Chapter 9 Solutions

Andrew Wu
Wasserman: All of Statistics

April 13, 2025

Problem 9.1. Let $X_1, \dots, X_n \sim \text{Gamma}(\alpha, \beta)$. Find the method of moments estimator for α and β.

Solution. Let $\theta = (\alpha, \beta)$. We must equate theoretical moments with sample moments. For the theoretical moments, we have $\mathbb{E}_\theta(X) = \alpha\beta$ and

$$\begin{aligned} \mathbb{E}_\theta(X^2) &= \mathbb{V}_\theta(X) + \mathbb{E}_\theta(X)^2 \\ &= \alpha\beta^2 + \alpha^2\beta^2 \\ &= \alpha\beta^2(1+\alpha). \end{aligned}$$

Thus we must solve the system

$$\begin{aligned} \widehat{\alpha}\widehat{\beta} &= \frac{1}{n}\sum_{i=1}^n X_i = \overline{X}_n \\ \widehat{\alpha}\widehat{\beta}^2(1+\widehat{\alpha}) &= \frac{1}{n}\sum_{i=1}^n X_i^2. \end{aligned}$$

So thus

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n X_i^2 &= \widehat{\alpha}\widehat{\beta}(\widehat{\beta} + \widehat{\alpha}\widehat{\beta}) \\ &= \overline{X}_n(\widehat{\beta} + \overline{X}_n) \end{aligned}$$

and it follows that

$$\widehat{\beta} = \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \overline{X}_n^2}{\overline{X}_n}.$$

Then

$$\begin{aligned} \widehat{\alpha} &= \frac{\overline{X}_n}{\widehat{\beta}} \\ &= \frac{\overline{X}_n^2}{\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \overline{X}_n^2}. \end{aligned}$$

□

Problem 9.2. Let $X_1, \dots, X_n \sim \text{Uniform}(a, b)$ where a and b are unknown parameters and $a < b$.

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

b) Find the MLE $\widehat{a}$ and $\widehat{b}$.

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

d) Let $\widehat{\tau}$ be the MLE of τ. Let $\widetilde{\tau}$ be the nonparametric plug-in estimator of $\tau = \int x dF(x)$. Suppose that $a = 1, b = 3, n = 10$. Find the MSE of $\widehat{\tau}$ by simulation. Find the MSE of $\widetilde{\tau}$ analytically. Compare.

Solution. As usual, set $\theta = (a, b)$.

a) The first and second theoretical moments are

$$\mathbb{E}_\theta(X) = \frac{a+b}{2}$$

and

$$\begin{aligned}\mathbb{E}_\theta(X^2) &= \mathbb{V}_\theta(X) + \mathbb{E}_\theta(X)^2 \\ &= \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} \\ &= \frac{a^2+ab+b^2}{3}.\end{aligned}$$

Thus we create the system

$$\begin{aligned}\frac{\widehat{a}+\widehat{b}}{2} &= \frac{1}{n}\sum_{i=1}^n X_i \\ \frac{\widehat{a}^2+\widehat{a}\widehat{b}+\widehat{b}^2}{3} &= \frac{1}{n}\sum_{i=1}^n X_i^2.\end{aligned}$$

Through substituting $\widehat{b} = 2\overline{X}_n - \widehat{a}$, we obtain the equation

$$\widehat{a}^2 + \widehat{a}(2\overline{X}_n - \widehat{a}) + (2\overline{X}_n - \widehat{a})^2 = \frac{3}{n}\sum_{i=1}^n X_i^2$$

which can be rearranged into the quadratic

$$\widehat{a}^2 - 2\overline{X}_n\widehat{a} + \left(4\overline{X}_n^2 - \frac{3}{n}\sum_{i=1}^n X_i^2\right) = 0.$$

Via the quadratic formula, we obtain

$$\begin{aligned}\widehat{a} &= \frac{2\overline{X}_n \pm \sqrt{4\overline{X}_n^2 - 4\left(4\overline{X}_n^2 - \frac{3}{n}\sum_{i=1}^n X_i^2\right)}}{2} \\ &= \overline{X}_n \pm \sqrt{-3\overline{X}_n^2 + \frac{3}{n}\sum_{i=1}^n X_i^2} \\ &= \overline{X}_n \pm \sqrt{3}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2}.\end{aligned}$$

Let $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$ be the sample variance. We claim that $S^2 = \frac{1}{n}\left(\sum_{i=1}^n X_i^2\right) - \overline{X}_n^2$ also.

Note that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}_n+\overline{X}_n^2) \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2-\frac{1}{n}\sum_{i=1}^{n}2X_i\overline{X}_n+\frac{1}{n}\sum_{i=1}^{n}\overline{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2-\frac{2\overline{X}_n}{n}\left(\sum_{i=1}^{n}X_i\right)+\overline{X}_n^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}_n^2
\end{aligned}$$

and thus it follows that

$$\widehat{a}=\overline{X}_n\pm\sqrt{3}S$$

and

$$\widehat{b}=\overline{X}_n+\sqrt{3}S.$$

b) To find the MLE $\widehat{a}$ and $\widehat{b}$, we find the likelihood function.

We have

$$\begin{aligned}
\mathcal{L}_n(\theta) &= \prod_{i=1}^{n}f(X_i;\theta) \\
&= \prod_{i=1}^{n}\frac{1}{b-a} \\
&= \frac{1}{(b-a)^n}.
\end{aligned}$$

We can maximize the likelihood directly in this case; simply take $\widehat{b}=\max(X_1,\ldots,X_n)$ and $\widehat{a}=\min(X_1,\ldots,X_n)$.

c) Since we have $\tau=\frac{a+b}{2}$, the MLE of τ is simply $\frac{\widehat{a}+\widehat{b}}{2}$, so $\frac{\min(X_i)+\max(X_i)}{2}$.

d) The nonparametric plug-in estimator of $\tau=\int xdF(x)$ is $\tilde{\tau}=\int xd\widehat{F}_n(x)=\sum_x xf(x)$, where $f(x)=\frac{1}{n}$ for $x=X_1,\ldots,X_n$ and 0 otherwise. That is, $\tilde{\tau}=\overline{X}_n$.

To find the MSE of $\tilde{\tau}$, we note first that the estimator is unbiased (i.e. $\mathbb{E}(\tilde{\tau})=\mathbb{E}(\overline{X}_n)=\frac{1}{n}\cdot n\cdot\mathbb{E}(X)=\frac{a+b}{2}=\tau$.) Therefore the MSE is just $\mathbb{V}(\widehat{\tau})$. But we have

$$\begin{aligned}
\mathbb{V}(\widehat{\tau}) &= \mathbb{V}(\overline{X}_n) \\
&= \frac{1}{n^2}\cdot n\cdot\mathbb{V}(X) \\
&= \frac{(b-a)^2}{12n} \\
&= \frac{4}{120}=\frac{1}{30}.
\end{aligned}$$

The rest is in a separate file.

□

Problem 9.3. Let $X_1,\ldots,X_n\sim N(\mu,\sigma^2)$. Let τ be the .95 percentile; i.e. $\mathbb{P}(X<\tau)=.95$.

a) Find the MLE of τ.

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

Solution. As usual, set $\theta = (\mu, \sigma)$.

a) We have $\mathbb{P}(X < \tau) = .95$, so therefore

$$\begin{aligned} .95 &= \mathbb{P}\left(\frac{X-\mu}{\sigma} < \frac{\tau-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{\tau-\mu}{\sigma}\right) \end{aligned}$$

and it follows that

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

Therefore

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

where $\widehat{\sigma}$ and $\widehat{\mu}$ are the MLEs for σ and μ. We previously computed that $\widehat{\mu} = \overline{X}_n$ and $\widehat{\sigma} = S$, where $S^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$, so the MLE is

$$\widehat{\tau} = \Phi^{-1}(.95)\cdot S + \overline{X}_n.$$

b) We will use the multiparameter delta method, with $g(\theta) = g(\mu,\sigma) = \Phi^{-1}(.95)\cdot\sigma + \mu = \tau$.

In Problem 9.8, we found that the Fisher Information matrix is $I_n(\mu,\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2}\end{pmatrix}$. Therefore

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

so

$$\widehat{J}_n = \begin{pmatrix} \frac{\widehat{\sigma}^2}{n} & 0 \\ 0 & \frac{\widehat{\sigma}^2}{2n}\end{pmatrix}.$$

Now, we have

$$\begin{aligned} \nabla g &= \begin{pmatrix} \frac{\partial g}{\partial \mu} \\ \frac{\partial g}{\partial \sigma}\end{pmatrix} \\ &= \begin{pmatrix} 1 \\ \Phi^{-1}(.95)\end{pmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} \widehat{\text{se}}(\widehat{\tau}) &= \sqrt{(\widehat{\nabla} g)^T \widehat{J}_n (\widehat{\nabla} g)} \\ &= \sqrt{\begin{pmatrix} 1 & \Phi^{-1}(.95)\end{pmatrix}\begin{pmatrix} \frac{\widehat{\sigma}^2}{n} & 0 \\ 0 & \frac{\widehat{\sigma}^2}{2n}\end{pmatrix}\begin{pmatrix} 1 \\ \Phi^{-1}(.95)\end{pmatrix}} \\ &= \frac{\widehat{\sigma}}{\sqrt{n}}\sqrt{1 + \frac{1}{2}\Phi^{-1}(.95)^2}. \end{aligned}$$

A $1-\alpha$ confidence interval for τ is thus

$$(\widehat{\tau} - z_{\alpha/2}\widehat{\text{se}}(\widehat{\tau}), \widehat{\tau} + z_{\alpha/2}\widehat{\text{se}}(\widehat{\tau})).$$

$\square$

Problem 9.4. Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$. Show that the MLE is consistent.

Solution. First, we find the MLE.

We have

$$\begin{aligned}
\mathcal{L}_n(\theta) &= \prod_{i=1}^n f(X_i;\theta) \\
&= \prod_{i=1}^n \frac{1}{\theta} \\
&= \frac{1}{\theta^n}
\end{aligned}$$

so to maximize the likelihood we want to minimize θ, so $\widehat{\theta} = \max(X_i)$ is the MLE.

Now we want to show that $\widehat{\theta} \xrightarrow{P} \theta$, where θ is the true value of the parameter. That is, we want to show that $\mathbb{P}(|\widehat{\theta} - \theta| < \epsilon) \to 0$ as $n \to \infty$.

We have

$$\begin{aligned}
\mathbb{P}(|\widehat{\theta} - \theta| > \epsilon) &= \mathbb{P}(X_1 \le \theta - \epsilon)\mathbb{P}(X_2 \le \theta - \epsilon) \cdots \mathbb{P}(X_n \le \theta - \epsilon) \\
&= \left(\frac{\theta - \epsilon}{\theta}\right)^n \to 0
\end{aligned}$$

as $n \to \infty$, as desired. □

Problem 9.5. Let $X_1, \ldots, X_n \sim \text{Pois}(\lambda)$. Find the method of moments estimator, the maximum likelihood estimator, and the Fisher information $I(\lambda)$.

Solution. We have $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ for all nonnegative integers x.

The first moment is simply $\mathbb{E}_\lambda(X) = \lambda$, and the first sample moment is $\frac{1}{n}\sum_{i=1}^n X_i$. So the method of moments estimator would be $\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$.

The likelihood function is

$$\begin{aligned}
\mathcal{L}_n(\theta) &= \prod_{i=1}^n f(X_i;\theta) \\
&= \prod_{i=1}^n e^{-\theta}\frac{\theta^{x_i}}{x_i!} \\
&= e^{-n\theta} \cdot \frac{\theta^{\sum x_i}}{\prod x_i!}.
\end{aligned}$$

We now consider the log-likelihood

$$\begin{aligned}
\log \mathcal{L}_n(\theta) = \ell_n(\theta) &= \log\left(e^{-n\theta} \cdot \frac{\theta^{\sum x_i}}{\prod x_i!}\right) \\
&= \log e^{-n\theta} + \log \theta^{\sum x_i} - \log \prod x_i! \\
&= -n\theta + \log\theta \cdot \left(\sum_{i=1}^n x_i\right) - \left(\sum_{i=1}^n \log x_i!\right)
\end{aligned}$$

and setting the derivative equal to 0, we obtain

$$0 = \frac{d}{d\theta}\ell_n(\theta) = -n + \frac{1}{\theta}\sum_{i=1}^n x_i.$$

Thus we obtain $\theta = \frac{1}{n}\sum_{i=1}^n X_i$, so the maximum likelihood estimator is also $\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^n X_i$.

Note first that $\log f(x;\lambda) = -\lambda + x\log\lambda - \log(x!)$. The Fisher information is

$$\begin{aligned}
I(\lambda) &= -\mathbb{E}_\lambda\left(\frac{\partial^2 \log f(X;\lambda)}{\partial\lambda^2}\right) \\
&= -\mathbb{E}_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\left(-\lambda + x\log\lambda - \log(x!)\right)\right) \\
&= -\mathbb{E}_\lambda\left(-\frac{x}{\lambda^2}\right) \\
&= \sum_{x=0}^{\infty}\frac{x}{\lambda^2}\cdot e^{-\lambda}\cdot\frac{\lambda^x}{x!} \\
&= \frac{e^{-\lambda}}{\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = \frac{1}{\lambda}.
\end{aligned}$$

We are done. $\square$

Problem 9.6. Let $X_1,\dots,X_n \sim N(\theta,1)$. Define

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

Let $\psi = \mathbb{P}(Y_1 = 1)$.

a) Find the maximum likelihood estimator $\widehat{\psi}$ of ψ.

b) Find an approximate 95 percent confidence interval for ψ.

c) Define $\widetilde{\psi}_n = \frac{1}{n}\sum_i Y_i$. Show that $\widetilde{\psi}$ is a consistent estimator of ψ.

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

e) Suppose that the data are not really normal. Show that $\widehat{\psi}$ is not consistent. What, if anything, does $\widehat{\psi}$ converge to?

Solution. a) We have

$$\begin{aligned}
\psi &= \mathbb{P}(Y_1 = 1) \\
&= \mathbb{P}(X_1 > 0) \\
&= \mathbb{P}(X_1 - \theta > -\theta) \\
&= 1 - \Phi(-\theta) = \Phi(\theta).
\end{aligned}$$

Now we'll find the maximum likelihood estimator $\widehat{\theta}$. We have

$$\begin{aligned}
\mathcal{L}_n(\theta) &= \prod_{i=1}^{n} f(X_i;\theta) \\
&= \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}(x_i-\theta)^2\right] \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[\sum_{i=1}^{n} -\frac{1}{2}(x_i-\theta)^2\right]
\end{aligned}$$

so

$$\ell_n(\theta) = n\log\left(\frac{1}{\sqrt{2\pi}}\right) - \sum_{i=1}^{n}\frac{1}{2}(x_i-\theta)^2.$$

Then

$$\begin{aligned}\ell_n'(\theta) &= -\sum_{i=1}^{n} \frac{1}{2} \cdot 2 \cdot (x_i - \theta) \cdot (-1) \\ &= \sum_{i=1}^{n} (x_i - \theta)\end{aligned}$$

and thus setting $\ell_n'(\theta) = 0$ yields $\widehat{\theta} = \overline{X}_n$. So $\widehat{\psi} = \Phi(\widehat{\theta}) = \Phi(\overline{X}_n)$.

b) We know that $\psi = \Phi(\theta)$ and thus by the Delta method a 95 percent confidence interval for ψ is

$$C_n = \left(\widehat{\psi} - z_{\alpha/2}\widehat{\text{se}}(\widehat{\psi}), \widehat{\psi} + z_{\alpha/2}\widehat{\text{se}}(\widehat{\psi})\right)$$

where $\widehat{\text{se}}(\widehat{\psi}) = |\Phi'(\widehat{\theta})|\widehat{\text{se}}(\widehat{\theta})$.

We have $\widehat{\text{se}}(\widehat{\theta}) = \sqrt{\frac{1}{I_n(\widehat{\theta}_n)}}$. Moreover, to compute the Fisher information, we note also that $f(X;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x-\theta)^2\right)$ so $\log f(X;\theta) = \log(\frac{1}{\sqrt{2\pi}}) - \frac{1}{2}(x-\theta)^2$. But

$$\begin{aligned}I(\theta) &= -\mathbb{E}_\theta\left(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right) \\ &= -\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\left(-\frac{1}{2}(x-\theta)\cdot 2\cdot(-1)\right)\right) \\ &= -\mathbb{E}_\theta\left(-1\right) = 1.\end{aligned}$$

Thus $\widehat{\text{se}}(\widehat{\theta}) = \sqrt{\frac{1}{n}}$. Moreover, $|\Phi'(\widehat{\theta})| = \left|\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(\overline{X}_n)^2\right)\right|$. So

$$\widehat{\text{se}}(\widehat{\psi}) = \sqrt{\frac{1}{n}} \cdot \left|\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(\overline{X}_n)^2\right)\right|.$$

Thus a 95 percent confidence interval is

$$\Phi(\overline{X}_n) \pm z_{\alpha/2} \cdot \sqrt{\frac{1}{n}} \cdot \left|\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(\overline{X}_n)^2\right)\right|.$$

c) Showing $\widetilde{\psi}_n$ is a consistent estimator of ψ is equivalent to showing that $\widetilde{\psi}_n \xrightarrow{P} \psi$. But $\mathbb{E}(Y) = \mathbb{P}(Y = 1)\cdot 1 + \mathbb{P}(Y=0)\cdot 0 = \psi$. Thus, by the Weak Law of Large Numbers, $\overline{Y}_n \to \mathbb{E}(Y) = \psi$, as desired.

d) We want to find the asymptotic relative efficiency of $\overline{Y}_n$ to $\Phi(\overline{X}_n)$.

We know that $\Phi(\overline{X}_n) - \psi \xrightarrow{\text{d}} N(0, \widehat{\text{se}}(\widehat{\psi})^2)$. We also know that $\overline{Y}_n - \psi \xrightarrow{\text{d}} N(0, \mathbb{V}(\overline{Y}_n))$.

Therefore

$$\begin{aligned}\text{ARE}(\widetilde{\psi}, \widehat{\psi}) &= \frac{\mathbb{V}(\overline{Y}_n)}{\widehat{\text{se}}(\widehat{\psi})^2} \\ &= \frac{\frac{1}{n^2}\cdot n \cdot \mathbb{V}(Y)}{\frac{1}{n}\cdot\frac{1}{2\pi}\cdot\exp(-(\overline{X}_n)^2)} \\ &= \frac{\psi(1-\psi)}{\frac{1}{2\pi}\exp(-(\overline{X}_n)^2)} \\ &= \frac{\psi(1-\psi)}{\frac{1}{2\pi}\exp(-\theta^2)}.\end{aligned}$$

Note that the ARE is a property of estimators under the true model, so we can make the substitutions for θ and ψ.

e) Suppose the data are not really normal, but that we're still trying to use the estimator $\widehat{\psi} = \Phi(\overline{X}_n)$.

We have $\psi = \mathbb{P}(Y_1 = 1) = \mathbb{P}(X_1 > 0)$. We want to show that $\Phi(\overline{X}_n)$ does not converge in probability to $\mathbb{P}(X_1 > 0)$.

Suppose even if the data are not really normal that they still are drawn from a distribution with mean θ; then $\overline{X}_n \xrightarrow{P} \theta$ so $\widehat{\psi} = \Phi(\overline{X}_n) \xrightarrow{P} \Phi(\theta)$.

Thus we want to show that if the data are not normal, then $\Phi(\theta) \neq \mathbb{P}(X_1 > 0)$. But it was only because of our normality assumption that we could state $\mathbb{P}(X_1 > 0) = \mathbb{P}(X_1 - \theta > -\theta) = \mathbb{P}(Z > -\theta) = \Phi(\theta)$, so without the normality assumption, we don't have this equality.

$\square$

Problem 9.7. n_1 people are given treatment 1 and n_2 people are given treatment 2. Let X_1 be the number of people on treatment 1 who respond favorably to the treatment and let X_2 be the number of people on treatment 2 who respond favorably. Assume that $X_1 \sim \text{Binomial}(n_1, p_1)$ and $X_2 \sim \text{Binomial}(n_2, p_2)$. Let $\psi = p_1 - p_2$.

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

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

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

Solution. Let $\theta = (p_1, p_2)$.

a) We have

$$\mathcal{L}(\theta) = \binom{n_1}{x_1} p_1^{x_1}(1-p_1)^{n_1-x_1} \cdot \binom{n_2}{x_2} p_2^{x_2}(1-p_2)^{n_2-x_2}$$

so

$$\ell(\theta) = \log\binom{n_1}{x_1} + \log\binom{n_2}{x_2} + x_1\log(p_1) + x_2\log(p_2) + (n_1 - x_1)\log(1-p_1) + (n_2 - p_2)\log(1-p_2).$$

Then

$$\frac{\partial}{\partial p_1}\ell(\theta) = \frac{x_1}{p_1} - \frac{n_1 - x_1}{1 - p_1}$$

so setting this equal to 0 yields $\widehat{p}_1 = \frac{x_1}{n_1}$, and similarly we obtain $\widehat{p}_2 = \frac{x_2}{n_2}$. Thus via the equivariance property of the MLE we obtain that $\widehat{\psi} = \widehat{p}_1 - \widehat{p}_2 = \frac{x_1}{n_1} - \frac{x_2}{n_2}$.

b) The Fisher information matrix is

$$I(p_1, p_2) = -\begin{bmatrix} \mathbb{E}_\theta(H_{11}) & \mathbb{E}_\theta(H_{12}) \\ \mathbb{E}_\theta(H_{21}) & \mathbb{E}_\theta(H_{22}) \end{bmatrix}$$

with $H_{jj} = \frac{\partial^2 \ell_n}{\partial \theta_j^2}$ and $H_{jk} = \frac{\partial^2 \ell_n}{\partial\theta_j \partial\theta_k}$. Here, again, $\theta = (p_1, p_2)$, so $\theta_1 = p_1$ and $\theta_2 = p_2$.

Having previously computed $\ell(\theta)$ in part (a), we note that

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta_1^2} &= \frac{\partial}{\partial p_1}\left(\frac{x_1}{p_1} - \frac{n_1 - x_1}{1 - p_1}\right) \\ &= -\frac{x_1}{p_1^2} - \frac{n_1 - x_1}{(1-p_1)^2}\end{aligned}$$

and similarly

$$\frac{\partial^2 \ell}{\partial \theta_2^2} = \frac{x_2}{p_2^2} - \frac{n_2 - x_2}{(1-p_2)^2}.$$

Next, we have

$$\begin{aligned}\frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1} &= \frac{\partial}{\partial p_2}\left(\frac{x_1}{p_1} - \frac{n_1 - x_1}{1-p_1}\right)\\ &= 0.\end{aligned}$$

Similarly we note that $H_{12} = 0$. Then, we compute

$$\begin{aligned}\mathbb{E}_\theta(H_{11}) &= \mathbb{E}_\theta\left(-\frac{x_1}{p_1^2} - \frac{n_1 - x_1}{(1-p_1)^2}\right)\\ &= -\mathbb{E}_\theta\left(\frac{x_1}{p_1^2}\right) - \mathbb{E}_\theta\left(\frac{n_1 - x_1}{(1-p_1)^2}\right)\\ &= -\frac{n_1 p_1}{p_1^2} - \frac{n_1 - n_1 p_1}{(1-p_1)^2}\\ &= -\frac{n_1}{p_1} - \frac{n_1}{1-p_1}.\end{aligned}$$

Then analogously we get

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

c) With $\psi = g(p_1, p_2) = p_1 - p_2$, we have $\widehat{\text{se}}(\widehat{\psi}) = \sqrt{(\widehat{\nabla} g)^T \widehat{J}_n (\widehat{\nabla} g)}$.
Having derived the Fisher Information Matrix previously, we note that

$$I_n^{-1}(p_1, p_2) = J_n(p_1, p_2) = \begin{bmatrix} \frac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{n_2} \end{bmatrix}.$$

Then

$$\begin{aligned}J_n(\widehat{p}_1, \widehat{p}_2) &= \begin{bmatrix} \frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} & 0 \\ 0 & \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2} \end{bmatrix}\\ &= \begin{bmatrix} \frac{x_1(n_1-x_1)}{n_1^2} & 0 \\ 0 & \frac{x_2(n_2-x_2)}{n_2^2} \end{bmatrix}.\end{aligned}$$

Next, we have

$$\begin{aligned}\nabla g &= \begin{pmatrix} \frac{\partial g}{\partial p_1} \\ \frac{\partial g}{\partial p_2} \end{pmatrix}\\ &= \begin{pmatrix} 1 \\ -1 \end{pmatrix}\end{aligned}$$

so it follows that

$$\begin{aligned}\widehat{\text{se}}(\widehat{\psi}) &= \sqrt{\begin{pmatrix} 1 & -1 \end{pmatrix} J_n(\widehat{p}_1, \widehat{p}_2) \begin{pmatrix} 1 \\ -1 \end{pmatrix}}\\ &= \sqrt{\begin{pmatrix} \frac{x_1(n_1-x_1)}{n_1^2} & -\frac{x_2(n_2-x_2)}{n_2^2} \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix}}\\ &= \sqrt{\frac{x_1(n_1 - x_1)}{n_1^2} + \frac{x_2(n_2 - x_2)}{n_2^2}}\end{aligned}$$

or alternatively,

$$\widehat{\text{se}}(\widehat{\psi}) = \sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{n_1} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n_2}}.$$

$\square$

Problem 9.8. Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Let $\tau = g(\mu, \sigma) = \sigma/\mu$. Find the Fisher Information matrix $I_n(\mu, \sigma)$.

Solution. Let $\theta = (\mu, \sigma)$ as usual.

We previously computed (in a textbook example) that the MLEs of μ and σ are $\mu = \overline{X}_n$ and $\sigma = S = \sqrt{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2}$, with

$$\ell_n(\theta) = -n \log \sigma - \frac{nS^2}{2\sigma^2} - \frac{n(\overline{X} - \mu)^2}{2\sigma^2}.$$

Then

$$\begin{aligned} H_{11} &= \frac{\partial^2 \ell_n}{\partial \mu^2} \\ &= \frac{\partial}{\partial \mu}\left(\frac{n \cdot 2(\overline{X} - \mu)}{2\sigma^2}\right) \\ &= -\frac{n}{\sigma^2}. \end{aligned}$$

Also,

$$\begin{aligned} H_{22} &= \frac{\partial^2 \ell_n}{\partial \sigma^2} \\ &= \frac{\partial}{\partial \sigma}\left(-\frac{n}{\sigma} + \frac{nS^2}{\sigma^3} + \frac{n(\overline{X} - \mu)^2}{\sigma^3}\right) \\ &= \frac{n}{\sigma^2} - \frac{3nS^2}{\sigma^4} - \frac{3n(\overline{X} - \mu)^2}{\sigma^4}. \end{aligned}$$

Next,

$$\begin{aligned} H_{12} &= \frac{\partial^2 \ell_n}{\partial \mu \partial \sigma} \\ &= \frac{\partial}{\partial \mu}\left(-\frac{n}{\sigma} + \frac{nS^2}{\sigma^3} + \frac{n(\overline{X} - \mu)^2}{\sigma^3}\right) \\ &= -\frac{n \cdot 2(\overline{X} - \mu)}{\sigma^3} \end{aligned}$$

and in this case, the order of the partials doesn't matter, so $H_{21} = -\frac{2n(\overline{X} - \mu)}{\sigma^3}$ too.

Now we have $\mathbb{E}_\theta(H_{11}) = -\frac{n}{\sigma^2}$. As $\mathbb{E}(\overline{X}) = \mu$, we have $\mathbb{E}_\theta(H_{12}) = \mathbb{E}_\theta(H_{21}) = 0$. Finally, we have

$$\begin{aligned} \mathbb{E}_\theta(H_{22}) &= \mathbb{E}_\theta\left(\frac{n}{\sigma^2} - \frac{3nS^2}{\sigma^4} - \frac{3n(\overline{X} - \mu)^2}{\sigma^4}\right) \\ &= \mathbb{E}_\theta\left(\frac{n}{\sigma^2}\right) - \mathbb{E}_\theta\left(\frac{3nS^2}{\sigma^4}\right) - \mathbb{E}_\theta\left(\frac{3n(\overline{X} - \mu)^2}{\sigma^4}\right) \\ &= \frac{n}{\sigma^2} - \frac{3n}{\sigma^4}\mathbb{E}_\theta(S^2) - \frac{3n}{\sigma^4}\mathbb{E}((\overline{X} - \mu)^2). \end{aligned}$$

But

$$\begin{aligned} \mathbb{E}_\theta(S^2) &= \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2\right) \\ &= \mathbb{E}_\theta\left(\frac{n-1}{n} \cdot S_n^2\right) \\ &= \frac{n-1}{n} \cdot \sigma^2 \end{aligned}$$

where S_n^2 is the sample variance.

Moreover, we have

$$\begin{aligned}
\mathbb{E}((\overline{X}-\mu)^2) &= \mathbb{E}(\overline{X}^2 - 2\mu\overline{X} + \mu^2) \\
&= \mathbb{E}(\overline{X}^2) - \mu^2 \\
&= \mathbb{V}(\overline{X}) + \mathbb{E}(\overline{X})^2 = \mu^2 \\
&= \frac{\sigma^2}{n} + \mu^2 - \mu^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Making these substitutions, we obtain

$$\mathbb{E}_\theta(H_{22}) = \frac{n}{\sigma^2} - \frac{3(n-1)}{\sigma^2} - \frac{3}{\sigma^2} = -\frac{2n}{\sigma^2}.$$

Therefore the Fisher Information matrix is

$$I_n(\mu,\theta) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix}.$$

$\square$

Problem 9.9. Let $X_1,\ldots,X_n \sim N(\mu,1)$. Let $\theta = e^\mu$ and let $\widehat{\theta} = e^{\overline{X}}$ be the MLE. Use the delta method to get $\widehat{\text{se}}$ and a 95 percent confidence interval for θ.

Solution. Here we have that $\theta = e^\mu$, so $\widehat{\text{se}}(\widehat{\theta}) = |g'(\widehat{\mu})|\widehat{\text{se}}(\widehat{\mu})$ with $g(x) = e^x$.

To find $\widehat{\text{se}}(\widehat{\mu})$, we must compute the Fisher information $I_n(\widehat{\mu})$. Noting that

$$\log f(X;\mu) = \log\left[\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(x-\mu)^2\right)\right] = \log\frac{1}{\sqrt{2\pi}} - \frac{1}{2}(x-\mu)^2,$$

we have

$$\begin{aligned}
I(\mu) &= -\mathbb{E}_\mu\left(\frac{\partial^2 \log f(X;\mu)}{\partial\mu^2}\right) \\
&= -\mathbb{E}_\mu\left(\frac{\partial}{\partial\mu}\left(-\frac{1}{2}\cdot 2\cdot(x-\mu)\cdot(-1)\right)\right) \\
&= 1.
\end{aligned}$$

Thus $\widehat{\text{se}}(\widehat{\mu}) = \sqrt{1/I_n(\widehat{\mu})} = \sqrt{\frac{1}{n}}$ so

$$\begin{aligned}
\widehat{\text{se}}(\widehat{\theta}) &= |g'(\widehat{\mu})|\widehat{\text{se}}(\widehat{\mu}) \\
&= |e^{\widehat{\mu}}|\cdot\sqrt{\frac{1}{n}} \\
&= e^{\overline{X}}\cdot\sqrt{\frac{1}{n}}.
\end{aligned}$$

Thus a 95 percent confidence interval is

$$C_n = \left(\widehat{\theta} - z_{0.025}e^{\overline{X}}\cdot\sqrt{\frac{1}{n}}, \widehat{\theta} + z_{0.025}e^{\overline{X}}\cdot\sqrt{\frac{1}{n}}\right).$$

$\square$