# Likelihood ratio tests

# Recap: likelihood ratio test

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

$$\frac{\sup\limits_{\theta \in \Theta_1} L(\theta | \mathbf{X})}{\sup\limits_{\theta \in \Theta_0} L(\theta | \mathbf{X})} > k,$$

where $k$ is chosen such that $\sup\limits_{\theta \in \Theta_0} \beta_{LR}(\theta) \leq \alpha$.

power of LRT

# Example: 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 $(H_A)$: $Y_i = \beta^T X_i + \varepsilon_i$

$SSE_{full} = \sum_{i=1}^{n} (Y_i - \hat{\beta}_{full}^T X_i)^2$ $\qquad \hat{\beta}_{full} = (X^T X)^{-1} X^T Y$

MLE for $\sigma^2$: $\hat{\sigma}^2_{full} = \frac{1}{n} SSE_{full}$

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

$SSE_{reduced} = \sum_{i=1}^{n} (Y_i - \hat{\beta}_{(1)\,reduced}^T X_{i(1)})^2$ $\qquad$ MLE for $\sigma^2$: $\frac{1}{n} SSE_{reduced}$

$\hat{\beta}_{(1)\,reduced} = (X_{(1)}^T X_{(1)})^{-1} X_{(1)}^T Y$

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

$q =$ # elements in $\beta_{(2)}$ $\qquad p =$ # elements in $\beta$

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

$$\iff \frac{L(\hat\beta_{full}, \hat\sigma^2_{full} \mid X,Y)}{L(\hat\beta_{reduced}, \hat\sigma^2_{reduced} \mid X,Y)} > K \iff \log L(\hat\beta_{full}, \hat\sigma^2_{full} \mid X,Y) - \log L(\hat\beta_{red.}, \hat\sigma^2_{red.} \mid X,Y) > \log(K)$$

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

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

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

$$\iff \frac{(SSE_{reduced} - SSE_{full})/q}{SSE_{full}/(n-p)} > \left(\exp\left(\frac{2\log K}{n}\right) - 1\right)\left(\frac{n-p}{q}\right)$$

$\sup_{\lambda \neq \lambda_0} L(\lambda|x) = \sup_{\lambda} L(\lambda|x)$

MLE: $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$

# Example: Poisson sample

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

> Write down the LRT statistic, and simplify as much as possible.

$$L(\lambda|x) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_i x_i} e^{-n\lambda}}{\prod_i x_i!}$$

$$\sup_{\lambda=\lambda_0} L(\lambda|x) = L(\lambda_0|x) \qquad \sup_{\lambda} L(\lambda|x) = L(\hat{\lambda}|x)$$

LRT rejects when
$$\frac{\hat{\lambda}^{\sum_i x_i} e^{-n\hat{\lambda}}}{\lambda_0^{\sum_i x_i} e^{-n\lambda_0}} > k$$

$$\left(\frac{\hat{\lambda}}{\lambda_0}\right)^{\sum_i x_i} e^{n(\lambda_0 - \hat{\lambda})} > k$$

$\hat{\theta} = \text{MLE}$

# Asymptotics of the LRT

Suppose we observe iid data $X_1, \ldots, X_n$ and want to test

$H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$ $(\theta \in \mathbb{R})$

Under $H_0$,

$$\underbrace{-2\left(\log L(\theta_0 \mid X) - \log L(\hat{\theta} \mid X)\right)}_{= 2 \log\left(\frac{L(\hat{\theta} \mid X)}{L(\theta_0 \mid X)}\right)} \xrightarrow{d} \chi^2_1$$

<u>Proof</u>: (let $\ell(\theta) = \log L(\theta \mid X)$)

① using Taylor expansion: $2\ell(\hat{\theta}) - 2\ell(\theta_0) \approx -\ell''(\hat{\theta})(\hat{\theta} - \theta_0)^2$

② $-\frac{1}{n}\ell''(\hat{\theta}) \xrightarrow{P} ?$, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} ?$

③ Apply Slutsky's & continuous mapping theorem

# Generalization to higher dimensions