Quiz Week 5 - Statistical Inference - Questions

FIT5197 teaching team

Question 1

For sample $\vec{x}$ of size n distributed as $N(\mu, \sigma)$ the sum of squared errors (SSE) of mean estimate μ is given by

$$\mathrm{SSE}(\mu) = \sum_{i=1}^{n} (x_i - \mu)^2$$

Demonstrate using differentiation that the SSE point estimate $\hat{\mu}$, corresponding to the value of μ that minimises the SSE, is equivalent to the sample mean

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Question 2

An alternative measure of goodness-of-fit to estimate the mean μ and standard deviation σ of a normally distributed random variable is maximum likelihood estimation. For sample $\mathbf{y} = (y_1, \ldots, y_n)$ drawn from the normally distributed random variables, $Y_i \sim N(\mu, \sigma)$, the likelihood is given by

$$p(\mathbf{y}|\mu, \sigma) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mu - y_i)^2 \right)$$

(a) Why does this expression involve the product of functions of y_i? Why is the likelihood a measure of goodness-of-fit to estimate the mean μ and standard deviation σ?

(b) Using the fact that $e^{-a} e^{-b} = e^{-a-b}$ for arbitrary variables a and b, Show the negative log-likelihood has the following form:

$$L_{-}(\mathbf{y}|\mu, \sigma) = -\log p(\mathbf{y}|\mu, \sigma) = \frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \quad (1)$$

(c) Is maximising the likelihood the same as minimising the negative log-likelihood? Why? What is the difference between maximising the log-likelihood and minimising the negative log-likelihood? Why bother minimise the negative log-likelihood, why not just use likelihood?

(d) Show the MLE estimate of the mean μ is

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

(e) Show the MLE estimate of the standard deviation σ is

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\mu})^2}$$

(f) Show that the bias of the MLE estimate of the mean, $\hat{\mu}$, given in question 2(d) above is equal to zero. Note that bias is given by $\boldsymbol{B}_\mu = E[\hat{\mu}] - \mu$

(g) Show that the variance of the MLE estimate of the mean, $\hat{\mu}$, given in question 2(d) above is equal to $\frac{\sigma^2}{n}$. Note the variance is given by $V[\hat{\mu}] = E[(\hat{\mu} - E[\hat{\mu}])^2]$.

Question 3

For a Poisson random variable X the probability mass function is given by:

$$P(X = x) = \begin{cases} exp(-\lambda)\frac{1}{x!}\lambda^x, & if\ x \in \mathbb{Z}_+ \\ 0, & if\ x \notin \mathbb{Z}_+ \end{cases}$$

where the parameter λ is the mean of the distribution. For sample $\mathbf{x} = (x_1, \dots, x_n)$ drawn from the Poisson distributed random variables, $X_i \sim Pois(\lambda)$,

(a) Derive the likelihood function, $p(\mathbf{x}|\lambda)$.

(b) Derive the negative log-likelihood function, $L_{-}(\mathbf{y}|\lambda)$.

(c) Derive the MLE estimate of the mean, $\hat{\lambda}$.

(d) Derive the bias of the MLE estimate of the mean, $\hat{\lambda}$.

(e) Derive the variance of the MLE estimate of the mean, $\hat{\lambda}$.

(f) Is the MLE estimate of the mean, $\hat{\lambda}$, consistent?

Question 4: R code hackers mega-mini challenge

For sample $\mathbf{x} = (x_1, \ldots, x_n)$ drawn from the Bernoulli distributed random variables, $X_i \sim Ber(\theta)$, it can be shown that the log-likelihood function obeys:

$$L_+(\mathbf{x}|\theta) = ylog(\theta) + (n - y)log(1 - \theta)$$

where $y = \sum_{i=1}^{n} x_i$. (For kicks you might derive this from the Bernoulli PMF.)

Using R, create a function for the log-likelihood that takes as input a Bernoulli random sample, $\mathbf{x}$, and the probability of success parameter θ. Now in R generate a Bernoulli random sample with $n = 20$ samples and $\theta = 0.5$. Then using the R built-in function called 'optimize' and your function for the log-likelihood compute the MLE estimate of the probability of success, $\hat{\theta}$. Explain your reasoning behind the inputs you entered into the 'optimize' function. See what happens to the sum of squares error of your estimates if you increase the number of samples n or vary the true value for θ. (Note this might require a bit of self study to get 'optimize' working, but don't fret, we will provide a solution).