ANSWER SHEET 2

Assignment 1. (a) Call g the function that maps (X,Y) onto $(U,V)=(X+Y,X-Y)$. g is a differentiable bijection whose inverse g^{-1} sends (U,V) into

$$\left(\frac{U+V}{2},\frac{U-V}{2}\right)$$

and has Jacobian

$$J=\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

The transformation theorem for random variable gives the joint density $f_{U,V}(u,v)$ as

$$\begin{aligned} f_{U,V}(u,v) &= f_{x,y}\left(\frac{u+v}{2},\frac{u-v}{2}\right)\cdot|det(J)| = \frac{1}{2}f_X\left(\frac{u+v}{2}\right)f_Y\left(\frac{u-v}{2}\right) = \\ &= \frac{1}{2}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(\frac{u+v}{2})^2\right\}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}(\frac{u-v}{2})^2\right\} = \\ &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{u^2}{2}\right\}\cdot\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\frac{v^2}{2}\right\}. \end{aligned}$$

(b) Observe that (Slide 66)

$$X+Y, X-Y\sim\mathcal{N}(0,2) \qquad \Longrightarrow \qquad \frac{X+Y}{2},\frac{X-Y}{2}\sim\mathcal{N}\left(0,\frac{1}{2}\right),$$

and in particular $f_{U,V}(u,v)=f_U(u)f_V(v)$, proving independence.

Assignment 2. (a)

$$\begin{aligned} Var(X) &= \mathbb{E}(X^2)-\mathbb{E}(X)^2=\mathbb{E}(X^2) \quad \text{hence}\mathbb{E}(X^2)=1 \\ \mathbb{E}(X^3) &= \int_{\mathbb{R}} x^3\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\mathrm{dx}=0 \\ \mathbb{E}(X^4) &= M_X^{(4)}(0)=3 \end{aligned}$$

where the fact that $\mathbb{E}(X^3)=0$ follows from antisimmetry around zero.

(b) Putting all the previous results toghether we have :

$$\begin{aligned} \mathrm{Cov}(X,X^2) &= \mathbb{E}(X\cdot X^2)-\mathbb{E}(X)\mathbb{E}(X^2)=0 \\ \mathrm{Corr}(X,X^2) &= \frac{\mathrm{Cov}(X,X^2)}{\mathrm{Var}(X)\mathrm{Var}(X^2)}=0 \end{aligned}$$

This exercise gives another example of how uncorrelation does not imply independence. (2) There is an exact relation between X and $Y=X^2$ given by the parabola. The sample correlation between the sample from X and X^2 decreases as the sample size increases (a consequence of the Law of Large Numbers).

Assignment 3. (a) We use the convention that $\binom{n}{m}=0$ if $m>n$. Then $\mathbb{P}(Y=m|X=n)=\binom{n}{m}p^m(1-p)^{n-m}$ and so $(n,m=0,1,2,\dots)$

$$\mathbb{P}(Y=m,X=n)=\mathbb{P}(Y=m|X=n)\mathbb{P}(X=n)=e^{-\lambda}\binom{n}{m}p^m(1-p)^{n-m}\frac{\lambda^n}{n!}.$$

(b) Using (a) and the law of total probability

$$\mathbb{P}(Y=m)=\sum_{n=0}^{\infty}\mathbb{P}(Y=m,X=n)=e^{-\lambda}\sum_{n=m}^{\infty}\frac{p^m(1-p)^{n-m}\lambda^n}{m!(n-m)!}=\frac{p^m\lambda^m}{m!}\sum_{k=0}^{\infty}e^{-\lambda}\frac{[(1-p)\lambda]^k}{k!}.$$

We identify the elements of a Poisson($\lambda[1-p]$) distribution in the sum :

$$\sum_{k=0}^{\infty}e^{-\lambda}\frac{[(1-p)\lambda]^k}{k!}=\sum_{k=0}^{\infty}e^{-\lambda p}e^{-\lambda[1-p]}\frac{[(1-p)\lambda]^k}{k!}=e^{-\lambda p}.$$

Thus $\mathbb{P}(Y=m)=(p\lambda)^m e^{-\lambda p}/m!$ for all m, and therefore $Y\sim Poisson(p\lambda)$.
In words : *a conditional-upon-Poisson binomial is again Poisson with a smaller parameter.*
(c) By the formulae for binomial distributions we have $\mathbb{E}[Y|X]=Xp$ and $\mathrm{Var}[Y|X]=Xp(1-p)$. Since X and Y are Poisson this gives

$$\mathbb{E}(\mathrm{Var}[Y|X])+\mathrm{Var}[\mathbb{E}(Y|X)]=\lambda p(1-p)+\lambda p^2=\lambda p=\mathrm{Var}\,Y.$$

(d) The moment generating function of $X+X'$ at $t\in\mathbb{R}$ is

$$M_X(t)M_{X'}(t)=\exp(\lambda[e^t-1])\exp(\mu[e^t-1])=\exp([\lambda+\mu][e^t-1]).$$

This is the moment generating function of a Poisson($\lambda+\mu$) random variable. (Direct calculation of $\mathbb{P}(X+X'=k)$ is also possible.)
(e) If $X+X'=k$ and $X=m$ then X' must equal $k-m$. Thus

$$\mathbb{P}(X=m|X+X'=k)=\frac{\mathbb{P}(X=m,X'=k-m)}{\mathbb{P}(X+X'=k)}=\frac{e^{-\lambda}\lambda^m}{m!}\frac{e^{-\mu}\mu^{k-m}}{(k-m)!}\Bigg/\frac{e^{-[\lambda+\mu]}(\lambda+\mu)^k}{k!}$$
$$=\binom{k}{m}\frac{\lambda^m\mu^{k-m}}{(\lambda+\mu)^k}.$$

This is reminiscent of the binomial distribution, and indeed, it equals

$$=\binom{k}{m}\frac{\lambda^m\mu^{k-m}}{(\lambda+\mu)^m(\lambda+\mu)^{k-m}}=\binom{k}{m}q^m(1-q)^{k-m},\qquad q=\frac{\lambda}{\lambda+\mu}.$$

We see that $X|X+X'=k$ is Binom($k,\lambda/(\lambda+\mu)$). In words, *a Poisson conditioned on its sum with an independent Poisson is binomial.*
(f) The black and red points are very close to each other. This means that the corresponding binomial and Poisson distributions are very similar. The approximation becomes better as n increases and worse as n decreases (try $n=7,8,9$). When $n<7$ there is an error because the success probability of the binomial is larger than one.
(g) We have for $x=\lambda(e^t-1)$

$$M_{B_n}(t)=(1-\lambda/n+\lambda e^t/n)^n=\left(1+\frac{\lambda(e^t-1)}{n}\right)^n\to\exp(\lambda[e^t-1]),\qquad n\to\infty.$$

The right-hand side is the moment generating function of a Poisson distribution function. This means that the sequence of distributions Binom($n,\lambda/n$) *converge* to the Poisson(λ) distribution as $n\to\infty$ in a sense that will be made precise later on in the course.

Assignment 4. (a) Let $X \sim \text{Geom}(p)$ and remember that

$$\sum_{i=0}^{n-1} a^i = \left(\frac{1-a^n}{1-a}\right).$$

$$\begin{aligned}\mathbb{P}(X \geq k) = 1 - \mathbb{P}(X < k) = 1 - \mathbb{P}(X \leq k-1) = 1 - \sum_{i=0}^{k-1}(1-p)^i p = \\ = 1 - p\sum_{i=0}^{k-1}(1-p)^i = 1 - p\frac{1-(1-p)^k}{p} = \\ = (1-p)^k.\end{aligned}$$

(b)

$$\begin{aligned}\mathbb{P}(X \geq k+m | X \geq k) &= \frac{\mathbb{P}(X \geq k+m, X \geq k)}{\mathbb{P}(X \geq k)} = \\ &= \frac{\mathbb{P}(X \geq k+m)}{\mathbb{P}(X \geq k)} = \frac{(1-p)^{k+m}}{(1-p)^k} = (1-p)^m = \mathbb{P}(X \geq m).\end{aligned}$$

(c) Rewrite the lack of memory property as

$$\mathbb{P}(Y \geq n+m) = \mathbb{P}(Y \geq m)\mathbb{P}(Y \geq n). \tag{1}$$

Let us prove by induction that

$$\mathbb{P}(Y \geq n) = \mathbb{P}(Y \geq 1)^n.$$

Substituting $n = 0$ into (1) we have $\mathbb{P}(Y \geq 0) = 1$, hence

$$\mathbb{P}(Y \geq n+1) = \mathbb{P}(Y \geq 1)\mathbb{P}(Y \geq n) = \mathbb{P}(Y \geq 1) \cdot \mathbb{P}(Y \geq 1)^n = \mathbb{P}(Y \geq 1)^{n+1}.$$

Now,

$$\begin{aligned}\mathbb{P}(Y = k) = \mathbb{P}(Y \geq k) - \mathbb{P}(Y \geq k+1) = \mathbb{P}(Y \geq 1)^k - \mathbb{P}(Y \geq 1)^{k+1} = \\ = \mathbb{P}(Y \geq 1)^k(1 - \mathbb{P}(Y \geq 1)) = (1-p)^k p\end{aligned}$$

where $p = 1 - \mathbb{P}(Y \geq 1)$. In particular $Y \sim \mathit{Geom}(p)$.

Assignment 5. (a) We need to find $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{\mathbf{X}}(x_1, x_2)dx_2$. In order to compute the integral, first we adjust the expression in the exponential so as to get a square form in x_2 as follows.

$$\begin{aligned}&\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) \\ = \ &(1-\rho^2)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \rho^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 \\ = \ &(1-\rho^2)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left\{\left(\frac{x_2-\mu_2}{\sigma_2}\right) - \rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right\}^2 \\ = \ &(1-\rho^2)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \frac{\{x_2 - A_1(x_1)\}^2}{\sigma_2^2}, \qquad \text{say,}\end{aligned}$$

where

$$A_1(x_1) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1).$$

Plugging-in the above expression in the integral, we get

$$\begin{aligned}
f_{\mathbf{X}}(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\, exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}\, exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left[x_2 - A_1(x_1)\right]^2\right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1}\, exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\, exp\left\{-\frac{\left[x_2 - A_1(x_1)\right]^2}{2\sigma_2^2(1-\rho^2)}\right\}. \quad (2)
\end{aligned}$$

Thus,

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1}\, exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\, exp\left\{-\frac{\left[x_2 - A_1(x_1)\right]^2}{2\sigma_2^2(1-\rho^2)}\right\}dx_2.$$

We can now identify the integrand above (as a function of x_2 for a fixed value of x_1) as the density of a Normal distribution with mean $A_1(x_1)$ and variance $\sigma_2^2(1-\rho^2)$. Thus, the value of the above integral is one. Hence,

$$f_{X_1}(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1}\, exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}, \quad x_1 \in \mathbb{R}.$$

(b) The calculations/integration in (a) can be done with respect to x_1 in the same way (by symmetry), and this results in the distribution of X_2 being

$$f_{X_2}(x_2) = \frac{1}{\sqrt{2\pi}\sigma_2}\, exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\}, \quad x_2 \in \mathbb{R}.$$

(c) The marginal density of X_i is Normal with mean μ_i and variance σ_i^2 for $i = 1, 2$.
(d) Looking at the factorization of the joint density, namely, equation (2), done when calculating the marginal density in part (a), and given that the marginal density of X_1 is the first part of the equation (2), it now follows that the conditional density of $X_2 \mid X_1 = x_1$ is given by

$$\begin{aligned}
f_{X_2|X_1=x_1}(x_2) &= \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_1}(x_1)} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\, exp\left\{-\frac{\left[x_2 - A_1(x_1)\right]^2}{2\sigma_2^2(1-\rho^2)}\right\}, \quad x_2 \in \mathbb{R}.
\end{aligned}$$

Similarly, the conditional density of $X_1 \mid X_2 = x_2$ is given by

$$\begin{aligned}
f_{X_1|X_2=x_2}(x_1) &= \frac{f_{\mathbf{X}}(x_1, x_2)}{f_{X_2}(x_2)} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}\, exp\left\{-\frac{\left[x_1 - A_2(x_2)\right]^2}{2\sigma_1^2(1-\rho^2)}\right\}, \quad x_1 \in \mathbb{R},
\end{aligned}$$

where

$$A_2(x_2) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2).$$

(e) Yes.
(f) $\mathbb{E}[X_1 \mid X_2] = A_2(X_2)$ and $\mathbb{E}[X_2 \mid X_1] = A_1(X_1)$.
(g) Note that

$$\mathbb{E}\{\mathbb{E}[X_1 \mid X_2]\} = \mathbb{E}\left\{\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)\right\} = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}\,\mathbb{E}(X_2 - \mu_2) = \mu_1 = \mathbb{E}(X_1),$$

since $\mathbb{E}(X_2 - \mu_2) = 0$. This is because the previous expectation is taken with respect to the unconditional distribution of X_2, and the mean of this unconditional distribution is μ_2.
(h) We know that $\mathrm{Cov}(X_1, X_2) = \mathbb{E}(X_1X_2) - \mathbb{E}(X_1)\,\mathbb{E}(X_2)$. Now,

$$\begin{aligned}
\mathbb{E}(X_1X_2) &= \mathbb{E}[\mathbb{E}(X_1X_2 \mid X_2)] \\
&= \mathbb{E}[X_2\,\mathbb{E}(X_1 \mid X_2)] \\
&= \mathbb{E}[X_2A_2(X_2)] \;=\; \mathbb{E}\left[X_2\left\{\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)\right\}\right] \\
&= \mu_1\,\mathbb{E}(X_2) + \rho\frac{\sigma_1}{\sigma_2}\,\mathbb{E}[X_2(X_2 - \mu_2)] \\
&= \mu_1\mu_2 + \rho\frac{\sigma_1}{\sigma_2}\left\{\mathbb{E}\left[X_2^2\right] - \mu_2\,\mathbb{E}(X_2)\right\} \\
&= \mu_1\mu_2 + \rho\frac{\sigma_1}{\sigma_2}\left\{\mathrm{Var}(X_2) + [\mathbb{E}(X_2)]^2 - \mu_2^2\right\} \\
&= \mu_1\mu_2 + \rho\frac{\sigma_1}{\sigma_2}\sigma_2^2 \;=\; \mu_1\mu_2 + \rho\sigma_1\sigma_2.
\end{aligned}$$

Thus, $\mathrm{Cov}(X_1, X_2) = \rho\sigma_1\sigma_2$.
(i) The mean vector of $\mathbf{X}$ is $\mu = (\mu_1, \mu_2)^T$, and the covariance matrix of $\mathbf{X}$ is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

(j) $\mathrm{Var}[X_1 \mid X_2] = \sigma_1^2(1-\rho^2)$ and $\mathrm{Var}[X_2 \mid X_1] = \sigma_2^2(1-\rho^2)$.
(k) Clearly, $\mathbb{E}[\mathrm{Var}[X_2 \mid X_1]] = \sigma_2^2(1-\rho^2)$. Also,

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}[X_2 \mid X_1]) &= \mathrm{Var}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X_1 - \mu_1)\right) \\
&= \rho^2\frac{\sigma_2^2}{\sigma_1^2}\mathrm{Var}(X_1 - \mu_1) \\
&= \rho^2\frac{\sigma_2^2}{\sigma_1^2}\sigma_1^2 = \rho^2\sigma_2^2.
\end{aligned}$$

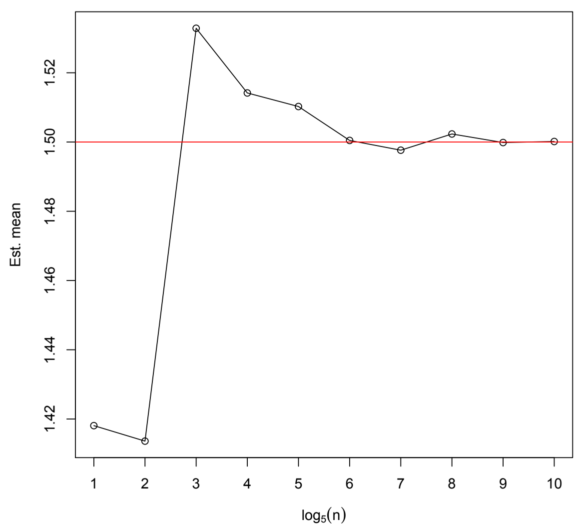
Thus, $\mathbb{E}[\mathrm{Var}[X_2 \mid X_1]] + \mathrm{Var}(\mathbb{E}[X_2 \mid X_1]) = \sigma_2^2(1-\rho^2) + \rho^2\sigma_2^2 = \sigma_2^2 = \mathrm{Var}(X_2)$.

Assignment 6. (a) The value of mean_est = 1.987.
(b) The value of mean_out = 1.999907. When M is set to 100, the value of mean_out changes very slightly, and the new value is 2. Since the code is computing an approximation of the expected value of a Poisson(2) distribution (since one cannot in practice compute an infinite sum), the increase in the value of M implies that the approximation is better. In fact, the values of `j*dpois(j,2)` are negligible for $j \geq 100$ so that the sum upto the first 100 terms gives the true expected value.

(c) The value of mean_out1 = 1.999998. It is very close to mean_out and mean_est.
(d) Since $\mathbb{P}[Y > k] = \sum_{j=k+1}^{\infty} \mathbb{P}[Y = j]$, we have

$$\begin{aligned}
\sum_{k=0}^{\infty} \mathbb{P}[Y > k] &= \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \mathbb{P}[Y = j] \\
&= \sum_{j=1}^{\infty} \sum_{k=0}^{j-1} \mathbb{P}[Y = j] \\
&= \sum_{j=1}^{\infty} j\mathbb{P}[Y = j] = \sum_{j=0}^{\infty} j\mathbb{P}[Y = j] = \mathbb{E}[Y].
\end{aligned}$$

(f) Yes. The value of mean_out = 1.504216. The expected value of a random variable having a Gamma(3,2) distribution = 3/2 = 1.5.
(g) For smaller values of the sample size $n = 5^j$, the difference between the true expected value (namely, 1.5) and the value of mean_outs[j] is greater compared to that for larger values of the sample size. In fact, for $n = 5^{10}$, the two values are almost the same. This indicates that the sample mean is a good estimator of the true expected value and becomes closer to it as the sample size grows.



(i) The value of mean_new = 1.5 and the absolute error in computation is $< 4.8 \times 10^{-5}$. This value is exactly equal to the true expected value modulo the absolute error. Since the code computes the integral of the survival function (namely, 1 - c.d.f.) over $(0, \infty)$, this indicates that the expected value of the Gamma(3,2) distribution can also be computed in this way.