University of Technology Sydney
School of Mathematical and Physical Sciences

Mathematical Statistics (37262) –
Quiz 2
SOLUTIONS

Question 1.

i) We know that, if $W_1, W_2, ..., W_r$ are independent $Geo(p)$ variables, then $X = \sum_{i=1}^{r} W_i \sim NegBin(r,p)$. Since we know that each $E(W_i) = \frac{1}{p}$ and $E(X) = \sum_{i=1}^{r} E(W_i)$ we have that $E(X) = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$.

ii) The first sample moment is $s_1 = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Matching this to equal $E(X) = \frac{r}{p}$ gives $\frac{1}{n}\sum_{i=1}^{n} x_i = \frac{r}{p}$ and hence the method of moments estimates p by $\hat{p}_{MM} = \frac{nr}{\sum_{i=1}^{n} x_i}$.

iii)

$$L(\boldsymbol{X} \mid p) = \prod_{i=1}^{n} \frac{(x_i - 1)!}{(r-1)!(x_i - r)!} p^r (1-p)^{x_i - r} = p^{nr}(1-p)^{\sum_{i=1}^{n} x_i - nr} \prod_{i=1}^{n} \frac{(k-1)!}{(r-1)!(k-r)!}$$

.

iv) $L(\boldsymbol{X} \mid p) = p^{nr}(1-p)^{\sum_{i=1}^{n} x_i - nr} \prod_{i=1}^{n} \frac{(k-1)!}{(r-1)!(k-r)!}$ hence

$$\ell(\boldsymbol{X} \mid p) = nr\ln(p) + \left(\sum_{i=1}^{n} x_i - nr\right)\ln(1-p) + \ln\left(\prod_{i=1}^{n} \frac{(k-1)!}{(r-1)!(k-r)!}\right).$$

This gives $\frac{\partial}{\partial p}\ell(\boldsymbol{X} \mid p) = \frac{nr}{p} - \frac{\sum_{i=1}^{n} x_i - nr}{1-p}$.

Solving to find when $\frac{\partial}{\partial p}\ell(\boldsymbol{X} \mid p) = 0$ gives $\hat{p}_{MLE} = \frac{nr}{\sum_{i=1}^{n} x_i}$.

The maximum likelihood estimator for p is the same as the estimator derived by the method of moments.

(10 Marks)

Question 2.

i) $y = \dfrac{e^x}{1+e^x}$ hence $x = \ln\left(\dfrac{y}{1-y}\right)$ and $\dfrac{dy}{dx} = y(1-y)$.

$$\int_{-\infty}^{\infty} e^{-x^2}\cos(x)dx \approx \frac{1}{5}\left[\frac{e^{-\left(\ln\left(\frac{u_1}{1-u_1}\right)\right)^2}\cos(\ln\left(\frac{u_1}{1-u_1}\right))}{u_1(1-u_1)} + \ldots + \frac{e^{-\left(\ln\left(\frac{u_5}{1-u_5}\right)\right)^2}\cos(\ln\left(\frac{u_5}{1-u_5}\right))}{u_5(1-u_5)}\right]$$

$$\approx \frac{1}{5}\left(1.71694 + 3.75392 + 0.24439 + 3.40959 + 0.55092\right) \approx 1.935$$

ii) $e^{-x^2}\cos(x)$ is an even function hence $\int_{-\infty}^{\infty} e^{-x^2}\cos(x)dx = 2\int_{-\infty}^{0} e^{-x^2}\cos(x)dx$.

We can the evaluate $2\int_{-\infty}^{0} e^{-x^2}\cos(x)dx$ via the substitution $z = e^x$ since this change of variables maps the region of integration $(-\infty,0)$ to $[0,1]$

(8 Marks)

Question 3.

i)

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \text{ where } \boldsymbol{X} = \begin{pmatrix} 1 & 1 & 0 & 3 \\ 1 & 2 & 0 & 2 \\ 1 & 3 & 1 & 2 \\ 1 & 4 & 1 & 2 \\ 1 & 5 & 2 & 1 \end{pmatrix}, \boldsymbol{Y} = \begin{pmatrix} 7.6 \\ 7.3 \\ 7.7 \\ 8.2 \\ 8.9 \end{pmatrix} \text{ and } \boldsymbol{\beta} = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}.$$

ii) For this second dataset, $2x_3 = x_1$. This linear dependence between the two predictors means that we cannot uniquely determine a coefficient for each. This will also give $\boldsymbol{X}^t\boldsymbol{X}$ as a singular (non-invertible) matrix.

(6 Marks)

Question 4.

i) $y_i = 0.5 + \varepsilon_i$. The two *y* values (0 and 1) both observed at the two *x* values (0 and 1) hence the trendline will be horizontal (i.e. slope of 0). The intercept term will therefore be the mean of the observed *y*s i.e. 0.5 here.

ii) The sum of squares than the fitted model is exactly the same as the sum of squares around the mean hence $SST = SSE$ so the R-squared value is 0.

iii) Any such three or more points which lie on a straight line would give an R-squared value of one. One such option would be $(0,0)$, $(1,1)$, $(2,2)$ and $(3,3)$.

(6 Marks)