Lecture 25: Likelihood ratio tests

Another question

Let $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$. We wish to test $H_0 : \lambda = \lambda_0$ vs. $H_A : \lambda \neq \lambda_0$.

if we had $H_A : \lambda = \lambda_1$, N-P: $\dfrac{L(\lambda_1 | X)}{L(\lambda_0 | X)}$

idea: instead of looking at a single $\lambda_1 \in \mathbb{R}$, let's maximize $L(\lambda | X)$

Statistic: $\underbrace{\dfrac{\sup_{\lambda \neq \lambda_0} L(\lambda | X)}{L(\lambda_0 | X)}}$

reject when this is $> k$

choose to satisfy desired type I error

Likelihood ratio tests

Let $X_1, \ldots, X_n$ be a sample from a distribution with parameter $\theta \in \mathbb{R}^d$. We wish to test

$H_0: \theta \in \Theta_0$ vs. $H_A: \theta \in \Theta_1$

The likelihood ratio test (LRT) rejects H_0 when

$$\frac{\sup_{\theta \in \Theta_1} L(\theta \mid X)}{\sup_{\theta \in \Theta_0} L(\theta \mid X)} > k$$

where k is chosen so that $\sup_{\theta \in \Theta_0} B(\theta) \leq \alpha$

Back to the Poisson example

Let $X_1, \ldots, X_n \overset{iid}{\sim} Poisson(\lambda)$. We wish to test $H_0 : \lambda = \lambda_0$ vs. $H_A : \lambda \neq \lambda_0$.

$$\Lambda = \frac{\sup_{\lambda \neq \lambda_0} L(\lambda | X)}{\sup_{\lambda = \lambda_0} L(\lambda | X)} = \frac{\sup_{\lambda \neq \lambda_0} L(\lambda | X)}{L(\lambda_0 | X)} = \frac{\sup_{\lambda} L(\lambda | X)}{L(\lambda_0 | X)}$$

For Poisson: $\hat{\lambda}_{MLE} = \bar{X}$

$$= \frac{L(\hat{\lambda}_{MLE} | X)}{L(\lambda_0 | X)} = \frac{L(\bar{X} | X)}{L(\lambda_0 | X)}$$

$$= \frac{(\bar{X})^{\sum_i X_i} e^{-n\bar{X}}}{\lambda_0^{\sum_i X_i} e^{-n\lambda_0}} = \left(\frac{\bar{X}}{\lambda_0}\right)^{\sum_i X_i} e^{n(\lambda_0 - \bar{X})} > k$$

$$\Rightarrow \left(\sum_i X_i\right)\left(\log(\bar{X}) - \log(\lambda_0)\right) + n(\lambda_0 - \bar{X}) > \log(k)$$

(later, we'll talk about asymptotic approximations for getting the threshold)

$$\left(\sum_i X_i\right)\left(\log(\bar{X}) - \log(\lambda_0)\right) + n(\lambda_0 - \bar{X}) > \log(k)$$

$$\left(\sum_i X_i\right) \log\left(\frac{1}{n} \sum_i X_i\right) - \left(\sum_i X_i\right) \log(\lambda_0) - \left(\sum_i X_i\right) > \log(k) - n\lambda_0$$

Could be solved numerically, not sure about a closed form...

Linear regression with normal data

Suppose we observe $(X_1, Y_1), \ldots, (X_n, Y_n)$, where $Y_i = \beta^T X_i + \varepsilon_i$ and $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$. Partition $\beta = (\beta_{(1)}, \beta_{(2)})^T$. We wish to test $H_0 : \beta_{(2)} = 0$ vs. $H_A : \beta_{(2)} \neq 0$.

Full model (H_A): $Y_i = \beta^T X_i + \varepsilon_i$

Reduced model (H_0): $Y_i = \beta_{(1)}^T X_{i(1)} + \varepsilon_i$

length of $\beta_{(2)}$

Test statistic: $F = \dfrac{(SSE_{reduced} - SSE_{full})/q}{SSE_{full}/(n-p)}$ length of β

Under H_0: $F \sim F_{q, n-p}$

$SSE_{full} = \sum_i (Y_i - \hat{\beta}_{full}^T X_i)^2$ $\qquad$ $SSE_{reduced} = \sum_i (Y_i - \hat{\beta}_{(1)\,red}^T X_{i(1)})^2$

$\hat{\beta}_{full} = (X^T X)^{-1} X^T Y$ $\qquad$ $\hat{\beta}_{(1)\,red} = (X_{(1)}^T X_{(1)})^{-1} X_{(1)}^T Y$

MLE $\hat{\sigma}_{full} = \frac{1}{n} SSE_{full}$ $\qquad$ $\hat{\sigma}_{reduced} = \frac{1}{n} SSE_{reduced}$

LRT: rejects H_0 if $\dfrac{\sup_{\beta, \sigma^2} L(\beta, \sigma^2 \mid X, Y)}{\sup_{\substack{\beta, \sigma^2 \\ \beta_{(2)} = 0}} L(\beta, \sigma^2 \mid X, Y)} > k$

$$\iff \frac{L(\hat{\beta}_{full}, \hat{\sigma}^2_{full} \mid X, Y)}{L(\hat{\beta}_{red}, \hat{\sigma}^2_{red} \mid X, Y)} > k \iff \log L(\hat{\beta}_{full}, \hat{\sigma}^2_{full} \mid X, Y) - \log L(\hat{\beta}_{red}, \hat{\sigma}^2_{red} \mid X, Y) > \log(k)$$

$$\log L(\hat{\beta}, \hat{\sigma}^2 \mid X, Y) = \log\left(\left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2}}\right)^n \exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_i (Y_i - \hat{\beta}^T X_i)^2\right\}\right)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} SSE \quad (SSE = n\hat{\sigma}^2)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{n}{2}$$

So LRT rejects H_0 if $\frac{n}{2}\log(\hat{\sigma}^2_{reduced}) - \frac{n}{2}\log(\hat{\sigma}^2_{full}) > \log(k)$

$$\iff \frac{n}{2}\log\left(\frac{SSE_{red}}{SSE_{full}}\right) > \log(k)$$

$$\iff \frac{(SSE_{red} - SSE_{full})/q}{SSE_{full}/(n-p)} > \underbrace{\left(\exp\left(\frac{2\log k}{n}\right) - 1\right)\left(\frac{n-p}{q}\right)}_{\text{choose to be upper } \alpha \text{ quantile } F_{q,n-p}}$$

Asymptotics of the LRT