Section 10: Large sample theory

ARE 210

November 7, 2017

1) Derive the Fisher Information matrix of a $N(\mu, \sigma^2)$ random variable. Construct an asymptotically consistent 95% confidence interval for $\widehat{\mu}_{MLE}$ using $\widehat{\sigma}_{MLE}$. What is its asymptotic variance? How can we interpret it asymptotically?

$\log p(x;\mu,\sigma^2) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$. Differentiating with respect to the parameters yields the first order conditions (score) $s(x;\mu,\sigma^2) \equiv \begin{pmatrix} \frac{\partial \log p(x;\mu,\sigma^2)}{\partial\mu} \\ \frac{\partial \log p(x;\mu,\sigma^2)}{\partial\sigma^2} \end{pmatrix} = \begin{pmatrix} -\frac{x-\mu}{\sigma^2} \\ -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2\sigma^4} \end{pmatrix}$, where $\sigma^k = (\sigma^2)^{k/2}$. Applying the analogy principle to the population first order condition $\mathbf{E}[s(x;\mu,\sigma^2)] = 0$ yields $\frac{1}{n}\sum_{i=1}^n s(X_i;\hat{\mu}_{MLE},\hat{\sigma^2}_{MLE}) = 0$. Solvings this yields $\begin{pmatrix} \hat{\mu}_{MLE} \\ \hat{\sigma^2}_{MLE} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{1}{n}\sum_{i=1}^n (X_i - \hat{\mu}_{MLE})^2 \end{pmatrix}$.

Next, we can compute $I(\mu,\sigma^2) = V[s] = \mathbf{E}[ss']$, suppressing notation for the score $s(x;\mu,\sigma^2)$. Calculating this yields $I(\mu,\sigma^2) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$. Using results for maximum likelihood estimators, $\sqrt{n}\begin{pmatrix} \hat{\mu}_{MLE} - \mu \\ \hat{\sigma^2}_{MLE} - \sigma^2 \end{pmatrix} \xrightarrow{d} N(0, I(\mu,\sigma^2)^{-1})$. This gives us the asymptotically consistent 95% confidence interval for $\hat{\mu}_{MLE}$

$$\left(\hat{\mu}_{MLE} + \Phi^{-1}(.025)\sqrt{\frac{\hat{\sigma^2}_{MLE}}{n}}, \hat{\mu}_{MLE} + \Phi^{-1}(.975)\sqrt{\frac{\hat{\sigma^2}_{MLE}}{n}}\right)$$

where Φ is the normal CDF.

2) (Problem 1, PS5) Consider the parametric model $\{p(x,\theta) : \theta > 0\}$ where

$$p(x,\theta) = \theta x^{\theta-1} \quad x \in (0,1)$$

a. Suppose we observe an i.i.d. sample from this density. Find the Maximum Likelihood estimator of θ and calculate the Fisher Information.

First, $\log p(x,\theta) = \log\theta + (\theta-1)\log x$. This yields $s(x,\theta) = \frac{1}{\theta} + \log x$, where s is the score (derivative of log likelihood with respect to θ). Population moment is $\mathbf{E}[s(X,\theta)] = 0$, which yields the sample equivalent $\frac{1}{n}\sum_{i=1}^n s(X_i, \hat\theta_{MLE}) = 0$. Solving yields the MLE $\hat\theta_{MLE} = -\frac{1}{\overline{\log X}}$, where $\overline{\log X} = \frac{1}{n}\sum_{i=1}^n \log X_i$ the sample mean of $\log X$.

Next, we calculate $I(\theta) = V[s(X,\theta)] = \mathbf{E}[s(X,\theta)^2]$. Using the results that $\mathbf{E}[\log X] = -\frac{1}{\theta}$ (from the population moment condition for the score) and $\mathbf{E}[(\log X)^2] = \frac{2}{\theta^2}$ (by integration by parts) yields $I(\theta) = \frac{1}{\theta^2}$.

b. Show whether the the MLE is consistent for θ.

First, note that $\overline{\log x} \xrightarrow{p} -\frac{1}{\theta}$. This follows from the sample mean of the score is a consistent estimator of 0, which follows from the LLN (since the score has finite variance, since $\theta > 0$. Applying the CMT to $\hat\theta_{MLE}$ yields $\hat\theta_{MLE} \xrightarrow{p} -\frac{1}{-\frac{1}{\theta}} = \theta$, so $\hat\theta_{MLE}$ is consistent for θ.

c. Derive the limiting distribution of the MLE.

We can do this two ways. First, we can use results for MLE to state $\sqrt{n}(\hat\theta_{MLE}-\theta) \xrightarrow{d} N(0, I(\theta)^{-1})$, which yields $\sqrt{n}(\hat\theta_{MLE}-\theta) \xrightarrow{d} N(0,\theta^2)$.

Alternatively, applying the MVT to $\hat\theta_{MLE}$ yields $\hat\theta_{MLE} - \theta = \frac{1}{\tilde b_n^2}(\overline{\log X} - \mathbf{E}[\log X])$, for $\tilde b_n = \alpha\overline{\log X} + (1-\alpha)\mathbf{E}[\log X]$ for some $\alpha \in (0,1)$. Note that $\tilde b_n \xrightarrow{p} \mathbf{E}[\log X] = -\frac{1}{\theta}$, and applying a CLT to $\overline{\log X}$ yields $\sqrt{n}(\overline{\log x} - \mathbf{E}[\log X]) \xrightarrow{d} N(0, V[\log X])$, where $V[\log X] = \mathbf{E}[(\log X)^2] - \mathbf{E}[\log X]^2 = \frac{1}{\theta^2}$. This yields $\sqrt{n}(\hat\theta_{MLE}-\theta) \xrightarrow{d} N(0,\theta^2)$.

d. Find a Method of Moments estimator for θ and discuss its consistency.

First, note that MLE is a MOM estimator, using the moment restriction $\mathbf{E}[\log X] + \frac{1}{\theta} = 0$. Alternatively, we can use $\mathbf{E}[X] - \frac{\theta}{\theta+1} = 0$. The analogous estimator is defineed by $\overline{X} - \frac{\hat\theta_{MOM}}{\hat\theta_{MOM}+1} = 0$, which yields $\hat\theta_{MOM} = \frac{\overline{X}}{1-\overline{X}}$. Since $\mathbf{E}[X^2] = \frac{\theta}{\theta+2}$, X has a finite second moment, so by a LLN $\overline{X} \xrightarrow{p} \frac{\theta}{\theta+1}$. Applying Slutsky's Rule and CMT yields $\hat\theta_{MOM} \xrightarrow{p} \theta$, so it is consistent.

e. Does there exist a UMVUE for θ? If so, does it attain the Cramer-Rao lower bound?

First, the MLE is biased (based on calculating it's expectation in Wolfram Alpha for $n = 2$ and $\theta = 2$), as is the MOM estimator we constructed. I'm not sure how to construct an unbiased estimator for θ, so I don't know if UMVUE exists, nor if it attains the Cramer-Rao lower bound.

3) Prove the asymptotic normality of the GMM estimator.

Let $\hat{\theta}_n = \arg\max_{b\in\Theta} -\frac{1}{2}m_n(b)^{\mathrm{T}}S_n(W)m_n(b)$ be the GMM estimator, with everything defined as in the notes. The first order condition is $M_n(\hat{\theta}_n)S_n(W)m_n(\hat{\theta}_n) = 0$. Next, applying the MVT to $m_n(\hat{\theta}_n)$ yields $m_n(\hat{\theta}_n) = m_n(\theta) + M_n(\tilde{b}_n)^T(\hat{\theta}_n - \theta)$, for $\tilde{b}_n = \alpha\hat{\theta}_n + (1-\alpha)\theta$ for $\alpha \in (0,1)$. Substituting this into the first order condition, and solving for $\hat{\theta}_n - \theta$ yields

$$\sqrt{n}(\hat{\theta}_n - \theta) = -\left[M_n(\hat{\theta}_n)S_n(W)M_n(\tilde{b}_n)^{\mathrm{T}}\right]^{-1} M_n(\hat{\theta}_n)S_n(W)\sqrt{n}m_n(\hat{\theta}_n)$$

Next, by a ULLN, $M_n(\hat{\theta}_n) \xrightarrow{p} M_n(\theta)$, and $M_n(\tilde{b}_n) \xrightarrow{p} M_n(\theta)$. By assumption, $S_n(W) \xrightarrow{p} S$, and by a central limit theorem $\sqrt{n}m_n(\hat{\theta}_n) \xrightarrow{d} N(0, V[m(W,\theta)])$. Let $\Sigma = V[m(W,\theta)]$. Applying Slutsky's rule to the expression above yields

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N\left(0, (MSM^{\mathrm{T}})^{-1}MS\Sigma SM^{\mathrm{T}}(MSM^{\mathrm{T}})^{-1}\right)$$