Asymptotics

Ricardo Dahis

PUC-Rio, Department of Economics

Summer 2023

Contents

Contents

consistency

- calculations simplify greatly as the sample size grows, and hence asymptotic analyses are a powerful and general evaluation tool
- minimum requirement: $T_n \equiv T_n(\boldsymbol{X})$ is a consistent sequence of estimators of the parameter θ if $T_n \xrightarrow{p} \theta$ for every $\theta \in \Theta$. That is, for every $\epsilon > 0$ and $\theta \in \Theta$,

$$\lim_{n \to \infty} \mathbb{P}_\theta \left(|T_n - \theta| < \epsilon \right) \;=\; 1$$

- although we colloquially speak about consistent estimators, it is actually the sequence of estimators that converge in probability to the true parameter value

consistency of the sample mean

- example: letting $X_1, \ldots, X_n \sim$ i.i.d.$N(\mu, 1)$ yields $\bar{X}_n \sim N(\mu, 1/n)$ and so

$$\begin{aligned}
\mathbb{P}_\mu(|\bar{X}_n - \mu| < \epsilon) &= \int_{\mu-\epsilon}^{\mu+\epsilon} \sqrt{\frac{n}{2\pi}} \exp\left(-\frac{n(\bar{x}_n - \mu)^2}{2}\right) \mathrm{d}\bar{x}_n \\
&= \int_{-\epsilon\sqrt{n}}^{\epsilon\sqrt{n}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \\
&= \mathbb{P}(|Z| < \epsilon\sqrt{n}) \to 1 \quad \text{as } n \to \infty
\end{aligned}$$

- more generally, apply Chebychev's inequality to show that

$$\begin{aligned}
\mathbb{P}_\mu(|T_n - \theta| \geq \epsilon) &\leq \frac{1}{\epsilon^2} \mathbb{E}_\mu(T_n - \theta)^2 \\
&= \frac{1}{\epsilon^2} \left[\text{var}_\theta(T_n) + \text{bias}_\theta^2(T_n)\right]
\end{aligned}$$

  converges to zero if and only if $\text{var}_\theta(T_n) \to 0$ and $\text{bias}_\theta(T_n) \to 0$ for all θ

- example: $\mathbb{E}_\mu(\bar{X}_n) = \theta$ and $\text{var}_\mu(\bar{X}_n) = \frac{1}{n}$

consistency of ML estimators

- theorem: let $X_i \sim$ i.i.d. $f(x|\boldsymbol{\theta})$. Define the (rescaled) likelihood function

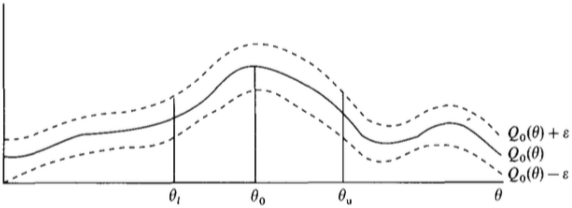
$$\hat{Q}_n(\theta) \;=\; n^{-1} \ln \ell(\boldsymbol{\theta}|\boldsymbol{x}) \;=\; n^{-1} \sum_{i=1}^{n} \ln f(x_i|\boldsymbol{\theta})$$

  Under mild regularity conditions, the maximum likelihood estimator $\hat{\boldsymbol{\theta}} = \arg\max \hat{Q}_n(\theta)$ is consistent, $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$

- this is an example of a extremum estimator: the proofs that follow do not require that $\hat{Q}_n(\theta)$ is a likelihood function, but rather that the estimator is the argument that maximizes some function that depends on parameters.
  - more applications of extremum estimators soon!

consistency of ML estimators

- why should this be the case? basic sketch of ideas:
  - as sample grows, $\hat{Q}_n(\theta) \xrightarrow{p} Q_0(\theta)$ for every θ
  - if $Q_0(\theta)$ is maximized uniquely at θ_0, the argmax of $\hat{Q}_n(\theta)$ should be close to θ_0
  - we need to ascertain that technical conditions are in place which allows us to exchange the limit of the maximum of $\hat{Q}_n(\theta)$ by the maximum of the limit $Q_0(\theta)$
  - if $\hat{Q}_n(\theta) \in [Q_0(\theta) - \varepsilon, Q_0(\theta) + \varepsilon]$, then $\hat{\theta} \in [\theta_l, \theta_u]$, and distance between θ_u and θ_l must be shrinking as $\varepsilon \to 0$



consistency of ML estimators

- definition: $\hat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$ if, and only if,

$$\sup_{\theta \in \Theta} |\hat{Q}_n(\theta) - Q_0(\theta)| \xrightarrow{p} 0.$$

- we prove consistency in the (more general) framework of extremum estimators.
- theorem: if there is a function $Q_0(\theta)$ such that:
  (i) $Q_0(\theta)$ is uniquely maximized at θ_0 (identification);
  (ii) Θ is compact;
  (iii) $Q_0(\theta)$ is continuous;
  (iv) $\hat{Q}_n(\theta)$ converges uniformly in probability to $Q_0(\theta)$.

  then $\hat{\theta} \xrightarrow{p} \theta_0$.

consistency of ML estimators

- proof: take an $\epsilon > 0$. Since $\hat{\theta}$ maximizes $\hat{Q}_n(\theta)$,

$$\hat{Q}_n(\hat{\theta}) \geq \hat{Q}_n(\theta_0) > \hat{Q}_n(\theta_0) - \frac{\epsilon}{3}$$

By uniform convergence of $\hat{Q}_n(\theta)$ to $Q_0(\theta)$, we also have that Q_0 and $\hat{Q}_n$ are arbitrarily close at any θ. So we can find an N such that $n \geq N$,

$$\begin{aligned} |Q_0(\theta) - \hat{Q}_n(\theta)| < \frac{\epsilon}{3} \quad &\Rightarrow \quad Q_0(\theta) - \hat{Q}_n(\theta) < \frac{\epsilon}{3} \\ &\Rightarrow \quad \hat{Q}_n(\theta) > Q_0(\theta) - \frac{\epsilon}{3} \end{aligned}$$

and

$$\begin{aligned} |Q_0(\theta) - \hat{Q}_n(\theta)| < \frac{\epsilon}{3} \quad &\Rightarrow \quad -Q_0(\theta) + \hat{Q}_n(\theta) < \frac{\epsilon}{3} \\ &\Rightarrow \quad Q_0(\theta) > \hat{Q}_n(\theta) - \frac{\epsilon}{3}. \end{aligned}$$

Since convergence is uniform, the above inequality holds for any $\theta \in \Theta$. In particular,

$$\begin{aligned} Q_0(\hat{\theta}) &> \hat{Q}_n(\hat{\theta}) - \frac{\epsilon}{3} \\ \hat{Q}_n(\theta_0) &> Q_0(\theta_0) - \frac{\epsilon}{3} \end{aligned}$$

consistency of ML estimators

- proof (cont'd): collecting inequalities,

$$
\begin{aligned}
Q_0(\hat{\theta}) &> \hat{Q}_n(\hat{\theta}) - \frac{\epsilon}{3} \\
\hat{Q}_n(\hat{\theta}) &> \hat{Q}_n(\theta_0) - \frac{\epsilon}{3} \\
\hat{Q}_n(\theta_0) &> Q_0(\theta_0) - \frac{\epsilon}{3}
\end{aligned}
$$

adding those inequalities, we have shown that for any $\epsilon > 0$, $Q_0(\hat{\theta}) > Q_0(\theta_0) - \epsilon$ with probability approaching 1.

Let $\mathcal{C}$ be any open subset of Θ containing θ_0. Then $\Theta \cap \mathcal{C}^c$ is compact. From the fact that $Q_0(\theta)$ is uniquely maximized at θ_0 and $Q_0(\theta)$ is continuous,

$$
\sup_{\theta \in \Theta \cap \mathcal{C}^c} Q_0(\theta) = Q_0(\theta^*) < Q_0(\theta_0)
$$

for some $\theta^* \in \Theta \cap \mathcal{C}^c$. Choosing $\epsilon = Q_0(\theta_0) - \sup_{\theta \in \Theta \cap \mathcal{C}^c} Q_0(\theta)$, it follows that

$$
Q_0(\hat{\theta}) > \sup_{\theta \in \Theta \cap \mathcal{C}} Q_0(\theta)
$$

and so $\hat{\theta} \in \mathcal{C}$. ■

consistency of ML estimators

- corollary: under conditions (i)-(iv), MLE is consistent.
- in particular, MLE satisfies the identification condition $Q_0(\theta)$ is uniquely maximized at θ_0.
- proof:

$$
\begin{aligned}
Q_0(\theta) - Q_0(\theta_0) &= \mathbb{E}\left(\ln \frac{f(x|\theta)}{f(x|\theta_0)}\right) \overset{\text{Jensen}}{<} \ln \mathbb{E}\left(\frac{f(x|\theta)}{f(x|\theta_0)}\right) \\
&= \ln \int \frac{f(x|\theta)}{f(x|\theta_0)} f(x|\theta_0) dx \\
&= \ln \int f(x|\theta) dx = \ln 1 = 0
\end{aligned}
$$

which implies that $Q_0(\theta) < Q_0(\theta_0)$ for any $\theta \neq \theta_0$ ∎

asymptotic distribution

- consistency says nothing about the asymptotic variance apart that it eventually converges to zero
- definition: the limiting variance τ^2 of the estimator T_n is given by

$$\lim_{n\to\infty} k_n \operatorname{Var} T_n \quad = \quad \tau^2 \quad < \quad \infty$$

  where k_n is a sequence of constants

- example: if $X_1, \ldots, X_n \sim$ i.i.d.$N(\mu, \sigma^2)$, then the limiting variance of $\bar{X}_n$ is $\sigma^2 = \lim_{n\to\infty} \sqrt{n} \operatorname{var} \bar{X}_n$ given that $\bar{X}_n \sim N(\mu, \sigma^2/n)$
- definition: for an estimator T_n, the asymptotic variance is σ^2 in

$$k_n \big(T_n - \tau(\theta) \big) \xrightarrow{d} N(0, \sigma^2)$$

  if such convergence exists

efficiency

- definition: a sequence of estimators T_n is asymptotically efficient for a parameter $\tau(\theta)$ if $\sqrt{n}\big(T_n - \tau(\theta)\big) \xrightarrow{d} N\big(0, \varsigma_\theta^2\big)$, with

$$\varsigma_\theta^2 \;=\; \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f(X|\theta)\right]^2} \qquad \text{(CR lower bound)}$$

- theorem: if $X_1, \ldots, X_n \sim$ iid $f(x|\theta)$, with $f(x|\theta)$ satisfying some mild regularity conditions, the ML estimator $\hat{\theta}_n$ is asymptotically efficient for θ, implying that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0)^{-1})$$

where $\mathcal{I}(\theta_0)$ is Fischer information matrix. That is, the MLE achieves the Cramér-Rao lower bound

asymptotic efficiency of MLE

- proof: under certain regularity conditions,
  (i) $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(X_i, \theta_0) \xrightarrow{d} N(0, \mathcal{I}(\theta_0))$, where $\mathcal{I}(\theta)$ is the Fischer information matrix

  (ii) $\frac{1}{n}\mathcal{H}(x_i, \theta_0) \xrightarrow{p} \mathbb{E}_\theta\left(\mathcal{H}(x, \theta_0)\right) = \mathcal{H}(\theta_0)$

  (iii) remember that $\mathcal{H}(\theta_0) = -\mathcal{I}(\theta_0)$

  then, Taylor-expanding the score, for some $\tilde{\theta} \in \left[\theta_0, \hat{\theta}_n\right]$,

$$
\begin{aligned}
0 &= \frac{1}{n} \sum_{i=1}^{n} s\left(X_i, \hat{\theta}_n\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} s\left(X_i, \theta_0\right) + \left(\frac{1}{n} \sum_{i=1}^{n} \mathcal{H}\left(x_i, \tilde{\theta}\right)\right) \left(\hat{\theta}_n - \theta_0\right)
\end{aligned}
$$

asymptotic efficiency of MLE

- proof (cont'd): therefore

$$
\begin{aligned}
(\hat{\theta}_n - \theta_0) &= -\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{H}(x_i,\tilde{\theta})\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}s\left(X_i,\theta_0\right) \\
\sqrt{n}(\hat{\theta}_n - \theta_0) &= \underbrace{-\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{H}(x_i,\tilde{\theta})\right)^{-1}}_{\xrightarrow{p}\mathcal{H}(\theta_0)+o_p(1)}\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}s\left(X_i,\theta_0\right)}_{\xrightarrow{d}N(0,\mathcal{I}(\theta_0))} \\
&\xrightarrow{d} N\left(0,\mathcal{I}(\theta_0)^{-1}\mathcal{I}(\theta_0)\mathcal{I}(\theta_0)^{-1}\right) \sim N\left(0,\mathcal{I}(\theta_0)^{-1}\right)
\end{aligned}
$$

that is, the MLE achieves the Cramér-Rao lower bound asymptotically ■

- procedure:
  - (i) calculate $\mathcal{I}(\theta)$ analytically
  - (ii) aproximate $\mathcal{I}(\theta_0)$ with $\mathcal{I}(\hat{\theta}_n)$, which should be a good approximation since $\hat{\theta}_n \xrightarrow{p} \theta_0$

comparisons

- definition: if two estimators W_n and V_n are such that

$$\begin{aligned}\sqrt{n}\,(W_n - \tau(\theta)) &\xrightarrow{d} N(0, \sigma_W^2)\\ \sqrt{n}\,(V_n - \tau(\theta)) &\xrightarrow{d} N(0, \sigma_V^2)\end{aligned}$$

then the asymptotic relative efficiency (ARE) is $ARE(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}$

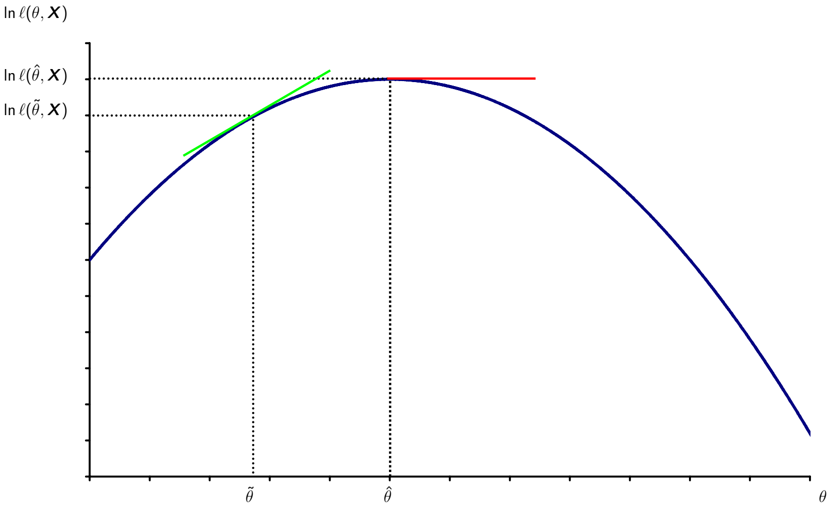
Contents

Contents

asymptotic tests

- as the sample size grows, the asymptotic approximation works better and we are able to derive tests even in complicated problems for which no optimal test exists
- **trinity of large-sample tests**
  (1) likelihood ratio tests: distance between log-likelihoods
  (2) Wald tests: distance between estimators
  (3) score tests (or LM tests): distance to zero score
- **differences**
  - LR tests estimate both restricted and unrestricted models
  - Wald tests estimate only unrestricted model (if simple null)
  - LM tests estimate only restricted model

trinity of tests



LR test, again

- it is one of the most useful methods for complicated problems because it gives not only an explicit definition of the test statistic, but also an explicit form for the rejection region

$$\text{reject } \mathbb{H}_0 \text{ if } \boldsymbol{x} \in \left\{ \boldsymbol{x} : \lambda(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\theta} \in \Theta_0} \ell(\boldsymbol{\theta}|\boldsymbol{x})}{\sup_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}|\boldsymbol{x})} \leq c \right\}$$

- even if we cannot obtain the two suprema analytically, we can usually compute them numerically

- to define a level α test, we choose c such that

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \mathbb{P}_{\boldsymbol{\theta}}\big(\lambda(\boldsymbol{X}) \leq c\big) \leq \alpha$$

asymptotic distribution of the LR test

- theorem: suppose that $X_1, \ldots, X_n \sim \text{iid} f(x|\theta)$, with the pdf satisfying the usual regularity conditions and consider testing the null $\mathbb{H}_0 : \theta = \theta_0$ versus the alternative $\mathbb{H}_1 : \theta \neq \theta_0$, then under $\mathbb{H}_0$,

$$-2 \ln \lambda(\boldsymbol{X}) \xrightarrow{d} \chi^2_1$$

under the null

- proof: Taylor expanding $\ln \ell(\theta|\boldsymbol{x})$ around $\hat{\theta}$ yields

$$\begin{aligned} \ln \ell(\theta|\boldsymbol{x}) &\cong \ln \ell(\hat{\theta}|\boldsymbol{x}) + \ln \ell'(\hat{\theta}|\boldsymbol{x})(\theta - \hat{\theta}) + \frac{1}{2} \ln \ell''(\hat{\theta}|\boldsymbol{x})(\theta - \hat{\theta})^2 \\ &\cong \ln \ell(\hat{\theta}|\boldsymbol{x}) + \frac{1}{2} \ln \ell''(\hat{\theta}|\boldsymbol{x})(\theta - \hat{\theta})^2 \end{aligned}$$

it then follows that

$$-2 \ln \lambda(\boldsymbol{x}) = 2\big[\ln \ell(\hat{\theta}|\boldsymbol{x}) - \ln \ell(\theta_0|\boldsymbol{x})\big] \cong -\ln \ell''(\theta_0|\boldsymbol{x})(\theta_0 - \hat{\theta})^2$$

completing the derivation as, under the null, $-\frac{1}{n} \ln \ell''(\hat{\theta}|\boldsymbol{x}) \xrightarrow{p} \mathcal{I}(\theta_0)$ and $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\big(0, \mathcal{I}(\theta_0)^{-1}\big)$ ■

LR test for Poisson intensity

- example: suppose that $X_1, \dots, X_n \sim \text{iid Poisson}(\lambda)$ and that the interest lies in testing $\mathbb{H}_0: \lambda = \lambda_0$ versus $\mathbb{H}_1: \lambda \neq \lambda_0$, then

$$-2\ln\lambda(\boldsymbol{x}) = -2\ln\left(\frac{e^{-n\lambda_0}\lambda_0^{n\bar{x}_n}}{e^{-n\hat{\lambda}}\hat{\lambda}^{n\bar{x}_n}}\right) = 2n\left[(\lambda_0 - \hat{\lambda}) - \hat{\lambda}\ln\left(\frac{\lambda_0}{\hat{\lambda}}\right)\right] > \chi^2_{1,\alpha}$$

  is the rejection region, where $\hat{\lambda} = \bar{x}_n$ is the ML estimator of λ

- accuracy of the asymptotic approximation

- simulation study with $\lambda_0 = 5$ and $n = 25$ (10,000 reps)

percentile	0.80	0.90	0.95	0.99
simulated distribution of the LR test	1.630	2.726	3.744	6.304
asymptotic approximation	1.642	2.706	3.841	6.635

extending the asymptotic theory...

- theorem: suppose that $X_1, \ldots, X_n \sim$ iid $f(x|\boldsymbol{\theta})$, with the pdf satisfying the usual regularity conditions and consider testing the null $\mathbb{H}_0 : \boldsymbol{\theta} \in \Theta_0$ versus the alternative $\mathbb{H}_1 : \boldsymbol{\theta} \in \Theta_0^c$. Then

$$-2 \ln \lambda(\boldsymbol{X}) \xrightarrow{d} \chi^2_d$$

under the null, where the degrees of freedom d is the difference between the number of free parameters in Θ and Θ_0

$$\text{reject } \mathbb{H}_0 \text{ if and only if } -2 \ln \lambda(\boldsymbol{X}) \geq \chi^2_{d,1-\alpha}$$

- note that the type I error probability will approach α if $\boldsymbol{\theta} \in \Theta_0$ only for large samples, and hence we say that the above rejection region yields an asymptotic size α test

LR test for multinomial probabilities

- example: suppose that $X_1, \ldots, X_n$ are iid discrete random variables with pmf $f(j|\boldsymbol{p}) = p_j$ for $j \in \{1, \ldots, 5\}$, then $\ell(\boldsymbol{p}|\boldsymbol{x}) = \prod_{i=1}^{n} p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4} p_5^{n_5}$, where n_j is the number of $x_1, \ldots, x_n$ equal to j

- test $\mathbb{H}_0 : \boldsymbol{p} \in \Theta_0$, where $\Theta_0 = \{\boldsymbol{p} : \ p_1 = p_2 = p_3 \text{ and } p_4 = p_5\}$

- full parameter space Θ has 4 free parameters, whereas only 1 free parameter remains after imposing the restrictions in Θ_0: $d = 3$

- unrestricted MLE: $\hat{p}_j = \frac{n_j}{n}$

Wald test

- large-sample test based on any asymptotically normal estimator

$$Z_n(\theta) = \frac{T_n - \theta}{\sigma(T_n)} \xrightarrow{d} N(0,1) \quad \text{for each fixed value of } \theta \in \Theta$$

- even if σ has to be estimated,

$$Z_n(\theta) = \frac{T_n - \theta}{\sigma(T_n)} = \frac{T_n - \theta}{\hat{\sigma}(T_n)} \frac{\hat{\sigma}(T_n)}{\sigma(T_n)} \xrightarrow{d} N(0,1)$$

  as long as $\hat{\sigma}(T_n) \xrightarrow{p} \sigma(T_n)$.

- example: consider testing $\mathbb{H}_0 : \theta = \theta_0$ versus $\mathbb{H}_1 : \theta \neq \theta_0$ using the fact that $Z_n(\theta_0) \xrightarrow{d} N(0,1)$ under the null $\mathbb{H}_0$
  - asymptotic size α requires to reject if $|Z_n(\theta_0)| > z_{1-\alpha/2}$
  - consistent because $\mathbb{P}_\theta \left(|Z_n(\theta_0)| > z_{1-\alpha/2}\right) \to 1$ for any $\theta \in \Theta_0^c$

Wald test for binomial probability

- example: suppose that $X_1, \ldots, X_n \sim$ iid Bernoulli(p) and that the interest lies in testing $\mathbb{H}_0 : p \leq p_0$ versus $\mathbb{H}_1 : p > p_0$, with $0 < p_0 < 1$

- $\bar{X}_n \sim$ MLE, with variance $\sigma^2(\bar{X}_n) = p(1-p)/n$

$$\mathrm{W}_n \;=\; Z_n(p_0)\,\frac{\sigma(\bar{X}_n)}{\hat{\sigma}(\bar{X}_n)} \;=\; \frac{\bar{X}_n - p_0}{\sigma(\bar{X}_n)}\,\frac{\sigma(\bar{X}_n)}{\sqrt{\bar{X}_n(1-\bar{X}_n)/n}} \;\xrightarrow{d}\; N(0,1)$$

- reject $\mathbb{H}_0 : p \leq p_0$ if $\mathrm{T}_n > z_{1-\alpha}$

- in the two-sided case with $\mathbb{H}_0 : p = p_0$, we can alternatively estimate $\sigma^2(\bar{X}_n) = p(1-p)/n$ by $p_0(1-p_0)/n$, yielding a more powerful test for some values of p

score test

- score statistic $S_\theta = \frac{\partial \ln \ell(\theta|\boldsymbol{X})}{\partial \theta}$ has mean zero and

$$\mathrm{var}_\theta(S_\theta) \;=\; \mathbb{E}_\theta\left[\frac{\partial \ln \ell(\theta|\boldsymbol{X})}{\partial \theta}\right]^2 \;=\; -\mathbb{E}_\theta\left[\frac{\partial^2 \ln \ell(\theta|\boldsymbol{X})}{\partial \theta^2}\right] \;=\; \mathcal{I}(\theta)$$

  for all θ, and hence

$$\mathrm{LM} = \frac{s(\boldsymbol{X}, \theta_0)}{\sqrt{\mathcal{I}(\theta_0)}} \xrightarrow{d} N(0,1)$$

- asymptotic level α score test rejects $\mathbb{H}_0 : \theta \leq \theta_0$ if $\mathrm{LM} > z_{1-\alpha}$

- if composite null, maximize restricted likelihood to obtain $\hat{\theta}_0$ (possibly by means of Lagrange multipliers)

score test for Bernoulli probability

- suppose that $X_1, \ldots, X_n \sim$ iid Bernoulli(p) and that the interest lies in testing $\mathbb{H}_0 : p = p_0$ versus $\mathbb{H}_1 : p \neq p_0$, then

$$\text{LM} \;=\; \frac{s_{p_0}}{\sqrt{\mathcal{I}(p_0)}} \;=\; \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)/n}} \;\xrightarrow{d}\; N(0,1)$$

  - reject $\mathbb{H}_0 : p = p_0$ if $|\text{LM}| > z_{1-\alpha/2}$
  - same test statistic than the alternative Wald test

Contents

Reference:

- Casella and Berger, Ch. 10
- Newey and McFadden, "Large Sample Estimation and Hypothesis Testing", Handbook of Econometrics, Ch. 36

Exercises:

- 10.1-10.10, 10.18-10.19, 10.22, 10.32-10.38, 10.40, 10.47