

## Statistics: Example Sheet 1 (of 3)

Comments and corrections to [s.pitts@statslab.cam.ac.uk](mailto:s.pitts@statslab.cam.ac.uk)

1. Ask your supervisor to test you on the sheet of common distributions handed out in lectures.
2. **(Probability review)** If  $X \sim \text{Exponential}(\lambda)$  and  $Y \sim \text{Exponential}(\mu)$  are independent, derive the distribution of  $\min(X, Y)$ . If  $X \sim \text{Gamma}(\alpha, \lambda)$  and  $Y \sim \text{Gamma}(\beta, \lambda)$  are independent, derive the distributions of  $X + Y$  and  $X/(X + Y)$ .
3. Suppose that  $(X, T)$  have a joint probability density function  $f_{X,T}(x, t; \theta)$ . Prove the factorisation criterion for the sufficiency of  $T$ .
4. In a genetics experiment, a sample of  $n$  individuals was found to include  $a, b, c$  of the three possible genotypes  $GG, Gg, gg$  respectively. The population frequency of a gene of type  $G$  is  $\theta/(\theta + 1)$ , where  $\theta$  is unknown, and it is assumed that the individuals are unrelated and that two genes in a single individual are independent. Show that the likelihood of  $\theta$  is proportional to  $\theta^{2a+b}/(1 + \theta)^{2a+2b+2c}$  and that the maximum likelihood estimate of  $\theta$  is  $(2a + b)/(b + 2c)$ .
5. (a) Let  $X_1, \dots, X_n$  be independent Poisson random variables, with  $X_i$  having mean  $i\theta$ , for some  $\theta > 0$ . Show that  $T = T(\mathbf{X}) = \sum_{i=1}^n X_i$  is a sufficient statistic for  $\theta$  and write down the distribution of  $T$ . Show that the maximum likelihood estimator  $\hat{\theta}$  of  $\theta$  is a function of  $T$ , and show that it is unbiased.  
(b) For some  $n > 2$ , let  $X_1, \dots, X_n$  be iid with  $X_i \sim \text{Exponential}(\theta)$ . Find a minimal sufficient statistic  $T$  and write down its distribution. Show that the maximum likelihood estimator  $\hat{\theta}$  of  $\theta$  is a function of  $T$ , and show that it is biased, but asymptotically unbiased. Find an injective function  $h$  on  $(0, \infty)$  such that, writing  $\psi = h(\theta)$ , the maximum likelihood estimator  $\hat{\psi}$  of the new parameter  $\psi$  is unbiased.
6. Suppose  $X_1, \dots, X_n$  are independent random variables with distribution  $\text{Bin}(1, p)$ .
  - (a) Show that a sufficient statistic for  $\theta = (1 - p)^2$  is  $T(\mathbf{X}) = \sum_{i=1}^n X_i$  and that the MLE for  $\theta$  is  $(1 - \frac{1}{n}T)^2$ .  
*Hint: use the chain rule,  $df/d\theta = (df/dp)(dp/d\theta)$ .*
  - (b) Show that the MLE is a biased estimator for  $\theta$ . Let  $\tilde{\theta} = 1(X_1 + X_2 = 0)$ . Show that  $\tilde{\theta}$  is unbiased for  $\theta$ . Use the Rao–Blackwell theorem to find a function of  $T$  which is an unbiased estimator for  $\theta$ .

7. For some  $n \geq 2$ , suppose that  $X_1, \dots, X_n$  are iid random variables uniformly distributed on  $[\theta, 2\theta]$  for some  $\theta > 0$ . Show that  $\tilde{\theta} = \frac{2}{3}X_1$  is an unbiased estimator of  $\theta$ . Show that  $T(\mathbf{X}) = (\min_i X_i, \max_i X_i)$  is a minimal sufficient statistic for  $\theta$ . Use the Rao–Blackwell theorem to find an unbiased estimator  $\hat{\theta}$  of  $\theta$  which is a function of  $T$  and whose variance is strictly smaller than the variance of  $\tilde{\theta}$  for all  $\theta > 0$ .
8. (a) Let  $X_1, \dots, X_n$  be iid with  $X_i \sim U[0, \theta]$ . Find the maximum likelihood estimator  $\hat{\theta}$  of  $\theta$ . Show that the distribution of  $R(\mathbf{X}, \theta) = \hat{\theta}/\theta$  does not depend on  $\theta$ , and use  $R(\mathbf{X}, \theta)$  to find a  $100(1 - \alpha)\%$  confidence interval for  $\theta$  for  $0 < \alpha < 1$ .
- (b) The lengths (in minutes) of calls to a call centre may be modelled as iid exponentially distributed random variables, and  $n$  such call lengths are observed. The original sample is lost, but the data manager has noted down  $n$  and  $t$  where  $t$  is the total length of the  $n$  calls in minutes. Derive a 95% confidence interval for the probability that a call is longer than 2 minutes if  $n = 50$  and  $t = 105.3$ .
9. Suppose that  $X_1 \sim N(\theta_1, 1)$  and  $X_2 \sim N(\theta_2, 1)$  independently, where  $\theta_1$  and  $\theta_2$  are unknown. Show that  $(\theta_1 - X_1)^2 + (\theta_2 - X_2)^2$  has a  $\chi_2^2$  distribution and that this is the same as  $\text{Exponential}(\frac{1}{2})$ , i.e., the exponential distribution with mean 2.

Show that both the square  $S$  and circle  $C$  in  $\mathbb{R}^2$ , given by

$$S = \{(\theta_1, \theta_2) : |\theta_1 - X_1| \leq 2.236; |\theta_2 - X_2| \leq 2.236\}$$

$$C = \{(\theta_1, \theta_2) : (\theta_1 - X_1)^2 + (\theta_2 - X_2)^2 \leq 5.991\}$$

are 95% confidence regions for  $(\theta_1, \theta_2)$ .

*Hint:*  $\Phi(2.236) = (1 + \sqrt{.95})/2$ , where  $\Phi$  is the distribution function of  $N(0, 1)$ . What might be a sensible criterion for choosing between  $S$  and  $C$ ?

10. Suppose that the number of defects on a roll of magnetic recording tape is modelled with a Poisson distribution for which the mean  $\lambda$  is known to be either 1 or 1.5. Suppose the prior mass function for  $\lambda$  is

$$\pi_\lambda(1) = 0.4, \quad \pi_\lambda(1.5) = 0.6 .$$

A random sample of five rolls of tape has  $\mathbf{x} = (3, 1, 4, 6, 2)$  defects respectively. Show that the posterior distribution for  $\lambda$  given  $\mathbf{x}$  is

$$\pi_{\lambda|\mathbf{X}}(1 | \mathbf{x}) = 0.012, \quad \pi_{\lambda|\mathbf{X}}(1.5 | \mathbf{x}) = 0.988 .$$

11. Suppose  $X_1, \dots, X_n$  are iid with (conditional) probability density function  $f(x | \theta) = \theta x^{\theta-1}$  for  $0 < x < 1$  (and is zero otherwise), for some  $\theta > 0$ . Suppose that the prior for  $\theta$  is  $\text{Gamma}(\alpha, \beta)$ ,  $\alpha > 0, \beta > 0$ . Find the posterior distribution of  $\theta$  given  $\mathbf{X} = (X_1, \dots, X_n)$  and the Bayesian estimator of  $\theta$  under quadratic loss.
12. For some  $n \geq 3$ , let  $\epsilon_1, \dots, \epsilon_n$  be iid with  $\epsilon_i \sim N(0, 1)$ . Set  $X_1 = \epsilon_1$  and  $X_i = \theta X_{i-1} + (1 - \theta^2)^{1/2} \epsilon_i$  for  $i = 2, \dots, n$  and some  $\theta \in (-1, 1)$ . Find a sufficient statistic for  $\theta$  that takes values in a subset of  $\mathbb{R}^3$ .