Problem 1

Let (Z, W) be Bivariate Normal, defined as

$$\begin{aligned} Z &= X, \\ W &= \rho X + \sqrt{1-\rho^2}Y, \end{aligned}$$

with X, Y i.i.d. $\mathsf{N}(0,1)$, and $-1 < \rho < 1$. Find $\mathsf{E}(W \mid Z)$ and $\mathsf{Var}(W \mid Z)$.

Solution: We have $W = \rho Z + \sqrt{1-\rho^2}Y$. By the linearity of conditional expectation, we have:

$$\mathsf{E}(W \mid Z) = \mathsf{E}(\rho Z \mid Z) + \sqrt{1-\rho^2}\mathsf{E}(Y \mid Z).$$

Since ρZ is a function of Z, $\mathsf{E}(\rho Z \mid Z) = \rho Z$. And, since Y and Z are independent, $\mathsf{E}(Y \mid Z) = \mathsf{E}(Y) = 0$. Thus:

$$\mathsf{E}(W \mid Z) = \rho Z.$$

We also write:

$$\begin{aligned} \mathsf{E}\left(W^2 \mid Z\right) &= \mathsf{E}\left(\rho^2 Z^2 + \left(1-\rho^2\right)Y^2 + 2\rho\sqrt{1-\rho^2}ZY \mid Z\right), \\ &= \rho^2\mathsf{E}\left(Z^2 \mid Z\right) + \left(1-\rho^2\right)\mathsf{E}\left(Y^2 \mid Z\right) + 2\rho\sqrt{1-\rho^2}\mathsf{E}\left(YZ \mid Z\right). \end{aligned}$$

We have:

$$\begin{aligned} \mathsf{E}\left(Z^2 \mid Z\right) &= Z^2, \\ \mathsf{E}\left(Y^2 \mid Z\right) &= \mathsf{E}\left(Y^2\right) = 1, \\ \mathsf{E}\left(YZ \mid Z\right) &= Z\mathsf{E}\left(Y \mid Z\right) = Z\mathsf{E}(Y) = 0. \end{aligned}$$

Therefore:

$$\mathsf{E}\left(W^2 \mid Z\right) = \rho^2 Z^2 + 1 - \rho^2.$$

From the definition of conditional variance we obtain:

$$\mathsf{Var}\left(W \mid Z\right) = \mathsf{E}\left(W^2 \mid Z\right) - \left(\mathsf{E}\left(W \mid Z\right)\right)^2 = 1 - \rho^2.$$

Problem 2

Let $\mathbf{X} = (X_1, X_2, X_3, X_4, X_5) \sim \mathsf{Mult}_5(n, \mathbf{p})$ with $\mathbf{p} = (p_1, p_2, p_3, p_4, p_5)$.

(a) Find $\mathsf{E}(X_1 \mid X_2)$ and $\mathsf{Var}(X_1 \mid X_2)$.

(b) Find $\mathsf{E}(X_1 \mid X_2 + X_3)$.

Solution: (a) From Theorem 7.4.5 (Multinomial conditioning), we know that:

$$(X_1, X_3, X_4, X_5) \mid X_2 \sim \mathsf{Mult}_4\left(n - X_2, \left(p_1', p_3', p_4', p_5'\right)\right),$$

where

$$p'_j = \frac{p_j}{p_1 + p_3 + p_4 + p_5}, \quad j = 1, 3, 4, 5.$$

From Theorem 7.4.3 (Multinomial marginals), we obtain the conditional distribution of X_1 given X_2:

$$X_1 \mid X_2 \sim \mathsf{Bin}\left(n - X_2, p'_1\right).$$

Thus

$$\begin{aligned}
\mathsf{E}(X_1 \mid X_2) &= (n - X_2)\, p'_1 = (n - X_2)\, \frac{p_1}{p_1 + p_3 + p_4 + p_5}, \\
\mathsf{Var}(X_1 \mid X_2) &= (n - X_2)\, p'_1 \left(1 - p'_1\right) = (n - X_2)\, \frac{p_1\,(p_3 + p_4 + p_5)}{(p_1 + p_3 + p_4 + p_5)^2}.
\end{aligned}$$

(b) From Theorem 7.4.4 (Multinomial lumping), we obtain:

$$(X_1, X_2 + X_3, X_4, X_5) \sim \mathsf{Mult}_4\,(n, (p_1, p_2 + p_3, p_4, p_5)).$$

From Theorem 7.4.5 (Multinomial conditioning), we get:

$$(X_1, X_4, X_5) \mid X_2 + X_3 \sim \mathsf{Mult}_3\left(n - (X_2 + X_3), \left(p''_1, p''_4, p''_5\right)\right),$$

where

$$p''_j = \frac{p_j}{p_1 + p_4 + p_5}, \quad j = 1, 4, 5.$$

From Theorem 7.4.3 (Multinomial marginals), it follows that:

$$X_1 \mid X_2 + X_3 \sim \mathsf{Bin}\left(n - (X_2 + X_3), p''_1\right).$$

Therefore

$$\mathsf{E}(X_1 \mid X_2 + X_3) = (n - (X_2 + X_3))\, p''_1 = (n - (X_2 + X_3))\, \frac{p_1}{p_1 + p_4 + p_5}.$$

Problem 3

Show that the following version of LOTP follows from Adam's law: for any event A and continuous random variable X with PDF f_X:

$$\mathsf{P}(A) = \int_{-\infty}^{\infty} \mathsf{P}(A \mid X = x) f_X(x)\, \mathrm{d}x.$$

Solution: We use the fundamental bridge:

$$\mathsf{P}(A) = \mathsf{E}\left(\mathsf{I}_A\right) = \mathsf{E}\left(\mathsf{E}\left(\mathsf{I}_A \mid X\right)\right) = \mathsf{E}\left(\mathsf{P}\left(A \mid X\right)\right) = \int_{-\infty}^{\infty} \mathsf{P}(A \mid X = x) f_X(x)\, \mathrm{d}x.$$

Problem 4

Let $N \sim \mathsf{Pois}(\lambda_1)$ be the number of movies that will be released next year. Suppose that for each movie the number of tickets sold is $\mathsf{Pois}(\lambda_2)$, independently.

(a) Find the mean and the variance of the number of movie tickets that will be sold next year.

(b) Use simulations in R (the statistical programming language) to numerically estimate mean and the variance of the number of movie tickets that will be sold next year assuming that the mean number of movies released each year in the US is 700, and that, on average, 800000 tickets were sold for each movie.

Solution: (a) Let X_j be the number of tickets sold for the jth movie released next year. The number of movie tickets that will be sold next year is $X_1 + \ldots + X_N$ where $X_1, \ldots, X_n$ are i.i.d. $\mathsf{Pois}(\lambda_2)$. From Theorem 4.8.1 (Sum of independent Poissons), we have

$$X_1 + \ldots + X_N \mid N \sim \mathsf{Pois}(N\lambda_2).$$

Thus

$$\mathsf{E}(X_1 + \ldots + X_N \mid N) = \mathsf{Var}(X_1 + \ldots + X_N \mid N) = N\lambda_2.$$

From Adam's law we obtain:

$$\mathsf{E}(X_1 + \ldots + X_N) = \mathsf{E}(\mathsf{E}(X_1 + \ldots + X_N \mid N)) = \mathsf{E}(N\lambda_2) = \mathsf{E}(N)\lambda_2 = \lambda_1\lambda_2.$$

From Eve's law we have:

$$\begin{aligned}\mathsf{Var}(X_1 + \ldots + X_N) &= \mathsf{E}(\mathsf{Var}(X_1 + \ldots + X_N \mid N)) + \mathsf{Var}(\mathsf{E}(X_1 + \ldots + X_N \mid N)), \\ &= \mathsf{E}(N\lambda_2) + \mathsf{Var}(N\lambda_2), \\ &= \lambda_1\lambda_2(1 + \lambda_2).\end{aligned}$$

(b) The code implementing the simulations is given in Listing 1. The output from this code is shown in Listing 2. The code prints out the ratio between the Monte Carlo estimates of the mean and the variance of the number of tickets sold and their theoretical values for $\lambda_1 = 700$ and $\lambda_2 = 800000$. Both ratios are very close to 1 indicating a very good approximation of the theoretical mean and variance with the Monte Carlo estimates.

```r
#set the seed
set.seed(0)

#average number of movies released
lambda1 = 700
#average number of tickets sold per movie
lambda2 = 800000

#number of simulations
nSimulations = 100000
```

```
ticketsSold = numeric(nSimulations)

for(i in seq_len(length(ticketsSold)))
{
  #sample the number of movies released next year
  N = rpois(n=1,lambda=lambda1)
  ticketsSold[i] = rpois(n=1,lambda=N*lambda2)
}

#Monte Carlo estimate of the mean number of tickets sold next year
mean(ticketsSold)
#ratio between the Monte Carlo estimate of the mean and the theoretical mean
cat("mean ratio = ",mean(ticketsSold)/(lambda1*lambda2),"\n")
#Monte Carlo estimate of the variance of the number of tickets sold next year
var(ticketsSold)
#ratio between the Monte Carlo estimate of the variance and the theoretical variance
cat("variance ratio = ",var(ticketsSold)/(lambda1*lambda2*(1+lambda2)))
```

Listing 1: Code implementing the simulations for Problem 4 part (b)

```
> mean(ticketsSold)
[1] 559896665
> #ratio between the Monte Carlo estimate of the mean and the theoretical mean
> cat("mean ratio = ",mean(ticketsSold)/(lambda1*lambda2),"\n")
mean ratio =  0.9998155
> #Monte Carlo estimate of the variance of the number of tickets sold next year
> var(ticketsSold)
[1] 4.478495e+14
> #ratio between the Monte Carlo estimate of the variance and the theoretical variance
> cat("variance ratio = ",var(ticketsSold)/(lambda1*lambda2*(1+lambda2)))
variance ratio =  0.9996629
```

Listing 2: Output from the code implementing the simulations for Problem 4 part (b)

Problem 5

Show that if $\mathsf{E}(Y \mid X) = c$ is a constant, then X and Y are uncorrelated. Hint: Use Adam's law to find $\mathsf{E}(Y)$ and $\mathsf{E}(XY)$.

Solution: From Adam's law we obtain:

$$
\begin{aligned}
\mathsf{E}(Y) &= \mathsf{E}\left(\mathsf{E}(Y \mid X)\right) = c, \\
\mathsf{E}(XY) &= \mathsf{E}\left(\mathsf{E}(XY \mid X)\right) = \mathsf{E}\left(X\mathsf{E}(Y \mid X)\right) = c\mathsf{E}(X).
\end{aligned}
$$

Thus:

$$\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y) = c\mathsf{E}(X) - c\mathsf{E}(X) = 0.$$

Therefore X and Y are uncorrelated.

Problem 6

Show that for any random variables X and Y,

$$\mathsf{E}(Y \mid \mathsf{E}(Y \mid X)) = \mathsf{E}(Y \mid X).$$

Hint: use Adam's law with extra conditioning.

Solution: We apply Theorem 9.3.8 (Adam's law with extra conditioning) with $Z = \mathsf{E}(Y \mid X)$, and obtain:

$$\mathsf{E}(Y \mid \mathsf{E}(Y \mid X)) = \mathsf{E}(\mathsf{E}(Y \mid X, \mathsf{E}(Y \mid X)) \mid \mathsf{E}(Y \mid X)).$$

From Definition 9.2.1 (Conditional expectation given a random variable), there exists a function $g(\cdot)$ such that $\mathsf{E}(Y \mid X) = g(X)$. Since conditioning on X and $g(X)$ is the same as conditioning on X, it follows that:

$$\mathsf{E}(Y \mid X, \mathsf{E}(Y \mid X)) = \mathsf{E}(Y \mid X).$$

Therefore:

$$\mathsf{E}(Y \mid \mathsf{E}(Y \mid X)) = \mathsf{E}(\mathsf{E}(Y \mid X) \mid \mathsf{E}(Y \mid X)) = \mathsf{E}(Y \mid X).$$

Problem 6

Let Y denote the number of heads in n flips of a coin, whose probability of heads is θ.

(a) Suppose θ follows a distribution $\mathsf{P}(\theta) = \mathsf{Beta}(a, b)$, and then you observe y heads out of n flips. Show algebraically that the mean $\mathsf{E}(\theta \mid Y = y)$ always lies between the mean $\mathsf{E}(\theta)$ and the observed relative frequency of heads:

$$\min\left\{\mathsf{E}(\theta), \frac{y}{n}\right\} \leq \mathsf{E}(\theta \mid Y = y) \leq \max\left\{\mathsf{E}(\theta), \frac{y}{n}\right\}.$$

Here $\mathsf{E}(\theta \mid Y = y)$ is the mean of the distribution $\mathsf{P}(\theta \mid Y = y)$, and $\mathsf{E}(\theta)$ is the mean of the distribution $\mathsf{P}(\theta) = \mathsf{Beta}(a, b)$.

(b) Show that, if θ follows a uniform distribution,

$$\mathsf{P}(\theta) = \mathsf{Unif}(0, 1),$$

we have

$$\mathsf{Var}(\theta \mid Y = y) \leq \mathsf{Var}(\theta).$$

Here $\mathsf{Var}(\theta \mid Y = y)$ is the variance of the distribution $\mathsf{P}(\theta \mid Y = y)$, and $\mathsf{Var}(\theta)$ is the variance of the distribution $\mathsf{P}(\theta) = \mathsf{Unif}(0, 1)$.

Solution: The sampling distribution for the data is $Y = y \sim \mathsf{Bin}(n, \theta)$. We use the Bayes' rule to obtain $\mathsf{P}(\theta \mid Y = y)$:

$$\begin{aligned} \mathsf{P}(\theta \mid Y = y) &\propto \mathsf{P}(Y = y \mid \theta)\mathsf{P}(\theta), \\ &\propto \theta^{a+y-1}(1-\theta)^{b+n-y-1}. \end{aligned}$$

This is the kernel of a Beta distribution, hence (see your textbook Story 8.3.3):

$$\mathsf{P}(\theta \mid Y = y) = \mathsf{Beta}(a + y, b + n - y).$$

It follows that posterior mean of θ is

$$\mathsf{E}(\theta \mid y) = \frac{a+y}{a+b+n} = \frac{a+b}{a+b+n}\frac{a}{a+b} + \frac{n}{a+b+n}\frac{y}{n}.$$

We let $c_1 = \frac{a+b}{a+b+n}$ and $c_2 = \frac{n}{a+b+n}$. Note that $c_1 + c_2 = 1$ and $c_1, c_2 \geq 0$. Moreover, the prior mean of θ is

$$\mathsf{E}(\theta) = \frac{a}{a+b}.$$

Thus:

$$\min\left\{\frac{a}{a+b}, \frac{y}{n}\right\} \leq \mathsf{E}(\theta \mid y) \leq \max\left\{\frac{a}{a+b}, \frac{y}{n}\right\}.$$

(b) We know from lecture notes that, if $\theta \sim \mathsf{Beta}(a, b)$, the mean and variance of θ are:

$$\mathsf{E}(\theta) = \frac{a}{a+b}, \mathsf{Var}(\theta) = \frac{\frac{a}{a+b} \cdot \frac{b}{a+b}}{a+b+1}.$$

These are the formulas we need to solve this question. A uniform prior distribution for θ is equivalent with $\theta \sim \mathsf{Beta}(a, b)$ such that $a = b = 1$. It follows that the prior mean of θ is $\mathsf{E}(\theta) = \frac{1}{2}$ and the prior variance for θ is $\mathsf{Var}(\theta) = \frac{1}{12}$. Furthermore, since the posterior for θ is $\mathsf{P}(\theta \mid Y = y) = \mathsf{Beta}(a + y, b + n - y)$, it follows that the posterior variance is

$$\mathsf{Var}(\theta \mid y) = \frac{\frac{y+1}{n+2} \cdot \frac{n-y+1}{n+2}}{n+3} = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}.$$

We must prove that

$$\mathsf{Var}(\theta \mid y) \leq \frac{1}{12}$$

for any $y \in \{0, 1, \ldots, n\}$. This is equivalent with

$$g(y) = (y+1)(n-y+1) \leq \frac{(n+2)^2(n+3)}{12}.$$

That is, the function $g(y)$, $y \in \{0, 1, \ldots, n\}$, has a maximum that is smaller than $\frac{(n+2)^2(n+3)}{12}$. We take the first and second derivatives of $g(y)$:

$$\frac{d}{dy}g(y) = n - 2y, \quad \frac{d^2}{dy^2}g(y) = -2$$

Since the second derivative is always negative, the function $g(y)$ is strictly concave and has a unique maximum. The value of y for which this maximum value is attained is obtained by solving the equation

$$\frac{d}{dy}g(y) = 0,$$

which gives $y_{max} = \frac{n}{2}$. Therefore $g(y) \leq g(y_{max}) = \left(\frac{n}{2}+1\right)^2$. It follows that, in order to show that $\mathsf{Var}(\theta \mid y) \leq \frac{1}{12}$, we must prove that

$$\left(\frac{n}{2}+1\right)^2 \leq \frac{(n+2)^2(n+3)}{12}$$

But this inequality is equivalent with

$$(n+2)^2 \leq (n+2)^2\frac{n+3}{3}$$

which holds because $n > 0$ (the data must contain at least one sample).

Problem 7

Let A, B and C be independent random variables with the following distributions:

$$\mathsf{P}(A=1) = 0.4,\ \mathsf{P}(A=2) = 0.6$$
$$\mathsf{P}(B=-3) = 0.25,\ \mathsf{P}(B=-2) = 0.25,\ \mathsf{P}(B=-1) = 0.25,\ \mathsf{P}(B=1) = 0.25$$
$$\mathsf{P}(C=1) = 0.5,\ \mathsf{P}(C=2) = 0.4,\ \mathsf{P}(C=3) = 0.1$$

(a) What is the probability that the quadratic equation

$$Ax^2 + Bx + C = 0$$

has two real roots that are different?

(b) What is the probability that the quadratic equation

$$Ax^2 + Bx + C = 0$$

has two real roots that are both strictly positive?

Solution: The roots of the equation

$$Ax^2 + Bx + C = 0$$

are:

$$X_1 = \frac{-B-\sqrt{\Delta}}{2A},\ X_2 = \frac{-B+\sqrt{\Delta}}{2A},$$

where

$$\Delta = B^2 - 4AC.$$

Since A, B and C are discrete random variables, X_1 and X_2 are also discrete random variables. Moreover, since A, B and C are independent, their joint PMF is

$$\mathsf{P}(A = a, B = b, C = c) = \mathsf{P}(A = a)\mathsf{P}(B = b)\mathsf{P}(C = c),$$

for $a \in \{1, 2\}$, $b \in \{-3, -2, -1, 1\}$, $c \in \{1, 2\}$.

(a) X_1 and X_2 are real if $\Delta \geq 0$, and they are different if $\Delta \neq 0$. Thus we need to find the probability:

$$\mathsf{P}(\Delta > 0) = \mathsf{P}(B^2 > 4AC) = \mathsf{E}\left[\mathsf{I}_{\{B^2>4AC\}}\right],$$

where in the last equality we used the fundamental bridge. As such, we need to calculate:

$$\mathsf{E}\left[\mathsf{I}_{\{B^2>4AC\}}\right] = \sum_{a\in\{1,2\}} \sum_{b\in\{-3,-2,-1,1\}} \sum_{c\in\{1,2\}} \mathsf{I}_{\{b^2>4ac\}}\mathsf{P}(A = a)\mathsf{P}(B = b)\mathsf{P}(C = c).$$

```
aVal = c(1,2)
aProb = c(0.4,0.6)

bVal = c(-3,-2,-1,1)
bProb = c(0.25,0.25,0.25,0.25)

cVal = c(1,2,3)
cProb = c(0.5,0.4,0.1)

prob8A = 0

for(i in seq_len(length(aVal)))
{
  for(j in seq_len(length(bVal)))
  {
    for(k in seq_len(length(cVal)))
    {
      if(bVal[j]*bVal[j] > 4*aVal[i]*cVal[k])
      {
        prob8A = prob8A + aProb[i]*bProb[j]*cProb[k]
      }
    }
  }
}

prob8A
```

Listing 3: Code for Problem 7 (a)

Based on this code, we obtain that $\mathsf{P}(\Delta > 0) = 0.165$.

(b) X_1 and X_2 are real if $\Delta \geq 0$. Since $X_1 \leq X_2$, the roots are both strictly positive when $X_1 > 0$. Thus we need to find the probability:

$$\mathsf{P}(\{\Delta \geq 0\} \cap \{X_1 > 0\}) = \mathsf{P}(\{B^2 \geq 4AC\} \cap \{-B - \sqrt{\Delta} > 0\}) = \mathsf{E}\left[\mathsf{I}_{\{B^2 \geq 4AC\}} \mathsf{I}_{\{B + \sqrt{\Delta} < 0\}}\right],$$

We need to calculate:

$$\mathsf{E}\left[\mathsf{I}_{\{B^2 \geq 4AC\}} \mathsf{I}_{\{B + \sqrt{\Delta} < 0\}}\right] = \sum_{a \in \{1,2\}} \sum_{b \in \{-3,-2,-1,1\}} \sum_{c \in \{1,2\}} \mathsf{I}_{\{b^2 \geq 4ac\}} \mathsf{I}_{\{b + \sqrt{b^2 - 4ac} < 0\}} \mathsf{P}(A = a)\mathsf{P}(B = b)\mathsf{P}(C = c).$$

```r
aVal = c(1,2)
aProb = c(0.4,0.6)

bVal = c(-3,-2,-1,1)
bProb = c(0.25,0.25,0.25,0.25)

cVal = c(1,2,3)
cProb = c(0.5,0.4,0.1)

prob8B = 0

for(i in seq_len(length(aVal)))
{
  for(j in seq_len(length(bVal)))
  {
    for(k in seq_len(length(cVal)))
    {
      if(bVal[j]*bVal[j] >= 4*aVal[i]*cVal[k])
      {
        if(bVal[j]+sqrt(bVal[j]*bVal[j] - 4*aVal[i]*cVal[k])<0)
        {
          prob8B= prob8B + aProb[i]*bProb[j]*cProb[k]
        }
      }
    }
  }
}

prob8B
```

Listing 4: Code for Problem 7 (b)

Based on this code, we obtain $\mathsf{P}(\{\Delta \geq 0\} \cap \{X_1 > 0\}) = 0.215$.