University of New South Wales

MATH 2901

Higher Theory of Statistics

Assignment 2

Keegan Gyoery (z5197058), John Lidbetter (z5160860),
Edward McInnes (z5162873), Alex Robinson (z5164884),
Ruby Smith (z5113171)

May 20, 2018

1. X_1 and X_2 have the following density functions:

$$\begin{aligned} f_{X_1}(x) &= \frac{1}{x\sqrt{2\pi}} e^{-(\ln x)^2/2} \qquad x > 0 \\ f_{X_2}(x) &= f_{X_1}(x)[1 + \sin(2\pi \ln x)] \qquad x > 0 \end{aligned}$$

(a) Graphs Here

(b)

$$\begin{aligned} \mathbb{E}\big[X_1^r\big] &= \int_{-\infty}^{\infty} x^r I_{(0,\infty)}(x) f_{X_1}(x) dx \\ &= \int_0^{\infty} x^r f_{X_1}(x) dx \\ &= \int_0^{\infty} \frac{x^r}{x\sqrt{2\pi}} e^{-(\ln x)^2/2} dx \\ &= \int_0^{\infty} \frac{x^{r-1}}{\sqrt{2\pi}} e^{-(\ln x)^2/2} dx \end{aligned}$$

Using the substitution $u = \ln x$, $x = e^u$, and $dx = e^u du$. The limits are now $-\infty$ and ∞.

$$\begin{aligned} &= \int_{-\infty}^{\infty} \frac{(e^u)^{r-1}}{\sqrt{2\pi}} e^{-u^2/2} e^u du \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{u(r-1)} e^{(-u^2+2u)/2} du \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\left[-u^2+2u+2u(r-1)\right]/2} du \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(-u^2+2ur)/2} du \\ &= \int_{-\infty}^{\infty} e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u^2-2ur+r^2)/2} du \\ &= \int_{-\infty}^{\infty} e^{r^2/2} \frac{1}{\sqrt{2\pi}} e^{-(u-r)^2/2} du \end{aligned}$$

Using the substitution $y = u - r$, $u = y + r$, and $du = dy$. The limits remain the same.

$$\begin{aligned} &= e^{r^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= e^{r^2/2}(1) \\ \therefore \mathbb{E}\big[X_1^r\big] &= e^{r^2/2} \end{aligned}$$

(c)

$$
\begin{aligned}
\mathbb{E}\left[X_2^r\right] &= \int_{-\infty}^{\infty} x^r I_{(0,\infty)}(x) f_{X_2}(x)dx \\
&= \int_0^{\infty} x^r f_{X_2}(x)dx \\
&= \int_0^{\infty} x^r f_{X_1}(x)[1+\sin(2\pi \ln x)]dx \\
&= \int_0^{\infty} x^r f_{X_1}(x) + x^r f_{X_1}(x)\sin(2\pi \ln x)dx \\
&= \int_0^{\infty} x^r f_{X_1}(x)dx + \int_0^{\infty} x^r f_{X_1}(x)\sin(2\pi \ln x)dx \\
&= \mathbb{E}\left[X_1^r\right] + \int_0^{\infty} x^r f_{X_1}(x)\sin(2\pi \ln x)dx \\
\therefore \mathbb{E}\left[X_2^r\right] &= \mathbb{E}\left[X_1^r\right] + \int_0^{\infty} x^r f_{X_1}(x)\sin(2\pi \ln x)dx
\end{aligned}
$$

(d)

2. A random variable X is said to follow a Pareto(α, k) distribution if the density function of X is:

$$
f_X(x) = \frac{\alpha k^\alpha}{x^{\alpha+1}} \qquad \alpha, k > 0 \text{ and } x > k
$$

Suppose, for $n > 2$, we have a sequence of iid Pareto(α, k) random variables $X_1, \ldots, X_n$

(a) In order to compute the MLE for k and α, we derive the likelihood function, and the log-likelihood function.

$$
\begin{aligned}
L(\alpha, k) &= \prod_{i=1}^{n} \frac{\alpha k^\alpha}{x_i^{\alpha+1}} \\
\therefore l(\alpha, k) &= \ln\left(\prod_{i=1}^{n} \frac{\alpha k^\alpha}{x_i^{\alpha+1}}\right) \\
&= \sum_{i=1}^{n} \ln\left(\frac{\alpha k^\alpha}{x_i^{\alpha+1}}\right) \\
&= \sum_{i=1}^{n} \left[\ln\left(\alpha k^\alpha\right) - \ln\left(x_i^{\alpha+1}\right)\right] \\
&= \sum_{i=1}^{n} \ln\left(\alpha k^\alpha\right) - \sum_{i=1}^{n} \ln\left(x_i^{\alpha+1}\right) \\
&= n\ln\left(\alpha k^\alpha\right) - \sum_{i=1}^{n} \ln\left(x_i^{\alpha+1}\right) \\
&= n\ln(\alpha) + n\alpha\ln(k) - (\alpha+1)\sum_{i=1}^{n}\ln(x_i)\ldots(1)
\end{aligned}
$$

Considering (1) as an equation in k, we note that (1) is increasing over all values of k, as $n, \alpha > 0$. Thus, $l(\alpha, k)$ is maximised when k takes its maximum value. As $k \leq x_i$, the maximum value that k can take is:

$$
\begin{aligned}
k &= \min(x_i) \\
\therefore \widehat{k} &= \min(X_i)
\end{aligned}
$$

Thus, $\widehat{k}$ is the MLE for k.

Now, considering (1) as an equation in α, we have:

$$\begin{aligned}
\therefore \frac{\partial l(\alpha,k)}{\partial \alpha} &= \frac{n}{\alpha} + n\ln(k) - \sum_{i=1}^{n}\ln(x_i) \\
\frac{\partial l(\alpha,k)}{\partial \alpha} &= 0 \\
0 &= \frac{n}{\alpha} + n\ln(k) - \sum_{i=1}^{n}\ln(x_i) \\
\frac{n}{\alpha} &= \sum_{i=1}^{n}\ln(x_i) - n\ln(k) \\
\therefore \alpha &= \frac{n}{\sum_{i=1}^{n}\ln(x_i) - n\ln(k)} \\
\therefore \widehat{\alpha} &= \frac{n}{\sum_{i=1}^{n}\ln(X_i) - n\ln\left(\widehat{k}\right)} \\
\therefore \widehat{\alpha} &= \frac{n}{\sum_{i=1}^{n}\left[\ln(X_i) - \ln(\min(X_i))\right]} \\
\frac{\partial^2 l(\alpha,k)}{\partial \alpha^2} &= \frac{-n}{\alpha^2} \\
\therefore \frac{\partial^2 l(\alpha,k)}{\partial \alpha^2} &< 0 \qquad \forall \alpha
\end{aligned}$$

Thus, $\widehat{\alpha}$ maximises the log-likelihood function, and thus maximises the likelihood function, and therefore is the MLE for α.

(b) The MLE of k is $\widehat{k} = \min(X_i)$. In order to derive the distribution of $\widehat{k}$, we must consider the CDF of a minimum of a sequence of random variables. Let $Y = \min(X_i)$.

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) \\
&= \mathbb{P}(\min(X_i) \le y) \\
&= 1 - \mathbb{P}(\min(X_i) > y) \\
&= 1 - \left([1 - F_{X_1}(y)][1 - F_{X_2}(y)]\dots[1 - F_{X_n}(y)]\right) \\
&= 1 - \left[1 - F_X(y)\right]^n \qquad \text{as } X_i\ \forall i \text{ are iid} \\
\therefore F_{\widehat{k}}(x) &= 1 - \left[1 - F_X(x)\right]^n \\
&= 1 - \left[1 - \left[1 - \left(\frac{k}{x}\right)^{\alpha}\right]\right]^n \\
\therefore F_{\widehat{k}}(x) &= 1 - \left(\frac{k}{x}\right)^{n\alpha}
\end{aligned}$$

Thus, $\widehat{k} \sim$ Pareto$(n\alpha, k)$

(c) The Bias of $\widehat{k}$ is given by:

$$\begin{aligned}
\text{Bias}\left(\widehat{k}\right) &= \mathbb{E}\left(\widehat{k}\right) - k \\
&= \frac{n\alpha k}{n\alpha - 1} - k \quad \text{as } n\alpha > 1 \\
&= \frac{n\alpha k}{n\alpha - 1} - \frac{(n\alpha - 1)k}{n\alpha - 1} \\
\text{Bias}\left(\widehat{k}\right) &= \frac{k}{n\alpha - 1} \\
\therefore \text{Bias}\left(\widehat{k}\right) &> 0 \quad \text{as } k > 0
\end{aligned}$$

Thus, the MLE $\widehat{k}$ is a biased estimator for k. An unbiased estimator for k requires:

$$\begin{aligned}
\left[\frac{n\alpha k}{n\alpha - 1}\right] C - k &= 0 \quad \text{for } C \text{ some constant} \\
\left[\frac{n\alpha k}{n\alpha - 1}\right] C &= k \\
\left[\frac{n\alpha}{n\alpha - 1}\right] C &= 1 \\
\therefore C &= \left[\frac{n\alpha - 1}{n\alpha}\right]
\end{aligned}$$

Thus, an MLE for k that is unbiased is:

$$\widehat{k} = \left[\frac{n\alpha - 1}{n\alpha}\right] \min(X_i)$$

3. Let $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$ be two independent random variables, and define $Z = \min(X, Y)$.

$$\begin{aligned}
\mathbb{P}(Z \leq z) &= \mathbb{P}(\min(X, Y) \leq z) \\
&= 1 - \mathbb{P}(\min(X, Y) > z) \\
&= 1 - \mathbb{P}(X > z, Y > z) \\
&= 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) \ldots (A) \quad \text{independence}
\end{aligned}$$

Now, considering Z^2

$$\begin{aligned}
\mathbb{P}(Z^2 \leq z) &= \mathbb{P}(-\sqrt{z} \leq Z \leq \sqrt{z}) \\
&= \mathbb{P}(Z \leq \sqrt{z}) - \mathbb{P}(Z \leq -\sqrt{z}) \\
&= 1 - \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y > \sqrt{z}) - [1 - \mathbb{P}(X > -\sqrt{z})\mathbb{P}(Y > -\sqrt{z})] \quad \text{from } (A) \\
&= \mathbb{P}(X > -\sqrt{z})\mathbb{P}(Y > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y > \sqrt{z}) \\
&= \mathbb{P}(X > -\sqrt{z})[1 - \mathbb{P}(Y \leq -\sqrt{z})] - \mathbb{P}(X > \sqrt{z})[1 - \mathbb{P}(Y \leq \sqrt{z})] \\
&= \mathbb{P}(X > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z}) - \mathbb{P}(X > -\sqrt{z})\mathbb{P}(Y \leq -\sqrt{z}) + \mathbb{P}(X > \sqrt{z})\mathbb{P}(Y \leq \sqrt{z})
\end{aligned}$$

Not entirely sure how the next bit works in terms of symmetry

$$\begin{aligned}
\therefore \mathbb{P}(Z^2 \leq z) &= \mathbb{P}(X > -\sqrt{z}) - \mathbb{P}(X > \sqrt{z}) \\
&= 1 - \mathbb{P}(X \leq -\sqrt{z}) - [1 - \mathbb{P}(X \leq \sqrt{z})] \\
&= \mathbb{P}(X \leq \sqrt{z}) - \mathbb{P}(X \leq -\sqrt{z}) \\
&= \mathbb{P}(-\sqrt{z} \leq X \leq \sqrt{z}) \\
\therefore \mathbb{P}(Z^2 \leq z) &= \mathbb{P}(X^2 \leq z)
\end{aligned}$$

I think we need to make a statement about convergence in distribution before the next statement

Since $X \sim \mathcal{N}(0, 1)$, $X^2 \sim \chi^2_1$ and thus $Z^2 \sim \chi^2_1$

4. (a)
   (b)
   (c)
   (d)
   (e)