

MATH562 – Fall 2025

Problem Set: Week 4

1. Y_1, \dots, Y_n is a random sample from the Lomax distribution with unknown θ and known $\alpha > 0$,

$$\Pr(Y \leq y) = \begin{cases} 1 - \frac{\theta^\alpha}{(\theta + y)^\alpha}, & y > 0, \\ 0, & y \leq 0, \end{cases}$$

where $\alpha, \theta > 0$. Find the expected information for θ and hence compare the maximum likelihood estimator with the moments estimator found on Sheet 2. **Solution:**

For the expected information we note that the log likelihood

$$\ell(\theta) = \log f(y; \theta) = (\alpha + 1) \log(\theta + y) - \log \alpha - \alpha \log \theta, \quad \theta > 0,$$

has

$$\frac{\partial \ell(\theta)}{\partial \theta^2} = -\frac{\alpha + 1}{(\theta + y)^2} + \frac{\alpha}{\theta^2},$$

and as

$$E\{(\theta + Y)^{-2}\} = \int_0^\infty \frac{1}{(\theta + y)^2} \frac{\alpha \theta^\alpha}{(\theta + y)^{\alpha+1}} dy = \frac{\alpha}{(\alpha + 2)\theta^2} \int_0^\infty \frac{(\alpha + 2)\theta^{\alpha+2}}{(\theta + y)^{\alpha+2+1}} dy$$

and the final integral here equals unity, the expected information for a single observation is

$$i_1(\theta) = \frac{\alpha}{\theta^2} - (\alpha + 1) \frac{\alpha}{(\alpha + 2)\theta^2} = \frac{\alpha}{(\alpha + 2)\theta^2}.$$

Provided that differentiation with respect to θ and integration with respect to y commute, which is the case here, the Cramèr–Rao lower bound says that no unbiased estimator can have a lower asymptotic variance than $1/\{n i_1(\theta)\}$. This applies to the method-of-moments estimator $\tilde{\theta}$ because we saw previously that $\text{Var}(\tilde{\theta}) = \alpha \theta^2 / \{n(\alpha - 2)\}$, provided $\alpha > 2$, and therefore

$$\frac{\text{Var}(\tilde{\theta})}{1/\{n i_1(\theta)\}} = \frac{\alpha^2}{(\alpha + 2)(\alpha - 2)} = \frac{\alpha^2}{\alpha^2 - 4} > 1,$$

with the ratio exploding as $\alpha \downarrow 2$, i.e., $\tilde{\theta}$ is increasingly inefficient as $\alpha \downarrow 2$, and actually has infinite variance when $0 < \alpha < 2$.

2. Suppose that a random sample Y_1, \dots, Y_n from the exponential density is rounded down to the nearest δ , giving δZ_j , where $Z_j = \lfloor Y_j / \delta \rfloor$. Then the loss of information due to rounding is the ratio of the Fisher information based on the rounded data to that based on the original data.

- (a) Show that the likelihood contribution from a rounded observation can be written as $(1 - e^{-\lambda \delta})e^{-Z\lambda \delta}$, and deduce that the Fisher information for λ based on the rounded sample is

$$i(\delta) = n \delta^2 \exp(-\lambda \delta) \{1 - \exp(-\lambda \delta)\}^{-2}, \quad \delta, \lambda > 0.$$

Solution: The variable $Z = \lfloor Y / \delta \rfloor$ is discrete, taking values in $\{0, 1, \dots\}$, and $\Pr(Z = z)$ equals

$$\Pr(\lfloor Y / \delta \rfloor = z) = \Pr\{Y < (z + 1)\delta\} - \Pr\{Y \leq z\delta\} = 1 - e^{-(z+1)\lambda \delta} - (1 - e^{-z\lambda \delta}) = (1 - e^{-\lambda \delta}) e^{-z\lambda \delta}.$$

As the Z_j are independent, their likelihood can be expressed as

$$L(\lambda) = f(z_1, \dots, z_n; \lambda) = \prod_{j=1}^n (1 - e^{-\lambda\delta}) e^{-z_j\lambda\delta}, \quad \lambda > 0,$$

and the log likelihood is

$$\ell(\lambda) = n \log(1 - e^{-\lambda\delta}) - \lambda\delta \sum_{j=1}^n z_j.$$

On differentiating this twice with respect to λ we obtain the given result; note that the z_j drop out, so $E(Z_j)$ is not required.

- (b) Show that $\iota(\delta)$ has limit n/λ^2 as $\delta \rightarrow 0$, and deduce that if $\lambda = 1$ the loss of information when data are rounded down to the nearest integer rather than recorded exactly is less than 10%. **Solution:** The limit is obtained because $1 - e^{-\lambda\delta} = \lambda\delta - (\lambda\delta)^2/2 + o(\delta^3)$. It is easy to check that the expected information based on the Y_j is n/λ^2 , so the ratio of information quantities (aka the relative efficiency) is

$$\frac{\iota(\delta)}{\iota(0)} = \frac{(\lambda\delta)^2 \exp(-\lambda\delta)}{\{1 - \exp(-\lambda\delta)\}^2},$$

which equals $e^{-1}/(1 - e^{-1})^2 = 0.920$ when $\lambda\delta = 1$.

- (c) Find the loss of information when $\lambda\delta = 0.1$, and comment briefly. **Solution:** The relative efficiency is 99.92% in this case. Hence there is essentially no loss of information on λ if the data are rounded to the nearest 0.1, when the mean is 1. This is slightly misleading, because in fact we should use the likelihood based on the Z_j , whereas in practice we would usually use the Z_j in the likelihood based on the Y_j .
- (d) By considering the resulting average log likelihood as $n \rightarrow \infty$, show that replacing Y_j by δZ_j in the usual likelihood leads to slight overestimation of λ in large samples. **Solution:** If we replace Y with $Z\delta$ in the usual log likelihood, we obtain average log likelihood

$$\bar{\ell}(\lambda') = n^{-1} \left(n \log \lambda' - \lambda' \sum_{j=1}^n \delta Z_j \right), \quad \lambda' > 0,$$

and as $n \rightarrow \infty$ this converges to $\log \lambda' - \lambda'\delta E(Z)$, which implies that the maximum likelihood estimator converges to $1/\{\delta E(Z)\}$. Now $1 + Z$ has a geometric distribution with success probability $1 - e^{-\lambda\delta}$, so $E(Z) = (1 - e^{-\lambda\delta})^{-1} - 1 = (e^{\lambda\delta} - 1)^{-1}$, implying that the maximum likelihood estimator based on the Z s converges to $(e^{\lambda\delta} - 1)/\delta = \lambda + \delta\lambda^2/2 + \delta^2\lambda^3/6 + \dots$; i.e., λ will be overestimated by $O(\delta)$.

3. A sample of $n = 16$ Vaudois number plates has maximum 523308 and average 320869. Suppose that these were sampled uniformly and independently on the interval $(0, \theta)$, where θ is the highest number plate in the canton.

- (a) A random sample $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} U(0, \theta)$ is available. Find the mean and variance of \bar{Y} and hence suggest an unbiased estimator of θ . Show that $Q = \bar{Y}/\theta$ is a pivot and find its approximate distribution. How could you find its exact distribution?

Solution: The mean and variance of \bar{Y} are respectively $\theta/2$ and $\theta^2/(12n)$, so $2\bar{Y}$ is an unbiased estimator of θ . If $\stackrel{d}{=}$ denotes equality in distribution then we can write $Y_j \stackrel{d}{=} \theta U_j$, where $U_j \sim U(0, 1)$ are independent for each j , and therefore

$$Q' = \bar{Y}/\theta \stackrel{d}{=} \theta\bar{U}/\theta = \bar{U}$$

is a function of the data and of θ with a distribution that is (in principle) known. Thus Q is a pivot. Its distribution could be estimated to arbitrary accuracy by simulation of U_1, \dots, U_n for any fixed n or we can note that it is symmetric about $1/2$ with variance $1/(12n)$, so a central limit theorem gives

$$\sqrt{12n}(Q - 1/2) \sim \mathbb{N}(0, 1).$$

This approximation will be good enough for most practical purposes when $n \geq 12$: taking $Z = \sum_{j=1}^{12} U_j - 1/2 \sim \mathbb{N}(0, 1)$ is an old fudge to get standard normal variables from $U(0, 1)$ ones.

- (b) Use the calculation in (a) to obtain an approximate 95% confidence interval for θ . **Solution:** We have

$$\Pr(z_{\alpha/2} \leq \sqrt{12n}(Q - 1/2) \leq z_{1-\alpha/2}) = 1 - \alpha,$$

and replacing Q by \bar{Y}/θ and inverting the pivot yields an interval with limits

$$L = \frac{\bar{Y}}{\frac{1}{2} + z_{1-\alpha/2}/(12n)^{1/2}}, \quad U = \frac{\bar{Y}}{\frac{1}{2} + z_{\alpha/2}/(12n)^{1/2}}.$$

Computing this with $n = 16$ and $\bar{y} = 320869$ gives the interval $(500225.9, 894902.8) \doteq (500226, 894903)$.

- (c) Compare your interval from (b) with that computed using the maximum as suggested in the lectures. Which interval is better? Justify your answer. **Solution:** The interval computed using the maximum is $(524136.7, 659001.1) = (524137, 659001)$. This is clearly better than that in (b), because (i) it is much shorter but both are 95% intervals, (ii) it lies wholly to the right of the observed maximum and therefore does not include values we already know to be impossible, and (iii) it is exact for any n . (The last reason is less important, because as mentioned above, the normal approximation to the distribution of \bar{Y} is excellent for $n = 16$.)

4. Consider a random sample of bivariate normal pairs (Y_j, X_j) ($j = 1, \dots, n$) with unknown mean vector (ψ, λ) , and known values of $\text{Var}(Y_j) = \sigma_1^2$, $\text{Var}(X_j) = \sigma_2^2$ and $\text{Corr}(X_j, Y_j) = \rho$. Suppose that independent observations X_{n+1}, \dots, X_{n+m} *i.i.d.* $\mathbb{N}(\lambda, \sigma_2^2)$ are also available. Such data might arise when measurements Y that are expensive or difficult to obtain are correlated with others, X , that are much cheaper or easier to obtain, so that it is only affordable to make n joint measurements but $m \gg n$ auxiliary measurements on X alone can also be provided. In the following, we investigate under what circumstances the auxiliary measurements aid in estimating ψ , and by how much.

- (a) Write down the log likelihood function for ψ and λ based on the data, and show that the maximum likelihood estimator is given by $\hat{\psi} = \bar{Y}_n + \rho\sigma_1(\bar{X}_{n+m} - \bar{X}_n)/\sigma_2$, where $n\bar{Y}_n = \sum_{j=1}^n Y_j$, $n\bar{X}_n = \sum_{j=1}^n X_j$ and $(n+m)\bar{X}_{n+m} = \sum_{j=1}^{n+m} X_j$. **Solution:** The negative log likelihood function is

$$\begin{aligned} -\ell(\psi, \lambda) &\equiv \frac{1}{2} \sum_{j=1}^n \begin{pmatrix} y_j - \psi & x_j - \lambda \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} y_j - \psi \\ x_j - \lambda \end{pmatrix} + \frac{1}{2\sigma_2^2} \sum_{j=n+1}^{n+m} (x_j - \lambda)^2, \\ &= \frac{1}{1-\rho^2} \sum_{j=1}^n \left\{ \frac{(y_j - \psi)^2}{2\sigma_1^2} - \frac{\rho(y_j - \psi)(x_j - \lambda)}{\sigma_1\sigma_2} + \frac{(x_j - \lambda)^2}{2\sigma_2^2} \right\} + \sum_{j=n+1}^{n+m} \frac{(x_j - \lambda)^2}{2\sigma_2^2}, \quad \psi, \lambda \in \mathbb{R}, \end{aligned}$$

and its first derivatives are

$$\begin{aligned} -\ell_\psi(\psi, \lambda) &= \frac{1}{1-\rho^2} \sum_{j=1}^n \left\{ \frac{\psi - y_j}{\sigma_1^2} + \frac{\rho(x_j - \lambda)}{\sigma_1\sigma_2} \right\} \propto \psi - \bar{y}_n + \sigma_1\rho(\bar{x}_n - \lambda)/\sigma_2, \\ -\ell_\lambda(\psi, \lambda) &= \frac{1}{1-\rho^2} \sum_{j=1}^n \left\{ \frac{\rho(y_j - \psi)}{\sigma_1\sigma_2} + \frac{\lambda - x_j}{\sigma_2^2} \right\} + \sum_{j=n+1}^{n+m} \frac{\lambda - x_j}{\sigma_2^2}. \end{aligned}$$

The first of these implies that if $\ell_\psi(\hat{\psi}, \hat{\lambda}) = 0$ then $\hat{\psi} = \bar{y}_n - \sigma_1 \rho (\bar{x}_n - \hat{\lambda}) / \sigma_2$. If we substitute this expression into the second equation, setting $\ell_\lambda(\hat{\psi}, \hat{\lambda}) = 0$, and simplify, we obtain $\hat{\lambda} = (n + m)^{-1} \sum_{j=1}^{n+m} x_j$ after a little struggle, which gives the desired formula for $\hat{\psi}$ on replacing the observations with the corresponding random variables.

- (b) Check that both \bar{Y}_n and $\hat{\psi}$ are unbiased estimators of ψ , find $\text{Var}(\bar{Y}_n)$, show that $\hat{\psi}$ has asymptotic variance $\sigma_1^2 \{1 - m\rho^2 / (n + m)\} / n$, and hence give the gain in relative efficiency due to the auxiliary data. Discuss how and why this changes when (i) $\rho = 0$, (ii) $m \rightarrow \infty$, and (iii) $\rho \rightarrow \pm 1$.

Solution: Linearity of the expectation operator implies that $E(\bar{Y}_n) = E(Y) = \psi$, $E(\bar{X}_n) = E(\bar{X}_{n+m}) = E(X) = \lambda$ and thus $E(\hat{\psi}) = E(\bar{Y}_n) + \rho\sigma_1 E(\bar{X}_{n+m} - \bar{X}_n) / \sigma_2 = \psi + \rho\sigma_1(\lambda - \lambda) / \sigma_2 = \psi$. Clearly $\text{Var}(\bar{Y}_n) = \sigma_1^2 / n$.

The Fisher information matrix is

$$i(\psi, \lambda) = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{n}{\sigma_1^2} & -\frac{n\rho}{\sigma_1\sigma_2} \\ -\frac{n\rho}{\sigma_1\sigma_2} & \frac{n}{\sigma_2^2} + \frac{m(1-\rho^2)}{\sigma_2^2} \end{pmatrix},$$

and the (ψ, ψ) corner of $i(\psi, \lambda)^{-1}$ is $\sigma_1^2 \{1 - m\rho^2 / (n + m)\} / n$, which is the asymptotic variance of $\hat{\psi}$; in fact here this variance is exact, as a direct computation of $\text{Var}(\hat{\psi})$ shows. Hence the relative efficiency is

$$\frac{\text{Var}(\bar{Y}_n)}{\text{Var}(\hat{\psi})} = \frac{1}{1 - m\rho^2 / (n + m)}.$$

(i) When $\rho = 0$, the relative efficiency is one, because X is then independent of Y and the auxiliary data give no information about ψ .

(ii) When $m \rightarrow \infty$, λ is known exactly from \bar{X}_{n+m} , so $\hat{\psi} = \bar{Y}_n + \rho\sigma_1(\lambda - \bar{X}_n) / \sigma_2$ and the relative efficiency becomes $1 / (1 - \rho^2)$.

(iii) When $\rho \rightarrow \pm 1$ the auxiliary observations X_{n+1}, \dots, X_{n+m} are perfectly informative about ψ , so the sample size is effectively $n + m$, and the relative efficiency is $(n + m) / n$.