Lecture 12: Likelihood Ratio Tests and Bayesian Tests

Mathematical Statistics II, MATH 60062/70062

Thursday March 3, 2022

Reference: Casella & Berger, 8.2.1-8.2.2

Recap: Likelihood ratio tests

Suppose that $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a random sample from $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$. The **likelihood function** is

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f_X(x_i \mid \boldsymbol{\theta}),$$

where $f_X(x \mid \boldsymbol{\theta})$ is the common population distribution.

The **likelihood ratio test (LRT) statistic** for testing

$$H_0 : \boldsymbol{\theta} \in \Theta_0 \quad \text{versus} \quad H_1 : \boldsymbol{\theta} \in \Theta_0^c$$

is

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta} \mid \boldsymbol{x})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta} \mid \boldsymbol{x})}$$

An LRT is any test that has a rejection region of the form $R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\}$, where c is any number satisfying $0 \leq c \leq 1$.

Normal LRT

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, 1)$, where $-\infty < \theta < \infty$ and σ_0^2 is known. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

The likelihood function is

$$L(\theta \mid \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(x_i-\theta)^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}.$$

Normal LRT

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, 1)$, where $-\infty < \theta < \infty$ and σ_0^2 is known. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

The likelihood function is

$$L(\theta \mid \boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(x_i-\theta)^2/2} = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2}.$$

Clearly,

$$\sup_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x}) = L(\theta_0 \mid \boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2}.$$

The MLE over the entire parameter space is $\hat{\theta} = \bar{X}$. Therefore,

$$\sup_{\theta \in \Theta} L(\theta \mid \boldsymbol{x}) = L(\bar{x} \mid \boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

The LRT statistic is

$$\lambda(\boldsymbol{x}) = \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\theta_0)^2}}{\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum_{i=1}^n (x_i-\bar{x})^2}} = e^{-\frac{1}{2}\left[\sum_{i=1}^n (x_i-\theta_0)^2 - \sum_{i=1}^n (x_i-\bar{x})^2\right]}.$$

Noting that

$$\sum_{i=1}^n (x_i-\theta_0)^2 = \sum_{i=1}^n (x_i-\bar{x})^2 + n(\bar{x}-\theta_0)^2,$$

the LRT statistic can be simplified as

$$\lambda(\boldsymbol{x}) = e^{-\frac{n}{2}(\bar{x}-\theta_0)^2}.$$

An LRT rejects the H_0 for small values of $\lambda(\boldsymbol{x})$, i.e., $\lambda(\boldsymbol{x}) \le c$.

To express the rejection region with a simpler statistic,

$$\begin{aligned}
\lambda(\boldsymbol{x}) = e^{-\frac{n}{2}(\bar{x}-\theta_0)^2} \leq c \iff & -\frac{n}{2}(\bar{x}-\theta_0)^2 \leq \log c \\
\iff & (\bar{x}-\theta_0)^2 \geq -\frac{2\log c}{n} \\
\iff & |\bar{x}-\theta_0| \geq \sqrt{-\frac{2\log c}{n}} = c'.
\end{aligned}$$

c is between 0 and 1, so logc is negative.

Therefore, the LRT rejection region can be written as

$$R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\} = \{\boldsymbol{x} : |\bar{x}-\theta_0| \geq c'\}.$$

To express the rejection region with a simpler statistic,

$$\begin{aligned}
\lambda(\boldsymbol{x}) = e^{-\frac{n}{2}(\bar{x}-\theta_0)^2} \le c \iff & -\frac{n}{2}(\bar{x}-\theta_0)^2 \le \log c \\
\iff & (\bar{x}-\theta_0)^2 \ge -\frac{2\log c}{n} \\
\iff & |\bar{x}-\theta_0| \ge \sqrt{-\frac{2\log c}{n}} = c'.
\end{aligned}$$

Therefore, the LRT rejection region can be written as

$$R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \le c\} = \{\boldsymbol{x} : |\bar{x}-\theta_0| \ge c'\}.$$

Normal Distribution: Normal LRT with known variance

Note that $\bar{X}$ is a sufficient statistic for $\mathcal{N}(\theta, 1)$.

Exponential LRT

Suppose $X_1, \ldots, X_n$ are iid $\text{Expo}(\theta)$ with PDF

$$f_X(x \mid \theta) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & x < \theta, \end{cases}$$

where $-\infty < \theta < \infty$. Consider testing

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0.$$

The likelihood function is

$$\begin{aligned} L(\theta \mid \boldsymbol{x}) &= \prod_{i=1}^{n} e^{-(x_i-\theta)} I(x_i \geq \theta) \\ &= \underbrace{e^{n\theta} I(x_{(1)} \geq \theta)}_{g(t|\theta)} \underbrace{e^{-\sum_{i=1}^{n} x_i}}_{h(\boldsymbol{x})}. \end{aligned}$$

By the Facorization Theorem, $T(\boldsymbol{X}) = X_{(1)}$ is a sufficient statistic.

We need to find the unrestricted MLE

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max}\, L(\theta \mid \boldsymbol{x}) = \underset{\theta \in \Theta}{\arg\max} \left\{ \prod_{i=1}^{n} e^{-(x_i - \theta)} I(x_i \geq \theta) \right\},$$

and the restricted MLE

$$\hat{\theta}_0 = \underset{\theta \in \Theta_0}{\arg\max}\, L(\theta \mid \boldsymbol{x}),$$

where $\Theta = \{\theta : -\infty < \theta < \infty\}$ and $\Theta_0 = \{\theta : -\infty < \theta \leq \theta_0\}$.

- When $\theta \leq x_{(1)}$, $L(\theta \mid \boldsymbol{x}) = e^{-\sum_{i=1}^{n} x_i + n\theta}$ is an increasing function of θ. When $\theta > x_{(1)}$, $L(\theta \mid \boldsymbol{x}) = 0$.

We need to find the unrestricted MLE

$$\hat{\theta} = \arg\max_{\theta \in \Theta} L(\theta \mid \boldsymbol{x}) = \arg\max_{\theta \in \Theta} \left\{ \prod_{i=1}^{n} e^{-(x_i - \theta)} I(x_i \geq \theta) \right\},$$

and the restricted MLE

$$\hat{\theta}_0 = \arg\max_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x}),$$

where $\Theta = \{\theta : -\infty < \theta < \infty\}$ and $\Theta_0 = \{\theta : -\infty < \theta \leq \theta_0\}$.

- When $\theta \leq x_{(1)}$, $L(\theta \mid \boldsymbol{x}) = e^{-\sum_{i=1}^{n} x_i + n\theta}$ is an increasing function of θ. When $\theta > x_{(1)}$, $L(\theta \mid \boldsymbol{x}) = 0$.
- The unrestricted MLE is $\hat{\theta} = X_{(1)}$.
- The restricted MLE is

$$\hat{\theta}_0 = \begin{cases} X_{(1)} & \text{if } \theta_0 \geq X_{(1)} \\ \theta_0 & \text{if } \theta_0 < X_{(1)}. \end{cases}$$

The LRT statistic is

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x})}{\sup_{\theta \in \Theta} L(\theta \mid \boldsymbol{x})} = \begin{cases} 1 & \text{if } \theta_0 \geq x_{(1)} \\ e^{-n(x_{(1)} - \theta_0)} & \text{if } \theta_0 < x_{(1)}. \end{cases}$$

An LRT rejects the H_0 for small values of $\lambda(\boldsymbol{x})$, i.e., $\lambda(\boldsymbol{x}) \leq c$.

$$\begin{aligned} \lambda(\boldsymbol{x}) = e^{-n(x_{(1)} - \theta_0)} \leq c &\iff -n(x_{(1)} - \theta_0) \leq \log c \\ &\iff x_{(1)} \geq \theta_0 - \frac{\log c}{n} = c'. \end{aligned}$$

Therefore, the LRT rejection region can be written as

$$R = \{\boldsymbol{x} : \lambda(\boldsymbol{x}) \leq c\} = \{\boldsymbol{x} : x_{(1)} \geq c'\}.$$

Sufficiency and LRT statistic

Suppose that $T(\boldsymbol{X})$ is a sufficient statistic for θ. If $\lambda^*(T(\boldsymbol{x})) = \lambda^*(t)$ is the LRT statistic based on T, and $\lambda(\boldsymbol{x})$ is the LRT statistic based on $\boldsymbol{X}$, then $\lambda^*(T(\boldsymbol{x})) = \lambda(\boldsymbol{x})$ for every $\boldsymbol{x} \in \mathcal{X}$.

Sufficiency and LRT statistic

Suppose that $T(\boldsymbol{X})$ is a sufficient statistic for θ. If $\lambda^*(T(\boldsymbol{x})) = \lambda^*(t)$ is the LRT statistic based on T, and $\lambda(\boldsymbol{x})$ is the LRT statistic based on $\boldsymbol{X}$, then $\lambda^*(T(\boldsymbol{x})) = \lambda(\boldsymbol{x})$ for every $\boldsymbol{x} \in \mathcal{X}$.

By the Factorization Theorem, we can write

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = g_T(t \mid \theta)h(\boldsymbol{x}),$$

where $g_T(t \mid \theta)$ is the PDF/PMF of T and $h(\boldsymbol{x})$ is free of θ. Thus,

$$\begin{aligned}
\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x})}{\sup_{\theta \in \Theta} L(\theta \mid \boldsymbol{x})} &= \frac{\sup_{\theta \in \Theta_0} g_T(t \mid \theta)h(\boldsymbol{x})}{\sup_{\theta \in \Theta} g_T(t \mid \theta)h(\boldsymbol{x})} \\
&= \frac{\sup_{\theta \in \Theta_0} g_T(t \mid \theta)}{\sup_{\theta \in \Theta} g_T(t \mid \theta)} \\
&= \frac{\sup_{\theta \in \Theta_0} L^*(\theta \mid t)}{\sup_{\theta \in \Theta} L^*(\theta \mid t)} = \lambda^*(T(\boldsymbol{x})),
\end{aligned}$$

where $L^*(\theta \mid t)$ is the likelihood function based on $T = t$.

Sufficiency and Normal LRT statistic

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, 1)$, where $-\infty < \theta < \infty$ and σ_0^2 is known. Consider testing

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

The sample mean $T(\boldsymbol{X}) = \bar{X}$ is a sufficient statistic for θ and it has a distribution of $\mathcal{N}(\theta, 1/n)$,

$$L^*(\theta \mid T(\boldsymbol{x})) = \frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(\bar{x}-\theta)^2}.$$

Therefore,

$$\lambda^*(T(\boldsymbol{x})) = \frac{\sup_{\theta \in \Theta_0} L^*(\theta \mid T(\boldsymbol{x}))}{\sup_{\theta \in \Theta} L^*(\theta \mid T(\boldsymbol{x}))} = \frac{\frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(\bar{x}-\theta_0)^2}}{\frac{1}{\sqrt{2\pi/n}} e^{-\frac{n}{2}(\bar{x}-\bar{x})^2}} = e^{-\frac{n}{2}(\bar{x}-\theta_0)^2},$$

same as the LRT statistic based on $\boldsymbol{X}$.

Normal LRT with unknown variance

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Both parameters are unknown. Consider testing

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0.$$

Set $\boldsymbol{\theta} = (\mu, \sigma^2)$. The likelihood function is

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}.$$

We need to find the unrestricted MLE

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\arg\max}\, L(\boldsymbol{\theta} \mid \boldsymbol{x}),$$

and the restricted MLE

$$\hat{\hat{\boldsymbol{\theta}}} = \underset{\boldsymbol{\theta} \in \Theta_0}{\arg\max}\, L(\boldsymbol{\theta} \mid \boldsymbol{x}),$$

where

$$\begin{aligned} \Theta &= \{\boldsymbol{\theta} = (\mu, \sigma^2) : -\infty < \mu < \infty, \sigma^2 > 0\} \\ \Theta_0 &= \{\boldsymbol{\theta} = (\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}. \end{aligned}$$

We call σ^2 a **nuisance parameter** because it is present in the model but is not of direct inferential interest (that is, H_0 vs. H_1).

The unrestricted MLE is

$$\hat{\boldsymbol{\theta}} = \begin{pmatrix} \bar{X} \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \end{pmatrix}.$$

The variance is 1/n because it is the solution of MLE function.

The restricted MLE is

$$\hat{\boldsymbol{\theta}}_0 = \begin{pmatrix} \mu_0 \\ \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 \end{pmatrix},$$

which maximizes $L(\boldsymbol{\theta} \mid \boldsymbol{x})$ over Θ_0.

Therefore, the LRT statistic is

$$\lambda(\boldsymbol{x}) = \frac{L(\hat{\boldsymbol{\theta}}_0 \mid \boldsymbol{x})}{L(\hat{\boldsymbol{\theta}} \mid \boldsymbol{x})} = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \mu_0)^2}\right]^{n/2}$$

Noting that $\sum_{i=1}^n (x_i - \mu_0)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2$, the rejection region

$$\lambda(\boldsymbol{x}) = \left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \mu_0)^2} \right]^{n/2} \leq c$$

can be written as

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu_0)^2} \leq c^{2/n},$$

which is equivalent to

$$\frac{1}{1 + \frac{n(\bar{x}-\mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \leq c^{2/n} \iff \frac{n(\bar{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \geq c^{-2/n} - 1.$$

By defining $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, the rejection region is equivalent to

$$\frac{(\bar{x}-\mu_0)^2}{s^2/n} = \frac{n(\bar{x}-\mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \geq (n-1)(c^{-2/n}-1)$$

T distribution: Normal LRT with unknown variance

That is,

$$\lambda(\boldsymbol{x}) \leq c \iff \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \geq c'$$

By defining $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, the rejection region is equivalent to

$$\frac{(\bar{x}-\mu_0)^2}{s^2/n} = \frac{n(\bar{x}-\mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} \geq (n-1)(c^{-2/n}-1)$$

That is,

$$\lambda(\boldsymbol{x}) \leq c \iff \left|\frac{\bar{x}-\mu_0}{s/\sqrt{n}}\right| \geq c'$$

This demonstrates that the Normal LRT is equivalent to a test based on Student's t statistic, known as the **one-sample t test**.

Bayesian tests

In a Bayesian framework, all inferences about θ are based on the posterior distribution $f(\theta \mid \boldsymbol{x})$, where

$$P(H_0 \text{ is true} \mid \boldsymbol{x}) = P(\theta \in \Theta_0 \mid \boldsymbol{x}) = \int_{\Theta_0} f(\theta \mid \boldsymbol{x}) d\theta$$
$$P(H_1 \text{ is true} \mid \boldsymbol{x}) = P(\theta \in \Theta_0^c \mid \boldsymbol{x}) = \int_{\Theta_0^c} f(\theta \mid \boldsymbol{x}) d\theta.$$

In classical statistics, θ is regarded as fixed and it makes no sense to assign probabilities to *non-random* events $\{\theta \in \Theta_0\}$ and $\{\theta \in \Theta_0^c\}$.

Bayesian tests

In a Bayesian framework, all inferences about θ are based on the posterior distribution $f(\theta \mid \boldsymbol{x})$, where

$$P(H_0 \text{ is true} \mid \boldsymbol{x}) = P(\theta \in \Theta_0 \mid \boldsymbol{x}) = \int_{\Theta_0} f(\theta \mid \boldsymbol{x}) d\theta$$
$$P(H_1 \text{ is true} \mid \boldsymbol{x}) = P(\theta \in \Theta_0^c \mid \boldsymbol{x}) = \int_{\Theta_0^c} f(\theta \mid \boldsymbol{x}) d\theta.$$

In classical statistics, θ is regarded as fixed and it makes no sense to assign probabilities to *non-random* events $\{\theta \in \Theta_0\}$ and $\{\theta \in \Theta_0^c\}$.

In a Bayesian test, the test statistic is $P(\theta \in \Theta_0^c \mid \boldsymbol{x})$ and the rejection region is $\{\boldsymbol{x} : P(\theta \in \Theta_0^c \mid \boldsymbol{x}) > c\}$, where $0 \leq c \leq 1$.

Normal Bayesian test

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, \sigma^2)$, where the prior distribution on θ is $\mathcal{N}(\mu, \tau^2)$. Assuming that σ^2, μ, and τ^2 are all known, then the posterior distribution of θ is a Normal, with mean

$$\frac{n\tau^2\bar{x} + \sigma^2\mu}{n\tau^2 + \sigma^2}$$

and variance

$$\frac{\sigma^2\tau^2}{n\tau^2 + \sigma^2}.$$

Consider testing

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0.$$

If we decide to accept H_0 if and only if

$$P(\theta \in \Theta_0 \mid \boldsymbol{x}) \geq P(\theta \in \Theta_0^c \mid \boldsymbol{x}),$$

then we will accept H_0 if and only if

$$\frac{1}{2} \leq P(\theta \in \Theta_0 \mid \boldsymbol{x}) \leq P(\theta \leq \theta_0 \mid \boldsymbol{x}).$$

\theta_0 is the hypothesis. P(\theta <= \theta_0)>=1/2 only when \theta_0 is greater than the posterior mean. The posterior mean is defined by x.

This means, we will accept H_0 if and only if the posterior mean is less than or equal to θ_0,

$$\implies \bar{x} \leq \theta_0 + \frac{\sigma^2(\theta_0 - \mu)}{n\tau^2}.$$