Tutorial 9

1. Let $X_1, \dots, X_n$ be a random sample from distribution:

$$g(x, \theta) = \begin{cases} \theta x^{\theta - 1} & 0 \le x \le 1 \qquad \theta > 0 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Find the MLE of $\frac{1}{\theta}$. Is it unbiased? Is it UMVU?

   (b) Show that $\overline{X}$ is an unbiased estimator of $\frac{\theta}{1+\theta}$. Is it UMVU?

2. Let X and Y be two discrete random variables with well defined expected values and variances. Prove that:

   (a) $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$

   (b) $\text{Var}(X) = \text{Var}(E[X|Y]) + \mathbb{E}[\text{Var}(X|Y)]$.

3. Let $X_1, \dots, X_{n+1}$ be independent Bernoulli(p) variables and let

$$h(p) = \mathbb{P}\left(\sum_{i=1}^{n} X_i > X_{n+1} \middle| \, p \right).$$

   (a) Show that

$$T(X_1, \dots, X_{n+1}) = \begin{cases} 1 & \sum_{j=1}^{n} X_j > X_{n+1} \\ 0 & \text{otherwise} \end{cases}$$

   is an unbiased estimator of $h(p)$.

   (b) Find the UMVUE of $h(p)$.

4. Let X be an observation from the probability with mass function:

$$p(-1, \theta) = \frac{\theta}{2}, \quad p(0, \theta) = 1 - \theta, \quad p(1, \theta) = \frac{\theta}{2} \qquad \theta \in [0, 1].$$

   (a) Find the maximum likelihood estimator of θ and show that it is unbiased.

   (b) Let

$$T(X) = \begin{cases} 2 & x = 1 \\ 0 & x = -1 \quad \text{or} \quad 0 \end{cases}$$

   Show that T is an unbiased estimator of θ.

   (c) Show that $\widehat{\theta}$ (maximum likelihood estimator) is minimal sufficient for θ and that $\mathbb{E}[T|\widehat{\theta}] = \widehat{\theta}$. Show that $\text{Var}(\widehat{\theta}) < \text{Var}(T)$.

5. Let X be the number of dots showing when a fair die is rolled; i.e.

$$p_X(x) = \frac{1}{6} \qquad x = 1, 2, 3, 4, 5, 6.$$

   Let Y be the number of heads obtained when X fair coins are tossed. Find

(a) The mean and variance of Y.

(b) The MSPE (mean squared prediction error) of the optimal linear predictor of Y based on X. The optimal linear predictor is the function $\widehat{Y} = aX + b$, where a and b are chosen such that $\mathbb{E}[\widehat{Y}] = \mathbb{E}[Y]$ and, subject to this constraint, to minimise $\mathrm{Var}(Y - \widehat{Y})$.

(c) The optimal linear predictor of Y given $X = x$ for $x = 1, 2, 3, 4, 5, 6$.

6. A person walks into a clinic at time t and is diagnosed with a certain disease. At the same time (t), a diagnostic indicator Z_0 of the severity of the disease (e.g. a blood cell count or a virus count) is obtained. Let S be the unknown date in the past when the subject was infected. We are interested in the time $Y_0 = t - S$ from infection until detection. Assume that the conditional density of Z_0 (the present condition) given $Y_0 = y$ is $N(\mu + \beta y_0, \sigma^2)$. Let

$$Z = \frac{Z_0 - \mu}{\sigma}, \qquad Y = \frac{\beta}{\sigma} Y_0.$$

(a) Show that the conditional density $p(z|y)$ of Z given $Y = y$ is $N(y, 1)$.

(b) Suppose that Y has exponential density $\pi(y) = \lambda \exp\{-\lambda y\}\mathbf{1}_{\{y>0\}}$ where $\lambda > 0$. Show that the conditional distribution of Y given $Z = z$ has density

$$\pi(y|z) = \frac{1}{(2\pi)^{1/2} c} \exp\left\{-\frac{1}{2}(y - (z - \lambda))^2\right\} \qquad y > 0$$

where c is a suitable constant (depending on z and λ). Compute c in terms of the c.d.f. Φ for a $N(0, 1)$ random variable.

(c) Find the conditional density $\pi_0(y_0|z_0)$ of Y_0 given $Z_0 = z_0$.

(d) Suppose it is known that $Z_0 = z_0$. Find an expression (in terms of the c.d.f for $N(0, 1)$ and its inverse) for $g(z_0)$, the best predictor of Y_0 given $Z_0 = z_0$ using mean *absolute* prediction error $\mathbb{E}\left[|Y_0 - g(Z_0)|\right]$.

(e) Let ϕ denote the density function for a $N(0, 1)$ random variable. Show that the best Mean Squared Prediction Error (MSPE) predictor of Y given $Z = z$ is:

$$\mathbb{E}[Y|Z = z] = \frac{1}{c}\phi(\lambda - z) - (\lambda - z).$$

Answers

1. (a) For $(x_1, \ldots, x_n) \in [0,1]^n$,

$$\log L(\theta; x_1, \ldots, x_n) = n \log \theta + (\theta - 1) \sum_{j=1}^{n} \log x_j$$

$$\frac{\partial}{\partial \theta} \log L(\theta) = \frac{n}{\theta} + \sum_{j=1}^{n} \log x_j$$

$$\frac{d^2}{d\theta^2} \log L(\theta) = -\frac{n}{\theta^2}$$

while $\log L(\theta) \overset{\theta \to 0, \theta \to +\infty}{\longrightarrow} -\infty$ hence unique maximiser which is $\widehat{\theta} = \frac{-1}{\sum_{j=1}^{n} \log x_j}$. Therefore:

$$\frac{1}{\widehat{\theta}_{ML}} = -\frac{1}{n} \sum_{j=1}^{n} \log X_j$$

$$\mathbb{E}_\theta \left[\frac{1}{\widehat{\theta}_{ML}} \right] = -\theta \int_0^1 x^{\theta-1} \log x dx = \theta \int_0^\infty e^{-\theta y} y dy = \frac{1}{\theta}$$

so $\frac{1}{\widehat{\theta}_{ML}}$ is an unbiased estimator of $\frac{1}{\theta}$.
To show that it is UMVU: this is an exponential family;

$$L(x_1, \ldots, x_n; \theta) = \frac{1}{\prod_{j=1}^{n} x_j} \prod_{j=1}^{n} \mathbf{1}_{[0,1]}(x_j) \exp \left\{ \theta \sum_{j=1}^{n} \log x_j + n \log \theta \right\}.$$

The sufficient statistic $T(x_1, \ldots, x_n) = \sum_{j=1}^{n} \log x_j$ is therefore *complete*. The UMVU estimator is therefore:

$$\mathbb{E} \left[\frac{1}{\widehat{\theta}_{ML}} | T(X) \right] = \frac{1}{\widehat{\theta}_{ML}}.$$

(b)

$$\mathbb{E}[\overline{X}] = \mathbb{E}[X] = \theta \int_0^1 x^\theta dx = \frac{\theta}{1+\theta}$$

so $\overline{X}$ is an unbiased estimator of $\frac{\theta}{1+\theta}$.
To show that it is not UMVU, $\mathbb{E}[\overline{X} | \sum_i \log X_i]$ is the unique UMVU estimator and this is not equal to $\overline{X}$ since with probability 1, $\overline{X}$ is not a function of $\sum_i \log X_i$.

2. (a)

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|Y]] &= \sum_y p_Y(y(\sum_x x p_{X|Y}(x|y) = \sum_{x,y} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \\ &= \sum_x x(\sum_y p_{X,Y}(x,y) = \sum_x x p_X(x) = \mathbb{E}[X]. \end{aligned}$$

(b)

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[\mathbb{E}[X^2|Y]] - \mathbb{E}[\mathbb{E}[X|Y]]^2 \\ &= \mathbb{E}[\mathrm{Var}(X|Y)] + \mathbb{E}[(\mathbb{E}[X|Y])^2] - \mathbb{E}[\mathbb{E}[X|Y]]^2 \\ &= \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y]).\end{aligned}$$

3. (a) Trivially clear from the definition: T is a binary variable taking values in $\{0,1\}$, therefore:

$$\mathbb{E}_p[T] = \mathbb{P}_p(T=1) = h(p).$$

(b) $\sum_{j=1}^{n+1} X_j$ is a complete sufficient statistic for p, hence unique UMVUE is

$$S := \mathbb{E}[T|\sum_{j=1}^{n+1} X_j].$$

Now, for each $y \in \{0,1,\ldots,n+1\}$:

$$\mathbb{E}[T|\sum_{j=1}^{n+1} X_j = y] = \mathbb{P}(T=1|\sum_{j=1}^{n+1} X_j = y) = \frac{\mathbb{P}(\sum_{j=1}^{n} X_j > X_{n+1}, \sum_{j=1}^{n+1} X_j = y)}{\mathbb{P}(\sum_{j=1}^{n+1} X_j = y)}.$$

The denominator is $\binom{n+1}{y} p^y (1-p)^{n+1-y}$; the numerator is:

$$\begin{cases} 0 & y=0 \\ \mathbb{P}(\sum_{j=1}^{n} X_j = 1, X_{n+1}=0) = np(1-p)^n & y=1 \\ \mathbb{P}(\sum_{j=1}^{n} X_j = 2, X_{n+1}=0) = \frac{n(n-1)}{2} p^2 (1-p)^{n-1} & y=2 \\ \mathbb{P}(\sum_{j=1}^{n+1} X_j = y) = \binom{n+1}{y} p^y (1-p)^{n+1-y} & y \geq 3 \end{cases}$$

Note: for $y \geq 3$, it always holds that $\sum_{j=1}^{n} X_j > X_{n=1}$. Putting this together gives:

$$\begin{cases} 0 & y=0 \\ \frac{n}{n+1} & y=1 \\ \frac{n-1}{n+1} & y=2 \\ 1 & y \geq 3 \end{cases}$$

4. (a)

$$L(\theta; x) = \frac{\theta}{2} \mathbf{1}_{\{-1,1\}}(x) + (1-\theta)\mathbf{1}_{\{0\}}(x)$$

Clearly this is maximised for: $\widehat{\theta}(1) = \widehat{\theta}(-1) = 1 \qquad \widehat{\theta}(0) = 0$.

To compute its expected value:

$$\mathbb{E}_\theta[\widehat{\theta}] = \frac{\theta}{2} + \frac{\theta}{2} = \theta.$$

(b) $\mathbb{E}[T(X)] = 2 \times \frac{\theta}{2} = \theta$ so it is unbiased.

(c) Note that $\widehat{\theta}(X) = |X|$. To show sufficiency:

$$\mathbb{P}_\theta(X = x||X| = y) = \frac{\mathbb{P}_\theta(X = x, |X| = y)}{\mathbb{P}_\theta(|X| = y)} = \begin{cases} 1 & x = 0, y = 0 \\ \frac{1}{2} & x = \pm 1, y = 1 \\ 0 & \text{other} \end{cases}$$

which does not depend on θ.

To prove *minimal* sufficiency: Any *reduction* is a function $S : S(|X|) = \text{constant}$ so that $\mathbb{P}_\theta(X \in .|S) = \mathbb{P}_\theta(X \in .)$ which does depend on θ. hence S is not sufficient. Therefore $\widehat{\theta}$ is *minimal* sufficient.

Clearly:

$$\mathbb{E}[T(X)||X|] = \begin{cases} 0 & X = 0 \\ 1 & X = \pm 1 \end{cases}$$

Finally: $\widehat{\theta} \sim Be(\theta)$ so that

$$\text{Var}(\widehat{\theta}) = \theta(1 - \theta).$$

while $T = 2Z$ for $Z \sim Be(\frac{\theta}{2})$ so that

$$\text{Var}(T) = 4\frac{\theta}{2}(1 - \frac{\theta}{2}) = 2\theta(1 - \frac{\theta}{2})$$

which is clearly greater.

5. (a) $\mathbb{E}[Y] = \frac{7}{4}$,

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X]) = \frac{1}{4}\mathbb{E}[X] + \frac{1}{4}\text{Var}(X) = \frac{3}{8} + \frac{12.5 + 4.5 + 0.5}{24} = \frac{26.5}{24} = 1\frac{5}{48}.$$

(b)

$$\widehat{Y} = aX + b$$

minimise

$$\text{Var}(Y - aX - b) = \text{Var}(Y) + a^2\text{Var}(X) - 2a\mathbf{C}(Y, X)$$

gives:

$$a = \frac{\mathbf{C}(Y, X)}{\text{Var}(X)}$$

We can show $\text{Cov}(X, Y) = \frac{1}{2}\text{Var}(X)$ as follows:

$$\mathbb{E}[XY] = \mathbb{E}[X\mathbb{E}[Y|X]] = \frac{1}{2}\mathbb{E}[X^2]$$

$$\mathbb{E}[Y] = \frac{1}{2}\mathbb{E}[X]$$

hence

$$\mathbf{C}(Y, X) = \frac{1}{2}\text{Var}(X) \Rightarrow a = \frac{1}{2}$$

$$\text{Var}(Y - \widehat{Y}) = \text{Var}(Y) - \frac{1}{4}\text{Var}(X) = \frac{1}{4}\mathbb{E}[X] = \frac{7}{8}.$$

(c)

$$\mathbb{E}[\widehat{Y}] = \mathbb{E}[Y] = \frac{1}{2}\mathbb{E}[X].$$

Now using $\widehat{Y} = aX + b$ with $a = \frac{1}{2}$ gives $b = 0$ so that

$$\widehat{Y} = \frac{1}{2}X.$$

6. (a) $Z \sim N(y, 1)$ follows directly from rescaling.

(b)

$$\begin{aligned}\pi(y|z) = \frac{\pi(y)p(z|y)}{p(z)} &\propto \lambda e^{-\lambda y}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-y)^2}\mathbf{1}_{\{y>0\}} \\ &= \frac{\lambda}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2} + y(z-\lambda) - \frac{z^2}{2}\right\}\mathbf{1}_{\{y\geq 0\}}\end{aligned}$$

so

$$\pi(y|z) = K\exp\left\{-\frac{1}{2}(y-(z-\lambda))^2\right\}\mathbf{1}_{\{y\geq 0\}}$$

$$1 = \sqrt{2\pi}K\int_{-(z-\lambda)}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \sqrt{2\pi}K\Phi(z-\lambda)$$

where Φ is the $N(0,1)$ c.d.f., hence

$$\pi(y|z) = \frac{1}{\sqrt{2\pi}\Phi(z-\lambda)}\exp\left\{-\frac{1}{2}(y-(z-\lambda))^2\right\}\mathbf{1}_{\{y\geq 0\}}$$

(c)

$$\pi_0(y_0|z_0) = \frac{\beta}{(2\pi)^{1/2}\sigma\Phi(\frac{z_0-\mu}{\sigma}-\lambda)}\exp\left\{-\frac{1}{2\sigma^2}(\beta^2 y_0 - (z_0-\mu-\lambda))^2\right\}\mathbf{1}_{\{y_0>0\}}$$

(d) First find $h(z)$, the best linear predictor of Y given $Z = a$. Then $g(z_0) = \frac{\beta}{\sigma}h(\frac{z_0-\mu}{\sigma})$. $h(z)$ is the value of h that minimises $\int_0^\infty |y-h|\pi(y|z)dy$ so that h satisfies:

$$\int_0^h \frac{1}{(2\pi)^{1/2}}\exp\{-\frac{1}{2}(y-(z-\lambda))^2\}dy = \int_h^\infty \frac{1}{(2\pi)^{1/2}}\exp\{-\frac{1}{2}(y-(z-\lambda))^2\}dy$$

giving

$$\begin{aligned}\Phi(h-(z-\lambda)) - \Phi(-(z-\lambda)) &= 1 - \Phi(h-(z-\lambda)), \\ \Phi(h-(z-\lambda)) &= \frac{1}{2}(1+\Phi(-(z-\lambda))) \\ h(z) &= (z-\lambda) + \Phi^{-1}\left(\frac{1}{2}(1+\Phi(\lambda-z))\right).\end{aligned}$$

(e) We want to find h which minimises

$$\int_0^\infty (y-h)^2\pi(y|z)dz$$

which is given by $h(z) = \mathbb{E}[Y|Z = z]$. This is:

$$\begin{aligned}
\mathbb{E}[Y|Z = z] &= (z - \lambda) + \frac{1}{(2\pi)^{1/2}c} \int_0^\infty (y - (z - \lambda))e^{-\frac{1}{2}(y-(z-\lambda))^2} dy \\
&= (z - \lambda) + \frac{1}{(2\pi)^{1/2}c} \int_0^{e^{-\frac{1}{2}(z-\lambda)^2}} dx \\
&= (z - \lambda) + \frac{1}{(2\pi)^{1/2}c} e^{-\frac{1}{2}(z-\lambda)^2}.
\end{aligned}$$