Solutions to selected exercises from Chapter 9 of *Wasserman — All of Statistics*

(1) The PDF is $f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$ for $x > 0$ and $f(x) = 0$ otherwise. The first moment is $\epsilon_1 = \alpha\beta$, as calculated in the text. The second moment is

$$\epsilon_2 = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha+1} e^{-x/\beta} dx.$$

Integration by parts yields

$$\int x^{\alpha+1} e^{-x/\beta} dx = -\beta x^{\alpha+1} e^{-x/\beta} + \beta(\alpha+1) \int x^\alpha e^{-x/\beta} dx$$

where we note that the first term is zero at $x = 0$ and approaches 0 as $x \to \infty$. Thus

$$\epsilon_2 = \beta(\alpha+1) \int_0^\infty \frac{x^\alpha e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} dx = \beta(\alpha+1)\alpha\beta = \beta^2(\alpha^2 + \alpha)$$

where again we have used the fact that the first moment is $\alpha\beta$.

The sample moments are

$$\hat{\epsilon}_1 = \frac{1}{n} \sum X_i = \overline{X_n} \text{ and } \hat{\epsilon}_2 = \frac{1}{n} \sum X_i^2.$$

Our system of equations for the method of moments estimators $\hat{\alpha}, \hat{\beta}$ is

$$\hat{\alpha}\hat{\beta} = \hat{\epsilon}_1, \quad \hat{\beta}^2(\hat{\alpha}^2 + \hat{\alpha}) = \hat{\epsilon}_2.$$

Setting $\hat{\beta} = \hat{\epsilon}_1/\hat{\alpha}$, substituting into the second equation, and solving for $\hat{\alpha}$ yields

$$\hat{\alpha} = \frac{\hat{\epsilon}_1^2}{\hat{\epsilon}_2 - \hat{\epsilon}_1^2}, \quad \hat{\beta} = \frac{\hat{\epsilon}_2 - \hat{\epsilon}_1^2}{\hat{\epsilon}_1}.$$

Estimating the mean as $\hat{\mu} = \hat{\epsilon}_1$ and the variance as $\hat{\sigma}^2 = \hat{\epsilon}_2 - \hat{\epsilon}_1^2$, we may rewrite this as

$$\hat{\alpha} = \frac{\hat{\mu}^2}{\hat{\sigma}^2}, \quad \hat{\beta} = \frac{\hat{\sigma}^2}{\hat{\mu}}.$$

(2) (a) The first and second moments are

$$\alpha_1(a,b) = \frac{1}{2} \frac{b^2 - a^2}{b - a} = \frac{1}{2}(a+b), \quad \alpha_2(a,b) = \frac{1}{3} \frac{b^3 - a^3}{b - a} = \frac{1}{3}(b^2 + ab + a^2).$$

The method of moments estimators satisfy the system

$$\frac{1}{2}(\hat{a} + \hat{b}) = \hat{\alpha}_1, \quad \frac{1}{3}(\hat{a}^2 + \hat{a}\hat{b} + \hat{b}^2) = \hat{\alpha}_2$$

where $\hat{\alpha}_1, \hat{\alpha}_2$ are the sample moments.

Solving the first equation for $\hat{b}$ yields $\hat{b} = 2\hat{\alpha}_1 - \hat{a}$. Plugging this into the second equation and simplifying yields

$$\hat{a}^2 - 2\hat{\alpha}_1\hat{a} + 4\hat{\alpha}_1^2 - 3\hat{\alpha}_2 = 0$$

which is a degree two polynomial in $\hat{a}$. Then $\hat{a}$ is one of the two roots $\hat{\alpha}_1 \pm \sqrt{3\hat{\alpha}_2 - 3\hat{\alpha}_1^2}$. Thus $\hat{b} = \hat{\alpha}_1 \mp \sqrt{3\hat{\alpha}_2 - 3\hat{\alpha}_1^2}$. However, $a < b$ so $\hat{b}$ must be the larger of the two roots and $\hat{a}$ must be the smaller. I.e.

$$\hat{a} = \hat{\mu} - \sqrt{3}\hat{\sigma}, \quad \hat{b} = \hat{\mu} + \sqrt{3}\hat{\sigma}$$

where $\hat{\mu} = \hat{\alpha}_1$ and $\hat{\sigma}^2 = \hat{\alpha}_2 - \hat{\alpha}_1^2$ are the sample mean and variance, respectively.

(b) Set $\theta = (a, b)$ and let $\hat{\theta} = (\hat{a}, \hat{b})$ be the MLE. The likelihood of θ is

$$\mathcal{L}_n(\theta) = \frac{1}{(b-a)^n} \prod \chi_{[a,b]}(X_i)$$

where $\chi_\cdot$ denotes an indicator function. This is equal to 0 if $X_i < a$ or $X_i > b$ for some i. Otherwise it is equal to $1/(b-a)^n$. Thus $\mathcal{L}_n(\theta)$ is maximized when b is as small as possible, without being less than any X_i and a is as large as possible, without being larger than any X_i. I.e.

$$\hat{a} = \min_i X_i, \quad \hat{b} = \max_i X_i.$$

(c) We have $\tau = \mathbb{E}(X_i) = (a+b)/2$. By equivariance of the MLE, the MLE of τ is

$$\hat{\tau} = \frac{\hat{a} + \hat{b}}{2} = \frac{\min X_i + \max X_i}{2}.$$

(d) See the Jupyter Notebook 2.ipynb.

(3) See the Jupyter Notebook 3.ipynb.

(4) The MLE for θ is $Y_n = \max\{X_1, \ldots, X_n\}$. For any $\epsilon > 0$,

$$\mathbb{P}(|\theta - Y_n| > \epsilon) < \mathbb{P}(Y_n < \theta - \epsilon) = \mathbb{P}(X_1 < \theta - \epsilon) \cdots \mathbb{P}(X_n < \theta - \epsilon) = \frac{(\theta - \epsilon)^n}{\theta^n}.$$

This converges to 0 as $n \to \infty$. So Y_n is a consistent estimator of θ.

(5) For the method of moments, there is only one moment to compute: the first moment. Thus the sample moment $\overline{X_n} = \frac{1}{n}\sum X_i$ must be equal to the first moment $\mathbb{E}_{\hat{\lambda}}(X) = \hat{\lambda}$. So the method of moments estimator $\hat{\lambda} = \overline{X_n}$.

For maximum likelihood, we have $f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!}$. The likelihood of λ is

$$\mathcal{L}_n(\lambda) = \frac{\lambda^{\sum X_i} e^{-\lambda}}{X_1! \cdots X_n!} \text{ so } \log \mathcal{L}_n(\lambda) = \left(\sum X_i\right) \log \lambda - n\lambda + c$$

where c is a constant and the derivative with respect to λ is $\sum X_i/\lambda - n$ and this is equal to 0 exactly at $\lambda = \sum X_i/n = \overline{X_n}$, which is the unique maximum. So the maximum likelihood estimator $\hat{\lambda} = \overline{X_n}$.

Finally, the score is

$$s(x;\lambda) = \frac{\partial \log f(x;\lambda)}{\partial \lambda} = \frac{\partial}{\partial \lambda}(x \log \lambda - \lambda - \log(x!)) = \frac{x}{\lambda} - 1 \text{ and } \frac{\partial s}{\partial \lambda} = -\frac{x}{\lambda^2}$$

The Fisher information is

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

(6) (a) The test statistics is $\psi = \mathbb{P}(Y_1 = 1) = \mathbb{P}(X_1 > 0) = 1 - \Phi(-\theta)$. The MLE for θ is $\overline{X_n}$. Hence the MLE for ψ is $\hat{\psi} = 1 - \Phi(-\overline{X_n})$.

(b) The standard error for $\overline{X_n}$ is $1/\sqrt{n}$. By the delta method, the standard error for $\hat{\psi}$ is approximately $\phi(-\overline{X}_n)\hat{\mathrm{se}}(\overline{X}_n) = \phi(-\overline{X}_n)/\sqrt{n}$ where ϕ is the PDF for the standard normal. An approximate 95% confidence interval is

$$1 - \Phi(-\overline{X_n}) \pm 2\frac{\phi(\overline{X}_n)}{\sqrt{n}}.$$

(c) By the Law of Large Numbers, $\tilde{\psi}_n = \frac{1}{n}\sum Y_i$ converges in probability to the mean of any Y_i, which is the probability that $Y_1 = 1$.

(d) The standard error of $\hat{\psi}$ is $\phi(-\overline{X}_n)/\sqrt{n}$. The standard error of $\tilde{\psi}$ is the square root of $\frac{1}{n^2}\sum \mathbb{V}(Y_i)$. Note Y_i is a Bernoulli(p) random variable with $p = 1 - \Phi(-\theta)$. Therefore

$$\mathrm{se}(\tilde{\psi}) = \frac{1}{\sqrt{n}}\Phi(-\theta)^{1/2}(1 - \Phi(-\theta))^{1/2}.$$

By asymptotic normality of the MLE, $\sqrt{n}(\hat{\psi} - \psi)$ converges in probability to $N(0, \phi(-\theta))$. By the Central Limit Theorem, $\sqrt{n}(\tilde{\psi} - \psi)$ converges in probability to $N(0, \Phi(-\theta)^{1/2}(1 - \Phi(-\theta))^{1/2})$. The asymptotic relative efficiency is

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

(7) (a) The joint PDF for X_1, X_2 is

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2) \text{ and } \log f_{X_1,X_2}(x_1, x_2) = \log f_{X_1}(x_1) + \log f_{X_2}(x_2).$$

This shows that the MLE for the pair (p_1, p_2) is $(\hat{p}_1, \hat{p}_2)$ where $\hat{p}_i$ is the MLE for p_i. So let's compute $\hat{p}_i$. If $X \sim \text{Binomial}(n, p)$ the likelihood is

$$\log f(X; p) = \log \binom{n}{X} + X \log p + (n - X) \log(1 - p)$$

and the derivative with respect to p is

$$\frac{X}{p} - \frac{n - X}{1 - p}.$$

Setting this equal to zero and solving for p yields the MLE $\hat{p} = \frac{X}{n}$. So $\hat{p}_i = \frac{X_i}{n_i}$. By equivariance

$$\hat{\psi} = \hat{p}_1 - \hat{p}_2 = \frac{X_1}{n_1} - \frac{X_2}{n_2}.$$

(b) We have that ℓ_2 is equal to

$$\log \binom{n_1}{X_1} + X_1 \log p_1 + (n_1 - X_1) \log(1 - p_1) + \log \binom{n_2}{X_2} + X_2 \log p_2 + (n_2 - X_2) \log(1 - p_2).$$

In the Fisher information matrix it is easy to see that $H_{12} = H_{21} = 0$ and

$$H_{11} = \frac{\partial}{\partial p_1}\left(\frac{X_1}{p_1} - \frac{n_1 - X_1}{1 - p_1}\right) = -\frac{X_1}{p_1^2} - \frac{n_1 - X_1}{(1 - p_1)^2}$$

and similarly for H_{22}. We have

$$\mathbb{E}(H_{11}) = -\frac{n_1 p_1}{p_1^2} - \frac{n_1 - n_1 p_1}{(1-p_1)^2} = \frac{-n_1}{p_1(1-p_1)}$$

and similarly for $\mathbb{E}(H_{22})$. Thus,

$$I_2(p_1, p_2) = \begin{pmatrix} \frac{n_1}{p_1(1-p_1)} & 0 \\ 0 & \frac{n_2}{p_2(1-p_2)} \end{pmatrix}$$

(c) We have $g(p_1, p_2) = \psi = p_1 - p_2$ so

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

By part (c),

$$J_2(p_1, p_2) = \begin{pmatrix} \frac{p_1(1-p_1)}{n_1} & 0 \\ 0 & \frac{p_2(1-p_2)}{n_2} \end{pmatrix}.$$

By Theorem 9.28, we may estimate the standard error of $\hat{\psi}$ as

$$\hat{\text{se}}(\hat{\psi}) = \sqrt{(\hat{\nabla} g)^T \hat{J}_n (\hat{\nabla} g)} = \left(\frac{\hat{p}_1(1-\hat{p}_1)}{\hat{n}_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{\hat{n}_2} \right)^{1/2}.$$

(d) See the Jupyter Notebook 7.ipynb.

(8) We did this as part of the solution to Exercise (3). The answer is

$$I_n(\theta) = \frac{n}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

See the solution to part b of Exercise (3).

(10) See the Jupyter Notebook 10.ipynb.