Diploma/MSc in Actuarial Mathematics and MSc in Financial Mathematics
Statistical Modelling
Tutorial 2 - solutions

1. (a) $x^T y = \sum_{k=1}^n x_k y_k$ and xy^T is the $n \times n$ matrix whose (i,j) entry is $x_i y_j$.

   (b) Without loss of generality, we can assume $E(X_i) = 0$ for all i, otherwise simply replace X_i by $X_i - E(X_i)$. In this case $\text{Cov}(X_i, X_j) = E(X_i X_j)$ and from (a), we can write $\Sigma = E(XX^T)$, where $X^T = (X_1, X_2, \ldots, X_n)$. But the matrix $E(XX^T)$ is positive definite because $y^T XX^T y = (y^T X)^2 = \Big(\sum_k y_k X_k\Big)^2 > 0$ with positive probability, so its expected value is positive. (Since the X_i are not constants with probability 1, it is not possible to find a $y \neq 0$ so that $\sum_k y_k X_k = 0$ with probability 1.)

   (c) Suppose Σ is not invertible. Then there is a solution $y \neq 0$ to the equation $\Sigma y = 0$, so $y^T \Sigma y = 0$ which contradicts the positive definiteness of Σ.

2. (a) Using $\text{Var}(\sum_i X_i) = \text{Cov}(\sum_i X_i, \sum_i X_i)$ and $\text{Cov}(X_i, X_j) = \rho\sigma^2$, we have

$$\text{Var}(\sum_i X_i) = \text{Cov}(\sum_i X_i, \sum_i X_i) = \sum_{i=1}^n \text{Cov}(X_i, X_i) + \sum_{i=1}^n \sum_{j \neq i} \text{Cov}(X_i, X_j)$$

$$= n\sigma^2 + 2(n-1)\rho\sigma^2.$$

   The desired result follows from $\text{Var}(\bar{X}) = n^{-2}\text{Var}(\sum_i X_i)$.

   (b) In the iid case, $\text{Var}(\bar{X}) = \sigma^2/n$. In the situation of (a), if $\rho > 0$, $\text{Var}(\bar{X}) > \sigma^2/n$ while if $\rho < 0$ $\text{Var}(\bar{X}) < \sigma^2/n$.

3. (a) Note that S_n is the number of X_i which take value 1, hence S_n has $\text{bin}(n,p)$ distribution.

   (b) Since $E(S_n) = np$ and $\text{Var}(S_n) = np(1-p)$, the Central Limit Theorem says that

$$\frac{S_n - np}{\sqrt{np(1-p)}} \sim N(0,1) \quad \text{approx. for large } n.$$

   Hence

$$P(S_n \leq k) = P(S_n < k + 1/2) = P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{k + 1/2 - np}{\sqrt{np(1-p)}}\Big)$$

$$\approx \Phi\Big(\frac{k + 1/2 - np}{\sqrt{np(1-p)}}\Big).$$

   The "continuity correction" of 1/2 could be omitted, in which case, the approximation which CLT gives would be

$$P(S_n \leq k) = P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\Big) \approx \Phi\Big(\frac{k - np}{\sqrt{np(1-p)}}\Big).$$

   We see that the difference between the 2 approximations is an extra term $(4np(1-p))^{-1/2}$ added to the argument of Φ. For large n, this difference is small, but the value of p is also significant in determining the relative accuracy of the 2 approximations.

4. Let X denote the number who do not turn up. Then $X \sim \text{bin}(310, 0.04)$ which is approx. $N(12.4, 11.904)$. Using the normal approximation gives

$$P(X \geq 10) = P(X > 9.5) \approx 1 - \Phi((9.5 - 12.4)/\sqrt{11.904}) = 1 - \Phi(-0.84) \approx 0.80.$$

If the continuity correction was omitted, the same calculation would give

$$P(X \geq 10) \approx 1 - \Phi((10 - 12.4)/\sqrt{11.904}) = 1 - \Phi(-0.70) \approx 0.76.$$

Even with a value of n as large as 310, the answers still differ by quite a lot. (This is because the value of p is quite small, so $(4np(1-p))^{-1/2}$ is still quite big.)

Alternatively, one could argue that $p = 0.04$ is sufficiently close to 0 to use the Poisson approximation, with mean 12.4 Therefore $P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.2092 \approx 0.79$, very similar to the answer obtained using the normal approximation with continuity correction. This example illustrates that there are some situations where both the Poisson and normal approximations are equally acceptable.

5. (a) Note that S_n has Poisson distribution with mean λn. Since $E(X_i) = \text{Var}(X_i) = \lambda$, the Central Limit Theorem says that

$$\frac{S_n - n\lambda}{\sqrt{n\lambda}} \sim N(0,1) \quad \text{approx.}$$

Hence (using the "continuity correction" as in Questions 2 and 3),

$$P(S_n \leq k) = P(S_n < k + 1/2) = P\Big(\frac{S_n - n\lambda}{\sqrt{n\lambda}} \leq \frac{k + 1/2 - n\lambda}{\sqrt{n\lambda}}\Big)$$

$$\approx \Phi\Big(\frac{k + 1/2 - n\lambda}{\sqrt{n\lambda}}\Big).$$

(b)

$$P(S_n = n) = P(n - 1/2 < S_n < n + 1/2)$$

$$= P\Big(\frac{n - 1/2 - n\lambda}{\sqrt{n\lambda}} < \frac{S_n - n\lambda}{\sqrt{n\lambda}} < \frac{n + 1/2 - n\lambda}{\sqrt{n\lambda}}\Big)$$

$$\approx \Phi\Big(\frac{n + 1/2 - n\lambda}{\sqrt{n\lambda}}\Big) - \Phi\Big(\frac{n - 1/2 - n\lambda}{\sqrt{n\lambda}}\Big).$$

(c) Setting $\lambda = 1$ in (b) gives

$$P(S_n = n) = \frac{n^n e^{-n}}{n!} \approx \Phi\Big(\frac{1}{2\sqrt{n}}\Big) - \Phi\Big(-\frac{1}{2\sqrt{n}}\Big).$$

But this last expression is the integral of the density of $N(0,1)$ over the interval $(-(2\sqrt{n})^{-1},\ (2\sqrt{n})^{-1})$, which is approximately the area of a rectangular strip with width $(\sqrt{n})^{-1}$ and height $(2\pi)^{-1/2}$. Hence

$$\frac{n^n e^{-n}}{n!} \approx (2\pi n)^{-1/2}.$$

The desired approximation follows by making $n!$ the subject of the above approximate identity.