COMP245: Probability and Statistics 2016 - Problem Sheet 8 Solutions

Estimation

S1) The pdf for each sample X_i is given by $f(x_i) = \lambda e^{-\lambda x_i}$, and hence the log-likelihood function is

$$\begin{aligned}\ell(\lambda) &= \sum_{i=1}^{n} \log\{f(x_i)\} \\ &= \sum_{i=1}^{n} \{\log(\lambda) - \lambda x_i\} = n\log(\lambda) - \lambda \sum_{i=1}^{n} x_i.\end{aligned}$$

To find the MLE for λ, we calculate the derivative of $\ell(\lambda)$ wrt λ,

$$\frac{d}{d\lambda}\ell(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i,$$

which is zero when $\lambda = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$. To check this is a maximum, we examine the second derivative of ℓ,

$$\frac{d^2}{d\lambda^2}\ell(\lambda) = -\frac{n}{\lambda^2}$$

which is in fact negative for any positive parameter λ, so $\hat{\lambda} = \frac{1}{\bar{x}}$ is the MLE.

S2) Let $(x_1, \dots, x_n)$ be the random sample from Poisson(λ). Then

$$\begin{aligned}L(\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i} \exp -\lambda}{x_i!} \\ \implies \ell(\lambda) &= \log(\lambda)\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n} \log(x_i!) \\ \frac{d\ell(\lambda)}{d\lambda} &= \frac{\sum_{i=1}^{n} x_i}{\lambda} - n \implies \hat{\lambda} = \frac{\sum_{i=1}^{n} x_i}{n}.\end{aligned}$$

Since the second derivative

$$\frac{d^2\ell(\lambda)}{d\lambda^2} = -\frac{\sum_{i=1}^{n} x_i}{\lambda^2}$$

is negative (provided not all the samples are zero) for all λ, then clearly the MLE for λ is given by the sample mean. (The case when all of the samples are zero is also easy to check — simply note that the first derivative is then negative everywhere and hence the likelihood is a decreasing function of λ.)

S3) (a) Let X be the number of individuals in a car. Then for $x \in \{1,2,3,4,5\}$ we have a contribution to the likelihood given by the pmf $p_x = p(x) = p(1-p)^{x-1}$; on the other hand, for the data for which we only know $x \geq 6$, these are observed with probability $p_6 = \mathrm{P}(X \geq 6) = (1-p)^{6-1}$. Combining all this together, we get a likelihood function for all the data

$$\begin{aligned} L(p) &= \prod_{i=1}^{6} p_i^{n_i} = p(1)^{n_1} p(2)^{n_2} p(3)^{n_3} p(4)^{n_4} p(5)^{n_5} \mathrm{P}(X \geq 6)^{n_6} \\ &= \prod_{i=1}^{5} \{p(1-p)^{i-1}\}^{n_i} (1-p)^{5n_6} \\ &= p^{n_1+n_2+n_3+n_4+n_5} (1-p)^{n_2+2n_3+3n_4+4n_5+5n_6} \end{aligned}$$

$$\Rightarrow \ell(p) = (n_1+n_2+n_3+n_4+n_5)\log(p) + (n_2+2n_3+3n_4+4n_5+5n_6)\log(1-p),$$

where $n_1, \ldots, n_5$ are the number of times we observed $1, \ldots, 5$ people in a car, n_6 is the number of times we observed at least 6 people in a car.

To find the maximum, we differentiate $\ell(p)$ wrt p and set equal to zero,

$$\begin{aligned} 0 = \frac{d}{dp}\ell(p) &= \frac{n_1+n_2+n_3+n_4+n_5}{p} - \frac{n_2+2n_3+3n_4+4n_5+5n_6}{1-p} \\ &\iff (n_1+n_2+n_3+n_4+n_5)(1-p) = (n_2+2n_3+3n_4+4n_5+5n_6)p \\ &\iff (n_1+n_2+n_3+n_4+n_5)(1-p) = (n_2+2n_3+3n_4+4n_5+5n_6)p \\ &\iff p = \frac{n_1+n_2+n_3+n_4+n_5}{n_1+2n_2+3n_3+4n_4+5n_5+5n_6}. \end{aligned}$$

Substituting in the values of the $\{n_i\}$ from our table, we get

$$\hat{p}_{MLE} = \frac{1464}{902 + 2\times 403 + 3 \times 106 + 4 \times 36 + 5 \times 16 + 5} = \frac{1464}{2278} = 0.643.$$

and this is a maximum since the second derivative of $\ell(p)$,

$$\frac{d^2}{dp^2}\ell(p) = -\frac{n_1+n_2+n_3+n_4+n_5}{p^2} - \frac{n_2+2n_3+3n_4+4n_5+5n_6}{(1-p)^2},$$

is negative everywhere.

(b) Substitute $p = \hat{p}_{MLE} = 0.643$ into the bin probability functions $p_i, i \in \{1, \ldots, 6\}$, from which the expected numbers are given as $E_i = 1469 \times p_i$. Using $O_i = n_i$, compute the χ^2 test statistic, and compare it with the 1% level of $\chi^2(6-1-1)$.

S4) (a) i. 95%. ii. 90%. iii. 99%. iv. 68%.

(b) 10.9375 ± 2.0094. We can be 95% confident that μ lies in this interval.

S5) The sample size $n = 100$ is quite large, so we can use the Normal distribution as an approximation to the t, so in both cases the confidence limits are $\bar{x} \pm 1.96\frac{s_{n-1}}{\sqrt{n}}$.

(a) $\bar{x} = \frac{250}{100} = 2.5$, $s_{n-1} = \sqrt{\frac{\sum_i x_i^2 - n\bar{x}^2}{n-1}} = \sqrt{\frac{725000 - 100 \times 2.5^2}{99}} = 85.539$. So the confidence limits are 2.5 ± 16.7656.

(b) 83.2 ± 1.254.

S6) In both cases we use the formula $\bar{x} \pm t_{n-1,\alpha}\frac{s_{n-1}}{\sqrt{n}}$ with n=8.

(a) $\alpha = 0.95$, C.I.: [3.83, 6.77].

(b) $\alpha = 0.995$, C.I.: [2.59, 8.01].