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.

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 $p_1 < U \le p_1 + p_2$ 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.

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 $X_2 + X_3$.

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 $1 - \rho^2$.

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

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 $\mathrm{Prob}[X_1 < X_2] = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. (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.

8. Let $X_1, X_2, \cdots, X_N$ be *iid* continuous random variables. We say a record has occurred at m $(1 \leq m \leq N)$ 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 $\sum_{m=1}^{k} \frac{m-1}{m^2}$.

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

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?

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.

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).