

# MATH562 – Fall 2025

## Problem Set: Week 2

1. Investigate exponential tilting of a baseline density function  $f_0(y)$  that is uniform on the interval  $\mathcal{Y} = (0, 1)$ , when the tilting functions are (a)  $s(y) = y$  and (b)  $s(y)^\top = (\log y, \log(1 - y))$ .

**Solution:** (a) In this case

$$k(\varphi) = \log \int_0^1 e^{\varphi y} dy = \log \left( \frac{e^\varphi - 1}{\varphi} \right), \quad \varphi \in \mathbb{N} = \mathbb{R},$$

with  $k(0) = 0$  defined by continuity as  $\varphi \rightarrow 0$ . Hence the resulting exponential family is

$$f(y; \varphi) = e^{\varphi y - k(\varphi)} = \frac{\varphi e^{\varphi y}}{e^\varphi - 1}, \quad y \in \mathcal{Y}, \varphi \in \mathbb{R}.$$

(b) In this case

$$k(\varphi) = \log \int_0^1 e^{\varphi_1 \log y + \varphi_2 \log(1-y)} dy = \log \int_0^1 y^{\varphi_1} (1-y)^{\varphi_2} dy$$

and we recognise this as a beta integral, defined for  $\varphi_1, \varphi_2 > -1$ , and then equal to  $\Gamma(1 + \varphi_1)\Gamma(1 + \varphi_2)/\Gamma(2 + \varphi_1 + \varphi_2)$ . The resulting density is therefore

$$f(y; \varphi) = \frac{\Gamma(2 + \varphi_1 + \varphi_2)}{\Gamma(1 + \varphi_1)\Gamma(1 + \varphi_2)} y^{\varphi_1} (1-y)^{\varphi_2}, \quad y \in \mathcal{Y}, \varphi = (\varphi_1, \varphi_2) \in (-1, \infty)^2,$$

which is just a rewriting of the beta density, for which  $\alpha = \varphi_1 + 1$  and  $\beta = \varphi_2 + 1$ , and  $\alpha, \beta > 0$ , in the usual notation.

- \*2. For  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2)$ , derive the limiting distribution of  $Y = 1/\bar{X}$  as  $n \rightarrow \infty$ . Why can the event  $\bar{X} = 0$  be neglected and what does the result tell us in practice?

*Hint:* In the derivation, the cases  $\mu \neq 0$  and  $\mu = 0$  have to be treated differently.

3. In a simple model for the spread of an epidemic in a large closed population of  $n$  identical individuals, it can be shown that the expression  $1 - \tau = e^{-R_0 \tau}$  relates the basic reproduction number  $R_0$  (the number of susceptible persons infected by a single infective person at the start of the epidemic) to the ultimate fraction  $\tau$  of the population who are infected.

- (a) Show that if  $R_0 \leq 1$  then there is only one possible value for  $\tau$ , but that if  $R_0 > 1$  then there are two possible values, and explain this heuristically.

**Solution:** Clearly  $\tau = 0$  is one possible solution for any  $R_0 > 0$ . The slopes of  $1 - \tau$  and  $e^{-R_0 \tau}$  at  $\tau = 0$  are respectively  $-1$  and  $-R_0$ , and it is clear from a plot of these two functions against  $\tau$  that there is a second, positive, solution to the equation within the interval  $\tau \in (0, 1)$  if  $R_0 > 1$ . The intuition is that if  $R_0 \leq 1$ , then the epidemic is certain to be of negligible size, but that if  $R_0 > 1$  then it may be of negligible size (if by chance it dies out immediately) but otherwise will affect a positive fraction of the population.

- (b) If it is positive, the final proportion infected can be estimated by  $\hat{\tau}$ , which has an approximate normal distribution with mean  $\tau$  and variance  $\sigma^2/n$ , where

$$\sigma^2 = \frac{\tau(1-\tau)}{\{1 - (1-\tau)R_0\}^2} \{1 + c^2(1-\tau)R_0^2\},$$

where  $c = \text{Var}(T)^{1/2}/\text{E}(T)$  is the coefficient of variation of the infectious period for an individual,  $T$ . Hence obtain an estimator of  $R_0$ , and show that this is approximately normal with mean  $R_0$  and variance  $\sigma_R^2/n$ , where  $\sigma_R^2 = \{1 + c^2(1 - \tau)R_0^2\}/\{\tau(1 - \tau)\}$ .

**Solution:** The equation  $1 - \tau = e^{-R_0\tau}$  gives  $R_0 = -\tau^{-1}\log(1 - \tau) = g(\tau)$ , say, which is a smooth function of  $\tau$  for  $\tau \in (0, 1)$ . Hence the delta method applies, and  $\hat{R}_0 = g(\hat{\tau})$  is asymptotically normal with mean  $g(\tau) = R_0$  and with variance given using

$$g'(\tau) = \log(1 - \tau)/\tau^2 + 1/\{\tau(1 - \tau)\} = -R_0/\tau + 1/\{\tau(1 - \tau)\} = \{1 - R_0(1 - \tau)\}/\{\tau(1 - \tau)\};$$

the delta method variance formula  $\sigma_R^2 = g'(\tau)^2\sigma^2$  gives the required result.

- (c) Find  $c^2$  when  $T$  is constant, uniform on some interval, and exponential. Hence suggest why an upper bound for the variance might be obtained by setting  $c = 1$ .

**Solution:** The variance of a constant is zero, so if the constant itself is non-zero,  $c^2 = 0/T = 0$ . If  $T \sim U(0, d)$  then  $\text{E}(T) = d/2$  and  $\text{Var}(T) = d^2/12$ , leading to  $c^2 = 1/3$ . An exponential variable with mean  $\mu$  has variance  $\mu^2$ , giving  $c^2 = 1$ . The exponential distribution has high coefficient of variation (for example, the gamma distribution has  $c = \alpha^{-1/2}$ , where  $\alpha$  is the shape parameter, and taking  $\alpha > 1$  gives a unimodal distribution that seems potentially suitable for  $T$ ), so taking  $c = 1$  should give an upper bound for  $\sigma^2$  and hence for  $\sigma_R^2$ .

4. (a) The random variable  $Y$  follows a Lomax distribution, i.e.,

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

where  $\alpha, \theta > 0$  are unknown. Is this an exponential family distribution?

**Solution:** The density is

$$f(y) = \frac{\alpha\theta^\alpha}{(\theta + y)^{\alpha+1}} = \exp\{-(\alpha + 1)\log(\theta + y) + \log\alpha + \alpha\log\theta\}, \quad y > 0, \quad \alpha, \theta > 0.$$

This is not of exponential family form, because no term in the log density can be written as  $s(y)^\top\varphi(\theta)$  for some function  $s(y)$  that depends only on  $y$ .

From here on  $Y_1, \dots, Y_n$  represent independent identically distributed variables from the Lomax distribution with unknown  $\theta$  and known  $\alpha > 2$ .

- (b) Given that

$$\text{E}(Y) = \frac{\theta}{\alpha - 1}, \quad \text{Var}(Y) = \frac{\alpha\theta^2}{(\alpha - 1)^2(\alpha - 2)},$$

use  $Y_1, \dots, Y_n$  to obtain a method-of-moments estimator  $\tilde{\theta}$  of  $\theta$ , and compute  $\text{Var}(\tilde{\theta})$ . Is  $\tilde{\theta}$  biased? **Solution:**

The method-of-moments estimator satisfies  $\bar{Y} = \tilde{\theta}/(\alpha - 1)$ , i.e.,  $\tilde{\theta} = (\alpha - 1)\bar{Y}$ , and this is easily checked to be unbiased using the given formula for  $\text{E}(Y)$  and the fact that  $\text{E}(\bar{Y}) = \text{E}(Y)$ .

Moreover  $\text{Var}(\bar{Y}) = \text{Var}(Y)/n$ , so

$$\text{Var}(\tilde{\theta}) = (\alpha - 1)^2\text{Var}(\bar{Y}) = \frac{(\alpha - 1)^2}{n}\text{Var}(Y) = \frac{\theta^2\alpha}{n(\alpha - 2)}.$$

- (c) Another estimator of  $\theta$  is  $\tilde{\theta}_c = c\bar{Y}$ , where  $\bar{Y} = n^{-1}(Y_1 + \dots + Y_n)$  and  $c > 0$ . Compute the bias and variance of  $\tilde{\theta}_c$ . What value of  $c$  minimises the mean square error of  $\tilde{\theta}_c$ ?

**Solution:**

The bias of  $\tilde{\theta}_c$  is

$$b(\tilde{\theta}_c; \theta) = E(\tilde{\theta}_c) - \theta = c \frac{\theta}{\alpha - 1} - \theta = \frac{(\alpha - 1 - c)\theta}{\alpha - 1},$$

and  $\text{Var}(\tilde{\theta}_c) = c^2 \text{Var}(\bar{Y})$ , so (after a little algebra) the mean squared error of  $\tilde{\theta}_c$  is

$$b(\tilde{\theta}_c; \theta)^2 + c^2 \text{Var}(\bar{Y}) = \frac{\theta^2}{n(\alpha - 1)^2(\alpha - 2)} \{n(\alpha - 2)(\alpha - 1 - c)^2 + c^2\alpha\}.$$

This is minimised by differentiating the expression  $\{\cdot\}$  here with respect to  $c$ , giving

$$\frac{d\{\cdot\}}{dc} = -2n(\alpha - 2)(\alpha - 1 - c) + 2c\alpha = 0,$$

which results in

$$c = \frac{n(\alpha - 1)(\alpha - 2)}{n(\alpha - 2) + \alpha} \rightarrow \alpha - 1, \quad n \rightarrow \infty.$$

Note that for any finite  $n$  the estimator is biased, but it is asymptotically unbiased, and that the second derivative  $d^2\{\cdot\}/dc^2 = 2n(\alpha - 1)(\alpha - 2) + 2\alpha$  is positive — which is obvious because the MSE is a sum of positive quadratic expressions in  $c$ .