E1 222 Stochastic Models and Applications
Problem Sheet 3.6

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.

Hint: In the third step of the algorithm, if $Y = j$ then it is accepted if $U < (p_j/Cq_j)$. U is uniform and by the condition on C, $(p_j/Cq_j) \le 1$. Hence, $P[Y \text{ is accepted} \mid Y = j] = (p_j/Cq_j)$. Now you get the second part of the question by noting that $p[Y = j] = q_j$. Summing $P[Y \text{ is accepted},\ Y = j]$ over j you get $P[Y \text{ is accepted}] = 1/C$. $X = j$ can happen by exiting the loop with $Y = j$. Exiting the loop on the n^{th} time with $Y = j$, for different n, constitute mutually exclusive events and the union of all these is the event of $X = i$ and thus you get the mass function of X as $f_X(i) = p_i$

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. Suppose X is a discrete rv taking values $\{x_1, x_2, \cdots, x_m\}$ with probabilities $p_1, \cdots p_m$. The usual method of simulating such a rv is as follows. We divide the $[0,\ 1]$ interval into bins of length p_1, p_2 etc. Then we generate a rv, uniform over $[0,\ 1]$ and depending on the bin it falls in, we decide on the value for X. That is, if $U \le p_1$ we assign $X = x_1$; if

$p_1 < U \leq p_1 + p_2$ then we assgn $X = x_2$ and so on.
Suppose X is a discrete random variable taking values $1, 2, \cdots, 10$. Its mass function is: $f_X(1) = 0.08, f_X(2) = 0.13, f_X(3) = 0.07, f_X(4) = 0.15, f_X(5) = 0.1, f_X(6) = 0.06, f_X(7) = 0.11, f_X(8) = 0.1, f_X(9) = 0.1, f_X(10) = 0.1$. Can you use the result of previous problem to suggest an efficient method for simulating X.

Hint: Here take the given distribution of X as the p_i's in the previous problem. Take $q_i = 0.1,\ 1 \leq i \leq 10$. Now $p_i \leq Cq_i,\ \forall i$ is satisfied with $C = 1.5$. Generating Y is very simple here: take Y as the integer part of $10U$ where U is uniform over $[0,\ 1]$. What role does C have in determining the efficiency of this method?

3. Let X_1, X_2, X_3 be independent random variables with finite variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ respectively. Find the correlation coefficient of $X_1 - X_2$ and $X_2 + X_3$.

Answer: One can show this through straight-forward algebra:

$$\begin{aligned}
\mathrm{Cov}(X_1 - X_2, X_2 + X_3) &= E[(X_1 - X_2)(X_2 + X_3)] - E[(X_1 - X_2)]E[(X_2 + X_3)] \\
= E[X_1X_2 + X_1X_3 - X_2^2 - X_2X_3] &- [EX_1EX_2 + EX_1EX_3 - (EX_2)^2 - EX_2EX_3] \\
&= \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, X_3) - \mathrm{Var}(X_2) - \mathrm{Cov}(X_2, X_3) \\
&= -\sigma_2^2
\end{aligned}$$

where we used the fact that X_1, X_2, X_3 are independent and hence uncorrelated.
Since the random variables are uncorrelated, $\mathrm{Var}(X_1 - X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \sigma_1^2 + \sigma_2^2$. Similarly $\mathrm{Var}(X_2 + X_3) = \sigma_2^2 + \sigma_3^2$. Now you can calculate the correlation coefficient.

Comment: By its definition, covariance satisfies: $\mathrm{cov}(kX, Y) = k\ \mathrm{cov}(X, Y)$, where k is a real constant, $\mathrm{cov}(X, Y) = \mathrm{cov}(Y, X)$ and $\mathrm{cov}(X, Y + Z) = \mathrm{cov}(X, Y) + \mathrm{cov}(X, Z)$. (Covariance is like an inner product). This can be used to directly deduce $\mathrm{Cov}(X_1 - X_2, X_2 + X_3) = \mathrm{Cov}(X_1, X_2) + \mathrm{Cov}(X_1, X_3) - \mathrm{Cov}(X_2, X_2) - \mathrm{Cov}(X_2, X_3) = -\sigma_2^2$, because these random variables are uncorrelated.

4. Let X and Y be random variables having mean 0, variance 1, and correlation coefficient ρ. Show that $X - \rho Y$ and Y are uncorrelated, and that $X - \rho Y$ has mean 0 and variance $1 - \rho^2$.

Hint: $E[X - \rho Y] = 0$ because $EX = EY = 0$. Hence, variance of $X - \rho Y$ is $E[X - \rho Y]^2 = 1 - \rho^2$ because $EX^2 = EY^2 = 1$ and $EXY = \rho$. (Since means are zero and variances 1, $\text{Cov}(X, Y) = \rho_{XY} = EXY$). For uncorrelatedness, $E[(X - \rho Y)Y] = EXY - \rho EY^2 = 0$

Comment: As we discussed we can think of all mean-zero random variables to be vectors in a vector space with covariance as the inner product. $\rho_{XY}\, Y$ can be thought of as projection of X on Y and hence the 'residual', $X - \rho Y$ is 'orthogonal' to Y.

5. 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).

6. Let X be a random variable with mass function given by

$$\begin{aligned} f_X(x) &= \frac{1}{18}, \quad x = 1, 3 \\ &= \frac{16}{18}, \quad x = 2. \end{aligned}$$

Show that there exists a δ such that $P[|X - EX| \geq \delta] = \text{Var}(X)/\delta^2$. This shows that the bound given by Chebyshev inequality cannot, in general, be improved.

Hint: Take $\delta = 1$ and calculate the probability on LHS.

7. Let $X_1, \cdots, X_n$ be independent random variables with X_i being exponential with parameter λ_i, $i = 1, \cdots, n$. (i). Show that $\text{Prob}[X_1 < X_2] = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. (ii). Let $Z = \min(X_1, \cdots, X_n)$. Find $E[Z]$. (iii). Let J be a random variable defined by: $J = k$ if X_k happens to be the minimum among X_1 to X_n. (That is, $J = \arg\min_i\{X_i\}$). Find distribution of J.

Hint: By now you should be an expert in calculating $P[X_1 < X_2]$. For example, it can be written as $\int_{-\infty}^{\infty} \int_x^{\infty} f_{X_1X_2}(x, y)\, dy\, dx$. It is a simple integral (recall that for exponential rv, $P[X > x] = e^{-\lambda x}$) and you can show it to be equal to $\frac{\lambda_1}{\lambda_1+\lambda_2}$. For the second part see if you can calculate $P[Z > a]$ and hence figure out its density and hence its expectation. (See if Z is an exponential rv). For the last part: the event $[J = k]$ is same as the event $[X_k < W]$ where $W = \min(X_1, \cdots, X_{k-1}, X_{k+1}, \cdots, X_n)$ and note that X_k and W are independent and use first part of the problem.

8. Let $X_1, X_2, \cdots, X_N$ be *iid* continuous random variables. We say a record has occurred at m $(1 \le m \le N)$ if $X_m > \max(X_{m-1}, \cdots, X_1)$. Show that (i). Probability that a record has occurred at m is equal to $\frac{1}{m}$. (ii). The expected number of records till k is $\sum_{m=1}^{k} \frac{1}{m}$. (iii). The variance of the number of records till k is $\sum_{m=1}^{k} \frac{m-1}{m^2}$.

Hint: Let I_k be the indicator random variable denoting whether or not a record has occurred at k. Then $I_k = 1$ if X_k is the largest of $X_1, X_2, \cdots, X_k$. Since these are iid continuous random variables, all orderings are equally likely (this is to be separately established) and hence $P[I_k = 1] = \frac{1}{k}$. Number of records till n can be expressed as a sum of such indicator random variables. The expectation would be the sum of the expectations. For the last part, argue that these indicator random variables are independent. Knowing that X_3 is the largest of X_1, X_2, X_3 does not tell you anything about whether or not X_2 is the largest of X_1, X_2.

9. Let X be a binomial random variable with parameters n and p. Let $Y = \max(0,\ X - 1)$. Show that $EY = np - 1 + (1 - p)^n$.

Hint: If $X = 0$ then $Y = 0$, otherwise $Y = X - 1$. Hence $EY = \sum_{k=1}^{n}(k - 1) f_X(k)$.

10. Let f be a density function with a parameter θ. (For example, f could be Gaussian with mean θ). Let $X_1, X_2, \cdots, X_n$ be iid with density f. These are said to be an iid sample from f or said to be iid realizations of X which has density f. Any function $T(X_1, \cdots, X_n)$ is called a statistic. Any estimator for θ is such a statistic. We choose a function based on what we think is the best guess for θ based on the sample. An estimator $T(X_1, \cdots, X_n)$ is said to be unbiased if $E[T(X_1, \cdots, X_n)] = \theta$. Let us write $\mathbf{X}$ for $(X_1, \cdots, X_n)$ and $T(\mathbf{X})$ for any statistic.
Suppose θ is the mean of the density f. Show that $T_1(\mathbf{X}) = (X_2 + X_5)/2$, $T_2(\mathbf{X}) = X_1$, $T_3(\mathbf{X}) = (\sum_{i=1}^{n} X_i)/n$ are all unbiased estimators

for θ. If T is an estimator for θ, the mean square error of the estimator is $E(T-\theta)^2$. Show that if T is unbiased then the mean square error is equal to the variance of the estimator. Among the the three estimators T_1, T_2, T_3 for the mean, listed earlier, which one has least mean square error?

Hint: If T is unbiased, $ET = \theta$ and hence $\mathrm{Var}(T) = E(T-\theta)^2$. We know that average of n iid rv has variance equal to $(1/n)$ times the variance of X_1. Hence, T_3 has least variance and hence least mean-square error (because all of them are unbiased).

11. Let $X_1, \cdots, X_n$ be iid with mean μ and variance σ^2. Let $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Show that

$$E\left(\sum_{k=1}^{n} (X_k - \bar{X})^2\right) = (n-1)\sigma^2.$$

(Hint: Write $(X_k - \bar{X}) = (X_k - \mu) - (\bar{X} - \mu)$ and note that $(\bar{X} - \mu) = \sum_k (X_k - \mu)/n$ and that $E(X_k - \mu)(X_j - \mu) = 0$ for $k \neq j$).
Based on this, suggest an unbiased estimator for the variance.
Let $S^2 = \sum_{k=1}^{n} (X_k - \bar{X})^2$. Suppose the first and third moments of X_i are zero. Find the covariance between $\bar{X}$ and S^2.

Hint: The first part involves straight-forward algebra. By using the hint in the problem, you should get it. This means that $\frac{1}{n-1} \sum_{k=1}^{n} (X_k - \bar{X})^2$ is an unbiased estimator of variance from n iid samples. This is an important result.

For the second part. Since μ, first moment of X_i, is given to be zero, $E[\bar{X}] = 0$ and hence covariance is $E[\bar{X}S^2]$. Multiply the two and then argue that it gives an expression where every term contains either X_j^3 or $X_j^2 X_k$ or $X_i X_j X_k$. Thus expectation is zero because first and third moments are zero and X_i are independent. Thus, $\bar{X}$ and S^2 are uncorrelated.

12. Let $X_1, X_2, \cdots, X_n$ be iid random variables with mean μ and variance σ^2. Let $\bar{X} = (\sum_{i=1}^{n} X_i)/n$ and $S^2 = \sum_{k=1}^{n} (X_k - \bar{X})^2/(n-1)$ be the sample mean and sample variance respectively. As we have seen, these are unbiased estimators of mean and variance. Show that $\mathrm{cov}(\bar{X},\ X_i - \bar{X}) = 0,\ i = 1, 2, \cdots, n$. (Hint: Note that $X_i\bar{X}$ can be

written as sum of terms like X_iX_j; note that $EX_iX_j = \mu^2$ if $i \neq j$ and is $\mu^2 + \sigma^2$ if $i = j$; note also that you know mean and variance of $\bar{X}$). Now suppose that the iid random variables X_i have normal distribution. Show that $\bar{X}$ and S^2 are independent random variables. (Hint: Try to use the result that for jointly Gaussian random variables, uncorrelatedness implies independence).

Hint: Since $E[X_i - \bar{X}] = 0$, we only need to show $E[\bar{X}(X_i - \bar{X})] = 0$. Now, $E[(\bar{X})^2] = \mathrm{Var}(\bar{X}) + (E[\bar{X}])^2 = \frac{\sigma^2}{n} + \mu^2$. We have $X_i\bar{X} = \frac{1}{n}(X_i^2 + \sum_{j\neq i} X_iX_j$. Now it is easy to compute expectation of this and hence show that $\bar{X}$ and $X_i - \bar{X}$ are uncorrelated.

For the second part first show that $\bar{X}$ and $X_i - \bar{X}$ are jointly Gaussian for each i. This can be done by writing this as a 2-D vector that can be written as a $2 \times n$ matrix multiplied by the n-dimensional vector with components X_i and noting that X_i are ind and Gaussian and hence jointly Gaussian. Since $\bar{X}$ and $X_i - \bar{X}$ are jointly Gaussian and uncorrelated (from the first part of the theorem), $\bar{X}$ is independent of $X_i - \bar{X}$ for each i and hence is independent of any function of them.