IV

Likelihood Ratio Tests (LRT)

Def: The likelihood ratio test statistic for testing

$H_0: \theta \in \Theta_0$; $H_1: \theta \in \Theta_0^c$

is

$$\lambda(\underline{x}) \equiv \frac{\sup_{\Theta_0} L(\theta|\underline{x})}{\sup_{\Theta} L(\theta|\underline{x})}$$

Def: A likelihood ratio test (LRT) is any hypothesis test that has a rejection region of the form $\{\underline{x}: \lambda(\underline{x}) \le c\}$, where $0 \le c \le 1$.

Note: 1° Related to the maximum likelihood estimators, LRTs are as widely applicable as maximum likelihood estimation.

2° The numerator of $\lambda(\underline{x})$ indicates the maximum prob. of the observed sample, computed over parameters in H_0. The LRT statistic $\lambda(\underline{x})$ is small if there are parameter points in H_1 for which the observed sample much more likely than any θ in H_0. Hence LRT says H_0 should be rejected if $\lambda(\underline{x})$ is small.

E.g.: 1° (Gaussian LRT)

Given $X \sim N(\theta, 1)$, $(\Theta = \mathbb{R})$

We have $H_0: \theta = \theta_0$; $H_1: \theta \ne \theta_0$.

LRT states that, rejection region $R = \{\underline{x}: \lambda(\underline{x}) \le c\}$.

where the LRT stat. $\lambda(\underline{x}) = \dfrac{L(\theta_0|\underline{x})}{L(\bar{x}|\underline{x})} = \exp\{-\frac{n}{2}(\bar{x}-\theta_0)^2\}$.

Hence the rejection region is simplified to $R = \{\underline{x}: |\bar{x}-\theta_0| \ge \sqrt{-\frac{2}{n}\ln c}\}$.

2° (Exponential LRT)

Given population $X \sim e^{-(x-\theta)} I(x \geq \theta)$, $\Theta = \mathbb{R}$.

Hypotheses are: $H_0: \theta \leq \theta_0$; $H_1: \theta > \theta_0$

The LRT stat. $\lambda(\underline{x}) = \dfrac{\sup_{\Theta_0} L(\theta|\underline{x})}{L(x_{(1)}|\underline{x})}$,

where the likelihood fn. $L(\theta|\underline{x}) = e^{-\sum x_i + n\theta} I(x_{(1)} \geq \theta)$

Hence, $\lambda(\underline{x}) = \begin{cases} 1 , & x_{(1)} \leq \theta_0 \\ e^{-n(x_{(1)}-\theta_0)} , & x_{(1)} > \theta_0 \end{cases}$.

LRT is thus simplified to state that

rejection region $R = \{\underline{x} : x_{(1)} \geq \theta_0 - \frac{1}{n} \ln c \}$.

Note: 1° The typical process to find the LRT is:

a) find the expression for $\lambda(\underline{x})$;

b) simplify the rejection region R to an expression involving a simpler statistic.

Thm: If $T(\underline{X})$ is a sufficient stat. for θ, and $\lambda^*(t)$ and $\lambda(\underline{x})$ are the LRT stat. based on T and $\underline{X}$ resp.,

then $\lambda^*(T(\underline{x})) = \lambda(\underline{x})$, $\forall \underline{x}$ in the sample space.

proof: $f(\underline{x}|\theta) = g(\underline{x})\, h(T(\underline{x})|\theta)$ is what sufficiency is all about.

Here we chose $h(T(\underline{x})|\theta)$ to be the pdf/pmf of $T(\underline{X})$.

Thus $\lambda^*(t) = \dfrac{\sup_{\Theta_0} h(t|\theta)}{\sup_{\Theta} h(t|\theta)}$,

$\lambda(\underline{x}) = \dfrac{\sup_{\Theta_0} f(\underline{x}|\theta)}{\sup_{\Theta} f(\underline{x}|\theta)}$.

The thm is obvious. $\square$.

Note: 1° In finding LRT, we simplify the rejection region R to an expression only depending on $T(\underline{X})$, a sufficient stat. for θ.

Eg: 1° (Gaussian LRT)

Given population $X \sim N(\theta, 1)$, $\Theta = \mathbb{R}$.

Hypotheses are: $H_0: \theta = \theta_0$; $H_1: \theta \neq \theta_0$.

$\overline{X}$ is a sufficient stat. for θ, and $\overline{X} \sim N(\theta, \frac{1}{n})$.

Using the thm., LRT states that,

rejection region $R = \{\underline{x} : \lambda^*(\bar{x}) \leq c\}$

where the LRT stat. based on $\overline{X}$, $\lambda^*(\bar{x}) = \dfrac{L(\theta_0|\bar{x})}{L(\bar{x}|\bar{x})}$.

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

Hence the rejection region is simplified to $R = \{\underline{x} : |\bar{x} - \theta_0| \geq \sqrt{-\frac{2}{n}\ln c}\}$

2° (Exponential LRT)

Given population $X \sim e^{-(x-\theta)} I(x \geq \theta)$, $\Theta = \mathbb{R}$

Hypotheses are $H_0: \theta \leq \theta_0$; $H_1: \theta > \theta_0$

$X_{(1)}$ is a sufficient stat. for θ, and $X_{(1)} \sim n e^{-n(z-\theta)} I(z \geq \theta)$

Using the thm., LRT states that,

rejection region $R = \{\underline{x} : \lambda^*(x_{(1)}) \leq c\}$

where the LRT stat. based on $X_{(1)}$, $\lambda^*(x_{(1)}) = \dfrac{\sup_{\Theta_0} L(\theta|x_{(1)})}{L(x_{(1)}|x_{(1)})}$

i.e. $\lambda^*(x_{(1)}) = \exp\{-n(x_{(1)} - \min\{\theta_0, x_{(1)}\})\}$

$$= \begin{cases} 1, & x_{(1)} \leq \theta_0 \\ e^{-n(x_{(1)}-\theta_0)}, & x_{(1)} > \theta_0. \end{cases}$$

Hence the rejection region is simplified to $R = \{\underline{x} : x_{(1)} \geq \theta_0 - \frac{1}{n}\ln c\}$.

Def: <u>Nuisance ~~para~~ parameters</u> are parameters that are present in a model but are not of direct inferential interest.

E.g.: 1° (Gaussian LRT with unknown σ^2)

Given population $X \sim N(\mu, \sigma^2)$, $(\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+)$

Hypotheses are: $H_0: \mu \le \mu_0$; $H_1: \mu > \mu_0$

Here σ^2 is a nuisance parameter.

The LRT stat.
$$\lambda(\underset{\sim}{x}) = \frac{\sup\limits_{\substack{\mu \in (-\infty, \mu_0] \\ \sigma^2 \in \mathbb{R}^+}} L(\mu, \sigma^2 \mid \underset{\sim}{x})}{\sup\limits_{\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+} L(\mu, \sigma^2 \mid \underset{\sim}{x})}$$
$$= \begin{cases} 1, & \hat{\mu} \le \mu_0 \\ \dfrac{L(\mu_0, \hat{\sigma}_0^2 \mid \underset{\sim}{x})}{L(\hat{\mu}, \hat{\sigma}^2 \mid \underset{\sim}{x})}, & \hat{\mu} > \mu_0 \end{cases}$$

where $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ are the MLE of μ and σ^2.

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

It can be shown that the test is equivalent to a test based on Student's t stat.