

MATH562 – Fall 2025

Problem Set: Week 7

1. If $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Poiss}(\lambda)$, justify why $S = \sum_{j=1}^n Y_j$ is a complete minimal sufficient statistic for λ , and state its distribution. **Solution:** The Poisson model is a $(1, 1)$ exponential family, so S is a complete minimal sufficient statistic. It is a sum of independent Poisson variables, so $S \sim \text{Poiss}(n\lambda)$.

- (a) What is the conditional distribution of Y_1, \dots, Y_n given that $S = s$?

Solution: The required conditional distribution is obtained by setting

$$f(y_1, \dots, y_n | s) = \frac{f(y_1, \dots, y_n)f(s | y_1, \dots, y_n)}{f(s)} = \frac{\prod_{j=1}^n \lambda^{y_j} e^{-\lambda} / y_j! \times I(s = y_1 + \dots + y_n)}{(n\lambda)^s e^{-n\lambda} / s!}$$

and this can be written as

$$f(y_1, \dots, y_n | s) = \frac{s!}{\prod_{j=1}^n y_j!} \prod_{j=1}^n n^{-y_j} \times I(s = y_1 + \dots + y_n), \quad y_1, \dots, y_n \in \{0, 1, \dots\},$$

which is multinomial with denominator s and probability vector $(1/n, \dots, 1/n)$.

- (b) One strategy to find an optimal unbiased estimator of some function $\psi(\lambda)$ is to find any unbiased estimator T of $\psi(\lambda)$, and then to compute $h(S) = E(T | S)$. Another strategy is to find the function $h(s)$ that satisfies $E\{h(S)\} = \psi(\lambda)$ for all λ . Will these give the same estimator?

Solution: Yes, because any function $h(S)$ that is unbiased for $\psi(\lambda)$ must be unique, by completeness of S .

- (c) Find minimum variance unbiased estimators of (i) $e^{-\lambda}$ and (ii) $e^{-2n\lambda}$. Do you think these are reasonable? If not, suggest better estimators.

Solution: (i) Either strategy could be used. For the first, note that as $T = I(Y_1 = 0)$ is unbiased for $e^{-\lambda}$, we need to compute $E(T | S)$. The multinomial distribution in (a) implies that $Y_1 | S = s \sim B(s, 1/n)$, so

$$E(T | S = s) = E\{I(Y_1 = 0) | S = s\} = \Pr(Y_1 = 0 | S = s) = (1 - 1/n)^s,$$

i.e., $(1 - 1/n)^S$ must be the optimal unbiased estimator. To check this, note that

$$E\{(1 - 1/n)^S\} = \sum_{s=0}^{\infty} (1 - 1/n)^s (n\lambda)^s e^{-n\lambda} / s! = e^{-n\lambda} \sum_{s=0}^{\infty} (n-1)^s \lambda^s / s! = e^{-n\lambda} e^{(n-1)\lambda} = e^{-\lambda},$$

as required.

For the second strategy, we seek to solve $E\{h(S)\} = e^{-\lambda}$, i.e.,

$$\sum_{s=0}^{\infty} h(s) (n\lambda)^s e^{-n\lambda} / s! = e^{-\lambda} \implies \sum_{s=0}^{\infty} h(s) (n\lambda)^s / s! = e^{(n-1)\lambda},$$

and taking $h(s) = (n-1)^s / n^s$ achieves this, and gives the same estimator as the first strategy.

(ii) Here it's not clear what unbiased estimator to start from, so we use the second strategy. We need to solve $E\{h(S)\} = e^{-2n\lambda}$, giving

$$\sum_{s=0}^{\infty} h(s) (n\lambda)^s e^{-n\lambda} / s! = e^{-2n\lambda} \implies \sum_{s=0}^{\infty} h(s) (n\lambda)^s / s! = e^{-n\lambda},$$

so we must choose

$$h(s) = \begin{cases} -1, & s \text{ odd,} \\ 1, & s \text{ even.} \end{cases}$$

The estimator in (i) seems reasonable, but that in (ii) is not, as it can be negative when estimating a positive quantity, and ± 1 is unlikely to be close to the target (a small positive number, most likely). An obvious alternative would be $\exp(-2S)$, as $\hat{\lambda} = S/n$ is the maximum likelihood estimator; this estimator is biased but asymptotically has the smallest variance (by the Cramèr–Rao lower bound).

2. Let $M = \max(Y_1, \dots, Y_n)$ and \bar{Y} be the maximum and average of $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} U(0, \theta)$, respectively. Recall from the lectures that M is a complete minimal sufficient statistic.

(a) Show that $U = 2\bar{Y}$ is unbiased for θ and compute its variance.

Solution: We have $E(Y_j) = \theta/2$ and $\text{Var}(Y_j) = \theta^2/12$ so the unbiasedness of U is immediate and its variance is $\theta^2/(3n) = O(n^{-1})$.

(b) Use the Rao–Blackwell theorem to get a better unbiased estimator. Compute its variance.

Solution: The Rao–Blackwell theorem tells us that $E(U | M)$ will be unbiased with a smaller variance, and we can write

$$U = \frac{2}{n}(M + Y'_1 + \dots + Y'_{n-1}),$$

where Y'_1, \dots, Y'_{n-1} are the values of the original sample that are not the maximum. Each of these must lie in the interval $(0, M)$ and we saw in the notes that conditional on $M = m$ they are independent with $U(0, m)$ distributions, so $E(Y'_j | M = m) = m/2$. Hence

$$E(U | M = m) = \frac{2}{n} \{m + (n-1)m/2\} = \frac{m}{n}(2 + n - 1) = (n+1)m/n,$$

so the unbiased estimator based on M is $\tilde{\theta} = (n+1)M/n$. As $E(M^r) = n\theta^r/(n+r)$, we have $\text{Var}(M) = n\theta^2/\{(n+2)(n+1)^2\}$, so

$$\text{Var}(\tilde{\theta}) = \frac{(n+1)^2}{n^2} \text{Var}(M) = \frac{\theta^2}{n(n+2)} = O(n^{-2}),$$

i.e., Rao–Blackwellisation has given an unbiased estimator with variance of lower order in n , which is therefore much better than U .

For an alternative argument not involving order statistics, we might write

$$E(U | M) = \frac{2}{n} E \left\{ \sum_{j=1}^n Y_j | M = m \right\} = \frac{2}{n} \sum_{j=1}^n E \{ Y_j I(Y_j = M) + Y_j I(Y_j < M) | M = m \}$$

and then note that

$$E \{ Y_j I(Y_j = M) | M = m \} + E \{ Y_j I(Y_j < M) | M = m \} = m \frac{1}{n} + \frac{m}{2} \frac{n-1}{n}$$

giving

$$E(U | M) = \frac{2}{n} \sum_{j=1}^n \left(m \frac{1}{n} + \frac{m}{2} \frac{n-1}{n} \right) = (n+1)m/n.$$

*3. Observations $\dots, Y_1, \dots, Y_n, \dots$ arise in time order.

(a) Starting from

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{Y_n | Y_1, \dots, Y_{n-1}}(y_n | y_1, \dots, y_{n-1}) f_{Y_1, \dots, Y_{n-1}}(y_1, \dots, y_{n-1}),$$

establish the *prediction decomposition*

$$f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f_{Y_1}(y_1) \prod_{j=2}^n f_{Y_j | Y_1, \dots, Y_{j-1}}(y_j | y_1, \dots, y_{j-1}).$$

Solution: Replacing n in the first expression by $n - 1$ allows the term $f_{Y_1, \dots, Y_{n-1}}(y_1, \dots, y_{n-1})$ to be written in terms of the conditional density of Y_{n-1} given Y_1, \dots, Y_{n-2} , and then we simply iterate this, finishing by writing $f_{Y_1, Y_2}(y_1, y_2) = f_{Y_2 | Y_1}(y_2 | y_1) f_{Y_1}(y_1)$.

(b) A stationary first-order Gaussian autoregressive process satisfies

$$Y_j | Y_1 = y_1, \dots, Y_{j-1} = y_{j-1} \sim \mathbb{N}\{\mu + \alpha(y_{j-1} - \mu), \sigma^2\}, \quad j = 1, \dots, n,$$

where $|\alpha| < 1$, $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Find the log likelihood for data y_0, y_1, \dots, y_n from this model if the initial value y_0 is treated (i) as a known constant and (ii) as coming from the stationary distribution, $\mathbb{N}\{\mu, \sigma^2/(1 - \alpha^2)\}$.

Solution: This is a Markov process because the density of Y_j given the past depends only on Y_{j-1} , so we obtain

$$\begin{aligned} f_{Y_0, \dots, Y_n}(y_0, \dots, y_n) &= f_{Y_0}(y_0) \prod_{j=1}^n f_{Y_j | Y_{j-1}}(y_j | y_{j-1}) \\ &= f(y_0) \prod_{j=1}^n \frac{\exp\{-\{y_j - \mu - \alpha(y_{j-1} - \mu)\}^2 / (2\sigma^2)\}}{(2\pi\sigma^2)^{1/2}}, \end{aligned}$$

and the log likelihood is

$$-2\ell(\mu, \alpha, \sigma^2) \equiv -2 \log f(y_0) + n \log \sigma^2 + \frac{1}{\sigma^2} \sum_{j=1}^n \{y_j - \mu - \alpha(y_{j-1} - \mu)\}^2,$$

with (i) $f(y_0) = 1$ and (ii) $-2 \log f(y_0) \equiv \log \sigma^2 - \log(1 - \alpha^2) + (y_0 - \mu)^2(1 - \alpha^2)/\sigma^2$.

(c) Give a minimal sufficient statistic in (b). Is the model an exponential family?

Solution: The quadratic form in (b) can be expanded as

$$\frac{1}{\sigma^2} \left\{ \sum_{j=1}^n y_j^2 - 2\alpha \sum_{j=1}^n y_j y_{j-1} + \alpha^2 \sum_{j=1}^n y_{j-1}^2 + 2(\alpha - 1)\mu \sum_{j=1}^n y_j - 2\alpha(\alpha - 1)\mu \sum_{j=0}^{n-1} y_j + n(\alpha - 1)^2 \mu^2 \right\},$$

so if (i) y_0 is constant this is a (5, 3) exponential family (i.e. with natural parameter dimension 5 but only 3 parameters) with parameter vector $\theta = (\mu, \sigma^2, \alpha)$, sufficient statistic

$$s(y)^\top = \left(\sum_{j=1}^n y_j^2, \sum_{j=1}^n y_{j-1}^2, \sum_{j=1}^n y_j y_{j-1}, \sum_{j=1}^n y_j, \sum_{j=0}^{n-1} y_j \right),$$

and

$$\varphi(\theta)^\top = -2(1/\theta_2, \theta_3^2/\theta_2, -2\theta_3/\theta_2, 2(\theta_3 - 1)\theta_1/\theta_2, 2\theta_3(\theta_3 - 1)\theta_1/\theta_2), \quad k(\varphi) = -\frac{n(\theta_3 - 1)^2 \theta_1^2}{2\theta_2} - \frac{n}{2} \log \theta_2,$$

and (ii) y_0 must be added to the sufficient statistic, making the model a $(6, 3)$ exponential family.

It remains to show that $s(y)$ is also minimal, for which the example on minimal sufficiency in (k, k) exponential families cannot be directly applied, as the parameters do not contain an open set in \mathbb{R}^6 the natural parameter space. Accordingly, we consider the ratio $f(y | \theta)/f(y' | \theta)$ with y, y' vectors of variables of length n , restricting to (i), to apply the corresponding theorem:

$$\log \frac{f(y | \theta)}{f(y' | \theta)} = (s(y) - s(y'))^\top \varphi(\theta)$$

First, it follows immediately that “if $s(y) = s(y')$, $f(y | \theta)/f(y' | \theta)$ does not depend on θ ”, and it remains to show the converse. So, assume the ratio does not depend on θ . Since $\varphi(\theta) = \varphi((\mu, 1, \alpha)^\top)/\sigma^2$, this means that $(s(y) - s(y'))^\top \varphi(\theta) = 0$ for all θ . Choose $\vartheta_1^\top = (0, 1, \frac{9}{10})$, $\vartheta_2^\top = (1, 1, 0)$, $\vartheta_3^\top = (-1, 1, 0)$, $\vartheta_4^\top = (1, 1, \frac{1}{2})$, $\vartheta_5^\top = (-1, 1, \frac{1}{2})$, then $\Phi = (\varphi(\vartheta_1), \dots, \varphi(\vartheta_5))$ has full rank and is invertible. Hence, $(s(y) - s(y'))^\top \Phi = 0$ yields $s(y) = s(y')$, and minimal sufficiency follows from the theorem seen in the lecture.

4. Consider discrete data Y with density $f(y; \theta)$ defined for $y \in \mathcal{Y}$ and let $T = t(Y)$ be a statistic based on Y . Define the sets $\mathcal{T} = \{t(y) : y \in \mathcal{Y}\}$ and $\mathcal{C}_s = \{y \in \mathcal{Y} : t(y) = s\}$ for $s \in \mathcal{T}$.

- (a) Show that the phrase ‘ $y \sim y'$ if and only if $y, y' \in \mathcal{C}_s$ ’ defines an equivalence relation, and that the same equivalence relation is given by taking any bijective function of $t(y)$. Deduce that the equivalence classes form a partition \mathcal{P}_T of \mathcal{Y} .

Solution: Obviously (i) $y \sim y$ (reflexivity) and (ii) $y' \sim y$ is equivalent to $y' \sim y$ (symmetry). Equally obvious is that (iii) $y' \sim y$ and $y'' \sim y'$ implies that $y'' \sim y$ (transitivity). Hence the relation \sim is an equivalence relation, which implies that the equivalence classes \mathcal{C}_s for $s \in \mathcal{T}$ form a partition of \mathcal{Y} . If $g(s)$ is a bijective function of s , then an inverse function g^{-1} exists and so $\mathcal{C}_h = \{y : g\{t(y)\} = h\} = \{y : g^{-1}[g\{t(y)\}] = g^{-1}(h)\} = \{y : t(y) = g^{-1}(h)\} = \mathcal{C}_{g^{-1}(h)}$, so the equivalence classes defined using g are the same as those defined using t .

- (b) If T is sufficient, show that the conditional distribution of Y given that $Y \in \mathcal{C}_s$ does not depend on θ ; then \mathcal{P}_T is called a *sufficient partition*.

Solution: We have

$$\Pr(Y = y | Y \in \mathcal{C}_s; \theta) = \frac{\Pr(Y = y; \theta) \Pr(Y \in \mathcal{C}_s | Y = y; \theta)}{\Pr(Y \in \mathcal{C}_s; \theta)} = \frac{\Pr(Y = y; \theta)}{\Pr(Y \in \mathcal{C}_s; \theta)} I\{t(y) = s\},$$

and the factorisation theorem applies and the last expression here is, as required,

$$\frac{g\{t(y); \theta\} h(y)}{\sum_{y' \in \mathcal{C}_s} g\{t(y'); \theta\} h(y')} I\{t(y) = s\} = \frac{g(s; \theta) h(y)}{\sum_{y' \in \mathcal{C}_s} g(s; \theta) h(y')} = \frac{h(y)}{\sum_{y' \in \mathcal{C}_s} h(y')}.$$

- (c) If Y consists of n independent Poisson variables with mean θ , show that $T = (Y_1, Y_2 + \dots + Y_n)$ is sufficient and give \mathcal{Y}, \mathcal{T} and the \mathcal{C}_s . Find a coarser sufficient partition, and check if it is minimal.

Solution: The joint density is

$$\prod_{j=1}^n \frac{\theta^{y_j}}{y_j!} e^{-\theta} = \theta^{y_1 + (y_2 + \dots + y_n)} \exp(-n\theta) \times h(y),$$

where $h(y) = 1/\prod y_j!$, so the factorisation theorem implies that $(T_1, T_2) = (Y_1, Y_2 + \dots + Y_n)$ is sufficient, with $\mathcal{Y} = \{(y, \dots, y_n) : y_1, \dots, y_n \in \{0, 1, \dots\}\}$ and

$$\mathcal{T} = \{(t_1, t_2) : t_1, t_2 \in \{0, 1, \dots\}\}, \quad \mathcal{C}_s = \{y \in \mathcal{Y} : y_1 = s_1, \sum_{j=2}^n y_j = s_2\}, \quad s \in \mathcal{T}.$$

Hence the elements of this partition are determined by points (s_1, s_2) in the positive quadrant with values in $\{0, 1, 2, \dots\}$. It is clear from the form of the joint density that $T_1 + T_2 = Y_1 + \dots + Y_n$ is also sufficient, and the corresponding sufficient partition consists of the sets $\mathcal{C}'_t = \{(s_1, s_2) \in \mathcal{T} : s_1 + s_2 = t\}$ for $t \in \{0, 1, \dots\}$, which lie on diagonals in the positive quadrant. This partition is minimal because if y_1, \dots, y_n and z_1, \dots, z_m are both Poisson samples, their likelihood ratio is

$$\frac{\theta^{s_z} \exp(-m\theta) / \prod z_j!}{\theta^{s_y} \exp(-n\theta) / \prod y_j!} \propto \theta^{s_z - s_y} \exp\{\theta(n - m)\},$$

and this is independent of θ iff $n = m$ and $s_z = \sum_{j=1}^m z_j = s_y = \sum_{j=1}^n y_j$. If the sample size n is fixed, the minimal sufficient statistic is $S = Y_1 + \dots + Y_n$.