SDS 383C - Statistical Modeling 1: Homework 1

Rahul Nandakumar
Graduate Student, ORIE Program (E-ID: rn9355)

September 11, 2022

1. Prove Slutsky's theorem (the version taught in class)

Solution:

Theorem 1 *Let $\{x_n\}_{n=1}^{\infty}$ and $\{z_n\}_{n=1}^{\infty}$ be a sequence of random variables; x, z be random variables; and $x_0, z_0 \in \mathbb{R}$ be constants. Then,*

- $x_n \xrightarrow{P} x \Rightarrow x_n \xrightarrow{D} x,$
- $x_n \xrightarrow{D} x_0 \Rightarrow x_n \xrightarrow{P} x_0,$
- $x_n \xrightarrow{D} x$ *and* $z_n \xrightarrow{P} z_0 \Rightarrow$
  $x_n + z_n \xrightarrow{D} x + z_0$, $x_n z_n \xrightarrow{D} x z_0$, $x_n/z_n \xrightarrow{D} x/z_0 (z_0 \neq 0)$

Proof: From the statements given to us, we can write the following.

$$F_{x_n+z_n}(t) = P[x_n + z_n \leq t, z_n \geq z_0 - \epsilon] + P[x_n + z_n \leq t, z_n < z_0 - \epsilon]$$

Let t be a point of continuity of F_{x+z_0}. Because a distribution function has atmsot countable many points of discontinuity, we can choose ϵ positive for any t such that $t + \epsilon$ are both points of continuity of F_{x+z_0}. We now get

$$F_{x_n+z_n}(t) \leq P[x_n \leq t - z_0 + \epsilon] + P[|z_n - z_0| > \epsilon]$$

Furthermore, we have

$$P[x_n \leq t - z_0 + \epsilon] = F_{x_n+z_0}(t + \epsilon)$$

Now because $F_{x_n+z_0}(t) = P[x_n \leq t - z_0] = F_{x_n}(t - z_0)$, we must have $x_n + z_0 \xrightarrow{D} x + z_0$. This leads to the following relationship,

$$\limsup_n F_{x_n+z_n}(t) \geq \lim_n F_{x_n+z_0}(t + \epsilon) + \lim_n P[|z_n - z_0| \geq \epsilon] = F_{(x+z_0)}(t + \epsilon)$$

Similarly,

$$1 - \limsup_n F_{x_n+z_n}(t) = P[x_n + z_n > t] \leq P[x_n > t - z_0 - \epsilon] + P[|z_n - z_0| > \epsilon]$$

and hence, we get

$$\liminf_n F_{x_n+z_n}(t) \geq \lim_n F_{x_n+z_0}(t - \epsilon) = F_{x+z_0}(t - \epsilon)$$

Therefore, we can say that

$$F_{x+z_0}(t-\epsilon) \leq \liminf_n F_{x_n+z_n}(t) \leq \limsup_n F_{x_n+z_n}(t) \leq F_{(x+z_0)}(t+\epsilon)$$

From the above equation, since ϵ can reach 0 and F_{x+z_0} is continuous at t, we can conclude that $x_n + z_n \xrightarrow{D} x + z_0$. A simplified version of these statements is that if x_n converges in distribution to x and z_n converges in probability to z_0, then their joint vector (x_n, z_n) converges in distribution to (x, z_0). By applying continuous mapping further, where $g(x, y) = x + y$, we get

$$x_n + z_n \xrightarrow{D} x + z_0$$

Furthermore, if $g(x, y) = xy$, we get

$$x_n z_n \xrightarrow{D} x z_0$$

And if $g(x, y) = x/y$,

$$x_n/z_n \xrightarrow{D} x/z_0$$

(Source: Mathematical Statistics: Basic Ideas and Selected Topics Volume 1 by Bickel and Doksum)

2. For $y \sim$ Poisson(λ), show that $\mathbb{E}(y) = var(y) = \lambda$. Method of moments suggests $\bar{y}_n$ the sample mean, as well as s_n^2, the sample variance, could both be reasonable estimators of λ. Which one would you prefer? Why?

Solution: We are given, $y \sim$ Poisson(λ). The PMF of this distribution is given by $P(Y = y) = \frac{e^{-\lambda}\lambda^y}{y!}$ Now, let us find the Moment Generating Function of the Poisson Distribution.

$$\begin{aligned} M_y(t) &= \mathbb{E}(e^{ty}) \\ &= e^{t.0}(P(y=0)) + e^{t.1}(P(y=1)) + \dots \\ &= 1.e^{\lambda} + e^t.e^{-\lambda}.\frac{\lambda}{1!} + e^{2t}.e^{-\lambda}.\frac{\lambda^2}{2!} + \dots \\ &= e^{-\lambda}.e^k, \text{ where } k = e^t\lambda \\ &= e^{-\lambda}e^{et\lambda} \\ &= e^{\lambda(e^t-1)} \end{aligned} \tag{1}$$

Using the result in (1), we can calculate the first and second order moments as follows. The first order moment can be calulated as

$$\begin{aligned} M^1(t) &= \frac{dM(t)}{dt} \\ &= \frac{d}{dt}\left[e^{\lambda(e^t-1)}\right] \\ &= e^{\lambda(e^t-1)} \times (\lambda e^t) \end{aligned}$$

The second order moment can be calculated as follows.

$$\begin{aligned}
M^2(t) &= \frac{d^2M(t)}{dt^2} \\
&= \frac{d}{dt}\left[\frac{d}{dt}\left[e^{\lambda(e^t-1)}\right]\right] \\
&= e^{\lambda(e^t-1)}(\lambda e^t) + (\lambda e^t)\left[e^{\lambda(e^t-1)}.\lambda e^t\right] \\
&= (e^{\lambda(e^t-1)}\lambda e^t)\left[1+\lambda e^t\right]
\end{aligned}$$

Now, we can calculate the first order and second order moment values at $t = 0$ to get $\mathbb{E}(y)$ and $\mathbb{E}(y^2)$ respectively. Therefore,

$$\begin{aligned}
\mathbb{E}(y) &= M^1(t)|_{t=0} \\
&= e^{\lambda(e^t-1)} \times (\lambda e^t)\Big|_{t=0} \\
&= e^{\lambda(e^0-1)} \times (\lambda e^0) \\
&= \lambda
\end{aligned} \tag{2}$$

$$\begin{aligned}
\mathbb{E}(y^2) &= M^2(t)|_{t=0} \\
&= (e^{\lambda(e^t-1)}\lambda e^t)\left[1+\lambda e^t\right]|_{t=0} \\
&= (e^{\lambda(e^0-1)}\lambda e^0)\left[1+\lambda e^0\right] \\
&= \lambda(1+\lambda) = \lambda + \lambda^2
\end{aligned} \tag{3}$$

Now, we know that $var(y) = \mathbb{E}(y^2) - (\mathbb{E}(y))^2$. Therefore, from (2) and (3), we can write

$$var(y) = \mathbb{E}(y^2) - (\mathbb{E}(y))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \tag{4}$$

From (4), we have shown that for a poission distribution with shape parameter λ, $\mathbb{E}(y) = var(y) = \lambda$. We can see that by using the method of moments, both the sample mean ($\bar{y}_n$) and sample variance (s_n^2) are reasonable estimators of λ. In this case, we can say that the $\mathbb{E}(y)$ is a better estimator of λ because we know that for any distribution, $\mathbb{E}(y)$ is an unbiased estimate of population mean μ (or in this case λ). Also, we find that for a Poission Distribution, the sample mean is a sufficient statistic and is therefore a much better estimator for λ. It is also much easier to compute $\mathbb{E}(y)$.

3. For $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \overset{iid}{\sim} f_{x,y}$ with finite second order moments. Show that the sample correlation coefficient r_n converges in probability to the population correlation coefficient ρ.

Solution: For any $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \overset{iid}{\sim} f_{x,y}$ with finite second order moments, the sample correlation coefficient is given by

$$r_n = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2\}^{1/2}}$$

Now, we divide and multiply r_n with n. We get,

$$r_n = \frac{\frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{n}}{\{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\}^{1/2}}$$

For the numerator, we have

$$N = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x})}{n}$$

Now, as $n \to \infty$, $N \xrightarrow{P} \sigma_{xy}$. Similarly, we have

$$\begin{aligned} D &= \{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\}^{1/2} \\ &= D_1 \times D_2 \end{aligned}$$

Where $D_1 = \{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}\}^{1/2}$ and $D_2 = \{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\}^{1/2}$ As $n \to \infty$, we have $D_1 \xrightarrow{P} \sigma_y$ and $D_2 \xrightarrow{P} \sigma_x$. Therefore, we can say that $D \xrightarrow{P} \sigma_x\sigma_y$.

We make use of the property of continuous mapping theorem, which states that if $x_n \xrightarrow{P} x$, then $g(x_n) \xrightarrow{P} g(x)$ for $\{x_n\}_{n=1}^{\infty}$ a sequence of random variables and g be any function with D_g points of discontinuity.

Now, we have

$$\begin{aligned} r_n &= \frac{N}{D} \\ &\xrightarrow{P} \frac{\sigma_{xy}}{\sigma_x\sigma_y} \\ &= \rho \end{aligned}$$

Therefore, we have shown that $r_n \xrightarrow{P} \rho$ as $n \to \infty$

4. Notations having their usual significance, for $y_1, \ldots, y_n \overset{iid}{\sim} Ga(\alpha, \beta)$, a method of moment estimator of α is $\hat{\alpha} = \bar{y_n}^2/s_n^2$. Using the multivariate delta method, show that $SE(\hat{\alpha}) = \sqrt{2\alpha(\alpha+1)/n}$ for large values of n.

Solution: Method of moments gives us an estimate of α, which is $\hat{\alpha} = \bar{y_n}^2/s_n^2$. We also know that $\hat{\beta} = \bar{y_n}/s_n^2$ Let $t_{1n} = \bar{y_n}$ and $t_{n2} = s_n^2$. We have proved in the class the following results.

$$\begin{aligned} var(\bar{y_n}) &= \frac{\alpha}{\beta^2 n} \\ var(s_n^2) &= \frac{6\alpha}{\beta^4 n} + \frac{2\alpha^2}{\beta^4 n} \\ cov(\bar{y_n}, s_n^2) &= \frac{2\alpha}{\beta^3 n} \end{aligned}$$

Using the conclusion from the multivariate delta method, which is stated as follows

$$\sqrt{n}\{h(\mathbf{t_n}) - h(\theta)\} \xrightarrow{D} \text{Normal}\{\mathbf{0}, h^{'}(\theta)^T \sum h^{'}(\theta)\}$$

Here, we have

$$\hat{\alpha_n} = \bar{y_n}^2/s_n^2$$
$$h(t_{1n}, t_{2n}) = \frac{t_{1n}^2}{t_{2n}}$$

Now, we can calculate $h^{'}(\theta)$ by taking the partial derivative of $h(t_{1n}, t_{2n})$ wrt. t_{1n} and t_{2n}. This yields.

$$\frac{\partial h}{\partial t_{1n}} = \frac{2t_{1n}}{t_{2n}}$$
$$\left.\frac{\partial h}{\partial t_{1n}}\right|_{\theta} = 2\beta; \text{ Since from the result of } \hat{\beta}_n$$
$$\frac{\partial h}{\partial t_{2n}} = \frac{-t_{1n}^2}{t_{2n}^2}$$
$$\left.\frac{\partial h}{\partial t_{2n}}\right|_{\theta} = -\beta^2; \text{ Since from the result of } \hat{\beta}_n$$

Therefore $h^{'}(\theta) = (2\beta, -\beta^2)$. Using this result, we can calculate $var(\bar{y}_n^2/s_n^2)$

$$\begin{aligned}
var(\bar{y}_n/s_n^2) &= h^{'}(\theta)^T \sum h^{'}(\theta) \\
&= (2\beta, -\beta^2) \begin{pmatrix} \frac{\alpha}{\beta^2 n} & \frac{2\alpha}{\beta^3 n} \\ \frac{2\alpha}{\beta^3 n} & \frac{6\alpha}{\beta^4 n} + \frac{2\alpha^2}{\beta^4 n} \end{pmatrix} \begin{pmatrix} 2\beta \\ -\beta^2 \end{pmatrix} \\
&= \begin{pmatrix} 0 & \frac{4\alpha}{\beta^2 n} - \frac{6\alpha}{\beta^2 n} - \frac{2\alpha^2}{\beta^2 n} \end{pmatrix} \begin{pmatrix} 2\beta \\ -\beta^2 \end{pmatrix} \\
&= \frac{2\alpha(\alpha+1)}{n}
\end{aligned}$$

Since we are asked to find $SE(\hat{\alpha}_n)$, we get

$$\begin{aligned}
SE(\hat{\alpha}_n) &= SE(\bar{y}_n/s_n^2) \\
&= \sqrt{var(\bar{y}_n/s_n^2)} \\
&= \sqrt{\frac{2\alpha(\alpha+1)}{n}}
\end{aligned}$$

5. Now fix the values of α and β. For your chosen values of α and β, draw a random sample of size of $n = 50$ from a $Ga(\alpha, \beta)$ distribution. Using method of moments and assuming α and β to now be unknown, estimate α, β. Plot the histogram of the samples, superimposed with the true density and the estimated density.
Repeat the above procudure B = 50, 500 and 1000 times. Plot a histogram of $\sqrt{n}\frac{(\hat{\alpha}_n - \alpha)}{\sqrt{2\hat{\alpha}_n(\hat{\alpha}_n+1)}}$ for each of the above values of B. In each case, superimpose a Normal (0,1) distribution over the histogram. For your final output, provide a very brief description of what you did, the plots, the codes and your general comments, if any.

Solution: Using method of moments, we arrive at the estimators of $\hat{\alpha}_n$ and $\hat{\beta}_n$ as follows.

$$\hat{\alpha}_n = \frac{\bar{y}_n^2}{{s_n}^2}$$
$$\hat{\beta}_n = \frac{\bar{y}_n}{{s_n}^2}$$

Now, we generate a random sample from a Gamma distribution with shape α and rate β with population size $n = 100000$. We assume $\alpha = 4$ and $\beta = 0.6$. We first generate a sample of 50 numbers from this population. We have 3 functions in our code, namely *estimators*, *plot*, *mvd_plotting*. The function *estimator* takes in the value of α and β, along with the sample size and generates a Gamma functions. Further, it estimates the value of α as $\hat{\alpha}$ and β as $\hat{\beta}$ and returns these values. The *plot* function takes in these values, and generates the estimated density, actual density and a histogram of the samples. For large values of n, we find that the method of moments way of estimating the parameters α and β is pretty accurate. This can be understood from the plot below.

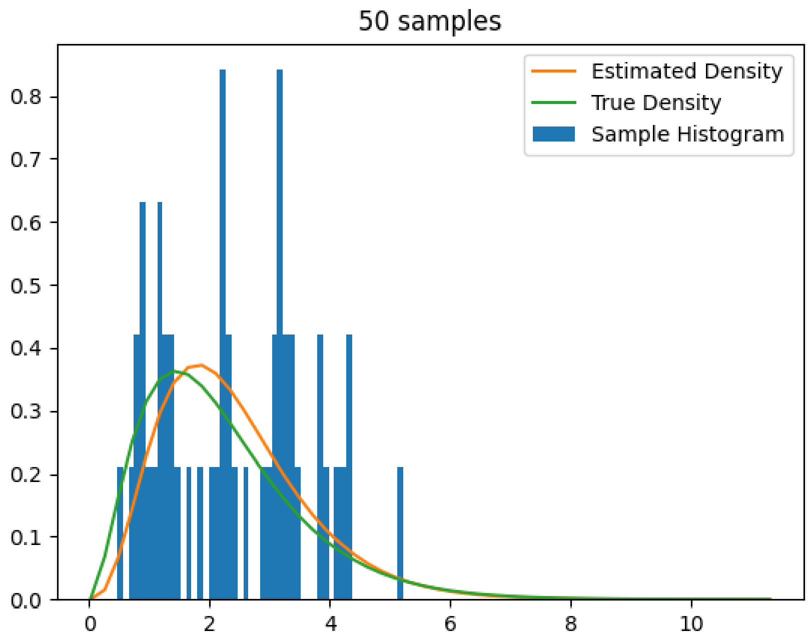


Figure 1: Estimated Density, True Density and sample histogram for a sample size of 50.

Now, we repeat the procudure of sampling from the gamma distribution and estimating α for $B = 50, 100, 500$ times. We calculate the value of $\sqrt{n}\frac{(\hat{\alpha}_n - \alpha)}{\sqrt{2\hat{\alpha}_n(\hat{\alpha}_n+1)}}$ and plot a histogram of it superimposed with a Normal(0,1) distribution.

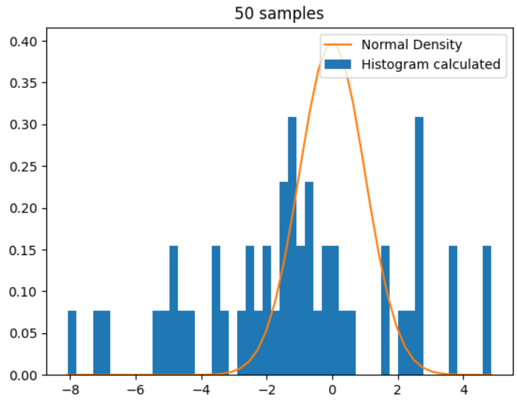


Figure 2: $B = 50$, plot of Normal(0,1) vs Histogram of calculated values.

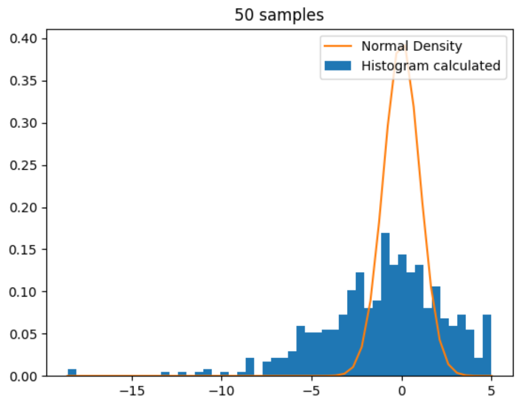


Figure 3: $B = 500$, plot of Normal(0,1) vs Histogram of calculated values.

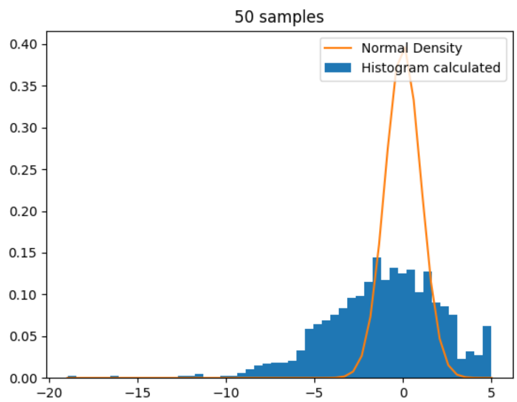


Figure 4: $B = 1000$, plot of Normal(0,1) vs Histogram of calculated values.

We get 3 results as illustrated above. We can thus conclude that for large values of sample numbers, the distribution of $\sqrt{n}\frac{(\hat{\alpha}_n - \alpha)}{\sqrt{2\hat{\alpha}_n(\hat{\alpha}_n+1)}}$ will converge to Normal(0, 1).

6. For a Normal(μ, σ^2) distribution, show that the MGF is $M(t) = exp(\mu t + \sigma^2 t^2/2)$

Solution: We know that the PMF of a Normal distribution with mean μ and standard deviation σ is given by

$$P(Y = y) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y-\mu)^2}$$

The moment generating function is given as $M(t) = \mathbb{E}(e^{ty})$. Therefore, we can write it as follows.

$$\begin{aligned}
M(t) &= \mathbb{E}(e^{ty}) \\
&= \int_{-\infty}^{\infty} e^{ty} f(y) dy \\
&= \int_{-\infty}^{\infty} e^{ty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-1}{2\sigma^2}(y-\mu)^2} dy \\
&= \frac{\sqrt{2}\sigma}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{t(\sqrt{2}z\sigma+\mu)-z^2} dz; \text{ Substituting } \frac{y-\mu}{\sqrt{2}\sigma} = z \\
&= \frac{e^{\mu t}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-(z-\frac{\sqrt{2}}{2}\sigma t)^2 + \frac{1}{2}\sigma^2 t^2} dz \\
&= \frac{e^{\mu t + \frac{1}{2}\sigma^2 t^2}}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-x^2} dx; \text{ Substituting } z - \frac{\sqrt{2}}{2}\sigma t = x \\
&= \frac{e^{(\mu t + \frac{1}{2}\sigma^2 t^2)}\sqrt{\pi}}{\sqrt{\pi}}; \text{ Since } \int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \\
&= e^{(\mu t + \frac{1}{2}\sigma^2 t^2)}
\end{aligned}$$

Therefore we have shown that for a Normal(μ, σ^2) distribution, the MGF is $M(t) = exp(\mu t + \sigma^2 t^2/2)$

7. Show that a binomial random variable R with denominator m and probability π has a cumulant generating function $K(t) = m\log(1 - \pi + \pi e^t)$. Find lim $k(t)$ as $m \to \infty, \pi \to 0$ in a way so that $m\pi \to \lambda > 0$. Show that

$$Pr(R = r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

and hence establish $R \xrightarrow{D}$ Poisson(λ). Using your favorite programming language, provide a numerical illustration of the result.

Solution: We are given that R is a binomial random variable with denominator m and probability π. Therefore, we can write the probability mass function as follows.

$$Pr(R = r) =^m C_r \pi^r (1 - \pi)^{m-r}$$

Now, we calculate the moment generating function of this probability distribution. We have,

$$
\begin{aligned}
M(t) &= \mathbb{E}(e^{tr}) \\
&= \sum_{r=0}^{m} \frac{m!}{(m-r!)r!} \pi^r (1-\pi)^{m-r}.e^{tr} \\
&= \sum_{r=0}^{m} \frac{m!}{(m-r!)r!} (\pi e^t)^r (1-\pi)^{m-r} \\
&= (1-\pi)^m + \frac{m}{1!}(1-\pi)^{m-1}(\pi e^t) + \frac{m(m-1)}{2!}(1-\pi)^{m-2}(\pi e^t)^2 + \cdots + (\pi e^t)^m \\
&= \left[\pi e^t + (1-\pi)\right]^m; \text{ By using the property of binomial expansion}
\end{aligned}
$$

Now, we have the moment generating function of a binomial distribution. Using this, we can find the cumulant generating function as follows.

$$
\begin{aligned}
K(t) &= \log M(t) \\
&= \log\left[\pi e^t + (1-\pi)\right]^m \\
&= m\log\left[\pi e^t + (1-\pi)\right]
\end{aligned}
$$

Thus, we have shown that the cumulant generating function for a binomial distribution is given by $m\log\left[\pi e^t + (1-\pi)\right]$. Now, we need to find the limit as asked.

$$\lim_{m\longrightarrow\infty,\pi\longrightarrow 0} K(t) = \lim_{m\longrightarrow\infty,\pi\longrightarrow 0} m\log\left[\pi e^t + (1-\pi)\right]$$

Let us assume $m\pi \to \lambda$. Therefore, $\pi = \frac{\lambda}{m}$. Therefore, we can write

$$
\begin{aligned}
\lim_{m\longrightarrow\infty,\pi\longrightarrow 0} K(t) &= \lim_{m\longrightarrow\infty,\pi\longrightarrow 0} m\log\left[\frac{\lambda}{m}e^t + (1-\frac{\lambda}{m})\right] \\
&= \lim_{m\longrightarrow\infty} \log\left(1 + \frac{\lambda(e^t-1)}{m}\right)^m \\
&= \log(e^{\lambda(e^t-1)}); \text{ Applying } \underset{m\longrightarrow\infty}{lim}(1+\frac{x}{n})^n = e^x \\
&= \lambda(e^t-1)
\end{aligned}
$$

Following this, we now try to establish that $R \xrightarrow{D} \text{Poisson}(\lambda)$

$$
\begin{aligned}
P(R=r) &= \frac{m!}{(m-r)!r!}\left(\frac{\lambda}{m}\right)^r\left(1-\frac{\lambda}{m}\right)^{m-r} \\
\lim_{m\longrightarrow\infty} P(R=r) &= \lim_{m\longrightarrow\infty} \frac{m!}{(m-r)!r!}\left(\frac{\lambda}{m}\right)^r\left(1-\frac{\lambda}{m}\right)^{m-r} \\
&= \frac{\lambda^r}{r!}\lim_{m\longrightarrow\infty} \frac{m!}{(m-r)!r!}\left(1-\frac{\lambda}{m}\right)^m\left(1-\frac{\lambda}{m}\right)^{-r}\left(\frac{1}{m}\right)^r
\end{aligned}
$$

We look at each part of the limit individually. First, we have

$$
\begin{aligned}
\lim_{m\longrightarrow\infty} \frac{m!}{(m-r)!r!}\left(\frac{1}{m^r}\right) &= \lim_{m\longrightarrow\infty} \frac{m(m-1)(m-2)\ldots(3)(2)(1)}{(m-r)(m-r-1)(m-r-2)\ldots(3)(2)(1)}\left(\frac{1}{m^r}\right) \\
&= \lim_{m\longrightarrow\infty} \frac{m(m-1)(m-2)\ldots(m-r+1)}{m^r}; \text{ Cancelling out denominator terms} \\
&= \lim_{m\longrightarrow\infty} \left(\frac{m}{m}\right)\left(\frac{m-1}{m}\right)\cdots\left(\frac{m-r+1}{m}\right) \\
&= 1
\end{aligned}
$$

Now for the second part, we have

$$\lim_{m \to \infty} \left(1 - \frac{\lambda}{m}\right)^m = e^{\lambda}; \text{ Here, } x = -\lambda \text{ and } \lim_{n \to \infty} \left(1 + \frac{x}{n}\right) = e^x$$

Also, we have

$$\lim_{m \to \infty} \left(1 - \frac{\lambda}{m}\right)^{-r} = 1; \text{ as } \frac{\lambda}{m} \to 0$$

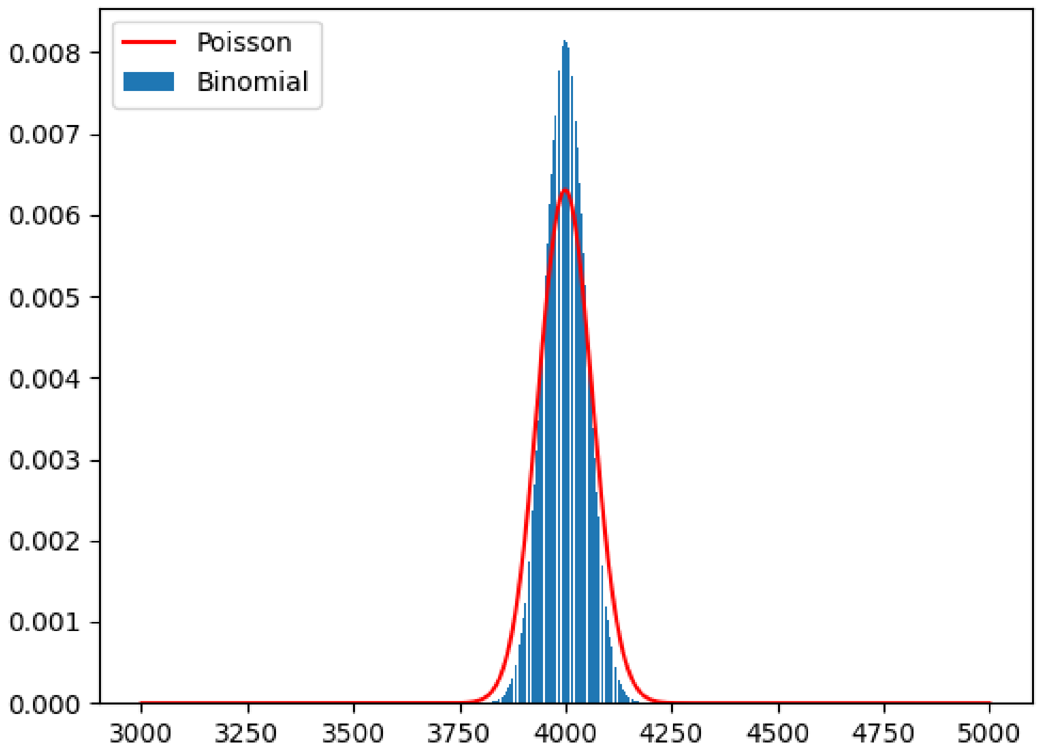
Therefore, we obtain the result,

$$\lim_{m \to \infty} \frac{m!}{(m-r)!r!} \left(\frac{\lambda}{m}\right)^r \left(1 - \frac{\lambda}{m}\right)^{m-r} = \frac{e^{-\lambda}\lambda^r}{r!}$$
$$\lim_{m \to \infty} \frac{m!}{(m-r!)r!} \pi^r (1-\pi)^{m-r} = \frac{e^{-\lambda}\lambda^r}{r!}$$

And we can say that

$$Pr(R = r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

Graphically, we arrive at the following result. For $m = 10000$ and $\pi = 0.4$, we get the following distribution



We find that as the sample size m is large, the Binomial distribution (in blue) converges to the Poisson Distribution (in red). The mean of the Poisson Distribution is $m \times \pi = 10000 \times 0.4 = 4000$. This is also in agreement with our mathematical derivation.

8. If $Z \sim \text{Normal}(0,1)$, derive the density of $Y = Z^2$. Although Y is determined by Z, show that they are uncorrelated.

Solution: It is given that Z is Normally distributed. We can write the moment generating function of Z^2 as

$$
\begin{aligned}
M_{Z^2}^t &= \mathbb{E}(tZ^2) \\
&= \int_{-\infty}^{\infty} e^{tZ^2} \frac{1}{\sqrt{2\pi}} e^{\frac{-Z^2}{2}} dZ \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{Z^2(t-\frac{1}{2})} dZ \\
&= \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\frac{1}{\sqrt{1-2t}}} e^{\frac{-Z^2}{2(\frac{1}{\sqrt{1-2t}})^2}} dZ \\
&= \frac{1}{\sqrt{1-2t}}.1; \text{ Since this is the pdf of a normal distribution with st.dev. } \frac{1}{\sqrt{1-2t}} \text{ and mean } 0 \\
&= \frac{1}{\sqrt{1-2t}}
\end{aligned}
$$

Now, we know that the moment generating function for a χ_k^2 distribution with k degrees of freedom is $(\frac{1}{\sqrt{1-2t}})^k$. By uniqueness theorem, we can say that since $M_{Z^2}^t = \frac{1}{\sqrt{1-2t}}$, this is the MGF of a χ_k^2 distribution with 1 degrees of freedom. Therefore, $Y \sim \chi_1^2$

Although we have $Y = Z^2$, we have the following relationship for the correlation coefficient.

$$
\begin{aligned}
\rho(X,Y) &= \frac{\text{Cov}(X,Y)}{\text{Var(X)}^{\frac{1}{2}}.\text{Var(Y)}^{\frac{1}{2}}} \\
\rho(Z,Z^2) &= \frac{\text{Cov}(Z,Z^2)}{\text{Var(Z)}^{\frac{1}{2}}.\text{Var}(Z^2)^{\frac{1}{2}}}
\end{aligned}
$$

We can calculate $\text{Cov}(Z,Z^2)$ as follows.

$$
\begin{aligned}
\text{Cov}(Z,Z^2) &= E(Z.Z^2) - E(Z).E(Z^2) \\
&= 0 - 0 \times 1 \\
&= 0.
\end{aligned}
$$

Therefore, we can say that although $Y = Z^2$ is determined by Z, they are uncorrelated.

9. Let $Y = X_1 + bX_2$, where X_j are independent normals with means μ_j and variances σ_j^2. Show that conditional on $X_2 = x$, the distribution of Y is normal with mean $\mu_1 + bx$ and variance σ_1^2. Hence establish that

$$
\int \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1-bx}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right)dx = \frac{1}{\sqrt{\sigma_1^2+b\sigma_2^2}}\phi\left(\frac{y-\mu_1-b\mu_2}{\sqrt{\sigma_1^2+b\sigma_2^2}}\right)
$$

Solution: It is given that $Y = X_1 + bX_2$. Let us try to find the MGF of this distribution.

$$
\begin{aligned}
M_Y(t) &= \mathbb{E}(e^{tY}) \\
&= \mathbb{E}(e^{t(X_1+bX_2))}) \\
&= \mathbb{E}(e^{tX_1}) \times \mathbb{E}(e^{tbX_2}) \\
&= M^t_{X_1} \times M^t_{bX_2} \\
&= e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \times e^{b(\mu_2 t + \frac{\sigma_2^2 t^2}{2})} \\
&= e^{(\mu_1+\mu_2 b)t + \frac{\sigma_1^2 + b\sigma_2^2}{2} t^2}
\end{aligned}
$$

From M^t_y, we can conclude that Y is a Normal Distribution with mean $\mu_1 + \mu_2 b$ and variance $\sigma_1^2 + b\sigma_2^2$. Therefore, $Y \sim \text{Normal}(\mu_1 + \mu_2 b, \sigma_1^2 + b\sigma_2^2)$

Although, it is given that we need to condition on $X_2 = x$. This implies that X_2 is no longer a distribution, but rather a single value. Therefore, $\mu_2 = x$ and $\sigma_2 = 0$. Substituting these values in the above equation, we get.

$$
\begin{aligned}
f(Y|X_2 = x) &\sim \text{Normal}(\mu_1 + \mu_2 b, \sigma_1^2 + b\sigma_2^2) \\
&\sim \text{Normal}(\mu_1 + bx, \sigma_1^2)
\end{aligned}
$$

This can also be written as

$$
f_Y(y) = \frac{1}{\sqrt{\sigma_1^2 + b^2\sigma_2^2}} \phi\left(\frac{y - \mu_1 - b\mu_2}{\sqrt{\sigma_1^2 + b^2\sigma_2^2}}\right)
$$

For the second part, let us assume $h_1(X_1, X_2) = X_1 + bX_2$, and let $h_2(X_1, X_2) = X_2$. Then, we get

$$
\begin{aligned}
|J(X_1, X_2)|^{-1} &= \frac{\partial h_1}{\partial x_1}\frac{\partial h_2}{\partial x_2} - \frac{\partial h_1}{\partial x_2}\frac{\partial h_2}{\partial x_1} \\
&= 1 \times 1 - b \times 0 \\
&= 1
\end{aligned}
$$

Using the property that random variates Y_1 and Y_2 are jointly continuous with the joint density function given by (where $x_1 = h_1(y_1, y_2)$, and $x_2 = h_2(y_1, y_2)$)

$$
f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(x_1, x_2)|J(x_1, x_2)|^{-1}
$$

Therefore, we get,

$$
\begin{aligned}
f_{Y,X_2} &= f_{X_1,X_2}|J(x_1, x_2)|^{-1} \\
&= f_{x_1,x_2} \\
&= \frac{1}{2\pi\sigma_1\sigma_2} e^{\frac{-1}{2}(\frac{x_1-\mu_1}{\sigma_1})^2} e^{\frac{-1}{2}(\frac{x_2-\mu_2}{\sigma_2})^2}
\end{aligned}
$$

Now, we find

$$
\begin{aligned}
f_{Y|X_2=x} &= \frac{f_{X_1,X_2}}{f_{X_2}} \\
&= \frac{\frac{1}{2\pi\sigma_1\sigma_2} e^{\frac{-1}{2}(\frac{x_1-\mu_1}{\sigma_1})^2} e^{\frac{-1}{2}(\frac{x_2-\mu_2}{\sigma_2})^2}}{\frac{1}{\sqrt{2\pi}\sigma_2} e^{\frac{-1}{2}(\frac{x_2-\mu_2}{\sigma_2})^2}} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} e^{\frac{-1}{2}(\frac{y-bx_2-\mu_1}{\sigma_1})^2} \\
&= \frac{1}{\sigma_1}\phi\left(\frac{y - \mu_1 - bx}{\sigma_1}\right)
\end{aligned}
$$

And,

$$f_{X_2} = \frac{1}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right)$$

By law of total probability, we get

$$\int \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1-bx}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{x-\mu_2}{\sigma_2}\right)dx = \frac{1}{\sqrt{\sigma_1^2+b\sigma_2^2}}\phi\left(\frac{y-\mu_1-b\mu_2}{\sqrt{\sigma_1^2+b\sigma_2^2}}\right)$$

10. Read pages 62-75 (Section 3.2: Normal Model) from AC Davison's Statistical Models.