MA20033 - Solution Sheet Two

Simon Shaw
s.c.shaw@maths.bath.ac.uk

2004/05 Semester I

1. **In Lecture 3, we showed that if $X_1, \ldots, X_n$ are independently and identically distributed as $Po(\lambda)$, and we observe $X_1 = x_1, \ldots, X_n = x_n$, then the maximum likelihood estimate of λ is $\hat{\lambda} = \overline{x}$.**

   (a) **Write down the probability mass function of X_i under the reparameterisation $\nu = 1/\lambda$ (i.e. change all the λ to $1/\nu$).**

   The probability mass function of a Poisson random quantity using the reparameterisation $\nu = 1/\lambda$ is

$$P(X = x|\nu) \quad = \quad \frac{(1/\nu)^x \exp(-1/\nu)}{x!} \quad x = 0, 1, \ldots$$

   (b) **Under this reparameterisation, what is the likelihood function, $L(\nu)$, given the data $X_1 = x_1, \ldots, X_n = x_n$?**

   Under the assumption of independence, the likelihood function is

$$\begin{aligned} L(\nu) &= \prod_{i=1}^{n} P(X_i = x_i|\nu) \\ &= \prod_{i=1}^{n} \frac{\nu^{-x_i} \exp(-1/\nu)}{x_i!} \\ &= \frac{\nu^{-\sum_{i=1}^{n} x_i} \exp(-n/\nu)}{\prod_{i=1}^{n} x_i!} \\ &= \frac{\nu^{-n\overline{x}} \exp(-n/\nu)}{\prod_{i=1}^{n} x_i!}. \end{aligned}$$

   (c) **Thus, find the maximum likelihood estimate, $\hat{\nu}$, of ν based on this data.**

   The log-likelihood is given by

$$l(\nu) \quad = \quad -n\overline{x} \log \nu - \frac{n}{\nu} - \sum_{i=1}^{n} \log x_i!$$

   Differentiating with respect to ν gives

$$l'(\nu) \quad = \quad -\frac{n\overline{x}}{\nu} + \frac{n}{\nu^2}.$$

Solving $l'(\nu) = 0$ gives $\hat{\nu} = 1/\overline{x}$. Note that $l''(\hat{\nu}) = -n\overline{x}^3 < 0$ so that this is the maximum likelihood estimate. Notice that for $X \sim Po(\lambda)$, with $P(X = x|\lambda) = \lambda^x \exp(-\lambda)/x!$, the maximum likelihood estimate is $\hat{\lambda} = \overline{x}$. Thus, $\hat{\nu} = 1/\hat{\lambda}$. This is an example of the invariance property of maximum likelihood estimates: if $\hat{\theta}$ is the maximum likelihood estimate of θ and $u(\theta)$ a function of θ, then $u(\hat{\theta})$ is the maximum likelihood estimate of $u(\theta)$.

2. **The independent observations $x_1, \ldots, x_n$ are assumed to come from a $N(\mu, \sigma^2)$ distribution.**

   (a) **In the case where σ^2 is known and μ is unknown, find the maximum likelihood estimator of μ. Is the estimator biased?**

   There is a single unknown parameter, μ, and the likelihood function is

$$L(\mu) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}.$$

   The log-likelihood is thus

$$l(\mu) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

   Differentiating with respect to μ gives

$$l'(\mu) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu).$$

   Solving $l'(\mu) = 0$ gives $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$. We may check that $l''(\overline{x}) = -n/\sigma^2 < 0$ so that this is indeed a maximum. The maximum likelihood estimator is thus $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Note that

$$\begin{aligned} b(\overline{X}) &= E(\overline{X}|\mu, \sigma^2) - \mu \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i|\mu, \sigma^2) - \mu \;=\; 0 \end{aligned}$$

   as $E(X_i|\mu, \sigma^2) = \mu$ for all i. Thus, $\overline{X}$ is an unbiased estimator for μ.

   (b) **In the case where μ is known and σ^2 is unknown, find the maximum likelihood estimator of σ^2. Is the estimator biased?**

   Once again there is a single unknown parameter, this time σ^2, and the likelihood function is

$$L(\sigma^2) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}.$$

   The log-likelihood is thus

$$l(\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Differentiating with respect to σ^2 gives

$$l'(\sigma^2) \quad = \quad -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Solving $l'(\sigma^2) = 0$ gives $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$. We may check that $l''(\hat{\sigma}^2) = -n/2(\sigma^2)^2 < 0$ so that this is indeed a maximum. The maximum likelihood estimator is thus $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. Note that

$$\begin{aligned}
b\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\right\} &= E\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \,\middle|\, \mu, \sigma^2\right\} - \sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} E\{(X_i - \mu)^2 | \mu, \sigma^2\} - \sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{Var(X_i - \mu | \mu, \sigma^2) + E^2(X_i - \mu | \mu, \sigma^2)\} - \sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} Var(X_i | \mu, \sigma^2) - \sigma^2 \ = \ 0,
\end{aligned}$$

as $Var(X_i|\mu, \sigma^2) = \sigma^2$ for all i. Thus, $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ is an unbiased estimator for σ^2.

(c) **In the case when μ and σ^2 are unknown, we showed, in Lecture 3, that the respective maximum likelihood estimators were $\overline{X}$ and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Are these estimators biased?**

We have showed in (a) that $\overline{X}$ is an unbiased estimator for μ. Now,

$$\begin{aligned}
b\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right\} &= E\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \,\middle|\, \mu, \sigma^2\right\} - \sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} Var(X_i - \overline{X} | \mu, \sigma^2) - \sigma^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\{Var(X_i|\mu, \sigma^2) + Var(\overline{X}|\mu, \sigma^2) - 2Cov(X_i, \overline{X}|\mu, \sigma^2)\} - \sigma^2.
\end{aligned}$$

Now,

$$\begin{aligned}
Cov(X_i, \overline{X}|\mu, \sigma^2) &= \frac{1}{n}\sum_{j=1}^{n} Cov(X_i, X_j|\mu, \sigma^2) \\
&= \frac{1}{n}\left\{Var(X_i|\mu, \sigma^2) + \sum_{j \neq i}^{n} Cov(X_i, X_j|\mu, \sigma^2)\right\} \ = \ \frac{\sigma^2}{n},
\end{aligned}$$

as $Cov(X_i, X_j|\mu, \sigma^2) = 0$ for $j \neq i$ as the individuals are independent. Also, as a consequence of this independence, $Var(\overline{X}|\mu, \sigma^2) = \sigma^2/n$. Hence,

$$b\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right\} \ = \ \frac{1}{n}\sum_{i=1}^{n}\left(\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right) - \sigma^2 \ = \ -\frac{1}{n}\sigma^2.$$

The estimator $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is biased for σ^2.

3. **A random quantity X has a uniform distribution on the interval $(0, \theta)$ so that its pdf is given by**

$$f(x|\theta) \;=\; \begin{cases} \frac{1}{\theta} & 0 < x \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

A random sample $X_1, \ldots, X_n$ is drawn from this distribution in order to learn about the value of θ.

(a) **Show that the joint pdf of the X_i is given by**

$$f(x_1, \ldots, x_n|\theta) \;=\; \begin{cases} \frac{1}{\theta^n} & 0 < m \leq \theta, \\ 0 & \text{otherwise,} \end{cases}$$

where $m = \max\{x_1, \ldots, x_n\}$.

The joint pdf is

$$f(x_1, \ldots, x_n|\theta) \;=\; \prod_{i=1}^{n} f(x_i|\theta) \;=\; \begin{cases} \frac{1}{\theta^n} & 0 < x_i \leq \theta, \; i = 1, \ldots, n, \\ 0 & \text{otherwise,} \end{cases}$$

But $\{0 < x_i \leq \theta, i = 1, \ldots, n\}$ is equivalent to $0 < m \leq \theta$, where $m = \max\{x_1, \ldots, x_n\}$, and the result follows.

(b) **Sketch the likelihood function, $L(\theta)$, of θ. Hence, explain why the maximum likelihood estimate of θ, $\hat{\theta} = m$.**

The likelihood function is

$$L(\theta) \;=\; f(x_1, \ldots, x_n|\theta) \;=\; \begin{cases} \frac{1}{\theta^n} & m \leq \theta, \\ 0 & \text{otherwise,} \end{cases}$$

We sketch $L(\theta)$ against θ and $L(\theta)$ is zero until $\theta = m$ and is a positive decreasing function, θ^{-n}, for $\theta \geq m$. Thus, m is the value of θ for which $L(\theta)$ attains its maximum.

(c) **Derive the exact sampling distribution of $M = \max\{X_1, \ldots, X_n\}$.**

For $X \sim U(0, \theta)$,

$$P(X \leq m|\theta) \;=\; \begin{cases} 0 & m < 0 \\ \frac{m}{\theta} & 0 < m \leq \theta \\ 1 & m > \theta. \end{cases}$$

Now,

$$P(M \leq m|\theta) \;=\; \prod_{i=1}^{n} P(X_i \leq m|\theta) \;=\; \begin{cases} 0 & m < 0 \\ \left(\frac{m}{\theta}\right)^n & 0 < m \leq \theta \\ 1 & m > \theta. \end{cases}$$

Hence,

$$f_M(m|\theta) \;=\; \frac{\partial}{\partial m} P(M \leq m|\theta) \;=\; \begin{cases} n\frac{m^{n-1}}{\theta^n} & m \leq \theta, \\ 0 & \text{otherwise.} \end{cases}$$

(d) **Thus, show that $E(M|\theta) = (1 - \frac{1}{n+1})\theta$. Is it surprising that the maximum likelihood estimate "under-estimates" θ? Provide an unbiased estimator of θ.**

$$E(M|\theta) \;=\; \int_0^\theta m\left(n\frac{m^{n-1}}{\theta^n}\right)dm \;=\; \left(1 - \frac{1}{n+1}\right)\theta.$$

We would expect the true value of θ to be larger than the largest observation, so the result is not surprising. An unbiased estimator of θ is $\frac{n+1}{n}M$, which is larger than M.

4. **The independent observations $x_1, \ldots, x_n$ are assumed to come from a $U(a, b)$ distribution. Find the maximum likelihood estimates of a and b.**

For this question, it is crucial to remember where the pdf of $X \sim U(a, b)$ is non-zero:

$$f(x|a, b) \;=\; \begin{cases} \frac{1}{b-a} & \text{if } a < x < b, \\ 0 & \text{if } x < a \text{ or } x > b. \end{cases}$$

Thus, the joint distribution of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n|a, b) \;=\; \begin{cases} (\frac{1}{b-a})^n & \text{if } a < x_i < b \;\forall x_i, \\ 0 & \text{if } x_i < a \text{ or } x_i > b \text{ for any } x_i, \end{cases}$$

so that the likelihood function is

$$L(a, b) \;=\; \begin{cases} (\frac{1}{b-a})^n & \text{if } a < \min_i x_i \text{ and } \max_i x_i < b, \\ 0 & \text{otherwise.} \end{cases}$$

To find the maximum likelihood estimates, we need to maximise $L(a, b)$. Clearly, this is achieved by minimising $b - a$ subject to the constraints that $a < \min_i x_i$ and $b > \max_i x_i$, so the maximum likelihood estimate of a is $\hat{a} = \min_i x_i$ and the maximum likelihood estimate of b is $\hat{b} = \max_i x_i$.

5. **Let $x_1, \ldots, x_n$ be a random sample from a geometric distribution**

$$P(X = x|p) \;=\; (1 - p)p^{x-1}, \; x = 1, 2, \ldots$$

where $p \in (0, 1)$ so that X is the number of trials until the first failure in a sequence of independent trials with success probability p.

(a) **Explain why the sample average $\overline{x}$ and sample size n are sufficient for calculating the likelihood function.**

The likelihood function is

$$L(p) \;=\; \prod_{i=1}^{n}(1 - p)p^{x_i-1} \;=\; (1 - p)^n p^{n\overline{x}-n}.$$

Hence, it is sufficient to know the values of n and $\overline{x}$ to compute $L(p)$.

(b) **Find the maximum likelihood estimator of p.**

The log-likelihood is

$$l(p) \quad = \quad n \log(1-p) + n(\overline{x}-1)\log p.$$

Differentiating with respect to p gives

$$l'(p) \quad = \quad -\frac{n}{1-p} + \frac{n}{p}.$$

Solving $l'(p) = 0$ gives $\hat{p} = \frac{\overline{x}-1}{\overline{x}}$. We may check that $l''(\hat{p}) = -n\overline{x}^3/(\overline{x}-1) < 0$ so that $\hat{p}$ is a maximum.

(c) **Show that $P(X > x|p) = p^x$.**

$$\begin{aligned} P(X > x|p) &= 1 - P(X \leq x|p) \\ &= 1 - \sum_{y=1}^{x} (1-p)p^{y-1} \\ &= 1 - \sum_{y=1}^{x} p^{y-1} + \sum_{y=1}^{x} p^y \ = \ p^x. \end{aligned}$$

6. **As part of a quality control procedure for a certain mass production process, batches containing very large numbers of components from the production are inspected for defectives. We will assume the process is in equilibrium so that each component is independent and is either acceptable, with probability p, or defective, with probability $q = 1 - p$.**

The inspection procedure is as follows. During each shift n batches are selected from the production and for each such batch components are inspected until a defective one if found, and the number of inspected components is recorded. At the end of the shift, there may be some inspected batches which have not yet yielded a defective component; and for such batches the number of inspected components is recorded.

Suppose that at the end of one such inspection shift, a defective component was detected in each of r of the batches, the recorded numbers of inspected batches being $x_1, \ldots, x_r$. Inspection of the remaining $s = n - r$ batches was incomplete, the recorded numbers of inspected components being $c_1, \ldots, c_s$.

(a) **By considering question 5, argue that the likelihood for $q = 1 - p$ based on these data is**

$$L(q) \quad = \quad q^r(1-q)^{x+c-r},$$

where $x = \sum_{i=1}^{r} x_i$ and $c = \sum_{i=1}^{s} c_i$.

The likelihood function is

$$\begin{aligned} L(p) &= P(X_1 = x_1, \ldots, X_r = x_r, X_{r+1} > c_1, \ldots, X_n > c_s|p) \\ &= \left\{ \prod_{i=1}^{r} P(X_i = x_i|p) \right\} \left\{ \prod_{j=1}^{s} P(X_{r+j} > c_j|p) \right\} \end{aligned}$$

$$
\begin{aligned}
&= \left\{\prod_{i=1}^{r}(1-p)p^{x_i-1}\right\}\left\{\prod_{j=1}^{s}p^{c_j}\right\} \\
&= (1-p)^r p^{\sum_{i=1}^{r} x_i - r} p^{\sum_{j=1}^{s} c_j}.
\end{aligned}
$$

Putting $q = 1 - p$, $x = \sum_{i=1}^{r} x_i$ and $c = \sum_{i=1}^{s} c_i$ gives the result.

(b) **Show that the maximum likelihood estimate of q is $\hat{q} = 1/a$, where $a = (x + c)/r$. Interpret a.**

The log-likelihood is

$$
l(q) = r \log q + (x + c - r) \log 1 - q.
$$

Differentiating with respect to q gives

$$
l'(q) = \frac{r}{q} - \frac{x + c - r}{1 - q}.
$$

Solving $l'(q) = 0$ gives $\hat{q} = 1/a$ where $a = (x + c)/r$. We may check that $l''(\hat{q}) = -(y + c)^3/\{r(y + c - r)\} < 0$ so that $\hat{q}$ is a maximum. a is the total number of components inspected divided by the number of components which failed. Thus, it is observed average number of components inspected until a failure so that $1/a$ intuitively makes sense as an estimate of q, the probability of a component failing.