Tutorial 5

1. Let $X_1, \ldots, X_n$ be i.i.d. $U(0, \theta)$ random variables where θ is an unknown parameter.

   (a) Let $T_1 = \frac{n+1}{n} \max_j X_j$. Compute $\mathbb{E}[T_1]$ and $\mathrm{Var}(T_1)$.

   (b) Let $T_2 = \frac{2}{n} \sum_{j=1}^n X_j$. Compute $\mathbb{E}[T_2]$ and $\mathrm{Var}(T_2)$.

   You should find that both T_1 and T_2 are unbiased estimators of θ (that is $\mathbb{E}[T_1] = \mathbb{E}[T_2] = \theta$, but that the variance of T_1 is substantially lower.

   (c) Show that $\max_j X_j$ is the maximum likelihood estimator of θ.

   (d) Show that $\frac{2}{n} \sum_{j=1}^n X_j$ is the Method of Moments estimator of θ (based on the first moment - the expectation).

2. Let X be a random variable with state space $\mathcal{X} = \{v_1, \ldots, v_k\}$, where $p_j = \mathbb{P}(X = v_j)$. Let $(X_1, \ldots, X_n)$ be a random sample from X. The *frequency plug-in principle* is simply the estimation procedure where $(p_1, \ldots, p_k)$ is estimated by $(\widehat{p}_1, \ldots, \widehat{p}_n) = \left(\frac{N_1}{n}, \ldots, \frac{N_k}{n}\right)$, where $N_j = \sum_{i=1}^n \mathbf{1}(X_i = v_j)$. The *extension* principle simply extends this, to estimating a continuous function $q(p_1, \ldots, p_k)$ by $q(\widehat{p}_1, \ldots, \widehat{p}_k)$, which is the *frequency substitution estimate.*

   Consider a population made up of three different types of individuals occurring in the Hardy-Weinberg proportions $\mathbb{P}_\theta(X = v_1) = \theta^2$, $\mathbb{P}_\theta(X = v_2) = 2\theta(1-\theta)$ and $\mathbb{P}_\theta(X = v_3) = (1-\theta)^2$ respectively.

   (a) Show that $T := \frac{N_1}{n} + \frac{N_2}{2n}$ is a frequency substitution estimate of θ.

   (b) Using part (a), find a frequency substitution estimate of the odds ratio $\frac{\theta}{1-\theta}$.

   (c) Suppose $v_1 = -1$, $v_2 = 0$ and $v_3 = 1$. By considering the first moment of X, show that T is a method of moment estimate of θ.

3. Let $X_1, \ldots, X_n$ be i.i.d., with $\mathrm{Beta}(\beta_1, \beta_2)$ distribution. Find the method of moments estimate of $\beta = (\beta_1, \beta_2)$ based on the first two moments.

4. Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli trials, each with success probability θ. Let $\psi : \mathbb{R}^n \times (0,1) \to \mathbb{R}$ be the function defined by:

$$\psi(X_1, \ldots, X_n, \theta) = \frac{1}{\theta} \sum_{j=1}^n X_j - \frac{1}{1-\theta} \left(n - \sum_{j=1}^n X_j \right).$$

   Compute

$$V(\theta_0, \theta) = \mathbb{E}_{\theta_0} \left[\psi(X_1, \ldots, X_n, \theta) \right]$$

   and show that θ_0 is the unique solution of $V(\theta_0, \theta) = 0$. Compute the estimating equation estimate of θ.

5. **General method of moment estimates** Suppose $X_1, \ldots, X_n$ are i.i.d., as $X \sim \mathbb{P}_\theta : \theta \in \Theta \subset \mathbb{R}^d$ and θ identifiable. Let $g_1, \ldots, g_d$ be linearly independent functions and set

$$\mu_j(\theta) := \mathbb{E}_\theta\left[g_j(X)\right] \qquad \widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} g_j(X_i)$$

The moment method estimates are $\widehat{\theta}$ such that $\mu_j(\widehat{\theta}) = \mu_j$ for each $j = 1, \ldots, d$. Furthermore, for the parameters of a canonical exponential family, the moment method estimator of the parameter vector η is the moment method estimator based on the sufficient statistic T. Recall that for an exponential family in its canonical coordinates

$$\dot{A}(\eta) = \mathbb{E}_\eta\left[T(X)\right]$$

where $\dot{A}$ denotes the vector of partial derivatives.

Suppose that $\{\mathbb{P}_\theta : \theta \in \Theta\}$ is a k-parameter exponential family generated by (h, T) where $T = (T_1, \ldots, T_k)$. Using $g_j = T_j$ in the above, find the method of moments estimates for the parameters in:

(a) The Rayleigh distribution:

$$p(x, \theta) = \left(\frac{x}{\theta^2}\right)\exp\left\{-\frac{x^2}{2\theta^2}\right\} \qquad x > 0, \quad \theta > 0,$$

(b) the Gamma distribution gamma(p, θ) where p is fixed.

6. When the data is not i.i.d., it may still be possible to express parameters as functions of moments and then use estimates based on replacing population moments with 'sample' moments.

Let $X_1, \ldots, X_n$ satisfy:

$$\begin{cases} X_i = \mu + e_i & i = 1, \ldots, n \\ e_i = \beta e_{i-1} + \epsilon_i & i = 1, \ldots, n, \qquad e_0 = 0. \end{cases}$$

where $\epsilon_1, \epsilon_2, \epsilon_3, \ldots$ are independent, identically distributed, $\mathbb{E}\left[\epsilon_j\right] = 0$ and $\mathrm{Var}(\epsilon_j) = \sigma^2$.

(a) Use $\mathbb{E}[X_i]$ to give a method of moments estimate of μ.

(b) Suppose $\mu = \mu_0$ and $\beta = b$ are fixed. Use $\mathbb{E}[U_i^2]$ where

$$U_i = \frac{X_i - \mu_0}{\left(\sum_{j=0}^{i-1} b^{2j}\right)^{1/2}}$$

to give a method of moments estimate of σ^2.

(c) If μ and σ^2 are fixed, can you give a method of moments estimate of β?

7. $X_1, \ldots, X_n$ random sample from distribution with density

$$f(x) = \theta x^{-\theta-1} \qquad x \geq 1$$

$\theta > 0$ unknown parameter

   (a) Compute $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$, the method of moments estimator.

   (b) Show that, for $\theta > 2$, $\hat{\theta}_n \to_p \theta$.

8. Let $\theta = (\theta_1, \theta_2)$ be a bivariate parameter. Suppose that X has state space $\mathcal{X}$, $T_1 : \mathcal{X} \to \mathbb{R}$ and $T_2 : \mathcal{X} \to \mathbb{R}$ are functions such that $T_1(X)$ is sufficient for θ_1 whenever θ_2 is fixed and known, whereas $T_2(X)$ is sufficient for θ_2 whenever θ_1 is fixed and known. Assume that $S = \{x | p(x, \theta) > 0\}$ does not depend on θ (p a density if X is a continuous variable, a probability function if it is discrete).

   (a) Show that if T_1 and T_2 do not depend on θ_2 and θ_1 respectively, then $(T_1(X), T_2(X))$ is sufficient for θ.

   (b) Give an example where $(T_1(X), T_2(X))$ is sufficient for θ, where $T_1(X)$ is sufficient for θ_1 whenever θ_2 is fixed and known, $T_2(X)$ is *not* sufficient for θ_2 whenever θ_1 is fixed and known.

9. **Censored geometric waiting times** Let

$$\mathbb{P}_\theta(X = k) = \theta^{k-1}(1 - \theta) \qquad k = 1, 2, \ldots$$

where $0 < \theta < 1$, where X is the time to failure. Suppose that we only record a time to failure if it occurs on or before time r, and record that it survives for longer than time r otherwise. Suppose we observe n individuals, m of which survive longer than time r and the failure times of the others are $Y_1, \ldots, Y_{n-m}$, where $1 \leq Y_j \leq r$ for $j = 1, \ldots, n-m$. Let $S = \sum_{j=1}^{n-m} Y_j$. Compute the maximum likelihood estimator of θ based on all the information.

10. **Maximum Likelihood (Normal)** Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, where both μ and σ are unknown. Compute the maximum likelihood estimators for the pair (μ, σ^2). Are the estimators of μ and σ^2 unbiased?

Short Answers

1. (a)

$$\mathbb{P}(\max_j X_j \le x) = \mathbb{P}(X_1 \le x)^n = \begin{cases} 0 & x < 0 \\ \frac{x^n}{\theta^n} & 0 \le x \le \theta \\ 1 & x > \theta \end{cases}$$

so, setting $Y = \max_j X_j$, the density is

$$p_Y(x;\theta) = \frac{nx^{n-1}}{\theta^n}\mathbf{1}_{[0,\theta]}(x)$$

so that

$$\mathbb{E}[Y] = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{n+1}\theta.$$

Since $T_1 = \frac{n+1}{n}Y$, therefore $\mathbb{E}_\theta[T_1] = \theta$.

$$\mathbb{E}_\theta[Y^2] = \frac{n}{\theta^n}\int_0^\theta x^{n+1} dx = \frac{n}{n+2}\theta^2$$

Hence

$$\mathrm{Var}_\theta(Y) = \frac{n}{n+2}\theta^2 - \frac{n^2}{(n+1)^2}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$$

so that

$$\mathrm{Var}_\theta(T_1) = \left(\frac{n+1}{n}\right)^2 \mathrm{Var}(Y) = \frac{1}{n(n+2)}\theta^2.$$

(b)

$$\mathbb{E}_\theta[T_2] = \frac{2}{n}\sum_{j=1}^n \mathbb{E}_\theta[X_j] = \frac{2}{n}\frac{\theta}{2} = \theta.$$

If $X \sim U(0,\theta)$ then $\mathrm{Var}_\theta(X) = \frac{\theta^2}{12}$ so that

$$\mathrm{Var}_\theta(T_2) = \frac{4}{n^2} \times n \times \frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

(c) That $\max_j X_j$ is the maximum likelihood estimator of θ is seen as follows:

$$L(\theta; x_1, \ldots, x_n) = \frac{1}{\theta^n}\mathbf{1}_{[\max_j x_j, +\infty)}(\theta)$$

which (as a function of θ) is maximised by $\widehat{\theta}_{ML} = \max_j x_j$.

(d) That T_2 is the method of moments estimator of θ is seen as follows: $\mathbb{E}_\theta[X_1] = \frac{\theta}{2}$ so that $\theta = 2\mathbb{E}_\theta[X_1]$ and hence $\widehat{\theta}_{MM} = 2\overline{X} = \frac{2}{n}\sum_{j=1}^n X_j$.

2. (a) $p_1 = \theta^2$, $p_2 = 2\theta(1-\theta) = 2\theta - 2\theta^2 = 2\theta - 2p_1$ so that

$$\theta = p_1 + \frac{p_2}{2}$$

hence a frequency substitution estimate is:

$$\widehat{\theta} = \frac{N_1}{n} + \frac{N_2}{2n}.$$

(b) There are several answers. One answer: from before, estimate is:

$$\frac{\widehat{\theta}}{1-\widehat{\theta}} = \frac{2N_1 + N_2}{2n - 2N_1 - N_2}.$$

Another answer:

$$\frac{\theta}{1-\theta} = \frac{\theta^2}{\theta(1-\theta)} = \frac{p_1}{p_2/2} \qquad \widehat{\frac{\theta}{1-\theta}} = \frac{2N_1}{N_2}.$$

(c) $\mathbb{E}_\theta[X] = -\theta^2 + (1-\theta)^2 = 1 - 2\theta$

$$\widehat{\mu} = -\frac{N_1}{n} + \frac{N_3}{n} = 1 - \frac{N_2}{n} - \frac{2N_1}{n}.$$

Hence moment method estimate (based on first moment) is:

$$1 - 2\widehat{\theta} = 1 - \frac{N_2}{n} - \frac{2N_1}{n} \Rightarrow \widehat{\theta} = \frac{N_1}{n} + \frac{N_2}{2n}$$

as required.

3.

$$f(x) = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} x^{\beta_1}(1-x)^{\beta_2} \qquad x \in (0,1)$$

$$\mathbb{E}[X] = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} \frac{\Gamma(\beta_1+1)\Gamma(\beta_2)}{\Gamma(\beta_1+\beta_2+1)} = \frac{\beta_1}{\beta_1+\beta_2}$$

$$\mathbb{E}[X^2] = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} \int_0^1 x^{\beta_1+1}(1-x)^{\beta_2-1} dx = \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)\Gamma(\beta_2)} \frac{\Gamma(\beta_1+2)\Gamma(\beta_2)}{\Gamma(\beta_1+\beta_2+2)}$$

$$\mathbb{E}[X^2] = \frac{(\beta_1+1)\beta_1}{(\beta_1+\beta_2+1)(\beta_1+\beta_2)}$$

Let $m_1 = \mathbb{E}[X]$ and $m_2 = \mathbb{E}[X^2]$. Then from the first equation

$$\beta_2 = \beta_1\left(\frac{1}{m_1} - 1\right)$$

and from the second equation

$$m_2 = \frac{\beta_1+1}{\beta_1+\beta_2+1} m_1 \Rightarrow m_2 = \frac{m_1^2(\beta_1+1)}{\beta_1+m_1} \Rightarrow \beta_1 = \frac{m_1(1-m_2)}{m_2-m_1^2}, \quad \beta_2 = \frac{(1-m_1)(1-m_2)}{m_2-m_1^2}$$

$$\widehat{\beta}_1 = \frac{\overline{x}(1-\overline{x^2})}{\overline{x^2}-\overline{x}^2} \qquad \widehat{\beta}_2 = \frac{(1-\overline{x})(1-\overline{x^2})}{\overline{x^2}-\overline{x}^2}.$$

4.
$$\mathbb{E}_{\theta_0}\left[\psi(X_1,\dots,X_n,\theta)\right] = n\left(\frac{\theta_0}{\theta} - \frac{1-\theta_0}{1-\theta}\right) = n\frac{\theta_0-\theta}{\theta(1-\theta)}$$

which is 0 if and only if $\theta = \theta_0$ as required.

Estimating equation estimate satisfies

$$\psi(X_1,\dots,X_n,\widehat{\theta}) = 0 = \frac{1}{\widehat{\theta}}\sum_{j=1}^n X_j - \frac{1}{1-\widehat{\theta}}\left(n - \sum_{j=1}^n X_j\right) \Rightarrow \widehat{\theta} = \frac{1}{n}\sum_{j=1}^n X_j.$$

5. (a) Set $\eta = \frac{1}{2\theta^2}$ then

$$p(x,\theta) = x\exp\left\{\eta(-x^2) - (-\log(2\eta))\right\}$$
$$T(x) = -x^2$$
$$A(\eta) = -\log(2\eta) \Rightarrow \frac{d}{d\eta}A(\eta) = -\frac{1}{\eta} = -2\theta^2 = -\mathbb{E}_\theta\left[X^2\right]$$

so

$$\widehat{\theta} = \left(\frac{1}{2n}\sum_{j=1}^n X_j^2\right)^{1/2}$$

(b)

$$p(x,\theta) = \frac{\theta^p}{\Gamma(p)}x^{p-1}e^{-\theta x} \qquad x > 0$$

set $\eta = \theta$ then

$$p(x,\eta) = \frac{x^{p-1}}{\Gamma(p)}e^{\eta(-x)-(-p\log(\eta))}$$
$$T(x) = x \qquad A(\eta) = -p\log(\eta)$$
$$\frac{d}{d\eta}A(\eta) = -\frac{p}{\eta}$$

so that

$$\mathbb{E}_\theta[T(X)] = -\mathbb{E}_\theta[X] = -\frac{p}{\theta}$$

so

$$\widehat{\theta} = \frac{p}{\overline{X}}$$

6. (a) $\mu = \mathbb{E}[X_i]$ so $\widehat{\mu} = \overline{X}$.

(b) Here $X_i = e_i$ and

$$\mathrm{Var}(e_i) = b^2\mathrm{Var}(e_{i-1}) + \sigma^2 \Rightarrow \mathrm{Var}(e_i) = \sigma^2\left(\sum_{i=0}^{j-1} b^{2i}\right)$$

It follows that $\mathbb{E}\left[U_i^2\right] = \sigma^2$ for each $i = 1,\dots,n$, hence

$$\widehat{\sigma}^2 := \frac{1}{n}\sum_{j=1}^n U_j^2$$

is a method of moments estimator of σ^2.

(c) For $i = 1, 2, \ldots, n$ (using $X_0 = \mu$ and $\mathbb{E}[X_i] = \mu$),

$$(X_i - \mu) = \beta(X_{i-1} - \mu) + \epsilon_i$$

so that

$$\mathrm{Cov}(X_i, X_{i-1}) = \beta \mathrm{Var}(X_{i-1})$$

hence

$$\sum_{i=1}^{n-1} \mathrm{Cov}(X_{i+1}, X_i) = \beta \sum_{i=1}^{n-1} \mathrm{Var}(X_i)$$

This leads to

$$\widehat{\beta} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)(X_{i+1} - \mu)}{\frac{1}{n}\sum_{i=1}^{n-1}(X_i - \mu)^2}$$

This method of moments estimator is consistent provided $|\beta| < 1$, so that

$$\mathrm{Var}(X_i) \to \frac{\sigma^2}{1-\beta^2} \qquad \mathrm{Cov}(X_i, X_{i+1}) \to \beta\frac{\sigma^2}{1-\beta^2}.$$

7. (a)

$$\mathbb{E}[X] = \theta \int_1^\infty x.x^{-\theta-1}dx = \theta \left[\frac{x^{1-\theta}}{1-\theta}\right]_1^\infty = \frac{\theta}{\theta - 1}$$

$$\frac{\widehat{\theta}_n}{\widehat{\theta}_n - 1} = \overline{X}$$

$$\widehat{\theta}_n = \frac{\overline{X}}{\overline{X} - 1}.$$

(b) To prove convergence,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{\overline{X}}{\overline{X}-1} - \theta\right| > \epsilon\right) &\leq \mathbb{P}\left(\left|\overline{X} - \frac{\theta}{\theta-1}\right| > \frac{\epsilon}{(\theta-1+\epsilon)(\theta-1)}\right) \\
&\leq \left(\frac{(\theta-1+\epsilon)(\theta-1)}{\epsilon}\right)^2 \mathrm{Var}(\overline{X}) \\
&= \left(\frac{(\theta-1+\epsilon)(\theta-1)}{\epsilon}\right)^2 \frac{1}{n}\mathrm{Var}(X_1) \stackrel{n\to+\infty}{\longrightarrow} 0
\end{aligned}$$

provided $\mathrm{Var}(X_1) < +\infty$, which requires $\theta > 2$. Techniques using the characteristic function enable this to be extended to $\theta > 1$; convergence can be proved using a characteristic function technique if $\mathbb{E}[|X_1|] < +\infty$.

8. (a) $T(X)$ sufficient if and only if there are functions g and h such that

$$p(x, \theta) = g(T(x), \theta)h(x).$$

The conditions give functions g_1, g_2, h_1 and h_2 such that

$$p(x, \theta) = g_1(T_1(x), \theta_1)h_1(x, \theta_2) = g_2(T_2(x), \theta_2)h_2(x, \theta_1).$$

It follows that

$$\frac{g_1(T_1(x), \theta_1)}{h_2(x, \theta_1)} = \frac{g_2(T_2(x), \theta_2)}{h_1(x, \theta_2)} = f(x)$$

and hence that

$$p(x, \theta) = g_1(T_1(x), \theta_1)g_2(T_2(x), \theta_2)f(x)$$

and hence $(T_1(X), T_2(X))$ is sufficient for (θ_1, θ_2).

(b) Sample $(X_1, \ldots, X_n)$ from $N(\mu, \sigma^2)$, statistics $T_1(X) = \sum_{j=1}^n X_j$ and $T_2(X) = \sum_{j=1}^n X_j^2$. Here $T_1(X)$ is sufficient for μ whether or not σ^2 is known. The statistic $\sum_{j=1}^n X_j^2$ is not sufficient for σ^2; $\sum_{j=1}^n X_j$ is also needed whether or not μ is known. This is seen from the factorisation: the formula is

$$p(x, \sigma^2) = \exp\left\{ -\frac{1}{2\sigma^2}\sum_{j=1}^n x_j^2 + \frac{\mu}{\sigma^2}\sum_{j=1}^n x_j - \frac{n\mu^2}{\sigma^2} - \frac{n}{2}\log(2\pi\sigma^2) \right\}$$

and, even if μ is fixed, both $\sum_{j=1}^n x_j^2$ and $\sum_{j=1}^n x_j$ give information about σ^2.

9. Firstly, we compute the probability that failure occurs after time r. It is:

$$\mathbb{P}_\theta(X \geq r+1) = (1-\theta)\sum_{k=r+1}^{\infty} \theta^{k-1} = (1-\theta)\theta^r \sum_{k=0}^{\infty} \theta^k = \theta^r$$

Now suppose the events $Y_1 \in A_1, \ldots, Y_n \in A_n$ are observed. The likelihood is:

$$L(\theta; Y_1 \in A_1, \ldots, Y_n \in A_n) = \prod_{j=1}^n L(\theta; Y_j \in A_j)$$

since these events are independent. If $A_j = \{r+1, r+2, \ldots\}$ we have $L(\theta, Y_j \in A_j) = \theta^r$. If $A_j = \{y_j\}$, we have $L(\theta; Y_j \in A_j) = \theta^{y_j - 1}(1-\theta)$ and multiplying these together gives:

$$L(\theta) = \theta^{S-(n-m)}(1-\theta)^{n-m}\theta^{rm}.$$

To get the maximum, it is probably easier to use logarithms:

$$\log L(\theta) = (S - n + (r+1)m)\log\theta + (n-m)\log(1-\theta)$$

Taking derivative to get critical points:

$$\frac{d}{d\theta}\log L(\theta) = \frac{S - n + (r+1)m}{\theta} - \frac{n-m}{1-\theta}$$

$$\widehat{\theta} = \frac{S + (r+1)m - n}{S + rm}$$

Maximum - clear -

$$\frac{d^2}{d\theta^2} = -\frac{S - n + (r+1)m}{\theta^2} - \frac{n-m}{(1-\theta)^2} < 0$$

hence strictly concave, $\log L(\theta) > +\infty$ and bounded above for $0 < \theta < 1$ and $\log L(0) = \log L(1) = -\infty$ so exists a maximum, which is unique and given at point where $\frac{d}{d\theta} \log L = 0$.

10. Let $v = \sigma^2$

$$\log L(\mu, v) = -\frac{n}{\log}(2\pi) - \frac{n}{2} \log v - \frac{1}{2v} \sum_{j=1}^{n} (x_j - \mu)^2$$

so the likelihood equations are:

$$\begin{cases} -\frac{n}{2v} + \frac{1}{2v^2} \sum_{j=1}^{n} (x_j - \mu)^2 = 0 \\ \sum_{j=1}^{n} (x_j - \mu) = 0 \end{cases}$$

It follows that $\widehat{\mu} = \frac{1}{n} \sum_{j=1}^{n} X_j = \overline{X}$. For the other equation,

$$v = \frac{1}{n} \sum_{j=1}^{n} (x_j - \mu)^2$$

so that $\widehat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} (X_j - \overline{X})^2$. This is an exponential family and conditions that MLE exists and is the unique solution of the likelihood equations are clearly satisfied. From earlier:

$$\mathbb{E}\left[\widehat{\mu}\right] = \mu, \qquad \mathbb{E}\left(\widehat{\sigma}^2\right) = \frac{n-1}{n} \sigma^2$$

so the variance estimator is biased.