ECON 6190

Problem Set 8

Gabe Sekeres

November 20, 2024

1. A Bernoulli random variable X is

$$\mathbb{P}(X = 0) = 1 - p$$
$$\mathbb{P}(X = 1) = p$$

We have a random sample $X_i, i = 1, \dots, n$ from X.

(a) Note that the PMF for some X_i is

$$f(x) = \mathbb{P}(X_i = x) = p^x(1-p)^{1-x}$$

for $x \in \{0, 1\}$. The likelihood function is

$$L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$

and so the log likelihood function is

$$\ell(p) = \sum_{i=1}^{n}(X_i \log p + (1 - X_i)\log 1 - p) = \log p \sum_{i=1}^{n} X_i + \log(1-p)\left(n - \sum_{i=1}^{n} X_i\right)$$

To find the MLE estimator, we find the first order condition, and get that

$$\frac{\partial \ell}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right) = 0$$

which, simplifying, gets us that

$$\hat{p}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

(b) Note that by inspection, $\mathbb{E}[X^2] < \infty$, as $\mathbb{E}[X], \mathbb{E}[X^2] \leq 1$. Note also that $\mathbb{E}[X] = p$. Thus, by the central limit theorem, we have that

$$\sqrt{n}(\bar{X}_n - \mathbb{E}[X]) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(X)) \implies \sqrt{n}(\hat{p}_{MLE} - p) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(X))$$

(c) Note that the asymptotic variance of $\hat{p}_{MLE}$ is the same as the variance of the random variable X. The estimator I propose for the asymptotic variance is

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

(d) To show that $\hat{\sigma}^2$ is consistent, we need that

$$\hat{\sigma}_n^2 \xrightarrow{p} \mathrm{Var}(X) \equiv \lim_{n\to\infty} \mathbb{E}[\sigma_n^2] = \mathrm{Var}(X)$$

Note that, since $X_i^2 = X_i$ for any outcome, we have that $\mathbb{E}[X^2] = \mathbb{E}[X] = p$, so

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1-p)$$

Thus, we have that fixing some n,

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}_n^2] &= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}[(X_i - \bar{X}_n)^2] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left[(X_i - p + p - \bar{X}_n)^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left[(X_i - p)^2\right] + 2\,\mathbb{E}[(X_i - p)\underbrace{(\bar{X}_n - p)}_{=0}] + \mathbb{E}[\underbrace{(p - \bar{X}_n)^2}_{=0}] \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}[(X_i - \mathbb{E}[X_i])^2] = \frac{n}{n-1}\,\text{Var}(X) \qquad \text{by IID}
\end{aligned}$$

Thus, as $n \to \infty$, $\mathbb{E}[\hat{\sigma}_n^2] \to \text{Var}(X)$, so $\hat{\sigma}_n^2$ is a consistent estimator.

(e) We have that the efficient score is

$$S = \frac{\partial}{\partial p}\log f(X \mid p) = \frac{\partial}{\partial p}\left[(X\log p + (1-X)\log 1 - p)\right] = \frac{X}{p} - \frac{1-X}{1-p}$$

which simplifies to

$$S = \frac{X-p}{p(1-p)}$$

Thus,

$$\text{Var}(S) = \text{Var}\left(\frac{X-p}{p(1-p)}\right) = \frac{\text{Var}(X)}{(p(1-p))^2} = \frac{1}{p(1-p)}$$

and

$$\mathscr{F}_p = \frac{1}{p(1-p)}$$

(f) We have that another measure of the information is the curvature of $\ell(p)$:

$$-\frac{\partial^2\ell(p)}{\partial p^2} = -\frac{\partial^2}{\partial p^2}\left[(X\log p + (1-X)\log 1 - p)\right] = -\frac{\partial}{\partial p}\left[\frac{X}{p} - \frac{1-X}{1-p}\right]$$

which, evaluating, returns

$$-\frac{\partial^2\ell(p)}{\partial p^2} = \left(\frac{X}{p^2} + \frac{1-X}{(1-p)^2}\right)$$

Taking the expectation, we get that:

$$\mathbb{E}\left[\frac{X}{p^2} + \frac{1-X}{(1-p)^2}\right] = \left(\frac{\mathbb{E}[X]}{p^2} + \frac{1-\mathbb{E}[X]}{(1-p)^2}\right) = \left(\frac{1}{p} + \frac{1}{1-p}\right) = \frac{1}{p(1-p)}$$

Which is the same as part (e)!

(g) The Cramer-Rao lower bound is

$$(n\mathscr{F}_p)^{-1} = \left(\frac{n}{p(1-p)}\right)^{-1} = \frac{p(1-p)}{n}$$

(h) Recall that $\hat{p}_{MLE}$ is the sample mean. From class, we know that the variance of the sample mean is

$$\text{Var}(\hat{p}_{MLE}) = \frac{\text{Var}(X)}{n} = \frac{p(1-p)}{n}$$

So the variance of the MLE estimator is the same as the CRLB!

(i) Since $p = \mathbb{E}[X]$, I propose the sample mean estimator as the method of moments estimator:

$$\hat{p}_{MME} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2. We have that $X \sim U[0, \theta]$ for some $\theta > 0$. Note that the density of X is

$$f(x \mid \theta) = \frac{1}{\theta} \cdot \mathbb{1}_{x \in [0,\theta]}$$

So the log density is

$$\log f(x \mid \theta) = \begin{cases} -\log(\theta) & 0 \le x \le \theta \\ -\infty & \text{otherwise} \end{cases}$$

We have that the log likelihood is

$$\ell(\theta) = \begin{cases} -n\log(\theta) & \max_i X_i \le \theta \\ -\infty & \text{otherwise} \end{cases}$$

Since this is always negative, it is maximized when θ is minimized in the finite region, meaning when $\theta = \max_i X_i$. Thus, the maximum likelihood estimator $\hat{\theta}_{MLE}$ is $\max_i X_i$.

3. We have that the log density is

$$\log f(x \mid \mu, \sigma^2) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right)\right) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}$$

so the log likelihood function is

$$\ell(\mu, \sigma^2) = \sum_{i=1}^{n} \log f(X_i \mid \mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2\sigma^2}$$

First, to find $\hat{\mu}_{MLE}$, we take first order conditions with respect to μ. We get that

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \frac{\sum_{i=1}^{n}(X_i - \mu)}{\sigma^2} = 0 \implies \sum_{i=1}^{n}(X_i - \mu) = 0 \implies \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Next, to find $\hat{\sigma}^2_{MLE}$, we take first order conditions with respect to σ^2, and get that

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

which implies that

$$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

4. We will prove the Information Matrix Equality, letting $f = f(x \mid \theta_0)$, ∇_j mean partial with respect to the jth element $\theta^{(j)}$, and ∇_{jk} mean second-order with respect to $\theta^{(j)}$ and $\theta^{(k)}$. Suppose we can exchange the integral $\int$ and derivatives ∇_j.

(a) We have that

$$\nabla_j \left[\int f dx \right] = \nabla_j[1] \Longrightarrow \int \nabla_j f(x \mid \theta_0) dx = 0$$

From the chain rule and the definition of expected value, we get that

$$0 = \int \nabla_j f(x \mid \theta_0) dx = \int f(x \mid \theta_0) \nabla_j \log f(x \mid \theta_0) dx = \mathbb{E}[\nabla_j \log f(x \mid \theta_0)]$$

(b) Differentiating both sides with respect to $\theta^{(k)}$ and using Leibniz rule, we get that

$$0 = \nabla_k \mathbb{E}[\nabla_j \log f(x \mid \theta_0)] = \mathbb{E}[\nabla_{jk} \log f] + \mathbb{E}[(\nabla_j \log f)(\nabla_k \log f)]$$

5. We have that $g(x)$ is the density of a random variable with mean μ and variance σ^2. We have that X is a random variable with density

$$f(x \mid \theta) = g(x)(1 + \theta(x - \mu))$$

We know all of $g(x)$, μ, and σ^2. The unknown parameter is θ, and we assume that X has bounded support so that $f(x \mid \theta) \geq 0$ for all x.

(a) We have that

$$\int_{-\infty}^{\infty} f(x \mid \theta) dx = \int_{-\infty}^{\infty} g(x) + \theta g(x)(x - \mu) dx = \int_{-\infty}^{\infty} g(x) dx + \theta \int_{-\infty}^{\infty} g(x)(x - \mu) dx$$

and since g is a density and from the definition of expectation, we have that

$$\int_{-\infty}^{\infty} f(x \mid \theta) dx = 1 + \theta \cdot 0 = 1$$

(b) We have that

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x \mid \theta) dx = \int_{-\infty}^{\infty} x g(x) + \theta x g(x)(x - \mu) dx$$

so

$$\mathbb{E}[X] = \mu + \theta \int_{-\infty}^{\infty} x^2 g(x) - \mu x g(x) dx = \mu + \theta \left(\int_{-\infty}^{\infty} x^2 g(x) dx - \mu \int_{-\infty}^{\infty} x g(x) dx \right)$$

Thus,

$$\mathbb{E}[X] = \mu + \theta \sigma^2$$

(c) We have that the log density is

$$\log f(x \mid \theta) = \log g(x) + \log(1 + \theta(x - \mu))$$

so the efficient score is

$$\frac{\partial}{\partial \theta} f(X \mid \theta) = \frac{\partial}{\partial \theta} [\log g(X) + \log(1 + \theta(X - \mu))] = \frac{X - \mu}{1 + \theta(X - \mu)}$$

and the Fisher Information is

$$\mathscr{F}_{\theta_0} = \mathbb{E}\left[\left(\frac{X - \mu}{1 + \theta_0(X - \mu)} \right)^2 \right]$$

(d) When $\theta_0 = 0$, this expression simplifies to

$$\mathscr{F}_{\theta_0} = \mathbb{E}\left[(X-\mu)^2\right] = \mathrm{Var}(X)$$

(e) We have that the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(X_i \mid \theta)$$

so the log likelihood function is

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(X_i \mid \theta) = \sum_{i=1}^{n} \log g(X_i) + \log(1 + \theta(X_i - \mu))$$

(f) The first order condition is

$$\frac{\partial}{\partial\theta}\ell(\theta) = \sum_{i=1}^{n} \frac{X_i - \mu}{1 + \theta(X_i - \mu)} = 0$$

(g) From the asymptotic properties of MLE estimators, we know that the unique MLE estimator $\hat{\theta}$ has the property of

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \mathscr{F}_{\theta_0}^{-1}) = \mathcal{N}\left(0, \left(\mathbb{E}\left[\left(\frac{X-\mu}{1+\theta_0(X-\mu)}\right)^2\right]\right)^{-1}\right)$$

(h) When $\theta_0 = 0$, we have that

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{Var}(X)^{-1}\right)$$

6. To complete the proof, note that the variance expanded is:

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0) - \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right]\right)\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0) - \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right]\right)'\right]$$

From the Analog Principle, we have that θ_0 maximizes the expected log likelihood function, meaning that using Liebniz integral rule, since θ_0 is a local maximum,

$$\mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right] = \frac{\partial}{\partial\theta}\mathbb{E}[\log f(X \mid \theta_0)] = 0$$

and thus,

$$\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right)\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right)'\right]$$

and by i.i.d.,

$$\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right)\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta_0)\right)'\right] = n\,\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x \mid \theta_0)\right)\left(\frac{\partial}{\partial\theta}\log f(x \mid \theta_0)\right)'\right] = n\mathscr{F}_{\theta_0}$$

7. From class, we have that the MME is

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \le x}$$

Note that each $\mathbb{1}_{X_i \le x}$ is a Bernoulli random variable with mean $F(x)$. We can view the empirical distribution function as the sample mean of a Bernoulli process, where $F(x)$ is the population mean. Cast this way, we have that by Central Limit Theorem

$$\sqrt{n}(\hat{F}_n(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}(\mathbb{1}_{X_i \le x}))$$

and from the properties of a Bernoulli random variable, $\mathbb{1}_{X_i \le x}$ has variance $p(1-p)$, where p is the population mean. Thus, we have that

$$\sqrt{n}(\hat{F}_n(x) - F(x)) \xrightarrow{d} \mathcal{N}(0, F(x)(1 - F(x)))$$

8. Let X follow an exponential distribution with pdf $f(x) = \theta \exp(-\theta x)$, $x \ge 0$, $\theta > 0$. The expected value of X is given by $\mathbb{E}[X] = \frac{1}{\theta}$.

   (a) We have that the efficient score is

   $$S = \frac{\partial}{\partial \theta} \log f(X \mid \theta) = \frac{\partial}{\partial \theta}\left[\log(\theta) - \theta X\right] = \frac{1}{\theta} - X$$

   Thus, the Fisher information is

   $$\mathbb{E}\left[\left(\frac{1}{\theta} - X\right)^2\right] = \mathbb{E}\left[\frac{1}{\theta^2} - \frac{2X}{\theta} + X^2\right] = \frac{1}{\theta^2} - \frac{2}{\theta^2} + \mathbb{E}[X^2]$$

   Using the definition of expected value, we have that

   $$\mathbb{E}[X^2] = \int x^2 f(x)dx = \int x^2 \theta \exp(-\theta x)dx = \frac{2}{\theta^2}$$

   Thus, the Fisher information simplifies to

   $$\mathscr{F}_\theta = \frac{1}{\theta^2} - \frac{2}{\theta^2} + \frac{2}{\theta^2} = \frac{1}{\theta^2}$$

   and the CRLB is

   $$(n\mathscr{F}_\theta)^{-1} = \frac{\theta^2}{n}$$

   (b) Note that since $\mathbb{E}[X] = \frac{1}{\theta}$, we have that defining the function $g(x) = x^{-1}$, $\mathbb{E}[g(X)] = \theta$, the MME for θ is

   $$\hat{\theta}_{MME} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{X_i}$$

   (c) Using Delta Method, since $\frac{1}{n}\sum_{i=1}^{n} X_i$ and $g(\cdot)$ are scalar-valued, we have that by the CLT,

   $$\sqrt{n}(\hat{\theta}_{MME} - \theta) = \sqrt{n}(g(\hat{\mu}) - g(\mu)) \xrightarrow{d} \mathcal{N}(0, (g'(u)\mid_{\mu})^2 \mathrm{Var}(X))$$

   We have that the variance is

   $$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

and that

$$g'(u)\mid_{\mu} = -\frac{1}{u^2}\bigg|_{\mu} = -\frac{1}{\mathbb{E}[X]^2} = -\theta^2$$

Thus, the asymptotic distribution of $\hat{\theta}_{MME}$ is

$$\sqrt{n}(\hat{\theta}_{MME} - \theta) \xrightarrow{d} \mathcal{N}\left(0, (-\theta^2)^2 \frac{1}{\theta^2}\right) = \mathcal{N}(0, \theta^2)$$