ECS 256 - Problem set 2

Olga Prilepova, Christopher Patton, Alexander Rumbaugh, John Chen, Thomas Provan

Problem 1.a

A coin is flipped k times with p probability of heads. For each head, the coin is flipped one additional time (a bonus flip). The number of bonus flips is referred to as Y and the total number of heads X

Var(X) can be found using the Law of Total Variance, and properties of binomial distributions. We will also need to use part of the derivation of EX:

$$\begin{aligned}
E(X|Y) &= E(X - Y + Y|Y) \\
&= E((X - Y)|Y) + E(Y|Y) && \text{(by 3.13)} \\
&= pY + Y && \text{(by 3.110)} \\
&= (1 + p)Y
\end{aligned}$$

$$\begin{aligned}
Var(X) &= E[Var(X|Y)] + Var[E(X|Y)] && \text{(by 9.8)} \\
&= E[Var(X|Y)] + Var[(1 + p)Y] && \text{(from above)} \\
&= E[Var(X|Y)] + (1 + p)^2 kp(1 - p) && \text{(by 3.34 and 3.109)} \\
&= E[Yp(l - p)] + (1 + p)^2 kp(1 - p) && \text{(by 3.111)} \\
&= kp^2(1 - p) + (1 + p)^2 kp(1 - p) && \text{(by 3.103)} \\
&= kp(1 - p)\left(p + (1 + p)^2\right)
\end{aligned}$$

Using p=0.5

$$\begin{aligned}
&= k(0.25)(0.5 + (1.5^2)) \\
&= 0.6875k
\end{aligned}$$

Problem 1.b

In the trapped miner example, a miner chooses between three doors with only one leading to safety after 2 hours. The other two doors lead back to the door room after 3 and 5 hours respectively.

We are interesting the variance of Y, the time it takes to escape the mine. We will build upon Ahmed Ahmedin's solution to EY, where N refers to the total attempts needed to escape and U_i refers to the time spent traveling on the i^{th} attempt.

$$\begin{aligned}
Var(Y) &= E[Var(Y|N)] + Var[E(Y|N)] && \text{(by 9.8)} \\
&= E[Var(Y|N)] + Var[4N - 2] && \text{(by 9.16)} \\
&= E[Var(Y|N)] + 16Var[N] && \text{(by 3.34 and 3.41)} \\
&= E[Var(Y|N)] + 16 \cdot \frac{1 - 1/3}{(1/3)^2} && \text{(by 3.93)} \\
&= E[Var(U_1 + U_2 + ... + U_n|N)] + 96 \\
&= E[Var(U_1|N) + ... + Var(U_{N-1}|N + Var(U_N|N)] + 96 && \text{(by 3.51)} \\
&= E[1 + 1 + ...1 + 0] + 96 \\
&= E[N - 1] + 96 \\
&= E[N] - 1 + 96 && \text{(by 3.17)} \\
&= 3 - 1 + 96 && \text{(by 3.92)} \\
&= 98
\end{aligned}$$

We know that Var(U_i—N) is independent because the miner's choice of door does not depend of a previous choice. Since we are conditioning this event on there being N attempts, the values of the first N-1 attempts will either be 3 or 5. So the variance of an individual attempt in this case is 1. The variance of the N^{th} attempt is 0 because that attempt always is the same tunnel.

Problem 2.a

For a vector Q of random variables $(Q_1, ..Q_n)$ we have:

$$Cov(Q) = E(QQ') - E(Q)E(Q') \quad \text{(by 13.53)}$$

Let $Q = Y|X$, where Y is vector valued. Then:

$$Cov(Y|X) = E\big((Y|X)(Y|X)'\big) - E(Y|X)E(Y|X)' \quad \text{(by 13.53)}$$

Taking expected value of both sides we have:

$$\begin{aligned} E\big(Cov(Y|X)\big) &= E\Big(E\big((Y|X)(Y|X)'\big) - E(Y|X)E(Y|X)'\Big) \\ &= E\Big(E\big((Y|X)(Y|X)'\big)\Big) - E\Big(E(Y|X)E(Y|X)'\Big) \\ &= E(YY') - E\Big(E(Y|X)E(Y|X)'\Big) && \text{(by Law of Tot. Expect.)} \end{aligned}$$

Now let $Q = E(Y|X)$, where Y is vector valued. Then:

$$\begin{aligned} Cov(E(Y|X)) &= E\big(E(Y|X)E(Y|X)'\big) - E\big(E(Y|X)\big)E\big(E(Y|X)\big)' && \text{(by 13.53)} \\ &= E\big(E(Y|X)E(Y|X)'\big) - E(Y)E(Y)' && \text{(by Law of Tot. Expect.)} \end{aligned}$$

Summing up the left sides and the right sides of these 2 equations we get:

$$\begin{aligned} E\big(Cov(Y|X)\big) + Cov(E(Y|X)) &= E(YY') - E\Big(E(Y|X)E(Y|X)'\Big) \\ &\quad + E\big(E(Y|X)E(Y|X)'\big) - E(Y)E(Y)' \\ E\big(Cov(Y|X)\big) + Cov(E(Y|X)) &= E(YY') - E(Y)E(Y)' \\ E\big(Cov(Y|X)\big) + Cov(E(Y|X)) &= Cov(Y) && \text{(by 13.53)} \end{aligned}$$

Problem 2.b

First, just an equation to remind us of what we're actually trying to find here, the correlation between X and Y.

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

From the problem **2.a**, we have the following.

$$
\begin{aligned}
Cov\big((X,Y)'\big) &= \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix} \\
&= \begin{pmatrix} E\big(Var(X|Y)\big) & E\big(Cov(X,Y|Y)\big) \\ E\big(Cov(X,Y|Y)\big) & E\big(Var(Y|Y)\big) \end{pmatrix} \\
&+ \begin{pmatrix} Var\big(EX|Y)\big) & Var\big(EX,EY|Y)\big) \\ Var\big(EX,EY|Y)\big) & Var\big(EY|Y)\big) \end{pmatrix}
\end{aligned}
$$

And since the summation of matrices is by element we can just focus on the following formula:

$$
Cov(X,Y) = E\big[Cov(X,Y|Y)\big] + Cov\big[E(X,Y|Y)\big]
$$

Let $B = X - Y$ and use that random variable in the following computations.

$$
Cov(B,Y) = E\big[Cov(B,Y|Y)\big] + Cov\big[E(B,Y|Y)\big]
$$

$$
\begin{aligned}
Var(B+Y) &= Var(B) + Var(Y) + 2Cov(B,Y) \\
Cov(B,Y) &= \big(Var(B+Y) - Var(B) - Var(Y)\big)/2
\end{aligned}
$$

$$
\begin{aligned}
Cov\big(B,Y|Y\big) &= \big(Var(B+Y|Y) - Var(B|Y) - Var(Y|Y)\big)/2 \\
Var(B+Y|Y) &= Var(B|Y) \\
Var(Y|Y) &= 0 \\
Cov\big(B,Y|Y\big) &= 0
\end{aligned}
$$

$$
\begin{aligned}
Cov\big(E(B,Y|Y)\big) &= \big(Var(EB+EY|Y) - Var(EB|Y) - Var(EY|Y)\big)/2 \\
&= \big(Var(1.5Y) - Var(0.5Y) - Var(Y)\big)/2 \\
&= \big(2.25Var(Y) - 0.25Var(Y) - Var(Y)\big)/2 \\
&= Var(Y)/2
\end{aligned}
$$

Hence:

$$
\begin{aligned}
Cov\big(B,Y\big) &= 0 + Var(Y) \\
&= Var(Y)/2 \\
&= kp(1-p)/2 \\
&= 0.25k/2 \\
&= 0.125k
\end{aligned}
$$

Now, let's see how we can get $Cov\big(X,Y\big)$ using that fact that we now know $Cov\big(B,Y\big)$. Remember, $X = B + Y$

$$
\begin{aligned}
Cov\big(X,Y\big) &= Cov\big(B+Y,Y\big) \\
Cov\big(B+Y,Y\big) &= Cov\big(B,Y\big) + Cov\big(Y,Y\big) \quad (13.2) \\
&= Cov\big(B,Y\big) + Var(Y)
\end{aligned}
$$

Now we have all the ingredients to find the correlation:

$$
\begin{aligned}
\rho(X,Y) &= \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \\
&= \frac{Cov\big(B,Y\big) + Var(Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \\
&= \frac{0.125k + 0.25k}{\sqrt{0.6875k}\sqrt{0.25k}} \\
&= \frac{0.375k}{k\sqrt{0.6875}\sqrt{0.25}} \\
&= \frac{0.375}{\sqrt{0.6875 * 0.25}} \\
&= 0.904
\end{aligned}
$$

Simulation proves this.

Problem 3 Let X_i denote the state of the machine at time index i, where $X_0 \sim \Pi$, the stationary distributions of the states. Finally, suppose there are n states, named numerically 1 to n. We want to compute

$$\rho(X_{i+j}, X_i) = \frac{\text{Cov}(X_{i+j}, X_i)}{\sqrt{\text{Var}(X_{i+j})} \cdot \sqrt{\text{Var}(X_i)}}$$

for all $1 \leq j \leq k$ and some k. We can simplify the denomonator by realizing that, since X_i and X_{i+j} are drawn from the same distribution, they have the same variance. The equation becomes:

$$\rho(X_{i+j}, X_i) = \frac{\text{Cov}(X_{i+j}, X_i)}{\text{Var}(X_i)}.$$

By definition, the expected value of X_i is

$$E(X_i) = \sum_{l=1}^{n} l \cdot \pi_l.$$

Thus,

$$\text{Var}(X_i) = E(X_i^2) - (EX_i)^2 = \sum_{l=1}^{n} (l^2 \pi_l) - \Big[\sum_{l=1}^{n} l\pi_l\Big]^2.$$

Now, let's derive $\text{Cov}(X_{i+j}, X_i)$. For this, we also need the expected value of X_{i+j}:

$$E(X_{i+j}|X_i) = \sum_{k=1}^{n} k \cdot m_{l,k}^{j}$$

where M^j is the transition matrix $M \cdot M \cdots M$ (j times), and $X_i = l$.

$$\text{Cov}(X_{i+j}, X_i) = E(X_{i+j}X_i) - E(X_{i+j})E(X_i)$$

Suppose $Q = X_{i,j}X_i$. By the law of total expectations, $E(Q) = E(E(Q|X_i))$. Thus,

$$\begin{aligned} E(X_{i+j}X_i) &= \sum_{l=1}^{n} \pi_l l \cdot E(X_{i+j}|X_i) \\ &= \sum_{l=1}^{n} \Bigg[\pi_l l \cdot \Big[\sum_{k=1}^{n} k \cdot m_{l,k}^{j}\Big]\Bigg]. \end{aligned}$$

Since X_{i+j} and X_i have the same distribution (Π), the covariance becomes

$$\mathrm{Cov}(X_{i+j}, X_i) = \sum_{l=1}^{n} \left[\pi_l l \cdot \left[\sum_{k=1}^{n} k \cdot m_{l,k}^{j} \right] \right] - (EX_i)^2.$$

We can now write code to calculate the correlation of X_i and X_{i+j}. The following function `mccor(tm, k)` returns a vector corresponding to $\rho(X_i, X_{i+1}) \ldots \rho(X_i, X_k)$.

```
# Calculate Pi distribution from a transition matrix.
findpis <- function(p) {
  n <- nrow(p)
  imp <- diag(n) - t(p)
  imp[n,] <- rep(1,n)
  rhs <- c(rep(0,n-1),1)
  solve(imp, rhs)
}

# Calculate correlations between the current state and
# the next K states, given transition matrix tm.
mccor <- function(tm, K) {

  Pi <- findpis(tm)
  n <- nrow(p)

  # mu = E[Xi] for all i.
  mu <- 0
  for (i in 1 : n) {
    mu <- mu + (i * Pi[i])
  }

  # denom = Var(Xi).
  denom <- 0
  for (i in 1 : n) {
    denom <- denom + (i^2 * Pi[i])
  }
  denom <- denom - mu^2

  # corr = Correlations.
  corr <- c()

  # Mj = transition matrix at time i+j.
```

```
  Mj <- tm

  i <- 1
  for(j in 1 : K) {

    outer_sum <- 0
    for (l in 1 : n) {
      inner_sum <- 0
      for (k in 1 : n) {
        inner_sum <- inner_sum + (k * Mj[l,k])
      }
      outer_sum <- outer_sum + (l * Pi[l] * inner_sum)
    }

    # numer = Cov(Xi, Xi+j).
    numer <- outer_sum - mu^2

    corr <- c(corr, c(numer / denom))
    Mj <- Mj %*% tm # Next Mj.
  }

  # Result is c(rho(Xi, Xi+1), rho(Xi, Xi+2) ... rho(Xi, Xi+k)).
  return (corr)
}
```