We reject $H_0$, at the asymptotic significance level $\alpha$ when

$$-2\log\Lambda \geq q_{\chi^2(1)}(1-\alpha).$$

Since $I(\theta)=1$, $\chi^2_W = \{\sqrt{n}(\hat{\theta}-\theta_0)\}^2$.

Furthermore,

$$\frac{\partial \log f(x_i-\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\left[\log\frac{1}{2} - |x_i-\theta|\right] = \operatorname{sgn}(x_i-\theta)$$

Finally,

$$\chi^2_R = \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\operatorname{sgn}(X_i-\theta_0)\right\}^2.$$

## 5. Likelihood Ratio Tests: multidimensional case

Let $X_1,\dots,X_n$ be a sample from $f(x,\underline{\theta})$, $\underline{\theta}=(\theta_1,\dots,\theta_p)\in \Theta \subset R^p$, $p>1$

The likelihood function has a form

$$L(\underline{\theta}) = \prod_{i=1}^{n} f(x_i,\underline{\theta}),$$

the log-likelihood

$$\ell(\underline{\theta}) = \log L(\underline{\theta}) = \sum_{i=1}^{n}\log f(x_i,\underline{\theta})$$

Let $\underline{\theta}_0$ be ~~the~~ three value of the parameter $\theta$.

Impose (additional) regularity conditions

(R6) There is an open subset $\Theta_0 \subseteq \Theta$, such that $\underline{\theta}_0\in\Theta_0$ and all third partial derivatives of $f(x,\underline{\theta})$ exist for all $\underline{\theta}\in\Theta_0$.

(R7) $$E_{\underline{\theta}}\left[\frac{\partial}{\partial\theta_j}\log f(x,\underline{\theta})\right]=0 \quad \text{for } j=1,\dots,p.$$

$$I_{jk}(\underline{\theta}) = \operatorname{Cov}\left(\frac{\partial\log f(X,\underline{\theta})}{\partial\theta_j}, \frac{\partial\log f(X,\underline{\theta})}{\partial\theta_k}\right) =$$

$$= -E_{\underline{\theta}}\left[-\frac{\partial^2\log f(X,\underline{\theta})}{\partial\theta_j\partial\theta_k}\right] \quad \text{for } j,k=1,\dots,p.$$

(R8) For all $\theta \in \Theta_0$,

$$I(\underline{\theta}) = \left[ I_{jk}(\underline{\theta}) \right]_{j,k=1}^{p}$$

is positive definite.

(R9) There exist ~~functio~~ functions $M_{jkl}(x)$, such that

$$\left| \frac{\partial^3 \log f(x, \underline{\theta})}{\partial \theta_j \, \partial \theta_k \, \partial \theta_l} \right| \leqslant M_{jkl}(x) \quad \text{for all } \underline{\theta} \in \Theta_0$$

and

$$\mathbb{E}_{\underline{\theta}_0} \left[ M_{jkl}(X) \right] < +\infty \quad \text{for all } j,k,l = 1,\dots,p.$$

Definition 1

The quantity

$$\hat{\underline{\theta}} = \underset{\underline{\theta} \in \Theta}{\operatorname{argmax}} \, L(\underline{\theta})$$

is called the maximum likelihood estimator of the parameter $\underline{\theta}$.

Remark 1

If $\hat{\underline{\theta}}$ is the MLE of $\underline{\theta}$, then $g(\hat{\underline{\theta}})$ is the MLE of $g(\underline{\theta})$

Example 1

$X_1, \dots, X_n$ i.i.d $X_1 \sim N(\mu, \sigma^2)$. Let $\underline{\theta} = (\mu, \sigma^2)$

and $\Theta = \mathbb{R} \times (0, +\infty)$. We have

$$L(\underline{\theta}) = L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{ -\frac{1}{2} \frac{(x_i - \mu)^2}{\sigma^2} \right\}$$

$$= \left( \frac{1}{2\pi\sigma^2} \right)^{n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\},$$

$$l(\underline{\theta}) = \log L(\underline{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Moreover,

$$\frac{\partial \ell(\mu,\sigma^2)}{\partial \mu} = \frac{1}{2\sigma^2}\, 2 \sum_{i=1}^{n}(x_i-\mu) = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} = 0$$

$$\frac{\partial \ell(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$$

$$\Leftrightarrow \; -\frac{n}{2\sigma^4}\left[\sigma^2 - \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right] = 0$$

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$$

$$\frac{\partial^2 \ell(\mu,\sigma^2)}{\partial \mu^2} = -\frac{n}{\sigma^2}\bigg|_{\sigma=\hat{\sigma}} = -\frac{n}{\hat{\sigma}^2}$$

$$\frac{\partial^2 \ell(\mu,\sigma^2)}{\partial \mu\, \partial \sigma^2} = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i-\mu) = \frac{\partial^2 \ell(\mu,\sigma^2)}{\partial \sigma^2\, \partial \mu}\bigg|_{\substack{\mu=\hat{\mu}\\ \sigma^2=\hat{\sigma}^2}} = 0$$

$$\frac{\partial^2 \ell(\mu,\sigma^2)}{\partial^2 \sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\mu)^2\bigg|_{\substack{\mu=\hat{\mu}\\ \sigma^2=\hat{\sigma}^2}} =$$

$$= \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\, n\hat{\sigma}^2 = \frac{n}{2\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^4} = -\frac{n}{2\hat{\sigma}^4}.$$

To sum up,

$$\frac{\partial^2 \ell(\mu,\sigma^2)}{\partial \cdot\, \partial \cdot} = \begin{bmatrix} -\frac{n}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{n}{2\hat{\sigma}^4} \end{bmatrix} \qquad -\frac{n}{\hat{\sigma}^2} < 0 \;\&\; \det|\;| > 0, \quad \det|\;| = \frac{n^2}{2\hat{\sigma}^6}$$

Indeed, $\cancel{\hat{\mu}=\bar{X}}$

$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}^2) = \left(\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right)$$

is the MLE of $\underline{\theta} = (\mu, \sigma^2)$.

Definition 2.

Let $\nabla \log f(X, \underline{\theta}) = \left( \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_1}, \dots, \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_p} \right)'$.

The matrix of the form

$$\underline{I}(\underline{\theta}) = \mathrm{Cov}\left( \nabla \log f(X, \underline{\theta}) \right)$$

is called the Fisher information matrix for the parameter $\underline{\theta}$. Obviously, $\underline{I}(\underline{\theta}) = [I_{j,k}]_{j,k=1}^{p}$, where

$$I_{j,k} = \mathrm{cov}\left( \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_j}, \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_k} \right), \quad j,k = 1, \dots, p.$$

Under the assumptions (R0 - R9),

$$0 = \int_{\mathbb{R}} \frac{\partial}{\partial \theta_j} f(x, \underline{\theta})\, dx = \int_{\mathbb{R}} \left[ \frac{\partial}{\partial \theta_j} \log f(x, \underline{\theta}) \right] f(x, \underline{\theta})\, dx =$$

$$= \mathbb{E}\left[ \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_j} \right] \Big/ \frac{\partial}{\partial \theta_k}$$

$$0 = \int_{\mathbb{R}} \left[ \frac{\partial^2 \log f(x, \underline{\theta})}{\partial \theta_j \partial \theta_k} \right] f(x, \underline{\theta})\, dx + \int_{\mathbb{R}} \frac{\partial}{\partial \theta_j} \log f(x, \underline{\theta}) \cdot \frac{\partial \log f(x, \underline{\theta})}{\partial \theta_k} f(x, \underline{\theta})\, dx.$$

Thus

$$\mathbb{E}\left[ \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_j} \frac{\partial \log f(X, \underline{\theta})}{\partial \theta_k} \right] = - \mathbb{E}\left[ \frac{\partial^2 \log f(X, \underline{\theta})}{\partial \theta_j \partial \theta_k} \right] = I_{jk}.$$

Since $X_1, \dots, X_n$ are independent

$$\mathrm{Cov}\left( \nabla \log \overset{L(\underline{\theta}, \underline{X})}{\cancel{f(X, \underline{\theta})}} \right) = \mathrm{Cov}\left( \sum_{i=1}^{n} \nabla \log f(X_i, \underline{\theta}) \right) = \sum_{i=1}^{n} \mathrm{Cov}\left( \nabla \log f(X_i, \underline{\theta}) \right) = n \underline{I}(\underline{\theta}).$$

Remark 2.

If $Y = u(X_1, \dots, X_n)$ is an unbiased estimate of the parameter $\theta_j$,

$$\mathrm{Var}(Y) \geq \frac{1}{n} \left[ \underline{I}^{-1}(\underline{\theta}) \right]_{jj}.$$

Example 2 (Fisher information matrix under normality)

We have $f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$, $\mu \in \mathbb{R}$, $\sigma > 0$.

Let $\underline{\theta} = (\mu, \sigma)$. We have

$$\log f(x,\mu,\sigma) = -\frac{1}{2}\log 2\pi - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \log f(x,\mu,\sigma)}{\partial\mu} = \frac{x-\mu}{\sigma^2}$$

$$\frac{\partial \log f(x,\mu,\sigma)}{\partial\sigma} = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}$$

$$\frac{\partial^2 \log f(x,\mu,\sigma)}{\partial\mu^2} = -\frac{1}{\sigma^2}$$

$$\frac{\partial^2 \log f(x,\mu,\sigma)}{\partial\sigma^2} = \frac{1}{\sigma^2} - \frac{3(x-\mu)^2}{\sigma^4}$$

$$\frac{\partial^2 \log f(x,\mu,\sigma)}{\partial\mu\,\partial\sigma} = -\frac{2(x-\mu)}{\sigma^3}$$

$$I_{11}(\mu,\sigma) = \frac{1}{\sigma^2}$$

$$I_{12}(\mu,\sigma) = -\mathbb{E}\left[-\frac{2(X-\mu)}{\sigma^3}\right] = 0 = I_{21}(\mu,\sigma)$$

$$I_{22}(\mu,\sigma) = -\mathbb{E}\left[\frac{1}{\sigma^2} - \frac{3(X-\mu)^2}{\sigma^4}\right] = -\frac{1}{\sigma^2} + \frac{3}{\sigma^4}\mathbb{E}[X-\mu]^2 = \frac{2}{\sigma^2}$$

To sum up,

$$\underline{I}(\mu,\sigma) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}.$$

## Theorem 1

Let $X_1, \dots, X_n$ be a sample with $f(x, \underline{\theta}), \underline{\theta} \in \Theta$.
Impose the regularity conditions (R0) - (R9). We have,

(i) the likelihood equation

$$\frac{\partial l(\underline{\theta})}{\partial \underline{\theta}} = \underline{0}$$

has a solution $\hat{\underline{\theta}}_n$ such that $\hat{\underline{\theta}}_n \xrightarrow{P} \underline{\theta}$,

(ii) for any sequence $\{\hat{\underline{\theta}}_n\}_{n \in \mathbb{N}}$ satisfying (i),

$$\sqrt{n}\,(\hat{\underline{\theta}}_n - \underline{\theta}) \xrightarrow{\mathcal{D}} N_p(\underline{0}, \underline{I}^{-1}(\underline{\theta})).$$

## Corollary 1

Under the assumptions of Theorem 1. Let $\{\hat{\underline{\theta}}_n\}_{n \in \mathbb{N}}$ be a sequence of consistent solutions of the equation (i). Then, $\hat{\underline{\theta}}_n$ is asymptotically efficient, i.e., for $j = 1, \dots, p$,

$$\sqrt{n}\,(\hat{\theta}_{n,j} - \theta_j) \xrightarrow{\mathcal{D}} N(0, I_{jj}^{-1}(\underline{\theta})).$$

Let $g : \Theta \to \mathbb{R}^k$, $1 \leq k \leq p$ be a function such that the $k \times p$ matrix of partial derivatives

$$\underline{B} = \left[\frac{\partial g_i}{\partial \theta_j}\right], \quad i = 1, \dots, k; \; j = 1, \dots, p$$

has continuous elements ~~which~~ and does not vanish in a neighborhood of $\theta$. Let $\hat{\underline{\eta}} = g(\hat{\underline{\theta}})$. Then $\hat{\underline{\eta}}$ is the MLE og $\underline{\eta} = g(\underline{\theta})$.

Furthermore,

$$\sqrt{n}\,(\hat{\underline{\eta}} - \underline{\eta}) \xrightarrow{\mathcal{D}} N_k(\underline{0}, \underline{B}\,\underline{I}^{-1}(\underline{\theta})\,\underline{B}').$$

Hence,

$$\underline{I}(\underline{\eta}) = (\underline{B}\,\underline{I}^{-1}(\underline{\theta})\,\underline{B}')^{-1}$$

is the Fisher information matrix for $\underline{\eta}$ provided that it exists.