Exercise Sheet 7

The exercise sheet is subdivided into three parts:

- Part A contains warm-up questions you should do in your own time.
- You should solve all questions of Part B and put your solutions in the locker in the undergraduate common room in the Department of Statistical Science (Room number 117) assigned to your tutor group by the deadline stated on the exercise sheet. Don't forget to mark your solutions with your name and student number.

  Reminder: If you miss the deadline for reasons outside your control, for example illness or bereavement, you ***must*** submit a claim for extenuating circumstances, normally within a week of the deadline. Your home department will advise you of the appropriate procedures. For Statistical Science students, the relevant information is on the DOSSSH Moodle page.

  Your tutor will then mark one question selected at random after you handed in your solutions (mark counts towards ICA) and will return your marked solutions to you at the next tutorial at which the solutions to the part B questions will be discussed. Marks available for each part question are indicated in square brackets, e.g. [3]. Very succinct solutions to these questions will usually be made available on Moodle at 9:15am on Thursday mornings, so don't hand in late!
- Solutions to Part C questions will be made available in time for exam revision. These questions will not be discussed at the tutorials.

A: Warm-up questions

1. Show that the pgf of the Geo(π) distribution is $\pi z\{1-(1-\pi)z\}^{-1}$. Deduce the mean and variance.

2. If X is Gam(ν, λ) show that $\log X$ has mgf $\lambda^{-s}\Gamma(s+\nu)/\Gamma(\nu)$.

B: Answers to hand in by Thursday, December 3rd, 9am

1. Let

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\mu_0, \Sigma_0\right),$$

where $\mu_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\Sigma_0 = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$.

   (a) Let $Y_1 = (X_1 + X_2)/2$. By writing $Y_1 = a^T X$ for a suitably chosen vector $a \in \mathbb{R}^2$, derive the distribution of $Y_1 = \frac{X_1+X_2}{2}$. [2]

   (b) Use moment generating functions to show that for any $n \in \mathbb{N}$, any random normal random vector $Z \sim N(\mu, \Sigma)$ with mean vector $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ and any matrix $A \in \mathbb{R}^{n\times n}$ we have that $Y = AZ \sim N(A\mu, A\Sigma A^T)$. Hence compute the joint distribution of Y_1 and Y_2, where Y_1 is as defined in part (a) and $Y_2 = X_2 - X_1$. [3]

   (c) Find an upper triangular[1] matrix $R \in \mathbb{R}^{2\times 2}$ such that $R^T R = \Sigma_0$. [2]

[1] 'Upper triangular' means that the elements below the diagonal are zero, i.e. here if $R = \begin{pmatrix} r_{1,1} & r_{1,2} \\ r_{2,1} & r_{2,2} \end{pmatrix}$ then $r_{2,1} = 0$.

(d) Find the eigenvalues $\lambda_1, \lambda_2 \in \mathbb{R}$ and corresponding eigenvectors $v_1, v_2 \in \mathbb{R}^2$ of Σ_0. Hence, verify that $\Sigma_0 = BDB^T$, where $D \in \mathbb{R}^{2\times 2}$ is diagonal with diagonal entries given by the eigenvalues and the columns of $B \in \mathbb{R}^{2\times 2}$ are corresponding unit-length eigenvectors of Σ_0, i.e. $\frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|}$ respectively. $\|x\|$ denotes the Euclidean length of a vector $x = (x_1, x_2)^T \in \mathbb{R}^2$, i.e $\|x\| = \sqrt{x_1^2 + x_2^2}$. Finally, verify that $S = B \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} B^T$ is symmetric and satisfies $SS = \Sigma_0$. [4]

(e) Obtain two independent, standard normally distributed random numbers with 6 significant digits from `https://www.random.org/gaussian-distributions/` and write them down denoting them as $z = (z_1, z_2)$. Use your result from part (c), to convert these two numbers from an $N(0, I)$ to an $N(0, \Sigma_0)$ distribution. Then use your result from part (d) to convert the same pair of numbers (z_1, z_2) to an $N(0, \Sigma_0)$ distribution. [4]

2. (a) During a lecture, the lecturer conducts an experiment whereby 36 student volunteers each select either a one penny or a two pence coin to put into a bag that is passed around. Once all 36 coins have been collected, the sum of money S is determined. Noting that $S - 36$ is the number of two pence coins in the sample, the experiment is modelled using a binomial distribution: $S - 36 \sim \text{Bin}(36, 0.5)$.

   i. Use the CLT to approximate the probability that $S \in [49, 55]$ without continuity correction, i.e. compute $P\left(\frac{49-\mu}{\sigma} < Z \leq \frac{55-\mu}{\sigma}\right)$, where μ and σ are the mean and standard deviation of S and Z follows a standard normal distribution. [2]

   ii. Use the CLT to approximate the probability that $S \in [49, 55]$ with continuity correction. [2]

   iii. Compare the two approximation results above to the true probability[2] $P(49 \leq S \leq 55) \approx 0.6585$. [1]

(b) Let the random variables $X_1, X_2, X_3, \ldots$ be independent and identically distributed. Furthermore, let $P(X_1 = 0) = 1 - \delta$ and $P(X_1 = \alpha) = \delta$, where $\alpha \in (0, \infty)$ and $\delta \in (0, 1)$ are parameters. The sample size is denoted as $n \in \mathbb{N} \setminus \{1\}$. You may use without proof that for the sample mean $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ we have $P(\bar{X}_n \leq 0) = (1-\delta)^n$, $\mathbb{E}[\bar{X}_n] = \alpha\delta$ and $\text{Var}(\bar{X}_n) = \alpha^2\delta(1-\delta)/n$.

   i. Use the CLT without continuity correction to approximate $P(\bar{X}_n \leq 0)$. Write down your final result in terms of Φ, the standard normal cdf. [2]

   ii. Use the CLT with continuity correction to approximate $P(\bar{X}_n \leq 0)$. Write down your final result in terms of Φ. [3]

   iii. Let $\delta = 1/n$ and $\alpha = \sqrt{\frac{n^3}{n-1}}$. Compute the limit of your approximating probabilities from parts bi and bii above when $n \to \infty$ and compare your results to the true limiting probability $\lim_{n\to\infty} P(\bar{X}_n \leq 0) = e^{-1}$. Explain why the CLT approximations both fail in this situation. [5]

[2] Take STAT7001 if you want to learn how to obtain such probabilities using a computer!