Chapter 10 Asymptotic Evaluations

- **Asymptotic evaluation of point estimators**
- **Asymptotic hypothesis testing**
- **Large sample confidence intervals**

Ch 10. Asymptotic Evaluation

Definition

Let $\{T_n = T_n(X_1, \cdots, X_n)\}$ be a sequence of estimators of a parametric function $\tau(\theta)$.

i) $\{T_n\}$ is said to be *consistent* estimator of $\tau(\theta)$ if for all $\epsilon > 0$,

$$P_\theta\{|T_n - \tau(\theta)| > \epsilon\} \to 0$$

as $n \to \infty$, for all $\theta \in \Theta$. $[T_n \xrightarrow{P} \tau(\theta)]$.

ii) $\{T_n\}$ is said to be *MSE consistent(MSEC)* to $\tau(\theta)$ if

$$E_\theta\{[T_n - \tau(\theta)]^2\} \to 0$$

as $n \to \infty$, for all $\theta \in \Theta$.

$\triangleleft$ Note: MSEC implies consistency

Definition (- Continued)

iii) $\{T_n\}$ is said to be strong consistent to $\tau(\theta)$ if for all $\epsilon > 0$,

$$T_n \overset{as}{\rightarrow} \tau(\theta)$$

as $n \rightarrow \infty$, for all $\theta \in \Theta$

iv) $\{T_n\}$ is said to be *Asymptotically Unbiased (AU)* to $\tau(\theta)$ if

$$E_\theta\{T_n\} \rightarrow \tau(\theta)$$

as $n \rightarrow \infty$, for all $\theta \in \Theta$.

Ch 10. Asymptotic Evaluation

Definition

Relative efficiency of an unbiased estimator T of $\tau(\theta)$ to another unbiased estimator T^* is given by

$$RE(T, T^*) = \frac{Var(T^*)}{Var(T)}.$$

Definition The efficiency of an unbiased estimator T of $\tau(\theta)$

is given by

$$\text{eff(T)=} \frac{[\tau'(\theta)]^2/nI_1(\theta)}{Var_\theta(T)}$$

The unbiased estimator T is efficient to $\tau(\theta)$ if eff(T)=1.

Asymptotic Efficiency

Definition: For an estimator W_n, if $k_n\left(W_n - \tau(\theta)\right) \to N\left(0, \sigma^2\right)$ in distribution, then σ^2 is called the asymptotic variance of the limiting distribution of W_n.

Definition: W_n is *asymptotically efficient* for $\tau(\theta)$ if

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \to N\left(0, v(\theta)\right)$$

with

$$v(\theta) = \frac{\left[\tau'(\theta)\right]^2}{E_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f\left(x|\theta\right)\right\}^2\right]} = \frac{[\tau'(\theta)]^2}{I_1(\theta)}.$$

that is the asymptotic variance of W_n achieves the Cramer-Rao Lower Bound.

Ch 10. Asymptotic Evaluation

Theorem (Consistency of MLE)

Let $X_1, \cdots, X_n \overset{iid}{\sim} f(x:\theta)$. Let $\hat{\theta}_n$ be the MLE of θ. Let $\tau(\theta)$ be a continuous parametric function of θ. Under regularity conditions (10.6.2), $\tau(\hat{\theta}_n)$ is a consistent estimator of $\tau(\theta)$.

Theorem (Asymptotic efficiency of MLE)

$X_1, \cdots, X_n \overset{iid}{\sim} f(x:\theta)$. Let $\hat{\theta}_n$ be the MLE of θ. Let $\tau(\theta)$ be a parametric function of θ. Under the regularity conditions, if $\tau'(\theta) \neq 0$, then $\tau(\hat{\theta}_n)$ is asymptotically efficient and

$$\sqrt{n}\left[\tau(\hat{\theta}_n) - \tau(\theta)\right] \xrightarrow{D} N\left(0,\ \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right)$$

$\lhd$ Note: It is often said that $\tau(\hat{\theta}_n)$ has an approximate normal distribution with mean $\tau(\theta)$ and variance $[\tau'(\theta)]^2/nI_1(\theta)$

Assumptions (Regularity Conditions)

(R0): The pdf is distinct, i.e. $\theta \neq \theta' \Rightarrow P\{f(X;\ \theta) \neq f(X;\ \theta')\} > 0$

(R1): The pdfs have common support S for all θ.

(R2): The point θ_0 is an interior point in Ω.

(R3): The pdf is twice differentiable as a function of θ.

(R4): The integral $\int f(x;\theta)dx$ can be differentiated twice under the integral sign as a function of θ,

$$\frac{\partial}{\partial\theta}\int f(x;\theta)dx = \int \frac{\partial f(x;\theta)}{\partial\theta}dx, \quad \frac{\partial^2}{\partial\theta^2}\int f\,dx = \int \frac{\partial^2 f}{\partial\theta^2}dx,$$

Proof (Consistency): $\forall\, \theta^* \neq \theta,$

By Jensen inequality and distinctness of the pdf

$$E_\theta\left[\ln \frac{f(X;\theta^*)}{f(X;\theta)}\right] < \ln\left[E_\theta \frac{f(X;\theta^*)}{f(X;\theta)}\right] = 0$$

$$\forall\, \delta > 0, \textit{such that } (\theta_0 - \delta, \theta_0 + \delta,) \subset \Omega,$$

$$E_{\theta_0}\left[\ln \frac{f(X;\theta_0 - \delta)}{f(X;\theta_0)}\right] < 0, \qquad E_{\theta_0}\left[\ln \frac{f(X;\theta_0 + \delta)}{f(X;\theta_0)}\right] < 0,$$

By Strong LLN (Law of Large Number) , n→∞,

$$\frac{1}{n}\left[\ln L(\theta_0 - \delta; x) - \ln L(\theta_0; x)\right] = \frac{1}{n}\sum_{i=1}^{n} \ln\left[\frac{f(x_i;\theta_0 - \delta)}{f(x_i;\theta_0)}\right]$$

$$\to E_{\theta_0}\left[\ln \frac{f(X;\theta_0-\delta)}{f(X;\theta_0)}\right] < 0, \quad a.s.$$

$$\frac{1}{n}\left[\ln L(\theta_0+\delta; x) - \ln L(\theta_0; x)\right] \to E_{\theta_0}\left[\ln \frac{f(X;\theta_0+\delta)}{f(X;\theta_0)}\right] < 0, a.s.$$

$\ln L(\theta, x)$ is continuous and differentiable as function of θ. Then there exists

$$\hat{\theta} \in (\theta_0-\delta, \theta_0+\delta), \textit{ such that } \quad \frac{\partial \ln L(\theta)}{\partial \theta} = 0, \, a.s.$$

$$P(|\hat{\theta}-\theta_0| < \delta) = 1, n \to \infty \qquad \hat{\theta} \xrightarrow{p} \theta_0$$

For condition (R1), the range of X cannot depend on θ.

Thus, for example, we cannot use this theorem to show that the MLE for θ for the $Uniform(0,\theta)$ distribution is consistent.

However, we can still show that $Y = X_{(n)}$ is consistent for θ.

We have that $EY = \int_0^\theta \frac{ny^n}{\theta^n} dy = \frac{n}{n+1}\theta \to \theta$ as $n \to \infty$.

Also $EY^2 = \int_0^\theta \frac{ny^{n+1}}{\theta^n} dy = \frac{n}{n+2}\theta^2$ and

$$VarY = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2 \theta^2$$

$$= \frac{n}{(n+2)(n+1)^2}\theta^2.$$

Thus $Bias\ X_{(n)} \to 0$ and $Var\ X_{(n)} \to 0$, which implies that $X_{(n)}$ is consistent for θ.

Proof (Asymptotic Efficiency): Let $l(\theta) = \ln L(\theta)$

$$0 = \frac{\partial l(\hat{\theta})}{\partial \theta} = \frac{\partial l(\theta_0)}{\partial \theta} + \frac{\partial^2 l(\theta_0)}{\partial \theta^2}(\hat{\theta} - \theta_0) + \frac{\partial^3 l(\theta_1)}{\partial \theta^3}\frac{(\hat{\theta} - \theta_0)^2}{2}$$

With θ_1 between θ_0 and $\hat{\theta}$, $and\ \theta_1 \xrightarrow{p} \theta.$

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \frac{\frac{1}{\sqrt{n}}\frac{\partial l(\theta_0)}{\partial \theta}}{-\frac{1}{n}\frac{\partial^2 l(\theta_0)}{\partial \theta^2} - \frac{1}{n}\frac{\partial^3 l(\theta_1)}{\partial \theta^3}\frac{(\hat{\theta} - \theta_0)}{2}}$$

ξ_n η_{1n} η_{2n}

$$E\left[\frac{\partial \ln f(X,\theta_0)}{\partial \theta}\right]=0,$$

$$\operatorname{var}\left[\frac{\partial \ln f(X,\theta_0)}{\partial \theta}\right]=E\left[\frac{-\partial^2 \ln f(X,\theta_0)}{\partial \theta^2}\right]=I(\theta_0)$$

By CLT (Central Limit Theorem) , when n$\rightarrow\infty$,

$$\frac{1}{\sqrt{nI(\theta_0)}}\frac{\partial l(\theta_0)}{\partial \theta}=\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln f(X_i,\theta)}{\partial \theta}}{\sqrt{I(\theta_0)/n}}\xrightarrow{d} N(0,\ 1)$$

$$\xi_n=\frac{1}{\sqrt{n}}\frac{\partial l(\theta_0)}{\partial \theta}\xrightarrow{d} N(0,\ I(\theta_0))$$

$$\eta_{1n} = -\frac{1}{n}\frac{\partial^2 l(\theta_0)}{\partial\theta^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{-\partial^2 \ln f(X_i,\theta_0)}{\partial\theta^2}$$

$$\xrightarrow{p} E\left[\frac{-\partial^2 \ln f(X,\theta_0)}{\partial\theta^2}\right] = I(\theta_0).$$

$$\left|\frac{1}{n}\frac{\partial^3 l(\theta_1)}{\partial\theta^3}\right| < \frac{1}{n}\sum_{i=1}^{n} H(X_i) \xrightarrow{p} E[H(X)] < \infty$$

$$\eta_{2n} = -\frac{1}{n}\frac{\partial^3 l(\theta_1)}{\partial\theta^3}\frac{(\hat{\theta}-\theta_0)}{2} \xrightarrow{p} 0$$

$$\eta_n = \eta_{1n} + \eta_{2n} \xrightarrow{p} I(\theta_0)$$

By Slutsky's Theorem , when n $\to\infty$,

$$\frac{\xi_n}{\eta_n / I(\theta_0)} \xrightarrow{p} N(0, I(\theta_0))$$

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) = \frac{\xi_n}{\eta_n} = \frac{1}{I(\theta_0)} \cdot \frac{\xi_n}{\eta_n / I(\theta_0)} \xrightarrow{d} N(0,\ I^{-1}(\theta_0))$$

$$\boxed{\hat{\theta} \dot{\sim} N(\theta_0, \frac{1}{nI(\theta_0)}) \hat{=} N(\theta_0, I_n^{-1}(\theta_0)), n \text{ is large}}$$

Theorem imply that under regularity conditions, MLEs $\widehat{\theta}_n$ and $g(\widehat{\theta}_n)$ for θ and $g(\theta)$ are asymptotically efficient, i.e.

$$\sqrt{n}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \frac{1}{I_1(\theta)}).$$

$$\sqrt{n}(g(\widehat{\theta}_n) - g(\theta)) \xrightarrow{\mathcal{D}} N(0, \frac{(g'(\theta))^2}{I_1(\theta)}).$$

Example $X_1, \ldots, X_n$ are i.i.d. Bernoulli(θ) random variables.

$$I_1(\theta) = \frac{1}{\theta(1-\theta)}$$

Theorem tells us that

$$\sqrt{n}(\bar{X} - \theta) \xrightarrow{\mathcal{D}} N(0, \theta(1-\theta)).$$

Suppose we wish to estimate $\tau(\theta) = \theta(1-\theta)$.

$$\sqrt{n}(\bar{X}(1-\bar{X}) - \tau(\theta)) \xrightarrow{\mathcal{D}} N(0, (1-2\theta)^2\theta(1-\theta)).$$

Definition: Let two estimators satisfy

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \to N\left(0, \sigma_W^2\right);$$

$$\sqrt{n}\left(V_n - \tau(\theta)\right) \to N\left(0, \sigma_V^2\right).$$

The a*symptotic relative efficiency* (ARE) of V_n with respect to W_n is

$$ARE\left(V_n, W_n\right) = \frac{\sigma_W^2}{\sigma_V^2}.$$

Often we talk about asymptotic relative efficiency with respect to the MLE $\hat{\tau}$. Since the MLE is asymptotically efficient, the ARE will be less than one.

Example

Let $X_1, \ldots X_n$ be i.i.d. $Gamma(\alpha, 1)$.

$$f(x|\alpha) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x},\ x \geq 0.$$

$$\log f(x|\alpha) = -\log\Gamma(\alpha) + (\alpha - 1)\log x - x.$$

So

$$\log f(\mathbf{x}|\alpha) = \sum_{i=1}^{n} \left[-\log\Gamma(\alpha) + (\alpha - 1)\log x_i - x_i\right]$$
$$= -n\log\Gamma(\alpha) + (\alpha - 1)\sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n} x_i.$$

So the MLE is the solution of

$$\frac{\partial}{\partial\alpha}\log\Gamma(\hat{\alpha}) = \frac{\sum_{i=1}^{n}\log x_i}{n}.$$

There is no closed form solution for $\hat{\alpha}$. We can obtain the MLE numerically. However, to conduct inference, we need the distribution of $\hat{\alpha}$.

We can obtain the asymptotic variance of the MLE however, and compare that variance with that of an estimator than is more easily computed (such as the method of moments estimator $\bar{X}$).

This is an exponential family, so there are no problems as far as satisfying the conditions for the asymptotic variance of the MLE to meet the Cramer-Rao Lower Bound.

We have that $\sqrt{n}\left(\hat{\alpha}-\alpha\right) \to N\left(0, \frac{1}{I\left(\alpha\right)}\right)$.

$$\log f\left(x|\alpha\right)=-\log\Gamma\left(\alpha\right)+\left(\alpha-1\right)\log x-x$$

$$\frac{\partial^2}{\partial\alpha^2}\log f\left(x|\alpha\right)=-\frac{\partial^2}{\partial\alpha^2}\log\Gamma\left(\alpha\right)=I\left(\alpha\right).$$

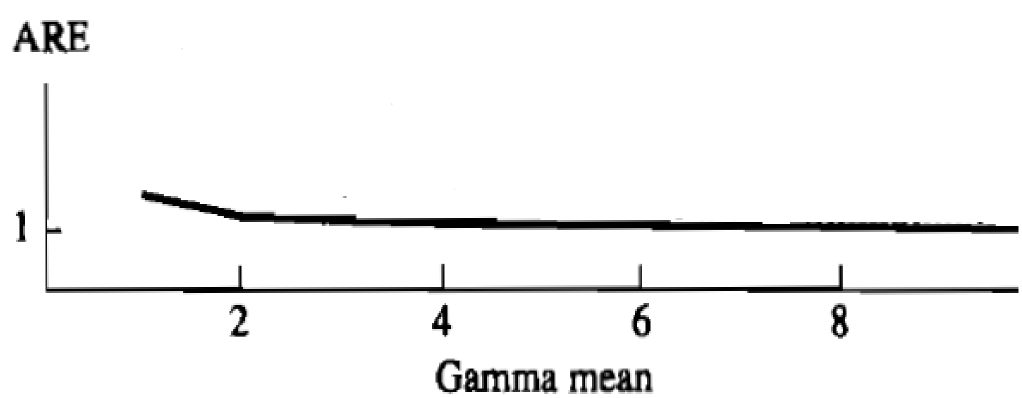
The method of moments estimator $\bar{X}$

$Var\left(\bar{X}\right)=\dfrac{\alpha}{n}$, and by the Central Limit Theorem,

$$\sqrt{n}\left(\bar{X}-\alpha\right)\to N\left(0,\alpha\right).$$

The ARE is then

$$\frac{-\dfrac{\partial^2}{\partial\alpha^2}\log\Gamma\left(\alpha\right)}{\alpha}.$$



ARE for MLE vs MME

Asymptotic hypothesis testing

We consider methods for deriving approximate large sample tests for situations when no optimal test exists.

Likelihood ratio tests

We can always compute the likelihood ratio though the MLE (restricted and-or unrestricted) may need to be obtained numerically.

But how do we determine the constant c to get a level α test?

Theorem

$X_1, \cdots, X_n \overset{iid}{\sim} f(x|\theta)$. $\lambda(\mathbf{x})$ *is a likelihood ratio for testing*

$$H_0 : \theta \in \Theta_0 \quad vs \quad H_1 : \theta \in \Theta_0^c$$

Then under the regularity conditions (CRLB) on $f(x|\theta)$ *and* H_0

$$-2\ln[\lambda(\mathbf{x})] \xrightarrow{D} \chi_k^2,$$

k =r-m

where $k =\#$ *of free parameters for* $\boldsymbol{\theta} \in \Theta$ *-* $\#$ *of free parameters for* $\boldsymbol{\theta} \in \Theta_0$*. This yields the approximate size* α *test*

$$\phi(\mathbf{x}) = \begin{cases} 1, & -2\ln[\lambda(\mathbf{x})] > \chi^2_{1-\alpha,k}, \\ \gamma, & -2\ln[\lambda(\mathbf{x})] = \chi^2_{1-\alpha,k}, \\ 0, & -2\ln[\lambda(\mathbf{x})] < \chi^2_{1-\alpha,k}. \end{cases}$$

Example

Let $X_1, \ldots, X_n$ be i.i.d. Poisson(λ). For testing $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$, we have

$$-2\log\lambda(\mathbf{x}) = -2\log\frac{e^{-n\lambda_0}\lambda_0^{\sum x_i}}{e^{-n\hat{\lambda}}\hat{\lambda}^{\sum x_i}}$$

$$= -2n\left(\lambda_0 - \hat{\lambda}\right) - \sum x_i \log\left(\lambda_0 / \hat{\lambda}\right)$$

$$= -2n\left[\left(\lambda_0 - \hat{\lambda}\right) - \hat{\lambda}\log\left(\lambda_0 / \hat{\lambda}\right)\right]$$

where $\hat{\lambda} = \overline{x}$ is the MLE of λ.

We would then reject H_0 at level α if $-2\log\lambda(\mathbf{x}) > \chi^2_{1,\alpha}$.

To get an idea of the accuracy of the approximation to the distribution of the test statistic $-2\log\lambda(\mathbf{x})$, the text presents a small simulation of the statistic.

For $\lambda_0 = 5$ and $n = 25$, 10,000 values of the test statistic were obtained. A comparison of the simulated (exact) and χ_1^2 cutoff points are given in the following table.

Percentile	.80	.90	.95	.99
Simulated	1.630	2.726	3.744	6.304
χ^2	1.642	2.706	3.841	6.635

- Example: $n = 60$.

A_i	1	2	3	4	5	6
Times	13	19	11	8	5	4

$H_0 : P\{A_i\} = 1/6.$

$p_{i0} = \frac{1}{6} \times 1 = 1/6,$

$np_{i0} = \frac{1}{6} \times 60 = 10.$

Example: *Goodness-of-fit test for a multinomial distribution*

Let $(X_1, \ldots, X_k)$ have a multinomial distribution based on n trials and parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$.

$$\Theta = \{\boldsymbol{\theta} : 0 < \theta_i,\ i = 1, \ldots, k; \sum_{i=1}^{k} \theta_i = 1\}$$

Want to test

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \qquad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0.$$

The distribution of $X_1, \ldots, X_k$ is

$$f(x_1, \ldots, x_k|\boldsymbol{\theta}) = \frac{n!}{x_1! \cdots x_k!} \theta_1^{x_1} \cdots \theta_k^{x_k}.$$

Under H_0,

$$\begin{aligned} f(x_1, \ldots, x_k|\boldsymbol{\theta}_0) &= \frac{n!}{x_1! \cdots x_k!} \theta_{10}^{x_1} \cdots \theta_{k0}^{x_k} \\ &= L(\Theta_0). \end{aligned}$$

The MLEs are

$\widehat{\theta}_i = X_i/n$, $i = 1, \ldots, k$. Therefore,

$$L(\Theta) = \frac{n!}{x_1! \cdots x_k!} \widehat{\theta}_1^{x_1} \cdots \widehat{\theta}_k^{x_k},$$

and

$$\lambda(x_1, \ldots, x_k) = \left(\frac{\widehat{\theta}_1}{\theta_{10}}\right)^{x_1} \cdots \left(\frac{\widehat{\theta}_k}{\theta_{k0}}\right)^{x_k}.$$

$$\lambda(x_1,\ldots,x_k)=\prod_{i=1}^{k}\left(\frac{x_i}{n\theta_{i0}}\right)^{x_i}$$

We reject H_0 when $\lambda(x_1,\ldots,x_k)>c$, or equivalently when $2\log\lambda(x_1,\ldots,x_k)>c_1$.

$$2\log\lambda(x_1,\ldots,x_k)=2\sum_{i=1}^{k}x_i\log\left(\frac{x_i}{n\theta_{i0}}\right)$$

In this case we have $m=0$ and $r=k-1$. Why?

When H_0 is true, the distribution of

$$2\log\lambda(X_1,\ldots,X_k)$$

is approximately χ^2_{k-1}. An approximate size α test is to reject H_0 iff

$$2\log\lambda(x_1,\ldots,x_k)\geq\chi^2_{k-1,1-\alpha}.$$

We may use this result to see a justification for Pearson's χ^2 goodness-of-fit test, which says to reject H_0 when

$$\sum_{i=1}^{k} \frac{(n\widehat{\theta}_i - n\theta_{i0})^2}{n\theta_{i0}} \geq \chi^2_{k-1,1-\alpha}.$$

We have

$$\begin{aligned} \log\left(\frac{x_i}{n\theta_{i0}}\right) &= \log 1 + \left(\frac{x_i}{n\theta_{i0}} - 1\right) \\ &\quad - \frac{1}{2}\left(\frac{x_i}{n\theta_{i0}} - 1\right)^2 \frac{1}{\eta_i^2}, \end{aligned}$$

where η_i is between $x_i/(n\theta_{i0})$ and 1.

Therefore,

$$\begin{aligned} 2\sum_{i=1}^{k} x_i \log\left(\frac{x_i}{n\theta_{i0}}\right) &= 2\sum_{i=1}^{k} x_i \left(\frac{\widehat{\theta}_i - \theta_{i0}}{\theta_{i0}}\right) \\ &\quad - \sum_{i=1}^{k} x_i \left(\frac{\widehat{\theta}_i}{\theta_{i0}} - 1\right)^2 \frac{1}{\eta_i^2}. \end{aligned}$$

The last expression is

$$2n\sum_{i=1}^{k}\left(\frac{\widehat{\theta}_i^2}{\theta_{i0}}-\widehat{\theta}_i\right)-n\sum_{i=1}^{k}\widehat{\theta}_i\left(\frac{\widehat{\theta}_i}{\theta_{i0}}-1\right)^2\frac{1}{\eta_i^2}=$$

$$2n\sum_{i=1}^{k}\frac{(\widehat{\theta}_i-\theta_{i0})^2}{\theta_{i0}}-n\sum_{i=1}^{k}\frac{\widehat{\theta}_i}{\theta_{i0}}\frac{1}{\eta_i^2}\frac{(\widehat{\theta}_i-\theta_{i0})^2}{\theta_{i0}}=$$

$$2n\sum_{i=1}^{k}\frac{(\widehat{\theta}_i-\theta_{i0})^2}{\theta_{i0}}-n\sum_{i=1}^{k}\frac{(\widehat{\theta}_i-\theta_{i0})^2}{\theta_{i0}}$$

$$+n\sum_{i=1}^{k}\left[1-\frac{\widehat{\theta}_i}{\theta_{i0}}\frac{1}{\eta_i^2}\right]\frac{(\widehat{\theta}_i-\theta_{i0})^2}{\theta_{i0}}=$$

$$\sum_{i=1}^{k}\frac{(n\widehat{\theta}_i-n\theta_{i0})^2}{n\theta_{i0}}+R_n.$$

The random variable R_n converges to 0 in probability when H_0 is true, and it follows that

$$\sum_{i=1}^{k} \frac{(n\widehat{\theta}_i - n\theta_{i0})^2}{n\theta_{i0}}$$

converges in distribution to χ^2_{k-1} under H_0.

- Example: $n = 60$.

A_i	1	2	3	4	5	6
Times	13	19	11	8	5	4

$$2\sum_{i=1}^{k} x_i \log\left(\frac{x_i}{n\theta_{i0}}\right)$$

=2*(13*ln 13/10+19*ln 19/10+11*ln 11/10+8*ln 8/10+5*ln 5/10+4*ln 4/10)=15.476

$$\sum_{i=1}^{k} \frac{(n\widehat{\theta}_i - n\theta_{i0})^2}{n\theta_{i0}} = \frac{(13-10)^2}{10} + \frac{(19-10)^2}{10} + \frac{(11-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(4-10)^2}{10} = \frac{91+65}{10} = 15.6 > 11.1$$

Example

Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, a\theta)$ and consider testing $H_0 : a = 1$ against $H_1 : a \neq 1$.

Then

$$f(\mathbf{x} | a, \theta) = (2\pi a\theta)^{-n/2} \exp\left[-\frac{\sum (x_i - \theta)^2}{2a\theta} \right].$$

So $$l(a, \theta | \mathbf{x}) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log a$$

$$-\frac{n}{2}\log\theta - \frac{\sum (x_i - \theta)^2}{2a\theta}.$$

$$\frac{\partial l}{\partial a} = -\frac{n}{2a} + \frac{\sum (x_i - \theta)^2}{2a^2\theta} = 0$$

$$\hat{a} = \frac{\sum\left(x_i - \hat{\theta}\right)^2}{n\hat{\theta}}.$$

$$\frac{\partial l}{\partial \theta} = -\frac{n}{2\theta} + \frac{\sum\left(x_i - \theta\right)^2}{2a\theta^2} + \frac{\sum\left(x_i - \theta\right)}{a\theta} = 0$$

Plugging in for a in the last equation, we have

$$-\frac{n}{2\theta} + \frac{n\theta}{2\theta^2} + \frac{n\theta\sum\left(x_i - \theta\right)}{\theta\sum\left(x_i - \theta\right)^2} = 0$$

Thus the unrestricted MLEs are $\hat{\theta} = \overline{x}$ and $\hat{a} = \frac{\sum\left(x_i - \overline{x}\right)^2}{n\overline{x}}$.

Under the null hypothesis, $a = 1$, and

$$f\left(\mathbf{x}|\theta\right) = \left(2\pi\theta\right)^{-n/2} \exp\left[-\frac{\sum\left(x_i - \theta\right)^2}{2\theta}\right].$$

So

$$l\left(\theta|\mathbf{x}\right) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\theta - \frac{\sum\left(x_i - \theta\right)^2}{2\theta}.$$

$$\frac{\partial l}{\partial\theta} = -\frac{n}{2\theta} + \frac{\sum\left(x_i - \theta\right)^2}{2\theta^2} + \frac{\sum\left(x_i - \theta\right)}{\theta} = 0$$

Multiplying through by $2\theta^2$,

$$-n\theta + \sum\left(x_i - \theta\right)^2 + 2\theta\sum\left(x_i - \theta\right) = 0$$

$$-n\theta + \sum x_i^2 - 2\theta\sum x_i + n\theta^2 + 2\theta\sum x_i - 2n\theta^2 = 0$$

$$-n\theta^2 - n\theta + \sum x_i^2 = 0 \text{ or } \theta^2 + \theta - \frac{\sum x_i^2}{n} = 0.$$

Then $\hat{\theta}_0 = \dfrac{-1 + \sqrt{1 + 4\dfrac{\sum X_i^2}{n}}}{2}$ (take the positive root since θ_0 is a variance).

Then

$$\lambda(\mathbf{x}) = \frac{\left(2\pi\hat{\theta}_0\right)^{-n/2} \exp\left[-\dfrac{\sum\left(x_i - \hat{\theta}_0\right)^2}{2\hat{\theta}_0}\right]}{\left(2\pi\hat{a}\overline{x}\right)^{-n/2} \exp\left[-\dfrac{\sum\left(x_i - \overline{x}\right)^2}{2\hat{a}\overline{x}}\right]}$$

$$= \left(\frac{\hat{\theta}_0}{\hat{a}\overline{x}}\right)^{-n/2} \exp\left[-\frac{\sum\left(x_i - \hat{\theta}_0\right)^2}{2\hat{\theta}_0} + \frac{n}{2}\right].$$

However, this expression is difficult to simplify to represent $\hat{\theta}_0$. Also, what is the distribution of $\hat{\theta}_0$?

But we do know that $-2\log\lambda(x) \approx \chi_1^2$, as there is one extra parameter in the unrestricted model.

So, to get a cutoff value or compute p-values for the test, we use the χ_1^2 table.

Another common method of constructing a large sample test is based on an estimator that has an asymptotic normal distribution. A *Wald test* is such a test.

In general, a Wald test is a test based on a statistic of the form

$$Z_n = \frac{W_n - \theta_0}{S_n},$$

where θ_0 is the hypothesized value of the parameter θ, the statistic W_n is a point estimator of θ, and S_n is an estimate of the standard deviation of W_n.

Example

Let $X_1, \ldots, X_n$ be a random sample from a $Bernoulli(p)$ distribution. Consider $H_0 : p \le p_0$ against $H_1 : p > p_0$, where $0 < p_0 < 1$.

The MLE of p is $\hat{p}_n = \bar{X}$. The Central Limit Theorem implies that for any $0 < p < 1$,

$$\frac{\hat{p}_n - p}{\sqrt{p(1-p)/n}}$$

converges to a standard normal random variable.

However, we do not know p , but a reasonable estimate is $S_n = \sqrt{\hat{p}_n\left(1-\hat{p}_n\right)/n}$, and it can be shown that $\sqrt{p\left(1-p\right)/n} / S_n$ converges in probability to 1.

Slutsky's Theorem implies that

$$\frac{\hat{p}_n - p}{\sqrt{\hat{p}_n\left(1-\hat{p}_n\right)/n}} \to N\left(0,1\right).$$

The Wald test statistic is defined by replacing p by p_0, and then the large sample Wald test rejects H_0 if $Z_n > z_\alpha$.

Another useful large sample test is a *score test*. The *score statistic* is defined to be

$$S(\theta) = \frac{\partial}{\partial\theta}\log f\left(\mathbf{X}\middle|\theta\right) = \frac{\partial}{\partial\theta}\log L\left(\theta\middle|\mathbf{X}\right).$$

We have shown that, for all θ, $E_\theta S(\theta) = 0$. In particular, when testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$, if H_0 is true, $E_\theta S(\theta_0) = 0$ Also,

$$Var_{\theta} S(\theta_0) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log L\left(\theta\middle|\mathbf{X}\right)\right)^2\right]$$

$$= -E_\theta\left[\left(\frac{\partial^2}{\partial\theta^2}\log L\left(\theta\middle|\mathbf{X}\right)\right)\right] = I_n(\theta).$$

The test statistic for the score statistic is

$$Z_S = S(\theta_0) / \sqrt{I_n(\theta_0)},$$

which, under H_0, has mean zero and variance 1, and converges to a standard normal.

Thus, the approximate level α score test rejects H_0 if $|Z_S| > z_{\alpha/2}$.

Example

Let $X_1, \ldots, X_n$ be a random sample from a $Bernoulli(p)$ distribution. Consider $H_0 : p = p_0$ against $H_1 : p \neq p_0$, where $0 < p_0 < 1$.

Here $f(\mathbf{x}|p) = p^{\sum x_i}(1-p)^{n-\sum x_i}$

$$= \left(\frac{p}{1-p}\right)^{\sum x_i}(1-p)^n.$$

Thus $S(p) = \dfrac{\partial}{\partial p}\log f(\mathbf{x}|p)$

$$= \frac{\partial}{\partial p}\left[\sum x_i \log p - \sum x_i \log(1-p) + n\log(1-p)\right]$$

$$= \frac{\sum x_i}{p} + \frac{\sum x_i}{1-p} - \frac{n}{1-p}$$

$$= \frac{\sum x_i (1-p) + \left(\sum x_i\right) p - np}{p(1-p)}$$

$$= \frac{\bar{x} - p}{p(1-p)/n};$$

$$I_n(p) = Var_p \frac{\bar{x} - p}{p(1-p)/n} = \frac{n}{p(1-p)}.$$

Hence, the score statistic is $Z_S = S(p_0)/\sqrt{I_n(p_0)} = \dfrac{\bar{x} - p_0}{\sqrt{p_0(1-p_0)/n}}$.

Large sample confidence intervals

Let T_n be an estimator of θ such that $\{T_n : n = 1, 2, \ldots\}$ is asymptotically normal, i.e.,

$$\frac{T_n - \theta}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1) \quad \forall\ \theta.$$

Suppose also that $\widehat{\sigma}_n$ is a consistent estimator of σ. Then

$$\frac{T_n - \theta}{\widehat{\sigma}_n/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1) \quad \forall\ \theta.$$

So,

$$P_\theta\left(-z_{\alpha/2} \le \frac{T_n - \theta}{\widehat{\sigma}_n/\sqrt{n}} \le z_{\alpha/2}\right) \approx 1 - \alpha$$

for all n sufficiently large, which implies that

$$P_\theta\left(T_n - z_{\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}} \le \theta \le T_n + z_{\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}\right) \approx 1 - \alpha.$$

So,

$$\left[T_n - z_{\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}, T_n + z_{\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}\right]$$

is an approximate $(1-\alpha)100\%$ c.i.

As in testing, MLEs can often be used to construct large sample confidence intervals.

Basically, we are inverting a Wald test.

Example 46 *Large sample c.i. for μ.* Suppose $X_1, \ldots, X_n$ is a random sample from a distribution with first four moments finite, and let

$$\mu = E(X_1) \quad \text{and} \quad \sigma^2 = \text{Var}(X_1).$$

Then, by the CLT,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1).$$

The weak law of large numbers (p. 232) implies that $S^2 = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{p} \sigma^2$, and so

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1).$$

An approximate $(1-\alpha)100\%$ c.i. for μ is thus

$$\left[\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right].$$

See also Examples 10.4.5 and 10.4.6.

Let $X_1,\ldots,X_n$ be a random sample from a *Bernoulli*(p) distribution.

$$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \xrightarrow{\mathcal{D}} N(0,1).$$

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

We could base a confidence interval on inverting a LRT (using an asymptotic χ^2 distribution).

For an asymptotic χ^2 test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, we do not reject if $-2\log\lambda(\mathbf{x}) < \chi^2_{k,\alpha}$, where k is the number of extra free parameters under the unrestricted model.

$$-2\log\frac{L(\theta_0)}{L(\hat{\theta}_{MLE})} < \chi^2_{k,\alpha}$$

$$\Rightarrow \log L(\theta_0) > \log L(\hat{\theta}_{MLE}) - \frac{\chi^2_{k,\alpha}}{2}.$$

So the confidence interval is

$$\left\{\theta : \log L(\theta) > \log L(\hat{\theta}_{MLE}) - \frac{\chi^2_{k,\alpha}}{2}\right\}.$$

Example 10.4.3 (Binomial LRT interval) For $Y = \sum_{i=1}^{n} X_i$, where each X_i is an independent Bernoulli(p) random variable, we have the approximate $1-\alpha$ confidence set

$$\left\{ p : -2 \log \left(\frac{p^y (1-p)^{n-y}}{\hat{p}^y (1-\hat{p})^{n-y}} \right) \leq \chi^2_{1,\alpha} \right\}.$$

We could base a confidence interval on inverting a score test too.
(using an asymptotic normal distribution).

Example 10.4.2 (Binomial score interval) Again using a binomial example, if $Y = \sum_{i=1}^{n} X_i$, where each X_i is an independent Bernoulli(p) random variable, we have

The score statistic is

$$Z_S = S(p) / \sqrt{I_n(p)} = \frac{\overline{x} - p}{\sqrt{p\,(1-p\,)/n}}.$$

An aproximate 1-a confidence interval

$$\left\{ p : \left| \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \right| \leq z_{\alpha/2} \right\}.$$

$$\left\{ p : \left(1 + \frac{z^2_{\alpha/2}}{n} \right) p^2 - \left(2\hat{p} + \frac{z^2_{\alpha/2}}{n} \right) p + \hat{p}^2 \leq 0 \right\}.$$

The 2 roots p1, p2

$$\frac{2\hat{p} + z^2_{\alpha/2}/n \pm \sqrt{(2\hat{p} + z^2_{\alpha/2}/n)^2 - 4\hat{p}^2(1 + z^2_{\alpha/2}/n)}}{2(1 + z^2_{\alpha/2}/n)},$$

CI: $\{p: \quad p1<p<p2\}$

Homework: p505~513

10.3, 10.9, 10.34(a), 10.36, 10.37, 10.38, 10.40