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

   $$F_X(x) = x, \quad \text{for } 0 \leq x \leq 1.$$

   Find the limiting distributions of Y_n and Z_n as $n \longrightarrow \infty$.

   (b) Suppose $X_1, \ldots, X_n$ have cdf

   $$F_X(x) = 1 - x^{-1}, \quad \text{for } x \geq 1.$$

   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

   $$F_X(x) = \frac{1}{1 + e^{-x}}, \quad \text{for } x \in \mathbb{R}.$$

   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

   $$F_X(x) = 1 - \frac{1}{1 + \lambda x}, \quad \text{for } x > 0.$$

   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 $n \longrightarrow \infty$.

(a) In the limit as $n \to \infty$ we have the limit for fixed *y as*

$$F_{Y_n}(y) = \{F_X(y)\}^n = y^n \to \begin{cases} 0, & y < 1 \\ 1, & y \geq 1 \end{cases}.$$

This is a step function with single step of size 1 at $y = 1$. Hence the limiting random variable Y is discrete with $\Pr(Y = 1) = 1$, *that is, the limiting distribution is* degenerate *at* 1. *Also in the limit as $n \to \infty$ we have the limit for* fixed *z as*

$$F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n = 1 - (1 - z)^n \to \begin{cases} 0, & z \leq 0 \\ 1, & z > 0 \end{cases}.$$

This is a step function with single step of size 1 at $z = 0$. Hence the limiting random variable Z is a discrete variable with $\Pr(Z = 0) = 1$, *that is, the limiting distribution is* degenerate *at* 0. *Note here that the limiting function is* not *a cdf as it is not* right-continuous, *but the limiting distribution still exists. The definition of convergence in distribution only* *refers to pointwise convergence* at points of continuity of the limit function*, and here the limit function is not continuous at zero.*

These results are intuitively reasonable. As the sample size gets increasingly large, we will very probably obtain a random variable arbitrarily close to each end of the range.

We have established convergence in distribution*, but we also have for* $1 > \varepsilon > 0$ *and as* $n \to \infty$*,*

$$\begin{aligned} \Pr(|Y_n - 1| < \varepsilon) &= \Pr(1 - Y_n < \varepsilon) = \Pr(1 - \varepsilon < Y_n) = 1 - \Pr(Y_n < 1 - \varepsilon) = 1 - \varepsilon^n \to 1, \\ \Pr(|Z_n - 0| < \varepsilon) &= \Pr(Z_n < \varepsilon) = 1 - (1 - \varepsilon)^n \to 1. \end{aligned}$$

So Y_n *and* Z_n converge in probability *to 1 and 0, respectively.*

(b) Recall that

$$F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n = 1 - \left(1 - \left(1 - \frac{1}{z}\right)\right)^n = 1 - \frac{1}{z^n}, \text{ for } z > 1.$$

In the limit as $n \to \infty$ *we have for* fixed z

$$F_{Z_n}(z) \to \begin{cases} 0, & z \leq 1 \\ 1, & z > 1 \end{cases}.$$

This is a step function with single step of size 1 at $z = 1$*. Hence the limiting random variable Z is a discrete variable with*

$$\Pr(Z = 1) = 1,$$

Again, the limiting function is not a cdf as it is not right continuous at one. This does not affect our conclusion since the limit function is not continuous at this point.

Setting $U_n = Z_n^n$*, we found that, for* $u > 1$*,*

$$F_{U_n}(u) = \Pr(U_n \leq u) = \Pr(Z_n^n \leq u) = \Pr\left(Z_n \leq u^{1/n}\right) = 1 - \frac{1}{\left(u^{1/n}\right)^n} = 1 - \frac{1}{u},$$

which does not depend on n*. Hence the limiting distribution of* U_n *is*

$$F_U(u) = 1 - \frac{1}{u}, \text{ for } u > 1.$$

For $u \leq 1$*,* $F_U(u) = 0$ *for all* n*.*

(c) Recall

$$F_{Y_n}(y) = \{F_X(y)\}^n = \left(\frac{1}{1 + e^{-y}}\right)^n, \qquad y \in \mathbb{R}.$$

In the limit as $n \to \infty$*, for* fixed y

$$F_{Y_n}(y) \to 0, \text{ for all } y.$$

Hence there is no limiting distribution*. Recall also that*

$$F_{U_n}(u) = F_{Y_n}(u + \log n) = \left(\frac{1}{1 + e^{-u - \log n}}\right)^n,$$

so that

$$F_{U_n}(u) = \left(\frac{1}{1 + \frac{e^{-u}}{n}}\right)^n = \left(1 + \frac{e^{-u}}{n}\right)^{-n} \to \exp\left\{-e^{-u}\right\}, \quad \text{as } n \to \infty,$$

which is a valid cdf. Hence the limiting distribution is

$$F_U(u) = \exp\left\{-e^{-u}\right\}, \quad u \in \mathbb{R}.$$

(d) *Recall*

$$F_{Y_n}(y) = \{F_X(y)\}^n = \left(\frac{\lambda y}{1+\lambda y}\right)^n, \textit{ for } y > 0,$$

and so as $n \to \infty$ *for* fixed y

$$F_{Y_n}(y) \to 0, \textit{ for all } y$$

and there is no limiting distribution. *In the limit as* $n \to \infty$ *for* fixed $z > 0$

$$F_{Z_n}(z) = 1 - \{1 - F_X(z)\}^n = 1 - \left(1 - \left(1 - \frac{1}{1+\lambda z}\right)\right)^n = 1 - \frac{1}{(1+\lambda z)^n} \to \begin{cases} 0, & z \leq 0 \\ 1, & z > 0 \end{cases}.$$

This is a step function with single step of size 1 at $z = 0$. *Hence the limiting random variable* Z *is a discrete variable with* $P(Z = 0) = 1$*: the limiting distribution is* degenerate *at* 0. *Again, the limiting function is not a cdf as it is not right continuous at zero, but this does not affect our conclusion, as the limit function is not continuous at this point.*
Recall that for $u > 0$,

$$F_{U_n}(u) = \Pr(U_n \leq u) = \Pr(Y_n/n \leq u) = \Pr(Y_n \leq nu) = F_{Y_n}(nu) = \left(\frac{\lambda nu}{1+\lambda nu}\right)^n,$$

so that

$$F_{U_n}(u) = \left(\frac{\lambda nu}{1+\lambda nu}\right)^n = \left(1 + \frac{1}{n\lambda u}\right)^{-n} \to \exp\left\{-\frac{1}{\lambda u}\right\}, \quad \textit{as } n \to \infty,$$

which is a valid cdf. Hence the limiting distribution is

$$F_U(u) = \exp\left\{-\frac{1}{\lambda u}\right\}, \textit{ for } u > 0.$$

Finally, recall that for $v > 0$,

$$F_{V_n}(v) = \Pr(V_n \leq v) = \Pr(nZ_n \leq v) = \Pr(Z_n \leq v/n) = F_{Z_n}(v/n) = 1 - \left(\frac{1}{1+\frac{\lambda v}{n}}\right)^n$$

so that

$$F_{V_n}(v) = 1 - \left(1 + \frac{\lambda v}{n}\right)^{-n} \to 1 - \exp\{-\lambda v\}, \quad \textit{as } n \to \infty,$$

which is a valid cdf. Hence the limiting distribution is

$$F_V(v) = 1 - \exp\{-\lambda v\}, \textit{ for } v > 0.$$

Hence the limiting distribution of V *is* Exponential(λ).

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.]*

By definition of mgfs for discrete variables, we can deduce immediately that since

$$M_X(t) = \sum_{x=-\infty}^{\infty} e^{tx} f_X(x),$$

$\Pr(X = x)$ *is just the coefficient of* e^{tx} *in the expression for* M_X*. Hence* $\Pr(X = 1) = 1/8$, $\Pr(X = 2) = 1/4$ *and* $\Pr(X = 3) = 5/8$*. Now* $\mathrm{E}(X^r) = M_X^{(r)}(0)$*, so that*

$$\begin{aligned} \mathrm{E}(X) &= M_X^{(1)}(0) = \tfrac{1}{8} + 2\tfrac{1}{4} + 3\tfrac{5}{8} = \tfrac{5}{2}, \\ \mathrm{E}(X^2) &= M_X^{(2)}(0) = \tfrac{1}{8} + 4\tfrac{1}{4} + 9\tfrac{5}{8} = \tfrac{27}{4}, \end{aligned}$$

so therefore

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - \{\mathrm{E}(X)\}^2 = \frac{1}{2}.$$

3. Suppose that X is a continuous random variable with pdf

$$f_X(x) = \exp\{-(x+2)\}\,, \quad \text{for} \quad -2 < x < \infty.$$

Find the mgf of X, and *hence* find the expectation and variance of X.

For this pdf,

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x = \int_{-2}^{\infty} e^{tx} e^{-(x+2)}\,\mathrm{d}x = e^{-2}\int_{-2}^{\infty} e^{-(1-t)x}\,\mathrm{d}x \\ &= \frac{e^{-2}}{1-t}\int_{-2(1-t)}^{\infty} e^{-y}\,\mathrm{d}y = \frac{e^{-2}}{1-t}\left[-e^{-y}\right]\Big|_{-2(1-t)}^{\infty} = \frac{e^{-2t}}{1-t}, \quad \textit{for } t < 1. \end{aligned}$$

Now

$$M_X^{(1)}(t) = \tfrac{e^{-2t}}{(1-t)^2}(2t-1), \qquad M_X^{(2)}(t) = \tfrac{e^{-2t}}{(1-t)^3}\left[1 + (2t-1)^2\right],$$

so that $M_X^{(1)}(0) = -1 = \mathrm{E}(X)$ *and* $M_X^{(2)}(0) = 2 = \mathrm{E}(X^2) \Longrightarrow$ *Var*$(X) = 1$.

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

   (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.

(a) *To calculate the mgf*

$$\begin{aligned} M_Z(t) &= \mathrm{E}(e^{tZ}) = \int_{-\infty}^{\infty} e^{zt} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2}\right\} \,\mathrm{d}z = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(z-t)^2}{2}\right\} \,\mathrm{d}z \\ &= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{u^2}{2}\right\} \,\mathrm{d}u = e^{t^2/2}, \end{aligned}$$

completing the square in z, and then setting $u = z - t$, as the integrand is a pdf.
Now, using the transformation theorem for univariate one-to-one transformations we have $X = \mu + \frac{1}{\lambda} Z$ implies $Z = \lambda(X - \mu)$, so

$$f_X(x) = f_Z(\lambda(x-\mu))\,\lambda = \frac{\lambda}{\sqrt{2\pi}} \exp\left\{-\frac{\lambda^2}{2}(x-\mu)^2\right\}, \qquad x \in \mathbb{R}.$$

To calculate the mgf of X,

$$M_X(t) = \mathrm{E}\left(e^{t(\mu + Z/\lambda)}\right) = e^{\mu t} M_Z(t/\lambda) = \exp\left\{\mu t + \frac{t^2}{2\lambda^2}\right\}.$$

(b) *Using the definition of expectation,*

$$\begin{aligned} \mathrm{E}(X) &= \int_{-\infty}^{\infty} x f_X(x) \,\mathrm{d}x = \int_{-\infty}^{\infty} x \left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda^2}{2}(x-\mu)^2\right\} \,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} \left(\mu + t\lambda^{-1}\right) \left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{t^2}{2}\right\} \lambda^{-1} \,\mathrm{d}t \;\; \textit{[with } t = \lambda(x-\mu)] \\ &= \mu \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{t^2}{2}\right\} \,\mathrm{d}t + \lambda^{-1} \int_{-\infty}^{\infty} t \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{t^2}{2}\right\} \,\mathrm{d}t \\ &= \mu, \end{aligned}$$

as the first integral is 1, and the second integral is zero, as the integrand is an odd function about zero. Hence

$$\mathrm{E}(X) = \mu.$$

Note that it is generally true that if a pdf is symmetric about a particular value, then that value is the expectation (if the expectation integral is finite). Alternately, we could use the mgf result

$$\mathrm{E}(X) = \frac{\mathrm{d}}{\mathrm{d}s}\{M_X(s)\}_{s=0} = M_X^{(1)}(0),$$

to compute

$$\mathrm{E}(X) = \frac{\mathrm{d}}{\mathrm{d}s}\left\{\exp\left\{\mu s + \frac{s^2}{2\lambda^2}\right\}\right\}_{s=0} = \left\{\left(\mu + \frac{s}{\lambda^2}\right) \exp\left\{\mu s + \frac{s^2}{2\lambda^2}\right\}\right\}_{s=0} = \mu.$$

The expectation of $g(X) = e^X$ is

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \,\mathrm{d}x = \int_{-\infty}^{\infty} e^x \left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda^2}{2}(x-\mu)^2\right\} \,\mathrm{d}x$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \exp\left\{\mu + t\lambda^{-1}\right\} \left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{t^2}{2}\right\} \lambda^{-1}\, \mathrm{d}t, \qquad [\textit{ setting } t = \lambda(x-\mu)] \\
&= \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \exp\left\{\mu + t\lambda^{-1} - \frac{t^2}{2}\right\} \mathrm{d}t \\
&= \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(t^2 - 2t\lambda^{-1} - 2\mu\right)\right\} \mathrm{d}t.
\end{aligned}
$$

Completing the square in the exponent, we have

$$
t^2 - 2t\lambda^{-1} - 2\mu = \left(t - \lambda^{-1}\right)^2 - \left(2\mu + \lambda^{-2}\right)
$$

and hence

$$
\begin{aligned}
\mathrm{E}\left[g(X)\right] &= \left(\frac{1}{2\pi}\right)^{1/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(t-\lambda^{-1}\right)^2 + \left(\mu + \frac{1}{2\lambda^2}\right)\right\} \mathrm{d}t \\
&= \exp\left\{\mu + \frac{1}{2\lambda^2}\right\} \int_{-\infty}^{\infty} \left(\frac{1}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}\left(t-\lambda^{-1}\right)^2\right\} \mathrm{d}t = \exp\left\{\mu + \frac{1}{2\lambda^2}\right\},
\end{aligned}
$$

as the integral is equal to 1 since it is the integral of a pdf for all choices of λ. Alternatively, simply note that $\mathrm{E}(e^X) \equiv M_X(1)$.

(c) If $Y = e^X$, the support of Y is $\mathbb{Y} = R^+$. From first principles

$$
F_Y(y) = \Pr\left(Y \le y\right) = \Pr\left(e^X \le y\right) = \Pr\left(X \le \log y\right) = F_X(\log y),
$$

so by differentiation

$$
f_Y(y) = f_X(\log y)\,\frac{1}{y}, \quad \textit{for } y > 0.
$$

Note that the function $g(t) = e^t$ is a monotone increasing function, with $g^{-1}(t) = \log t$, so that we can use the transformation result directly. That is,

$$
f_Y(y) = f_X(g^{-1}(y))\,|J(y)| \qquad \textit{where} \qquad |J(y)| = \left|\frac{\mathrm{d}}{\mathrm{d}t}\left\{g^{-1}(t)\right\}_{t=y}\right| = \left|\frac{\mathrm{d}}{\mathrm{d}t}\left\{\log t\right\}_{t=y}\right| = \frac{1}{y}.
$$

Hence

$$
f_Y(y) = \frac{1}{y}\left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda^2}{2}(\log y - \mu)^2\right\}, \quad \textit{for } y > 0.
$$

For the expectation, we have from first principles

$$
\begin{aligned}
\mathrm{E}\left(Y\right) &= \int_0^{\infty} y f_Y(y)\, \mathrm{d}y = \int_{-\infty}^{\infty} y\,\frac{1}{y}\left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda}{2}(\log y - \mu)^2\right\} \mathrm{d}y \\
&= \int_{-\infty}^{\infty} \left(\frac{\lambda^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda^2}{2}(t-\mu)^2\right\} e^t\, \mathrm{d}t \;=\; \exp\left\{\mu + \frac{1}{2\lambda^2}\right\},
\end{aligned}
$$

where $t = \log y$, as the integral is precisely the one carried out above. Note that the expectation could be written down immediately as $M_X(1)$. This illustrates the transformation/expectation result that, if $Y = g(X)$, then

$$
\mathrm{E}\left(Y\right) = \mathrm{E}\left[g(X)\right].
$$

(d) *If $T = Z^2$, then from first principles*

$$\begin{aligned} F_T(t) &= \Pr(T \le t) = \Pr(Z^2 \le t) = \Pr(-\sqrt{t} \le Z \le \sqrt{t}) \\ \Longrightarrow f_T(t) &= \tfrac{1}{2\sqrt{t}}\left[f_Z(\sqrt{t}) + f_Z(-\sqrt{t})\right] = \tfrac{1}{\sqrt{2\pi}} t^{-1/2} \exp\left\{-\tfrac{t}{2}\right\}, \quad t > 0, \end{aligned}$$

and hence

$$\begin{aligned} M_T(t) &= \mathrm{E}(e^{tT}) = \int_{-\infty}^{\infty} e^{tx} f_T(x)\,\mathrm{d}x = \int_0^{\infty} e^{tx} \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{x}{2}\right\}\,\mathrm{d}x \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi x}} \exp\left\{-\frac{(1-2t)x}{2}\right\}\,\mathrm{d}x \\ &= \left(\frac{1}{1-2t}\right)^{1/2} \int_0^{\infty} \frac{1}{\sqrt{2\pi y}} \exp\left\{-\frac{y}{2}\right\}\,\mathrm{d}y = \left(\frac{1}{1-2t}\right)^{1/2}, \end{aligned}$$

where $y = (1-2t)x$, as the integrand is a pdf.

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

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

We have $K_X(t) = \log M_X(t)$, hence

$$K_X^{(1)}(t) = \frac{\mathrm{d}}{\mathrm{d}s}\left\{K_X(t)\right\}_{s=t} = \frac{\mathrm{d}}{\mathrm{d}s}\left\{\log M_X(t)\right\}_{s=t} = \frac{M_X^{(1)}(t)}{M_X(t)} \Longrightarrow K_X^{(1)}(0) = \frac{M_X^{(1)}(0)}{M_X(0)} = \mathrm{E}(X),$$

as $M_X(0) = 1$. Similarly

$$K_X^{(2)}(t) = \frac{M_X(t) M_X^{(2)}(t) - \left\{M_X^{(1)}(t)\right\}^2}{\left\{M_X(t)\right\}^2},$$

and hence

$$K_X^{(2)}(0) = \frac{M_X(0) M_X^{(2)}(0) - \left\{M_X^{(1)}(0)\right\}^2}{\left\{M_X(0)\right\}^2} = \mathrm{E}(X^2) - \left\{\mathrm{E}(X)\right\}^2,$$

so that $K_X^{(2)}(0) = \mathrm{Var}(X)$.

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, $X \sim \mathrm{Negative\ Binomial}(n, \theta)$.

The key is to find iid random variables $X_1, \ldots, X_n$ such that

$$X = \sum_{i=1}^{n} X_i,$$

and then to use the Central Limit Theorem for large n:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{\mathcal{D}} Z \sim \text{Normal}(0,1), \quad \textit{so that } X = \sum_{i=1}^{n} X_i \overset{\text{approx}}{\sim} \text{Normal}(n\mu, n\sigma^2),$$

where μ =E(X_i) and σ^2 =Var(X_i).

(a) $X \sim \text{Binomial}(n,\theta) \implies X = \sum_{i=1}^{n} X_i$ *where* $X_i \sim \text{Bernoulli}(\theta)$*, so that* μ =*E*$(X_i) = \theta$ *and* σ^2 =*Var*$(X_i) = \theta(1-\theta)$*, and hence*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\theta}{\sqrt{n\theta(1-\theta)}} \sim \text{Normal}(0,1) \implies X \overset{\text{approx}}{\sim} \text{Normal}(n\theta, n\theta(1-\theta)).$$

(b) $X \sim \text{Poisson}(\lambda) \implies X = \sum_{i=1}^{n} X_i$ *where* $X_i \sim \text{Poisson}\,(\lambda/n)$*, so that* μ =*E*$(X_i) = \lambda/n$ *and* σ^2 =*Var*$(X_i) = \lambda/n$*, and hence*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\frac{\lambda}{n}}{\sqrt{n\,(\lambda/n)}} = \frac{\sum_{i=1}^{n} X_i - \lambda}{\sqrt{\lambda}} \xrightarrow{\mathcal{D}} \text{Normal}(0,1) \implies X \overset{\text{approx}}{\sim} \text{Normal}(\lambda, \lambda).$$

Note that this uses the result that the sum of independent Poisson variables also has a Poisson distribution.

(c) $X \sim \text{Negative Binomial}(n,\theta) \implies X = \sum_{i=1}^{n} X_i$ *where* $X_i \sim \text{Geometric}(\theta)$*, so that* μ =*E*$(X_i) = 1/\theta$ *and* σ^2 =*Var*$(X_i) = (1-\theta)/\theta^2$*, and hence*

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\frac{1}{\theta}}{\sqrt{n\left((1-\theta)/\theta^2\right)}} \xrightarrow{\mathcal{D}} \text{Normal}(0,1) \implies X \overset{\text{approx}}{\sim} \text{Normal}\left(\frac{n}{\theta}, \frac{n(1-\theta)}{\theta^2}\right).$$

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

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

(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 $M_n \le x$ 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.

(a) Since $\Pr(X_i \leq x) = x$ *for each* i*, and the variables are independent, we know that* $J_n \sim$ BINOMIAL$(n, p = x)$ *and* $\Pr(M_n \leq x) = \Pr(J_n \geq r)$*:*

$$F_{M_n}(x) = \Pr(J_n \geq r) = \sum_{j=r}^{n} \binom{n}{j} x^j (1-x)^{n-j}.$$

(b) By the normal approximation to the binomial, we know that as $n \to \infty$

$$\frac{J_n - nx}{\sqrt{nx(1-x)}} \xrightarrow{\mathcal{D}} \mathbb{N}(0,1).$$

Transforming $\Pr(J_n \geq r)$ *and using* $r = \frac{n+1}{2}$ *gives the result stated.*

(c) Considering the point of L_n *to which* $M_n = x$ *corresponds,*

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

Applying F_Z*, the cdf of a standard normal variable, then gives the result stated:*

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

(d)

$$F_{S_n}(s) = \Pr(S_n \le s) = \Pr\left(\left(M_n - \frac{1}{2}\right) n^p \le s\right) = \Pr\left(M_n \le \frac{1}{2} + sn^{-p}\right) = F_{M_n}\left(\frac{1}{2} + sn^{-p}\right)$$

(e) From the earlier representation in terms of J_n and L_n,

$$\lim_{n\to\infty} F_{S_n}(s) = \lim_{n\to\infty} \Pr\left(\frac{J_n - \frac{n}{2} - sn^{1-p}}{\sqrt{n(\frac{1}{2} + sn^{-p})(\frac{1}{2} - sn^{-p})}} \ge \frac{1 - 2sn^{1-p}}{2\sqrt{n(\frac{1}{2} + sn^{-p})(\frac{1}{2} - sn^{-p})}}\right) = \Pr(Z \ge c_n) \tag{1}$$

Where again $Z \sim \mathbb{N}(0,1)$ and

$$c_n = \frac{1 - 2sn^{1-p}}{2\sqrt{n(\frac{1}{2} + sn^{-p})(\frac{1}{2} - sn^{-p})}} = \frac{\frac{1}{2} - sn^{1-p}}{\sqrt{\frac{n}{4} - s^2 n^{1-2p}}}.$$

(f) To avoid a degenerate asymptotic distribution, we want to pick p so that $\lim_{n\to\infty} c_n = c$, where c is finite. By inspection try $p = \frac{1}{2}$:

$$c_n = \frac{\frac{1}{2} - s\sqrt{n}}{\sqrt{\frac{n}{4} - s^2}} \longrightarrow -2s \text{ as } n \to \infty.$$

(g)

$$\lim_{n\to\infty} F_{S_n}(s) = \Pr(Z \ge -2s) = \Pr\left(-\frac{Z}{2} \le s\right) = \Pr\left(\frac{Z}{2} \le s\right).$$

Thus, $S_n \xrightarrow{\mathcal{D}} N(0, \frac{1}{4})$ and and because $M_n = \frac{S_n}{\sqrt{n}} + \frac{1}{2}$, we have $M_n \overset{\text{approx}}{\sim} N(\frac{1}{2}, \frac{1}{4n})$.