(\theta, \beta | x) = n \log \beta + (\beta - 1) \sum \log x_i - n\beta \log \theta \]

\[ \frac{\partial \ell}{\partial \beta} = \frac{n}{\hat{\theta}} + \sum \log x_i - n \log \theta = 0 \]

\[ \hat{\beta} = \frac{n \log \hat{\theta} - \sum \log x_i}{n} \]

\[ = \frac{n x(n) - \sum \log x_i}{n} \]

(b) - size \( \alpha \) LRT

\[ \lambda(x) = \frac{\sup_{\theta \in \Omega_0} L(\hat{\theta} | x)}{\sup_{\theta \in \Omega} L(\hat{\theta} | x)} \]

\[ = \begin{cases} 1 & \theta_0 < x(n) \\ \frac{L(\theta_0 | x)}{L(x(n) | x)} & \theta_0 \geq x(n) \end{cases} \]

Therefore, the rejection region for size \( \alpha \) LRT is

\[ R = \left\{ x : x(n) \leq \theta_0 \alpha^{\frac{1}{n\beta_0}} \right\} \]

(c) - Upper \( 1 - \alpha \) confidence limit

The acceptance region of size \( \alpha \) LRT is

\[ A(\theta_0) = \left\{ x : x(n) > \theta_0 \alpha^{\frac{1}{n\beta_0}} \right\} \]

By inserting the acceptance region, the \( 1 - \alpha \) confidence interval becomes

\[ C(X) = \left\{ \theta : X(n) > \theta \alpha^{\frac{1}{n\beta_0}} \right\} \]

\[ = \left\{ \theta : \theta < X(n) \alpha^{-\frac{1}{n\beta_0}} \right\} \]

Therefore, the upper \( 1 - \alpha \) confidence limit is \( X(n) \alpha^{-\frac{1}{n\beta_0}} \).
Practice Problem 4

Problem
A random sample $X_1, \cdots, X_n$ is drawn from a population $N(\theta, \theta)$ where $\theta > 0$.

(a) Find the $\hat{\theta}$, the MLE of $\theta$
(b) Find the asymptotic distribution of $\hat{\theta}$.
(c) Compute $\text{ARE}(\hat{\theta}, \bar{X})$. Determine whether $\hat{\theta}$ is asymptotically more efficient than $\bar{X}$ or not.

You may use the following fact: $\text{Var}(X^2) = 4\theta^3 + 2\theta^2$.

(a) - MLE of $\theta$

$$L(\theta|x) = (2\pi)^{n/2} \exp \left[ -\frac{\sum_{i=1}^{n}(x_i - \theta)^2}{2\theta} \right]$$

$$l(\theta|x) = \frac{n}{2} \log(2\pi) + \frac{n}{2} \log \theta - \frac{\sum_{i=1}^{n}(x_i - \theta)^2}{2\theta}$$

$$= \frac{n}{2} \log(2\pi) + \frac{n}{2} \log \theta - \frac{\sum x_i^2}{2\theta} + \sum x_i - \frac{n\theta}{2}$$

$$\ell(\theta|x) = \frac{n}{2\theta} + \frac{\sum x_i^2}{2\theta^2} - \frac{n}{2} = n\theta - \frac{\sum x_i^2 - n\theta^2}{2\theta^2} = 0$$

$$n\theta^2 + n\theta - \sum x_i^2 = 0$$

$$\hat{\theta} = \frac{-1 + \sqrt{1 + 4 \sum x_i^2 / n}}{2}$$

$$\frac{1}{n} \sum x_i^2 = \hat{\theta}^2 + \hat{\theta}$$

(b) - Asymptotic distribution of MLE

By CLT, Let $W = \frac{1}{n} \sum X_i^2$, then

$$W \sim \mathcal{N}\left(\frac{\text{EX}^2}{n}, \frac{\text{Var}(X^2)}{n}\right) = \mathcal{N}\left(\theta + \theta^2, \frac{4\theta^3 + 2\theta^2}{n}\right)$$

The asymptotic distribution of MLE $\hat{\theta}$

$$\hat{\theta} \sim \mathcal{N}\left(\theta, \frac{\sigma^2(\theta)}{n}\right)$$

for some function $\sigma^2(\theta)$ and we would like to find $\sigma^2(\theta)$ using the asymptotic distribution of $W$.

(b) - Asymptotic distribution of MLE (cont’d)

Let $g(y) = y^2 + y$, then $g'(y) = (2y + 1)$ and $g(\hat{\theta}) = W$. Then by the Delta Method, the asymptotic distribution of $W$ can be written as

$$W = g(\hat{\theta}) \sim \mathcal{N}\left(g(\theta), g'(\theta)\frac{\sigma^2(\theta)}{n}\right)$$

$$= \mathcal{N}\left(\theta^2 + \theta, \frac{(2\theta + 1)^2 \sigma^2(\theta)}{n}\right)$$

$$= \mathcal{N}\left(\theta^2 + \theta, \frac{4\theta^3 + 2\theta^2}{n}\right)$$

$$\sigma^2(\theta) = \frac{4\theta^3 + 2\theta^2}{(2\theta + 1)^2} = \frac{2\theta^2(2\theta + 1)}{(2\theta + 1)^2} = \frac{2\theta^2}{2\theta + 1}$$
(b) - Asymptotic distribution of MLE (cont’d)

The asymptotic distribution of MLE \( \hat{\theta} \)

\[
\hat{\theta} \sim \mathcal{N} \left( \theta, \frac{\sigma^2(\theta)}{n} \right) \\
= \mathcal{N} \left( \theta, \frac{2\theta^2}{n(2\theta + 1)} \right)
\]

Note that you cannot use CR-bound for the asymptotic variance of MLE because the regularity condition does not hold (open set criteria).

(c) - ARE of MLE compared to \( \bar{X} \)

By CLT, the asymptotic distribution of \( \bar{X} \) is

\[
\bar{X} \sim \mathcal{N} \left( \theta, \frac{\theta}{n} \right)
\]

Then, ARE\( (\hat{\theta}, \bar{X}) \) is

\[
\text{ARE}(\hat{\theta}, \bar{X}) = \frac{\theta}{\frac{2\theta^2}{2\theta + 1}} = 1 + \frac{1}{2\theta} > 1
\]

Therefore, \( \hat{\theta} \) is more efficient estimator than \( \bar{X} \).

Wrapping Up

1. Many thanks for your attentions and feedbacks.
2. Please complete your teaching evaluations, which will be very helpful for further improvement in the next year.
3. Final exam will be Thursday April 25th, 4:00-6:00pm.
4. The last office hour will be held Wednesday April 24th, 4:00-5:00pm.
5. The grade will be posted during the weekend.
6. Don’t forget the materials we have learned, because they are the key topics for your candidacy exam.