Section 13: Large sample hypothesis testing

ARE 210

November 28, 2017

The section notes are available on the section Github at github.com/johnloeser/are210 in the "section13" folder.

1 Definitions

- Large sample hypothesis testing setup
  - $\{X_i\}_{i=1}^n$ iid, $X_i \sim P_0$, $\theta(P_0) \equiv \theta_0$ $d \times 1$ vector of parameters [write P_θ for some distribution in the probability space]
  - Null and alternative hypotheses $\left\{ \begin{array}{ll} H & : \theta_0 \in \Theta_H \equiv \{\theta \in \Theta : a(\theta) = 0\} \\ K & : \theta_0 \in \Theta \setminus \Theta_H \equiv \{\theta \in \Theta : a(\theta) \neq 0\} \end{array} \right\}$, where $a(\theta)$ is an $r \times 1$ vector of restrictions
  - Assume $A(\theta) \equiv \frac{\partial}{\partial \theta^{\mathrm{T}}} a(\theta)$, an $r \times d$ Jacobian, has full row rank
- Definitions
  - $\{\delta_n(X)\}_{n=1}^\infty \equiv \{\delta_n\}_{n=1}^\infty$ is **pointwise asymptotically level-α** if $\forall\ \theta \in \Theta_H$, $\limsup_{n\to\infty} \mathbf{E}_{P_\theta}[\delta_n] \leq \alpha$
  - $\lim_{n\to\infty} \sup_{\theta\in\Theta_H} \mathbf{E}_{P_\theta}[\delta_n]$ is the **limiting size of** δ_n
  - δ_n has **asymptotic significance level-α uniformly in** P_0 if the limiting size of δ_n is less than or equal to α
  - δ_n is **consistent** if $\lim_{n\to\infty} \mathbf{E}_P[\delta_n] = 1\ \forall\ \theta \in \Theta \setminus \Theta_H$.
- Estimator and more definitions
  - Assume $\theta_0 = \arg\min_{\theta\in\Theta} Q_0(\theta)$, construct a sequence $Q_n(\theta) \xrightarrow{p} Q_0(\theta)$, and define $\hat{\theta}_u \equiv \arg\min_{\theta\in\Theta} Q_n(\theta)$ and $\hat{\theta}_c \equiv \arg\min_{\theta\in\Theta_H} Q_n(\theta)$
  - Let $A_n = A(\hat{\theta}_c)$ and $A_0 = A(\theta_0)$
  - Assuming a "well behaved" Q_n, we can also write
    * $\frac{\partial Q_n(\hat{\theta}_u)}{\partial \theta} = 0$ in the unconstrained case
    * $\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta} + A(\hat{\theta}_c)^{\mathrm{T}} \lambda_n = 0$ subject to $a(\hat{\theta}_c) = 0$ in the constrained case
  - Assume $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} \xrightarrow{d} N(0, \Sigma)$
  - Let $\Psi_n = \frac{\partial^2 Q_n(\hat{\theta}_c)}{\partial \theta \partial \theta^{\mathrm{T}}}$, such that $\Psi_n \xrightarrow{p} \Psi \equiv \frac{\partial^2 Q_0(\theta_0)}{\partial \theta \partial \theta^{\mathrm{T}}}$

- Let $\Sigma_n = V_{\hat{\theta}_c}\left[\sqrt{n}\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta}\right]$, and assume $\Sigma_n \xrightarrow{p} \Sigma$.[1]

- Some asymptotic results under the null
  - $\sqrt{n}(\hat{\theta}_u - \theta_0) = -\Psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1) \xrightarrow{d} N(0, \Psi^{-1}\Sigma\Psi^{-1})$
  - $A_0\sqrt{n}\lambda_n + \Psi\sqrt{n}(\hat{\theta}_c - \theta_0) = -\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1)$
  - $\sqrt{n}\lambda_n = -[A_0\Psi^{-1}A_0^{\mathrm{T}}]^{-1}A_0\Psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1)$
  - $\sqrt{n}(\hat{\theta}_c - \theta_0) = -[\Psi^{-1} - \Psi^{-1}A_0^{\mathrm{T}}(A_0\Psi^{-1}A_0^{\mathrm{T}})^{-1}A_0\Psi^{-1}]\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1)$
  - $\sqrt{n}(\hat{\theta}_c - \hat{\theta}_u) = -\Psi^{-1}A_0^{\mathrm{T}}(A_0\Psi^{-1}A_0^{\mathrm{T}})^{-1}A_0\Psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta} + o_p(1)$
- Key result
  - $X \sim N(\mu, \Sigma)$ d-dimensional vector random variables $\Rightarrow (X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu) \sim \chi^2_d$
- Possible test statistics
  - Likelihood ratio test
    * $\mathrm{LR} \equiv 2n(Q_n(\hat{\theta}_u) - Q_n(\hat{\theta}_c))$
    * Use $Q_n(\hat{\theta}_u) - Q_n(\hat{\theta}_c) = \frac{1}{2}(\hat{\theta}_u - \hat{\theta}_c)^{\mathrm{T}}\frac{\partial^2 Q_n(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}(\hat{\theta}_u - \hat{\theta}_c)$ for some $\theta^* \in \left(\hat{\theta}_c, \hat{\theta}_u\right)$, with $\theta^*, \hat{\theta}_c, \hat{\theta}_u \xrightarrow{p} \theta_0$
    * Gives $\mathrm{LR} \approx -n(\hat{\theta}_u - \hat{\theta}_c)^{\mathrm{T}}\Psi(\hat{\theta}_u - \hat{\theta}_c)$
    * More algebra, $\mathrm{LR} \approx -(A_0\Psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta})^{\mathrm{T}}(A_0\Psi^{-1}A_0^{\mathrm{T}})^{-1}(A_0\Psi^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial \theta})$
    * $\Sigma = -\Psi \Rightarrow \mathrm{LR} \xrightarrow{d} \chi^2_r$
  - Wald test
    * $\mathrm{WS} = na(\hat{\theta}_u)^{\mathrm{T}}V_{\theta_0}[a(\hat{\theta}_u)]^{-1}a(\hat{\theta}_u)$
    * Since θ_0 is unobserved, we can estimate it using $\hat{\theta}_u$
    * $\mathrm{WS} \xrightarrow{d} \chi^2_r$
  - Lagrange multiplier/score test
    * $n(\lambda_n)^{\mathrm{T}}V_\theta[\lambda_n]^{-1}\lambda_n \xrightarrow{d} \chi^2_r$
    * When $\Psi = -\Sigma$, using $\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta} = -A_n^{\mathrm{T}}\lambda_n$, we can show $n(\frac{Q_n(\hat{\theta}_c)}{\partial \theta})^{\mathrm{T}}\Sigma_n^{-1}(\frac{Q_n(\hat{\theta}_c)}{\partial \theta}) \xrightarrow{d} \chi^2_r$

2 Useful tips

- One useful simplifying case is where we have the null $\theta_0 = 0$, in which case $A(\theta) \equiv \mathbf{I}_n$
- Recall that $\Psi = -\Sigma$ when we use maximum likelihood or efficient GMM to esti-

[1] Note that this notation is somewhat ambiguous - we're treating $\hat{\theta}_c$ like a parameter here, not like a random variable.

mate θ. When this does not hold, we can estimate these two matrices separately and relax the assumption that $\Psi = -\Sigma$.

3 Practice questions

1. Consider the linear regression model $Y_i|Z_i \sim N(\beta_0+\beta_1 Z_i, \sigma^2)$, where Z_i is a scalar, and consider the null hypothesis $\beta_1 = 0$.
   (a) What is θ_0? $Q_0(\theta)$? $Q_n(\theta)$? $\hat{\theta}_u$? $\hat{\theta}_c$? $a(\theta)$? $A(\theta)$? A_n? $\frac{\partial Q_0(\theta)}{\partial \theta}$? $\frac{\partial Q_n(\theta)}{\partial \theta}$? $\frac{\partial Q_n(\hat{\theta}_u)}{\partial \theta}$? $\frac{\partial Q_n(\hat{\theta}_c)}{\partial \theta}$? λ_n? Σ? Σ_n? Ψ? Ψ_n?
   (b) Derive the LR test and its asymptotic distribution.
   (c) Derive the Wald test and its asymptotic distribution.
   (d) Derive the score test and its asymptotic distribution.
   (e) Compare the tests, do they differ in power? Size? How do their sizes compare to the t-test we discussed previously?