1. (24 pts) Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$.

   (a) (6 pts) Find a sufficient statistic for the parameter β, assuming α is known. Is the statistic you found above minimally sufficient? Justify your answer.

   (b) (6 pts) Find the MLE of β assuming α is known.

   (c) (6 pts) Find the Fisher information and construct an approximate 95% normal-based confidence interval for β.

   (d) (6 pts) Find the asymptotic distribution for the MLE of β assuming α is known.

Solution:

(a) Assume α is known.

$$f(x_1, \ldots, x_n; \beta) = \prod_{i=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} X_i^{\alpha-1} e^{-\beta X_i}$$
$$= \underbrace{\prod_{i=1}^{n} X_i^{\alpha-1}}_{h(\mathbf{x})} \underbrace{\left(\frac{\beta^\alpha}{\Gamma(\alpha)}\right)^n \exp\left(-\beta \sum_{i=1}^{n} X_i\right)}_{g(T(\mathbf{x}),\beta)}.$$

Therefore a sufficient statistic for β is $\sum_{i=1}^n X_i$. For minimal sufficiency using the exponential family form:

$$f(x_1, \ldots, x_n; \beta) = \underbrace{\left[\prod_{i=1}^{n} X_i^{\alpha-1}\right] [\Gamma(\alpha)]^{-n}}_{h(\mathbf{x})} \exp\left(\underbrace{-\beta}_{\eta_1(\beta)} \underbrace{\sum_{i=1}^{n} X_i}_{T_1(\mathbf{x})} + \underbrace{\alpha n \log \beta}_{A(\beta)}\right).$$

Thus, $\sum_i X_i$ is minimally sufficient.

(b)

$$\frac{\partial f(x_1, \ldots, x_n; \beta)}{\partial \beta} = \frac{\log(\prod_{i=1}^n X_i^{\alpha-1}) + n\alpha \log \beta - n \log(\Gamma(\alpha)) - \beta \sum_{i=1}^n X_i}{\partial \beta}$$
$$= \frac{n\alpha}{\beta} - \sum_{i=1}^{n} X_i.$$

Setting equal to 0 and solving yields

$$\hat{\beta}_{MLE} = \frac{n\alpha}{\sum_i X_i}.$$

The second derivative is

$$\frac{\partial^2 f(x_1, \ldots, x_n; \beta)}{\partial \beta^2} = -\frac{n\alpha}{\beta^2} < 0.$$

(c) First, find the Fisher information.

$$\begin{aligned}
I(\beta) &= -E_\beta \left[\frac{\partial^2}{\partial \beta^2} \log f(x; \beta) \right] \\
&= -E_\beta \left[\frac{\partial^2}{\partial \beta^2} \log \left(\frac{\beta^\alpha}{\Gamma(\alpha) x^{\alpha-1} e^{-\beta x}} \right) \right] \\
&= -E_\beta \left[\frac{\partial}{\partial \beta} \left(\frac{\alpha}{\beta} - x \right) \right] \\
&= -E_\beta \left[-\frac{\alpha}{\beta^2} \right] = \frac{\alpha}{\beta^2}.
\end{aligned}$$

The asymptotic distribution is thus

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_{MLE} - \beta) \to N(0, 1/I_1(\beta)) \\
\implies &\sqrt{n} \left(\frac{n\alpha}{\sum_{i=1}^n X_i} - \beta \right) \to N(0, \hat{\beta}^2_{MLE}/\alpha).
\end{aligned}$$

An approximate 95% confidence interval is

$$\left| \frac{\sqrt{n}(\frac{n\alpha}{\sum_i X_i} - \beta)}{\sqrt{\frac{\hat{\beta}^2_{MLE}}{\alpha}}} \right| > z_{0.025}.$$

(d) The asymptotic distribution for the MLE of β assuming α is known is

$$\frac{\hat{\beta}_{MLE} - \beta}{\sqrt{\frac{\hat{\beta}^2_{MLE}}{\alpha}}} \to N(0, 1)$$

or equivalently

$$\hat{\beta}_{MLE} \to N \left(\beta, \frac{n\alpha}{(\sum_{i=1}^n X_i)^2} \right)$$

or

$$\hat{\beta}_{MLE} \to N \left(\beta, \sqrt{\frac{\beta^2}{n\alpha}} \right).$$

2. (12 pts) Let X and Y be two random variables with joint distribution F. Suppose we observe pairs $(x_1, y_1), \ldots, (x_n, y_n)$, a random sample from F.

   (a) (6 pts) Without making any assumptions about F, form a statistic for testing $H_0 : P(X > Y) = 0.5$.

   (b) (6 pts) How would you calculate the p-value?

Solution:

(a) Let $Z = X - Y$. Then $Z_i = X_i - Y_i$ for $i = 1, \ldots, n$. Let

$$B = \sum_{i=1}^{n} I(Z_i > 0) =: \sum_{i=1}^{n} b_i$$

where $B \sim \text{Bin}(n, p)$ and $p = P(Z > 0)$. Then $H_0 : p = 0.5$. For B, we know that the MLE for p is

$$\hat{p} = \frac{\sum_{i=1}^{n} I(Z_i > 0)}{n}.$$

The Wald test gives

$$\frac{\hat{p} - 0.5}{\hat{se}(\hat{p})} \to N(0, 1).$$

We also know that

$$\hat{se}(\hat{p}) = \hat{se}\left(\frac{\sum_{i=1}^{n} I(Z_i > 0)}{n}\right) = \hat{se}\left(\frac{B}{n}\right) = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}.$$

Therefore, a test statistic is

$$\left|\frac{\hat{p} - 0.5}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\right|.$$

(b) First, we should calculate the number of pairs (X_i, Y_i), where $X_i > Y_i$. Let that number be N. Then the p-value is given by

$$\sup_{\theta \in \Theta_0} P_\theta(T(B) \geq T(b))$$

where

$$T(B) = \left| \frac{\hat{p} - 0.5}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \right| \qquad T(b) = \left| \frac{\frac{N}{n} - 0.5}{\sqrt{\frac{\frac{N}{n}(1-\frac{N}{n})}{n}}} \right|$$

and we know that under H_0, $T(B) \to N(0,1)$. Therefore, the p-value is $2(1 - \Phi(T(b))$.

3. (6 pts) Let $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(u, 1)$ and let $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} \mathcal{N}(v, 4)$. X_i's and Y_i's are independent of each other. Find the Fisher information matrix for the parameter (u, v).

Solution:

$$\mathcal{L}_n(X, Y; u, v) = \prod_{i=1}^{n} (2\pi)^{-1/2} \exp\left(-\frac{1}{2}(X_i - u)^2\right) \prod_{i=1}^{n} (8\pi)^{-1/2} \exp\left(-\frac{1}{8}(Y_i - v)^2\right).$$

$$\ell_n(X, Y; u, v) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{n}(X_i - u)^2 - \frac{1}{2}n\log 8\pi - \frac{1}{8}\sum_{i=1}^{n}(Y_i - v)^2.$$

$$\frac{\partial \ell_n}{\partial u} = \sum_{i=1}^{n}(X_i - u)$$

$$\frac{\partial \ell_n}{\partial v} = \frac{1}{4}\sum_{i=1}^{n}(Y_i - v)$$

$$\frac{\partial^2 \ell_n}{\partial u^2} = \sum_{i=1}^{n}(-1) = -n$$

$$\frac{\partial^2 \ell_n}{\partial u \partial v} = \frac{\partial^2 \ell_n}{\partial v \partial u} = 0$$

$$\frac{\partial^2 \ell_n}{\partial v^2} = \frac{1}{4}\sum_{i=1}^{n}(-1) = -\frac{1}{4}n.$$

Thus the Fisher information matrix is given by

$$\begin{pmatrix} n & 0 \\ 0 & \frac{1}{4}n \end{pmatrix}.$$

4. (12 pts) (Cont'd with problem 3.) Now we consider a Bayesian setting. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(u, 1)$ and let $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \mathcal{N}(v, 4)$. X_i's and Y_i's are independent of each other.

   (a) (6 pts) Assume the conjugate prior distribution for u is $\mathcal{N}(a, b)$. Find the posterior distribution for u, conditioning on $X_1, \ldots, X_n$. It is fine to write the distribution family and specify the parameters; you do not need to write out the PDF or CDF.

   (b) (6 pts) Suppose we use squared error loss, and let $\hat{u}_n$ be the Bayes estimator based on observing $X_1, \ldots, X_n$. Find $\hat{u}_n$.

Solution:

(a)

$$
\begin{aligned}
f(u \mid X_1, \ldots, X_n) &\propto f(X_1, \ldots, X_n \mid u) f(u) \\
&= \prod_{i=1}^{n} (2\pi)^{-1/2} \exp\left(-\frac{1}{2}(X_i - u)^2\right) \cdot (2\pi b)^{-1/2} \exp\left(-\frac{1}{2b}(u-a)^2\right) \\
&\propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(X_i - u)^2 - \frac{1}{2b}(u-a)^2\right) \\
&= \exp\left(-\frac{b\sum_i X_i^2 - 2bu\sum_i X_i + nbu^2 + u^2 + a^2 - 2ua}{2b}\right) \\
&\propto \exp\left(-\frac{(nb+1)u^2 - 2(a + b\sum_i X_i)u}{2b}\right) \\
&\propto \exp\left(-\frac{\left(u - \frac{a + b\sum_i X_i}{nb+1}\right)^2}{2\frac{b}{nb+1}}\right)
\end{aligned}
$$

Therefore,

$$
u \mid X_1, \ldots, X_n \sim N\left(\frac{a + b\sum_{i=1}^{n} X_i}{nb+1}, \frac{b}{nb+1}\right).
$$

(b) The Bayes risk is

$$
r(f, \hat{u}) = E_X[r(\hat{u} \mid X^n)] = \int r(\hat{u} \mid x) f(x)\, dx.
$$

For every X^n, $r(\hat{u} \mid X^n)$ gives a single value so if we need to minimize Bayes risk, we just need to minimize $r(\hat{u} \mid X^n)$ for every X^n. Under squared error loss, the posterior mean minimizes $r(\hat{u} \mid X^n)$ so

$$\hat{u}_n = \frac{a + b\sum_{i=1}^n X_i}{nb + 1}.$$

5. (24 pts + 10 pts (extra)) Consider the regression model $Y_i = r(X_i) + \varepsilon_i$ for $i = 1, \ldots, n$, with $\varepsilon_1, \ldots, \varepsilon_n$ iid and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

   (a) (6 pts) Consider $X_1, \ldots, X_n$ as given. Find a Kernel estimator for $r(x)$. Specify what kernel function you use.

   (b) (10 pts) (Optional; for extra credit) Cont'd with (a). Is the above estimator unbiased at a given x? What is the variance of the above estimator at a given x? Explain the tradeoff between bias and variance.

   (c) (6 pts) Now consider $X_1, \ldots, X_n$ as given and $r(X_i) = X_i$. Find the MLE for σ.

   (d) (6 pts) Cont'd with (c). What is the MLE's asymptotic distribution? Approximate asymptotic variance?

   (e) (6 pts) Cont'd with (c). Now we assume that $n = 2m$, $\varepsilon_i \sim \mathcal{N}(0, \sigma_1^2)$ when $i = 1, \ldots, m$, and $\varepsilon_i \sim \mathcal{N}(0, \sigma_2^2)$ when $i = m+1, \ldots, 2m$. Carry out a test for testing $H_0 : \sigma_1 = \sigma_2$ vs. $H_1 : \sigma_1 \neq \sigma_2$.

Solution:

(a) A kernel estimator is given by

$$\hat{r}(x) = \sum_{j=1}^{n} \frac{w_j(x) Y_j}{\sum_{i=1}^{n} w_i(x)}$$

where $w_j(x) = K(\frac{x - X_j}{n})$ is a kernel for example $w_j(x) = \exp\{-\left(\frac{x-X_j}{n}\right)^2\}$.

(b)

(c)

$$\begin{aligned} f(y_1, \ldots, y_n; \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(y_i - r(x_i))^2\right) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i)^2\right). \end{aligned}$$

The log-likelihood is then given by

$$\ell(\theta) = C - n \log \sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i)^2$$

with derivative
$$\frac{\partial \ell(\theta)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^n (y_i - x_i)^2}{\sigma^3}.$$
Setting equal to 0 and solving gives
$$\hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^n (y_i - x_i)^2}{n}$$
and by equivariance,
$$\hat{\sigma}_{MLE} = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n}}.$$
We can check that the second derivative is negative:
$$\frac{\partial^2 \ell(\theta)}{\partial \sigma^2} = \frac{n}{\sigma^2} - \frac{3\sum_{i=1}^n (y_i - x_i)^2}{\sigma^4} = \frac{n\sigma^2 - 3\sum_{i=1}^n (y_i - x_i)^2}{\sigma^4} < 0.$$

(d) We can use the Fisher information to compute the asymptotic distribution and variance.
$$\begin{aligned} I_n(\hat{\sigma}_{MLE}) &= -E_\theta\left[\frac{\partial^2}{\partial \theta^2 \ell_n(\theta)}\right] \\ &= -E\left[\frac{n}{\sigma^2} - \frac{3\sum_{i=1}^n (y_i - x_i)^2}{\sigma^4}\right] \\ &= -\frac{n}{\sigma^2} + \frac{3n\sigma^2}{\sigma^4} \\ &= \frac{2n}{\sigma^2}. \end{aligned}$$
Thus, the variance is
$$\mathrm{Var}(\hat{\sigma}_{MLE}) = \frac{1}{I_n(\hat{\sigma}_{MLE})} = \frac{\sigma^2}{2n},$$
$$\hat{\mathrm{Var}}(\hat{\sigma}_{MLE}) = \frac{1}{I_n(\hat{\sigma}_{MLE})} = \frac{\hat{\sigma}^2_{MLE}}{2n}.$$
and the asymptotic distribution is
$$\frac{\hat{\sigma}_{MLE} - \sigma}{\hat{se}(\hat{\sigma}_{MLE})} \to N(0,1).$$

(e) We can perform a likelihood ratio test.

$$T(\mathbf{x}) = \frac{\sup_{\theta\in\Theta} L_n(\theta)}{\sup_{\theta\in\Theta_0} L_n(\theta)} = \frac{\sup_{\theta\in\Theta} L_n(\theta)}{L_n(\hat{\theta}_{\mathrm{MLE}})} = \frac{L_n(\hat{\sigma}_{1,\mathrm{MLE}}, \hat{\sigma}_{2,\mathrm{MLE}})}{L_n(\hat{\sigma}_{\mathrm{MLE}})}.$$

Under $H_0 : \sigma_1 = \sigma_2 = \sigma$.

$$\hat{\sigma}_{1,\mathrm{MLE}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - x_i)^2}, \qquad \hat{\sigma}_{2,\mathrm{MLE}} = \sqrt{\frac{1}{m}\sum_{i=m+1}^{2m}(y_i - x_i)^2},$$

$$\hat{\sigma}_{\mathrm{MLE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}.$$

Then, we can compute the test statistic

$$\begin{aligned}
T(X) &= \frac{\left(\frac{1}{\hat{\sigma}_{1,\mathrm{MLE}}}\cdot\frac{1}{\hat{\sigma}_{2,\mathrm{MLE}}}\right)^m}{\left(\frac{1}{\hat{\sigma}_{\mathrm{MLE}}}\right)^n} \exp\Bigg(- \frac{1}{2\hat{\sigma}^2_{1,\mathrm{MLE}}}\sum_{i=1}^{m}(y_i - x_i)^2 - \frac{1}{2\hat{\sigma}^2_{2,\mathrm{MLE}}}\sum_{i=m+1}^{2m}(y_i - x_i)^2 \\
&\quad + \frac{1}{2\hat{\sigma}^2_{\mathrm{MLE}}}\sum_{i=1}^{n}(y_i - x_i)^2\Bigg) \\
&= \frac{\hat{\sigma}^n_{MLE}}{(\hat{\sigma}_{1,MLE}\hat{\sigma}_{2,MLE})^m}\exp\left(-\frac{m}{2} - \frac{m}{2} + \frac{n}{2}\right) \\
&= \frac{\hat{\sigma}^n_{MLE}}{(\hat{\sigma}_{1,MLE}\hat{\sigma}_{2,MLE})^m}.
\end{aligned}$$

The asymptotic distribution is given by

$$\Lambda(\mathbf{x}) = 2\log T(\mathbf{x}) \sim \chi^2_1.$$

$$\text{Reject } H_0 \text{ if } \Lambda(\mathbf{x}) > \chi^2_{1,1-\alpha}.$$

6. (12 pts) Suppose $X_1, \ldots, X_n$ are i.i.d. Poisson variables with mean λ and we are interested in estimating $p = P(X_i = 0) = e^{-\lambda}$.

   (a) (6 pts) One estimator for p is the proportion of zeros in the sample, $\tilde{p} = \#\{i \le n : X_i = 0\}/n$. Find the limiting distribution for $\sqrt{n}(\tilde{p} - p)$.

   (b) (6 pts) Another estimator would be the maximum likelihood estimator $\hat{p}$. Give a formula for $\hat{p}$ and determine the limiting distribution for $\sqrt{n}(\hat{p} - p)$.

Solution:

(a) Note that $\bar{p}$ is the average of iid Bernoullis with success probability p. Then by the central limit theorem,

$$\sqrt{n}(\bar{p} - p) \sim \mathcal{N}(0, \bar{\sigma}^2)$$

with some variance $\bar{\sigma}^2$. What variance? Simply the variance of each Bernoulli variate, so

$$\bar{\sigma}^2 = p(1 - p) = e^{-\lambda}(1 - e^{-\lambda}).$$

(b) We may recall that the Poisson MLE is $\hat{\lambda} = \bar{X}$. (Of course, we can also derive it, but this has been done many times by now.) Then, by equivariance,

$$\hat{p} = e^{-\hat{\lambda}}.$$

By standard MLE theorems (or the central limit theorem), we also have

$$\sqrt{n}(\hat{\lambda} - \lambda) \sim \mathcal{N}(0, \gamma^2)$$

with some variance γ^2. What variance? You can use either the CLT or the MLE theorems. If you use the CLT, then $\gamma^2 = \text{Var}(X_1) = \lambda$, the variance of a single Poisson. On the other hand, the MLE theorem for iid data says $\gamma^2 = 1/I_1(\lambda)$, the inverse Fisher information of one observation. We can compute

$$I_1(\lambda) = -\mathbb{E}\big[\ell_1''(\lambda)\big] = -\mathbb{E}\Big[-\frac{X}{\lambda^2}\Big] = \frac{\mathbb{E}[X]}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda},$$

so again $\gamma^2 = \lambda$.

Finally, by the delta method, we have

$$\sqrt{n}(\hat{p} - p) \sim \mathcal{N}(0, \sigma^2)$$

with some variance σ^2. What variance? This is given by the delta method formula. Let $f(x) = e^{-x}$. Then $p = f(\lambda)$ and $\hat{p} = f(\lambda)$, and since $f'(\lambda) = -e^{-\lambda} = -p$, we can write

$$\sqrt{n}(\hat{p}-p) = \sqrt{n}\big(f(\hat{\lambda})-f(\lambda)\big) \xrightarrow{d} \mathcal{N}\big(0,\ (f'(\lambda))^2 \cdot \lambda\big) = \mathcal{N}\big(0,\ \lambda e^{-2\lambda}\big) = \mathcal{N}\big(0,\ \lambda p^2\big)\,,$$

showing us that $\sigma^2 = \lambda p^2 = \log(1/p) \cdot p^2$.

7. (10 pts) Consider a decision problem with possible states of nature θ_1 and θ_2. Let X be a random variable with probability function $p(x \mid \theta)$:

$P(X=0 \mid \theta_1) = 0.2, P(X=1 \mid \theta_1) = 0.8; P(X=0 \mid \theta_2) = 0.4, P(X=1 \mid \theta_2) = 0.6.$

Two non-randomized actions a_1 and a_2 are considered with the following loss function:

$$
\begin{aligned}
&L(\theta_1, a_1(0)) = 1, \quad L(\theta_1, a_1(1)) = 2, \quad L(\theta_1, a_2(0)) = 4, \quad L(\theta_1, a_2(1)) = 0;\\
&L(\theta_2, a_1(0)) = 3, \quad L(\theta_2, a_1(1)) = 1, \quad L(\theta_2, a_2(0)) = 1, \quad L(\theta_2, a_2(1)) = 4.
\end{aligned}
$$

(a) (5 pts) Suppose θ has the prior distribution $\Lambda(\theta)$ defined by $P(\theta = \theta_1) = 0.9$, $P(\theta = \theta_2) = 0.1$. What is the Bayes rule with respect to $\Lambda(\theta)$?

(b) (5 pts) Find the minimax rule(s).

Solution:

(a) First, we compute the risks of each action, state pair.

$$
\begin{aligned}
R(\theta_1, a_1) &= \mathbb{E}_X[L(\theta_1, a_1) \mid \Theta = \theta_1]\\
&= P(X=1 \mid \Theta=\theta_1)\, L(\theta_1, a_1(X=1)) + P(X=0 \mid \Theta=\theta_1)\, L(\theta_1, a_1(X=0))\\
&= 0.8 \times 2 + 0.2 \times 1 = 1.8,
\end{aligned}
$$

$$
\begin{aligned}
R(\theta_2, a_1) &= P(X=1 \mid \Theta=\theta_2)\, L(\theta_2, a_1(X=1)) + P(X=0 \mid \Theta=\theta_2)\, L(\theta_2, a_1(X=0))\\
&= 0.6 \times 1 + 0.4 \times 3 = 1.8,
\end{aligned}
$$

$$
\begin{aligned}
R(\theta_1, a_2) &= P(X=1 \mid \Theta=\theta_1)\, L(\theta_1, a_2(X=1)) + P(X=0 \mid \Theta=\theta_1)\, L(\theta_1, a_2(X=0))\\
&= 0.8 \times 0 + 0.2 \times 4 = 0.8,
\end{aligned}
$$

$$
R(\theta_2, a_2) = 0.4 \times 1 + 0.6 \times 4 = 2.8.
$$

Then,

$$
BR(a_1) = (0.9,\, 0.1) \begin{pmatrix} 1.8 \\ 1.8 \end{pmatrix} = 1.8,
$$

$$
BR(a_2) = (0.9,\, 0.1) \begin{pmatrix} 0.8 \\ 2.8 \end{pmatrix} = 1.0.
$$

Therefore the Bayes rule is a_2.

(b)

$$\sup_{\theta \in \{\theta_1, \theta_2\}} R(\theta, a_1) = 1.8,$$
$$\sup_{\theta \in \{\theta_1, \theta_2\}} R(\theta, a_2) = 2.8.$$

Therefore, the minimax rule is a_1.