STAT215: Homework 3
Due: Wednesday, Mar 06

1. (10 pt) Let $(X_1, \cdots, X_n)$ be a random sample of binary random variables with $P(X_1 = 1) = p$, where $p \in (0,1)$ is unknown. Let $\hat{\theta}$ be the MLE of $\theta = p(1 - p)$.

   (i) Show that $\hat{\theta}$ is asymptotically normal when $p \neq \frac{1}{2}$

   (ii) When $p = \frac{1}{2}$, derive a nondegenerated asymptotic distribution of $\hat{\theta}$ with an appropriate normalization.

solution:

(i) Since the sample mean $\bar{X}$ is the MLE of $p$, the MLE of $\theta$ is $\bar{X}(1 - \bar{X})$. From the central limit theorem, $\sqrt{n}((\bar{X} - p) \xrightarrow{d} N(0, \theta))$. Using the $\delta$--method with $g(x) = x(1 - x), g'(x) = 1 - 2x$, we obtain that $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, (1 - 2p)^2\theta)$. This asymptotic distribution is degenerate when $p = \frac{1}{2}$.

(ii) When $p = \frac{1}{2}$, $\sqrt{n}(\bar{X} - \frac{1}{2}) \xrightarrow{d} N(0, \frac{1}{4})$. Hence,

$$4n(\frac{1}{4} - \hat{\theta}) = 4n(\bar{X} - \frac{1}{2}) \xrightarrow{d} \chi^2_1$$

2. (10 pt) Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent and identically distributed random 2-vectors taking values in the unit square $[0,1] \times [0,1]$, with joint CDF

$$F(x, y) = P(X_1 \leq x, Y_1 \leq y) = x \cdot y \cdot (x \wedge y)^\theta$$

for $0 \leq x \leq 1, 0 \leq y \leq 1$, where $\theta \geq 0$ is unknown.

   (i) Obtain the likelihood function for $\theta$.

   (ii) Obtain the MLE $\hat{\theta}_n(x, y)$ of $\theta$.

   (iii) (Extra Credit) Can you find the asymptotic distribution of $\hat{\theta}_n(x, y)$?

solution:
(i) Note that $F(x, y)$ is differentiable when $x \neq y$ but not when $x = y$. When $x \neq y$, $(X_1, Y_1)$ has density

$$f_{\theta}(x, y) = \begin{cases} (\theta + 1)(1-x)^{\theta} & x > y \\ (\theta + 1)(1-y)^{\theta} & x < y \\ \theta(1-x)^{\theta+1} & x = y \end{cases}$$

Let $X = X_1$ and $Y = Y_1$, we have

$$(X > t, Y > t) = (1-t)^{\theta+1}$$

and

$$P(X > t, Y > t, X \neq Y) = 2P(X > t, Y > t, X > Y)$$

$$= 2(\theta + 1) \int_t^1 \int_t^x (1-y)^{\theta} dy dx$$

$$= \frac{2(1-t)^{\theta+2}}{\theta + 2}$$

$$\therefore$$

$$P(X > t, X = Y) = P(X > t, Y > t) - P(X > t, Y > t, X \neq Y)$$

$$= \frac{\theta(1-t)^{\theta+2}}{\theta + 2}$$

which means $(X, Y)$ has density $\theta(1-t)^{\theta+1}$ on the line $x = y$. Then the probability density of $(X, Y)$ is:

$$f_{\theta}(x, y) = \begin{cases} (\theta + 1)(1-x)^{\theta} & x > y \\ (\theta + 1)(1-y)^{\theta} & x < y \\ \theta(1-x)^{\theta+1} & x = y \end{cases}$$

(ii) Let $T$ be the number of $(X_i, Y_i)$'s with $X_i = Y_i$ and $V_i := (X_i \vee Y_i)$, the likelihood function can be rewritten as

$$l(\theta) = (\theta + 1)^{n-T} \theta^T \Pi_{i=1}^n (1-V_i)^{\theta} \Pi_{i:X_i=Y_i} (1-V_i)$$

Let $\frac{\partial \log l(\theta)}{\partial \theta} = 0$ we can find the unique solution (in the parameter space) for $\theta$ is

$$\hat{\theta} = \frac{\sqrt{(n-W)^2 + 4WT} -(n-W)}{2W}$$

, where $W = -\sum_{i=1}^n \log (1-V_i)$. Since $\frac{\partial^2 \log l(\theta)}{\partial \theta^2} = -\frac{n-T}{(\theta+1)^2} - \frac{T}{\theta} < 0$, $\hat{\theta}$ is the MLE of $\theta$.

(iii) $E(T) = \frac{n\theta}{\theta+2}$, we have

$$I(\theta) = -\frac{E[\frac{\partial^2 \log l(\theta)}{\partial \theta^2}]}{n} = \frac{\theta^2 + 4\theta + 1}{\theta(\theta+2)(\theta+1)^2}$$

$$\therefore$$

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow N(0, \frac{\theta(\theta+2)(\theta+1)^2}{\theta^2 + 4\theta + 1})$$
3. (10 pt) Find a function of $\theta$, $\eta(\theta)$, for which the amount of information is independent of $\theta$ (i.e., for which $I^\eta(\theta)$ is constant), when $P_\theta$ is

(a) the Poisson distribution with unknown mean $\theta > 0$;
(b) the binomial distribution with known size $n$ and unknown probability $\theta \in (0, 1)$;
(c) the gamma distribution with known shape parameter $\alpha$ and unknown rate parameter $\theta > 0$ (i.e., with mean $\alpha/\theta$).

solution:

(a) The Fisher information about $\theta$ is $I(\theta) = 1/\theta$. Let $\eta = \eta(\theta)$. If the Fisher information about $\eta$ is not depending on $\theta$, i.e.

$$\tilde{I}(\eta) = (\frac{d\theta}{d\eta})^2 I(\theta) = c$$

then $\frac{d\theta}{d\eta} = 1/\sqrt{c\theta}$. So $\eta(\theta) = 2\sqrt{\theta}/\sqrt{c}$

(b) Similar to above, choose $\eta(\theta) = \arcsin(\sqrt{\theta})$, then $\tilde{I}(\eta) = 4n$, which is independent of $\theta$

(c) Let $\eta(\theta) = \log \theta$, then $\tilde{I}(\eta) = \alpha$.

4. (10 pt) Suppose $X \sim \text{Po}(\theta)$ has a Poisson distribution with unknown mean $\theta$. Determine the Jeffreys prior $\pi_J$ for $\theta$. Find the Bayes estimate for $\theta$ under quadratic loss upon observing the random sample $\{x_1, \ldots, x_n\}$, both for prior distribution $\pi_J$ and also for the exponential prior, $\pi(\theta) := e^{-\theta}, \theta > 0$.

solution:

(a) The Poisson pmf is $p(x \mid \theta) = e^{-\theta} \frac{\theta^x}{x!}$, $x = 0, 1, \ldots$. The Fisher information is

$$I(\theta) = -E\left(\frac{\partial^2 \log f(x \mid \theta)}{\partial \theta^2}\right) = \frac{1}{\theta}$$

. It follows that the Jeffreys prior is:

$$\pi_J(\theta) \propto \theta^{-\frac{1}{2}}, \theta > 0$$
5. (10 pt) Let \( \{X_1, \ldots, X_n\} \overset{\text{iid}}{\sim} \text{Un}(0, \theta) \) be a random sample from the uniform distribution on the interval \( (0, \theta) \) with \( \theta > 0 \).

(i) Show that \( \frac{d}{d\theta} \int x f_\theta(x) \, dx = \frac{d}{d\theta} \int x f_\theta(x) \, dx \), where \( f_\theta \) is the density of \( X_{(n)} \), the largest order statistic.

(ii) Show that the Fisher information inequality does not hold for the uniformly minimum-variance unbiased estimator (UMVUE) of \( \theta \) [Hint: \( X_{(n)} \) is complete and sufficient for \( \theta \), so any function of \( X_{(n)} \) that is an unbiased estimator of \( \theta \) is the UMVUE].

solution:

(i) \( f_\theta(x) = n\theta^{-n} x^{n-1} I_{(0,\theta)}(x) \), we then have

\[
\int x \frac{d}{d\theta} f_\theta(x) \, dx = -\frac{n^2}{\theta^{n+1}} \int_0^\theta x^n \, dx = -\frac{n^2}{n+1}
\]

On the other hand,

\[
\frac{d}{d\theta} \int x f_\theta(x) \, dx = \frac{d}{d\theta} \left( \frac{n}{\theta^n} \int_0^\theta x^n \, dx \right) = \frac{n}{n+1} \neq \int x \frac{d}{d\theta} f_\theta(x) \, dx
\]

(ii) The UMVUE of \( \theta \) is \( \frac{(n+1)X_{(n)}}{n} \) with variance \( \frac{\theta^2}{n(n+2)} \). The Fisher information is \( I(\theta) = \frac{n}{\theta^2} \). Hence \([I(\theta)]^{-1} > \frac{\theta^2}{n(n+2)}\), i.e., the Fisher information inequality does not hold.

6. (10 pt) Let \( \{X_1, \ldots, X_m\} \) and \( \{Y_1, \ldots, Y_n\} \) be independent random samples from the \( \text{No}(\mu_x, \sigma_x^2) \) and \( \text{No}(\mu_y, \sigma_y^2) \) distributions, respectively. Consider the estimation of \( \Delta := \mu_y - \mu_x \) under squared-error loss.

(i) Show that \( \bar{Y} - \bar{X} \) is a minimax estimator of \( \Delta \) when \( \sigma_x \) and \( \sigma_y \) are known, where \( \bar{Y} \) and \( \bar{X} \) are sample means.

(ii) Show that \( \bar{Y} - \bar{X} \) is a minimax estimator of \( \Delta \) when \( \sigma_x \in (0, c_x] \) and \( \sigma_y \in (0, c_y] \), where \( c_x \) and \( c_y \) are constants.
solution:

(i) Let \( \pi_{x,j} = \text{No}(0, j) \) and \( \pi_{y,j} = \text{No}(0, j) \), \( j = 1, 2, \ldots \), and \( \pi_{x,j} \times \pi_{y,j} \) be the prior of \((\mu_x, \mu_y)\). The Bayes estimators for \( \mu_x \) and \( \mu_y \) are \( \frac{m_j}{m_j + \sigma^2_x} \bar{X} \) and \( \frac{n_j}{n_j + \sigma^2_y} \bar{Y} \). Hence the Bayes estimator for \( \Delta \) is

\[
\delta_j = \frac{n_j}{n_j + \sigma^2_y} \bar{Y} - \frac{m_j}{m_j + \sigma^2_x} \bar{X}
\]

with Bayes risk \( r_{\delta_j} = \frac{j\sigma^2_x}{n_j + \sigma^2_y} + \frac{j\sigma^2_y}{m_j + \sigma^2_x} \to \frac{\sigma^2_x}{n} + \frac{\sigma^2_y}{m} \), as \( j \to \infty \), which does not depend on \((\mu_x, \mu_y)\) and is equal to the risk of \( Y - X \). By Theorem 3.3.3(pg 174) in B&D, \( Y - X \) is minimax.

(ii) Let \( \Theta = \{(\mu_x, \mu_y, \sigma_x, \sigma_y) : \mu_x \in \mathbb{R}, \mu_y \in \mathbb{R}, \sigma_x \in (0, c_x], \sigma_y \in (0, c_y]\} \) and \( \Theta_0 = \{(\mu_x, \mu_y, \sigma_x, \sigma_y) : \mu_x \in \mathbb{R}, \mu_y \in \mathbb{R}, \sigma_x = c_x, \sigma_y = c_y\} \). As shown in (i), \( Y - X \) is minimax when \( \Theta_0 \) is considered. Since

\[
\sup_{\theta \in \Theta} R_{Y - X}(\theta) = \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m} = \sup_{\theta \in \Theta_0} R_{Y - X}(\theta)
\]

we have \( Y - X \) is minimax.

7. (10 pt) BONUS question:

Find the Jeffreys Prior for a one-dimensional location-scale family, i.e., one where

\[
f(x \mid a, b) = b \cdot f((x - a) \cdot b)
\]

for \( a \in \mathbb{R} \), \( b > 0 \) for some strictly positive p.d.f. \( f(z) > 0 \).

(Hint: Write \( f(z) = e^{-\phi(z)} \) and do everything in terms of \( \phi(z) \equiv -\log f(z) \); change variables in the expectation step to \( z := b \cdot (x - a) \).)

solution:

First we’ll find the information matrix \( I(a, b) = \begin{vmatrix} \frac{\partial^2 \ell}{\partial a \partial a} & \frac{\partial^2 \ell}{\partial a \partial b} \\ \frac{\partial^2 \ell}{\partial b \partial a} & \frac{\partial^2 \ell}{\partial b \partial b} \end{vmatrix} \), where \( \ell \) is the log likelihood function.

Introduce the function \( \phi \), with \( \phi(z) \equiv -\log f(z) \), and by consequence, \( f(z) = e^{-\phi(z)} \). Then we have

\[
\ell(a, b) \equiv \log g(x|a, b) \\
= \log[b \cdot f((x - a) \cdot b)] \\
= \log b + \log f((x - a) \cdot b) \\
= \log b - \phi((x - a) \cdot b)
\]
Note that one function $f$ from the original problem statement has been replaced $g$ for clarity. Now let’s work out the derivatives:

\[
\frac{\partial \ell}{\partial a} = b \phi'(x - a) b \\
\frac{\partial^2 \ell}{\partial a^2} = -b^2 \phi''((x - a)b) \\
\frac{\partial^2 \ell}{\partial b \partial a} = b \phi''((x - a)b)(x - a) + \phi'((x - a)b) \\
\frac{\partial \ell}{\partial b} = \frac{1}{b} - \phi'((x - a)b)(x - a) \\
\frac{\partial^2 \ell}{\partial b^2} = -\frac{1}{b^2} - \phi''((x - a)b)(x - a)^2 \\
\frac{\partial^2 \ell}{\partial a \partial b} = b \phi''((x - a)b)(x - a) + \phi'((x - a)b)
\]

The Information matrix $I(a, b)$ involves the expectations of each of these derivatives, so let’s work those out too.

\[
\mathbb{E} \left( \frac{\partial^2 \ell}{\partial a^2} | a, b \right) = \int_{x \in \mathbb{R}} -b^2 \phi''((x - a)b) \cdot g(x|a, b) dx \\
= \int_{x \in \mathbb{R}} -b^2 \phi''((x - a)b) \cdot be^{-\phi((x-a)b)} dx
\]

Now we use the (suggested) transformation of variables: $z = b(x - a)$, $dz = b \cdot dx$.

\[
\mathbb{E} \left( \frac{\partial^2 \ell}{\partial a^2} | a, b \right) = \int_{z \in \mathbb{R}} -b^2 \phi''(z) e^{-\phi(z)} dz \\
= b^2 \left[ -\int_{z \in \mathbb{R}} \phi''(z) e^{-\phi(z)} dz \right] \\
= b^2 \cdot K
\]

We note that the integral in this expression depends on neither $a$ nor $b$, and is therefore a constant. We’re calling that constant $K$.

The same transformation of variables from $x$ to $z$ is made in the remaining two expectations. (The two mixed partials are the same.) Two more constants $L$ and $M$ are defined.
\[ \mathbb{E}\left( \frac{\partial^2 \ell}{\partial \theta^2} \mid a, b \right) = \int_{x \in \mathbb{R}} \left[ -\frac{1}{b^2} - \phi''((x - a)b)(x - a)^2 \right] \cdot g(x \mid a, b) \, dx \]

\[ = \int_{x \in \mathbb{R}} \left[ -\frac{1}{b^2} - \phi''((x - a)b)(x - a)^2 \right] \cdot be^{-\phi((x-a)b)} \, dx \]

\[ = \int_{z \in \mathbb{R}} \left[ -\frac{1}{b^2} - \phi''(z) \left( \frac{z}{b} \right)^2 \right] \cdot e^{-\phi(z)} \, dz \]

\[ = \frac{1}{b^2} \left[ - \int_{z \in \mathbb{R}} (-1 - z^2 \phi''(z)) e^{-\phi(z)} \, dz \right] \]

\[ = \frac{1}{b^2} \cdot L \]

\[ \mathbb{E}\left( \frac{\partial^2 \ell}{\partial a \partial b} \mid a, b \right) = \int_{x \in \mathbb{R}} \left[ \phi''((x - a)b)(x - a) + \phi'((x - a)b) \right] \cdot g(x \mid a, b) \, dx \]

\[ = \int_{x \in \mathbb{R}} \left[ \phi''((x - a)b)(x - a) + \phi'((x - a)b) \right] \cdot be^{-\phi((x-a)b)} \, dx \]

\[ = \int_{z \in \mathbb{R}} \left[ z\phi''(z) + \phi'(z) \right] \cdot e^{-\phi(z)} \, dz \]

\[ = M \]

Thus the information matrix is given by

\[ I(a, b) = \left| \begin{array}{c} Kb^2 M \\ M \frac{1}{b^2} L \end{array} \right| \]

The determinant of this matrix is \((Kb^2)\left(\frac{1}{b^2}L\right) - M^2\), which is a constant in \(a\) and \(b\). Thus the square root of its absolute value is also a constant in \(a\) and \(b\), so the Jeffrey’s prior must be \(\pi(a, b) \propto 1\).