E1 222 Stochastic Models and Applications

1. Let $p_i, q_i,\ i = 1, \cdots, N$, be positive numbers such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i = 1$ and $p_i \le Cq_i,\ \forall i$ for some positive constant C. Consider the following algorithm to simulate a random variable, X:

   1. Generate a random number Y such that $P[Y = j] = q_j,\ j = 1, \cdots, N$. (That is, the mass function of Y is $f_Y(j) = q_j$).
   2. Generate U uniform over $[0,\ 1]$.
   3. Suppose the value generated for Y in step-1 is j. If $U < (p_j/Cq_j)$, then set $X = Y$ and exit; else go to step-1.

   On any iteration of the above algorithm, if condition in step-3 becomes true, we say the generated Y is accepted. Find the value of $P[Y \text{ is accepted} \mid Y = j]$. Show that $P[Y \text{ is accepted},\ Y = j] = p_j/C$. Now calculate $P[Y \text{ is accepted}]$. Use these to calculate the mass function of X.

Answer: If $Y = j$ then, by definition, Y is accepted if $U < (p_j/Cq_j)$ and U is uniform over $(0, 1)$. Hence,

$$P[Y \text{ is accepted}|Y = j] = P[U < (p_j/Cq_j)] = p_j/Cq_j$$

Hence,

$$P[Y \text{ is accepted},\ Y = j] = P[Y \text{ is accepted}|Y = j]P[Y = j] = \frac{p_j}{Cq_j} q_j = \frac{p_j}{C}$$

Now we have

$$P[Y \text{ is accepted}] = \sum_j P[Y \text{ is accepted},\ Y = j] = \sum_j \frac{p_j}{C} = \frac{1}{C}$$

Finally we get

$$\begin{aligned}
P[X = j] &= \sum_n P[X = j,\ \text{n times through the loop}] \\
&= \sum_n P[\text{(n-1) times } Y \text{ not accepted and } n^{th} \text{ time } Y = j \text{ and } Y \text{ is accepted}] \\
&= \sum_n \left(1 - \frac{1}{C}\right)^{n-1} \frac{1}{C}\, p_j = p_j
\end{aligned}$$

Comment: This is known as rejection-sampling method to generate a sample of the random variable X. Here, the distribution of Y is called the proposal distribution. If, generating Y is much simpler than generating X, this would be a useful method.

2. Let X, Y, Z be random variables having mean zero and variance 1. Let ρ_1, ρ_2, ρ_3 be the correlation coefficients between $X \& Y$, $Y \& Z$ and $Z \& X$, respectively. Show that

$$\rho_3 \geq \rho_1\rho_2 - \sqrt{1-\rho_1^2}\sqrt{1-\rho_2^2}.$$

(Hint: Write $XZ = [\rho_1 Y + (X - \rho_1 Y)][\rho_2 Y + (Z - \rho_2 Y)]$, and then use the previous problem and Cauchy-Schwartz inequality).

Answer: Since all random variable have mean zero and variance 1, $\rho_1 = EXY$, $\rho_2 = EYZ$ and $\rho_3 = EXZ$. Using the hint, we have

$$\rho_3 = EXZ = \rho_1\rho_2 E[Y^2] + \rho_1 E[Y(Z-\rho_2 Y)] + \rho_2 E[Y(X-\rho_1 Y)] + E[(X-\rho_1 Y)(Z-\rho_2 Y)]$$

From the previous problem we know Y and $(Z - \rho_2 Y)$ are uncorrelated. Hence, $E[Y(Z - \rho_2 Y)] = EYE(Z - \rho_2 Y) = 0$ because $EY = 0$. Similarly, $E[Y(X - \rho_1 Y)] = 0$. We also have $EY^2 = 1$. Hence we get

$$\rho_3 = EXZ = \rho_1\rho_2 + E[(X - \rho_1 Y)(Z - \rho_2 Y)]$$

By Cauchy-Schwartz inequality (and results of previous problem)

$$|E[(X-\rho_1 Y)(Z-\rho_2 Y)]| \leq \sqrt{\mathrm{Var}(X - \rho_1 Y)\mathrm{Var}(Z - \rho_2 Y)} = \sqrt{1-\rho_1^2}\sqrt{1-\rho_2^2}$$

Hence

$$-|E[(X - \rho_1 Y)(Z - \rho_2 Y)]| \geq -\sqrt{1-\rho_1^2}\sqrt{1-\rho_2^2}$$

Thus, we get

$$\begin{aligned} \rho_3 &= \rho_1\rho_2 + E[(X - \rho_1 Y)(Z - \rho_2 Y)] \\ &\geq \rho_1\rho_2 - |E[(X - \rho_1 Y)(Z - \rho_2 Y)]| \\ &\geq \rho_1\rho_2 - \sqrt{1-\rho_1^2}\sqrt{1-\rho_2^2} \end{aligned}$$

3. X_1, X_2, X_3 are iid uniform $(-1,\ 1)$ random variables. $Z = X_1 + X_2 + X_3$. Find density of Z

Answer: Let $W = X_1 + X_2$. Then we showed in class

$$f_W(w) = \begin{cases} \frac{w+2}{4} & \text{if } -2 < w < 0 \\ \frac{2-w}{4} & \text{if } 0 < w < 2 \end{cases}$$

Since W and X_3 are independent,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X_3}(t)\ f_W(z-t)\ dt$$

For $f_{X_3}(t)$ to be non-zero, we need $-1 \leq t \leq 1$. For $f_W(z-t)$ to be non zero, we need $-2 \leq z - t \leq 2$. This gives us $\max(-1, z-2) \leq t \leq \min(1, z+2)$. We also know that $-3 \leq Z \leq 3$.

This gives us $-1 \leq t \leq z+2$ for $-3 \leq z \leq -1$, $\quad -1 \leq t \leq 1$ for $-1 \leq z \leq 1$, $\quad$ and $z - 2 \leq t \leq 1$ for $1 \leq z \leq 3$.

To substitute for $F_W(z-t)$ we need to know whether $z-t < 0$ or not. In the integral, we anyway have $-1 \leq t \leq 1$. Hence, for $z \in [-3,\ -1]$ we know $z - t \leq 0$ and for $z \in [1,\ 3]$, $z - t > 0$. For the range $z \in [-1, 1]$, we need to consider the two ranges separately. Thus, we can calculate f_Z as follows

- for $-3 \leq z \leq -1$

$$\begin{aligned} f_Z(z) &= \int_{-1}^{z+2} \frac{1}{2}\ \frac{z-t+2}{4}\ dt \\ &= \frac{z^2}{16} + \frac{3z}{8} + \frac{9}{16} \end{aligned}$$

- for $-1 \leq z \leq 1$

$$\begin{aligned} f_Z(z) &= \int_{-1}^{z} \frac{1}{2}\ \frac{2-z+t}{4}\ dt + \int_{z}^{1} \frac{1}{2}\ \frac{z-t+2}{4}\ dt \\ &= -\frac{z^2}{8} + \frac{3}{8} \end{aligned}$$

- for $1 \leq z \leq 3$

$$\begin{aligned} f_Z(z) &= \int_{z-2}^{1} \frac{1}{2}\ \frac{2-z+t}{4}\ dt \\ &= \frac{z^2}{16} - \frac{3z}{8} + \frac{9}{16} \end{aligned}$$