MATH50010: Probability for Statistics
Problem Sheet 5

1. In question 6 of Problem Sheet 4, you derived the cdfs of a number of random variables involving the minimum or maximum of a random sample. In this problem we will derive the limiting distribution of these same random variables.

   Suppose $(X_1, \ldots, X_n)$ is a collection of independent and identically distributed random variables taking values on $\mathbb{X}$ with pmf/pdf f_X and cdf F_X, let Y_n and Z_n correspond to the *maximum* and *minimum* order statistics derived from $X_1, \ldots, X_n$.

   (a) Suppose $X_1, \ldots, X_n \sim \text{Unif}(0, 1)$, that is
   $$F_X(x) = x, \text{ for } 0 \leq x \leq 1.$$
   Find the limiting distributions of Y_n and Z_n as $n \longrightarrow \infty$.

   (b) Suppose $X_1, \ldots, X_n$ have cdf
   $$F_X(x) = 1 - x^{-1}, \text{ for } x \geq 1.$$
   Find the limiting distributions of Z_n and $U_n = Z_n^n$ as $n \longrightarrow \infty$.

   (c) Suppose $X_1, \ldots, X_n$ have cdf
   $$F_X(x) = \frac{1}{1 + e^{-x}}, \text{ for } x \in \mathbb{R}.$$
   Find the limiting distributions of Y_n and $U_n = Y_n - \log n$, as $n \longrightarrow \infty$.

   (d) Suppose $X_1, \ldots, X_n$ have cdf
   $$F_X(x) = 1 - \frac{1}{1 + \lambda x}, \text{ for } x > 0.$$
   Let $U_n = Y_n/n$ and $V_n = nZ_n$. Find the limiting distributions of Y_n, Z_n, U_n, and V_n as $n \longrightarrow \infty$.

2. Suppose that the random variable X has mgf, $M_X(t)$ given by
   $$M_X(t) = \frac{1}{8}e^t + \frac{2}{8}e^{2t} + \frac{5}{8}e^{3t}.$$
   Find the probability distribution, expectation, and variance of X.

   [Hint: Consider M_X and its definition.]

3. Suppose that X is a continuous random variable with pdf
   $$f_X(x) = \exp\{-(x + 2)\}, \text{ for } -2 < x < \infty.$$
   Find the mgf of X, and *hence* find the expectation and variance of X.

4. Suppose $Z \sim \text{N}(0, 1)$.

   (a) Find the mgf of Z, and also the pdf and the mgf of the random variable X, where
   $$X = \mu + \frac{1}{\lambda}Z,$$
   for parameters μ and $\lambda > 0$.

(b) Find the expectation of X, and the expectation of the function $g(X)$, where $g(x) = e^x$. Use both the definition of the expectation directly and the mgf and compare the complexity of your calculations.

(c) Suppose now Y is the random variable defined in terms of X by $Y = e^X$. Find the pdf of Y, and show that the expectation of Y is

$$\exp\left\{\mu + \frac{1}{2\lambda^2}\right\}.$$

(d) Let random variable T be defined by $T = Z^2$. Find the pdf and mgf of T.

5. Suppose that X is a random variable with pmf/pdf f_X and mgf M_X. The *cumulant generating function* of X, K_X, is defined by $K_X(t) = \log\left[M_X(t)\right]$. Prove that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{K_X(t)\right\}_{t=0} = \mathrm{E}(X), \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2}\left\{K_X(t)\right\}_{t=0} = \mathrm{Var}(X).$$

6. Using the central limit theorem, construct Normal approximations to random variables with each of the following distributions,

(a) Binomial distribution, $X \sim \mathrm{Binomial}(n, \theta)$;

(b) Poisson distribution, $X \sim \mathrm{Poisson}(\lambda)$;

(c) Negative Binomial distribution, $X \sim \text{Negative Binomial}(n, \theta)$.

For discussion

7. Suppose we observe a sequence of random variables from a uniform distribution, $X_i \overset{\text{iid}}{\sim} \text{UNIFORM}(0, 1)$, for $i = 1, 2, \ldots$. We wish to investigate the asymptotic distribution of the sample median of the first n variables in this sequence. We assume n is odd for simplicity; then M_n is the middle value in the ordered list of the first n variables. Let

$$\begin{aligned} M_n &= \mathrm{median}(X_1, \ldots, X_n), \text{ where } n \text{ is odd} \\ &= r^{\text{th}} \text{ order statistic with } r = (n+1)/2. \end{aligned}$$

(a) First, we will derive the CDF of M_n. Let J_n be the number of the $X_1, \ldots, X_n$ that are less than or equal to x. Explain why $M_n \leq x$ if and only if *at least* r of the first n of the X_i are less than or equal to x. What is the distribution of J_n?

(b) Show that

$$F_{M_n}(x) = \Pr\left(L_n \geq \frac{n + 1 - 2nx}{2\sqrt{nx(1-x)}}\right),$$

where L_n is a transformation of J_n that converges in distribution to $Z \sim N(0, 1)$ as $n \to \infty$.

(c) Show that M_n has a degenerate limit

$$\lim_{n\to\infty} F_{M_n}(x) = \begin{cases} 0 & \text{if } x < 1/2, \\ \frac{1}{2} & \text{if } x = 1/2, \\ 1 & \text{if } x > 1/2. \end{cases}$$

(d) As in the central limit theorem, we seek a rescaling of M_n that has a non-degenerate distribution. Consider the variable $S_n = (M_n - \frac{1}{2})n^p$, for some power p. First, write down F_{S_n} in terms of F_{M_n}.

(e) Show that

$$\lim_{n\to\infty} F_{S_n}(s) = \Pr\left(Z \geq \frac{\frac{1}{2} - sn^{1-p}}{\sqrt{\frac{n}{4} - s^2 n^{1-2p}}} \right),$$

where $Z \sim N(0,1)$.

(f) Find the value of p that gives rise to a non-degenerate distribution.

(g) Deduce that M_n has an approximate normal distribution as n becomes large, and state (in terms of n) its mean and variance.