Chapter 9: Parametric Inference

All of Statistics, Wasserman

1. Let $X_1, \ldots, X_n \sim \text{Gamma}(\alpha, \beta)$. Find the method of moments estimator for α and β.

Solution. By definition, we must solve the following system of equations for α and β:

$$\begin{aligned} \overline{X} &= E_{\alpha,\beta}(X) = \alpha\beta, \\ \overline{X^2} &= E_{\alpha,\beta}(X^2) = V_{\alpha,\beta}(X) + (E_{\alpha,\beta}(X))^2 = \alpha\beta^2 + \alpha^2\beta^2. \end{aligned}$$

Substituting $\alpha = \overline{X}/\beta$ in the second equation and solving for β yields

$$\alpha = \frac{\overline{X}^2}{\overline{X^2} - \overline{X}^2} \text{ and } \beta = \frac{\overline{X^2} - \overline{X}^2}{\overline{X}}.$$

◊

2. Let $X_1, \ldots, X_n \sim U(a, b)$ where a and b are unkown parameters and $a < b$.

(a) Find the method of moments estimators of a and b.

(b) Find the MLE of a and b.

(c) Let $\tau = \int x \, dF(x)$. Find the MLE of τ.

(d) Let $\widehat{\tau}$ be the MLE of τ. Let $\tilde{\tau}$ be the non-parametric plug-in estimator of τ. Find the MSE of $\tilde{\tau}$.

Solution. (a) By definition, we must solve the following system of equations for a and b:

$$\overline{X} = E_{a,b}(X) = \frac{a+b}{2}, \tag{1}$$

$$\overline{X^2} = E_{a,b}(X^2) = V_{a,b}(X) + (E_{a,b}(X))^2 = \frac{(b-a)^2}{12} + \left(\frac{a+b}{2}\right)^2. \tag{2}$$

Calculating $(2) - (1)^2$ yields $\overline{X^2} - \overline{X}^2 = \frac{(b-a)^2}{12}$. Substituting $a = 2\overline{X} - b$ in this equation, solving for b, and substituing the value obtained for b in the first equation yields

$$a = \overline{X} - \sqrt{3(\overline{X^2} - \overline{X}^2)} \text{ and } b = \sqrt{3(\overline{X^2} - \overline{X}^2)} + \overline{X}.$$

(b) Recall that $f(x; a, b) = \frac{1}{b-a} I_{(a,b)}(x)$. Hence,

$$L(a,b) = \frac{1}{(b-a)^n} \prod_{i=1}^n I_{(a,b)}(X_i) = \begin{cases} \frac{1}{(b-a)^n} & \text{if } a \leq \min_{i=1}^n X_i \text{ and } b \geq \max_{i=1}^n X_i \\ 0 & \text{otherwise.} \end{cases}$$

Its easy to see that $\widehat{a} = \min_{i=1}^n X_i$ and $\widehat{b} = \max_{i=1}^n X_i$ maximize L.

(c) Note that $\tau = E_{a,b}(X) = (a+b)/2$. Thus, by equivariance of the MLE, $\widehat{\tau} = (\widehat{a} + \widehat{b})/2$.

(d) By definition, $\tilde{\tau} = \int x \, d\widehat{F}(x) = \overline{X}$. Since $E\overline{X} = EX_1$ and $V\overline{X} = \frac{VX_1}{n}$,

$$\text{MSE}(\tilde{\tau}) = \text{MSE}(\overline{X}) = \text{bias}^2(\overline{X}) + V(\overline{X}) = 0 + \frac{(b-a)^2/12}{n} = \frac{(b-a)^2}{12n}.$$

$\Diamond$

3. Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$. Let τ be the 95 percentile, i.e., $P(X < \tau) = .95$.

(a) Find the MLE of τ.

(b) Find an expression for an approximate $1-\alpha$ confidence interval for τ.

Solution. (a) By definition,

$$.95 = P(X < \tau) = P\left(\frac{X-\mu}{\sigma} < \frac{\tau-\mu}{\sigma}\right) = P\left(Z < \frac{\tau-\mu}{\sigma}\right) = \Phi\left(\frac{\tau-\mu}{\sigma}\right)$$

Solving for τ yields

$$\tau = \sigma \cdot \Phi^{-1}(.95) + \mu.$$

Therefore, by equivariance of the MLE,

$$\widehat{\tau} = \widehat{\sigma} \cdot \Phi^{-1}(.95) + \widehat{\mu}$$

where $\widehat{\mu}$ and $\widehat{\sigma}$ are the MLE of μ and σ (cf. Example 9.11).

(b) Recall that

$$J_n = \frac{1}{n}\begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix}.$$

The gradient of $g(\mu, \sigma) = \tau$ is

$$\nabla g = \begin{pmatrix} 1 \\ \Phi^{-1}(.95) \end{pmatrix}$$

Thus,

$$\widehat{\text{se}}(\widehat{\tau}) = \sqrt{\begin{pmatrix} 1 & \Phi^{-1}(.95) \end{pmatrix} \begin{pmatrix} \frac{\hat{\sigma}^2}{n} & 0 \\ 0 & \frac{\hat{\sigma}^2}{2n} \end{pmatrix} \begin{pmatrix} 1 \\ \Phi^{-1}(.95) \end{pmatrix}} = \widehat{\sigma} \cdot \sqrt{\frac{1}{n} + \frac{\Phi^{-1}(.95)^2}{2n}}$$

An expression for an approximate $1-\alpha$ confidence interval for τ is

$$\widehat{\tau} \pm z_{\alpha/2} \cdot \widehat{\text{se}}(\widehat{\tau}).$$

$\Diamond$

4. Let $X_1, \dots, X_n \sim U(0, \theta)$. Show that the MLE is consistent.

Proof. The MLE is $\widehat{\theta} = \max_{i=1}^n X_i$ (cf. Exercise 2b). Calculating directly, for all $\epsilon > 0$

$$\begin{aligned} P\left(|\widehat{\theta} - \theta| > \epsilon\right) &= P\left(\max_{i=1}^n X_i - \theta > \epsilon\right) + P\left(\theta - \max_{i=1}^n X_i > \epsilon\right) \\ &= P\left(\theta - \max_{i=1}^n X_i > \epsilon\right) && \text{(since } X_i < \theta \text{ for all } i\text{)} \\ &= P\left(\max_{i=1}^n X_i < \theta - \epsilon\right) \\ &= \prod_{i=1}^n P(X_i < \theta - \epsilon) \\ &= \left(\frac{\theta - \epsilon}{\theta}\right)^n \xrightarrow{n \to \infty} 0. \end{aligned}$$

$\square$

5. Let $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$. Find the method of moments estimator, the MLE, and $I(\lambda)$.

Solution. Since $E_\lambda(X) = \lambda$ for $X \sim \text{Poisson}(\lambda)$, the method of moments estimator of λ is simply $\overline{X}$. Recall that the pmf is given by

$$f(x;\lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x \geq 0.$$

Hence, the likelihood function is

$$L(\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!} = e^{-n\lambda}\frac{\lambda^S}{\prod_{i=1}^n X_i!}$$

where $S = \sum_{i=1}^n X_i$. Then

$$\ell(\lambda) = -n\lambda + S\log\lambda - \log\prod_{i=1}^{n} X_i!$$

Differentiating yields

$$\ell'(\lambda) = -n + \frac{S}{\lambda}. \tag{3}$$

Equating to 0 and solving for λ yields the MLE

$$\widehat{\lambda} = \frac{S}{n} = \overline{X}.$$

Setting $n = 1$ and differentiating (3) yields

$$\frac{\partial^2 f(X;\lambda)}{\partial\lambda^2} = -\frac{X}{\lambda^2}$$

Finally,

$$I(\lambda) = -E_\lambda\left(-\frac{X}{\lambda^2}\right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}.$$

$\Diamond$

6. Let $X_1, \ldots, X_n \sim N(\theta, 1)$. Define

$$Y_i = \begin{cases} 1 & \text{if } X_i > 0, \\ 0 & \text{if } X_i \leq 0. \end{cases}$$

Let $\psi = P(Y_1 = 1)$.

(a) Find the MLE $\widehat{\psi}$ of ψ.

(b) Find an approximate 95 percent confidence interval of ψ.

(c) Define $\tilde{\psi} = \overline{Y}$. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

(d) Compute the asymptotic relative efficiency of $\tilde{\psi}$ to $\widehat{\psi}$.

(e) Suppose the data are not really normal. (Let $\widehat{\psi} := \Phi(\widehat{\theta})$ where $\widehat{\theta}$ is the MLE of θ.)[1] Show that $\widehat{\psi}$ is not consistent. What, if anything, does $\widehat{\psi}$ converge to?

[1] This sentence doesn't actually appear in the textbook, but seems prudent since otherwise one may think $\widehat{\psi}$ is the MLE of ψ under the new hypothesis. This is not the case since the MLE is always consistent.

Solution. (a) Note that

$$\psi = P(Y_1 = 1) = P(X_i > 0) = P(X_i - \theta > -\theta) = P(Z > -\theta) = P(Z < \theta) = \Phi(\theta) =: g(\theta).$$

where Φ is the cdf of $Z \sim N(0,1)$. By equivariance of the MLE, $\widehat{\psi} = \Phi(\widehat{\theta})$ where $\widehat{\theta}$ is the MLE of θ.

(b) Recall that

$$\widehat{\mathrm{se}}(\widehat{\theta}) = \widehat{\mathrm{se}}(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{n}}.$$

By the delta method,

$$\widehat{\mathrm{se}}(\widehat{\psi}) = |g'(\theta)|\widehat{\mathrm{se}}(\widehat{\theta}) = \phi(\widehat{\theta})\frac{1}{\sqrt{n}}$$

where $\phi = \Phi'$. Therefore,

$$\widehat{\psi} \pm z_{\alpha/2} \cdot \widehat{\mathrm{se}}(\widehat{\psi})$$

is a $1-\alpha$ percent confidence interval of ψ.

(c) Since $Y_i \sim \text{Bernoulli}(\psi)$, the LLN implies the desired result.

(d) Since $Y_i \sim \text{Bernoulli}(\psi)$,

$$V\tilde{\psi} = V\overline{Y} = \frac{\psi(1-\psi)}{n}.$$

Therefore,

$$\mathrm{ARE}(\tilde{\psi}, \widehat{\psi}) = \frac{\frac{\phi(\widehat{\theta})^2}{n}}{\frac{\psi(1-\psi)}{n}} = \frac{\phi(\widehat{\theta})^2}{\psi(1-\psi)}.$$

(e) Since $\widehat{\theta} \to \theta$ (the MLE is consistent) and convergence in probability is equivariant (cf. Theorem 5.5f),

$$\Phi(\widehat{\theta}) \to \Phi(\theta) \neq \psi.$$

since the data are not normal

◊

7. (Comparing two treatments.) Let $i = 1, 2$. n_i people are given treatment i. Let X_i be the number of people who respond favorably to treatment i. Assume that $X_i \sim \text{Binomial}(n_i, p_i)$. Let $\psi = p_1 - p_2$.

(a) Find the MLE $\widehat{\psi}$ of ψ.

(b) Find the Fisher information matrix $I(p_1, p_2)$.

(c) Use the multiparameter delta method to find the asymptoptic standard error of $\widehat{\psi}$.

Solution. (a) The likelihood function for p_i is given by

$$L(p_i) = \binom{n_i}{X_i} p_i^{X_i}(1-p_i)^{n_i - X_i}.$$

Recall we're assuming only one sample $X_i \sim \text{Binomial}(n_i, p_i)$. Moreover,

$$\ell(p_i) = \log\binom{n_i}{X_i} + X_i \log p_i + (n_i - X_i)\log(1-p_i).$$

Differentiating yields

$$\ell'(p_i) = \frac{X_i}{p_i} + \frac{n_i - X_i}{1 - p_i}(-1).$$

Equating to 0 and solving for p_i yields the MLE

$$\widehat{p_i} = \frac{X_i}{n_i}.$$

By equivariance of the MLE,

$$\widehat{\psi} = \widehat{p_1} - \widehat{p_2} = \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

(b) The multiparameter model to consider is

$$f\left((X_1, X_2); p_1, p_2\right) = f(X_1, p_1) \cdot f(X_2, p_2)$$

since $X_1 \perp X_2$. Calculating directly[2]

$$\frac{\partial^2 \ell}{\partial p_i^2} = -\frac{X_i}{p_i^2} - \frac{n_i - X_i}{(1 - p_i)^2} \qquad \text{and} \qquad \frac{\partial^2 \ell}{\partial p_i \partial p_j} = 0 \text{ if } i \neq j.$$

Moreover, since $E_{p_i}(X_i) = n_i p_i$,

$$E_{(p_1,p_2)}\left(\frac{\partial^2 \ell}{\partial p_i^2}\right) = -\frac{n_i p_i}{p_i^2} - \frac{n_i - n_i p_i}{(1 - p_i)^2} = -\frac{n_i}{p_i(1 - p_i)}$$

Therefore,

$$I(p_1, p_2) = -\begin{pmatrix} -\frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & -\frac{n_2}{p_2(1-p_2)} \end{pmatrix} = \begin{pmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{pmatrix}$$

(c) The inverse of $I(p_1, p_2)$ is

$$J(p_1, p_2) = \begin{pmatrix} \frac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{n_2} \end{pmatrix}$$

The transformation g is given by $g(p_1, p_2) = p_1 - p_2$, so

$$\nabla g = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Therefore,

$$\widehat{\text{se}}(\widehat{\psi}) = \sqrt{\begin{pmatrix} 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} & 0 \\ 0 & \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}} = \sqrt{\frac{\widehat{p}_1(1 - \widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1 - \widehat{p}_2)}{n_2}}$$

Furthermore, $\widehat{\psi} \pm z_{\alpha/2} \cdot \widehat{\text{se}}(\widehat{\psi})$ is a $1 - \alpha$ percent confidence interval of ψ.

◊

[2] The log-likelihood of (p_1, p_2) is the sum of the log-likelihoods, so differentiating w.r. to p_i, nullifies the log-likelihood of p_j.