E1 222 Stochastic Models and Applications
Problem sheet #3

1. Let X_1, X_2, X_3 be iid random variables having exponential distribution with mean 1. Let $Y_1 = X_1 + X_2 + X_3$, $Y_2 = \frac{X_1+X_2}{X_1+X_2+X_3}$, $Y_3 = \frac{X_1}{X_1+X_2}$. Find the joint density of Y_1, Y_2, Y_3. Are Y_1, Y_2, Y_3 independent?

2. Let X, Y be iid geometric random variables with parameter p. Let $Z = X - Y$ and $W = \min(X, Y)$. Find the joint mass function of Z, W. Show that Z, W are independent.

3. Let X be a random variable having Gaussian density with mean zero and variance 1. Show that $Y = X^2$ has gamma density with parameters $\frac{1}{2}$ and $\frac{1}{2}$. (Note: $\Gamma(0.5) = \sqrt{\pi}$).

4. Let $X_1, \cdots, X_n$ be iid random variables having Gaussian density with mean zero and variance σ^2. Show that $Y = \frac{X_1^2+\cdots+X_n^2}{\sigma^2}$ has Gamma density with parameters $\frac{n}{2}$ and $\frac{1}{2}$. (This rv, Y, is said to have chi-squared distribution with n degrees of freedom).

5. Let X, Y, Z be iid continuous random variables. Show that $P[X < Y] = 0.5$ irrespective of what is the common density function of these random variables. Now calculate $P[X < Y < Z]$ and show that its value is same irrespective of what is the common density function of these random variables. Based on all this, can you guess what is the value of $P[X < Y, Z < Y]$.

6. 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 $p_i \leq Cq_i,\ \forall i$, 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.

    (Note that we could have written step-3 above as: If $U < (p_Y/Cq_Y)$ then set $X = Y$; 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 $P[Y \text{ is accepted}|Y = j]$. 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.

7. 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. (Note that a simple minded way of simulating X is as follows. We generate a random number Y uniformly over $[0,\ 1]$. Then we compute X as: if $Y \leq 0.08$ then $X = 1$ else if $Y \leq 0.08 + 0.13$ then $X = 2$ else if $Y \leq 0.08 + 0.13 + 0.07$ then $X = 3$ else ... and so on. We are looking for a computationally more efficient way of generating X).

8. Let X, Y be independent Gaussian random variables with mean zero and variance unity. Define random variables D and θ by

$$D = X^2 + Y^2; \qquad \theta = \tan^{-1}(Y/X)$$

(where, by convention, we assume θ takes values in $[0, 2\pi]$; for this we first calculate $\tan^{-1}(|Y|/|X|)$ in the range $[0, \pi/2]$ and then put that angle in the appropriate quadrant based on signs of Y and X). Show that D and θ are independent with D being exponential with $\lambda = 0.5$ and θ being uniform over $[0,\ 2\pi]$. (Since the above transformation is invertible, you can use the theorem for calculating the joint density of D, θ).

Now let D be exponential with $\lambda = 0.5$ and let θ be uniform over $[0,\ 2\pi]$. Suppose D and θ are independent. Define random variables X, Y by

$$X = \sqrt{D}\ \cos\theta; \qquad Y = \sqrt{D}\ \sin\theta$$

What would be joint distribution of X, Y?

9. Consider the following algorithm for generating random numbers X and Y :

    1. Generate U_1 and U_2 uniform over $[0,\ 1]$.

2. Set $X = \sqrt{-2\log(U_1)}\cos(2\pi U_2)$ and $Y = \sqrt{-2\log(U_1)}\sin(2\pi U_2)$.

What would be the joint distribution of X and Y ?

10. Consider the following algorithm for generating random variables V_1 and V_2 :

    1. Generate X_1 and X_2 uniform over $[-1,\ 1]$.
    2. If $X_1^2 + X_2^2 > 1$ then go to step 1; else set $V_1 = X_1,\ V_2 = X_2$ and exit.

    What would be the joint distribution of V_1 and V_2 ?

11. Suppose we have access to a random number generator that can generate random numbers uniformly distributed over $(0, 1)$. Using the results of the previous problems, suggest a method for generating samples of X where X has Gaussian density with mean zero and variance unity.

12. 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 $1 - \rho^2$.

13. 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).

14. Suppose a random experiment has r possible outcomes which occur with probabilities $p_1, \cdots, p_r$. Suppose this experiment is repeated n times. Let X be the number of times first outcome occurs and let Y be the number of times second outcome occurs. Show that

$$\rho_{XY} = -\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}.$$

    (Hint: Define indicator functions I_i, J_i so that I_i is indicator of the event that the i^{th} trial results in outcome 1 and J_i is the indicator of the event that the i^{th} trial results in second outcome. Now you can write X, Y as sums of these indicators and XY will contains all terms of the type $I_i J_j$.)

15. 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.

16. A coin, with probability heads being p, is tossed repeatedly till we get r heads. Let N be the number of tosses needed. Calculate EN. (Hint: Try to express N as a sum of geometric random variables).

17. A fair dice is rolled repeatedly till each of the numbers $1, 2, \cdots, 6$, appears atleast once. Find the expected number of rolls needed.

18. A coin with probability of heads p is tossed n times. We say, a changeover occurs when the result of the previous toss is different from the current toss. (This is same as the problem in the earlier assignment except that now probability of heads is p rather than 0.5). Find the expected number of changeovers.

19. An interval of length 1 is broken at a point uniformly distributed over $(0, 1)$. Let c be a fixed point in $(0, 1)$. Find the expected length of the subinterval that contains the point c. Show that this probability is maximized when $c = 0.5$.

20. Let X have exponential distribution with mean $1/\lambda$. Find $E[X|X > 1]$.

21. Suppose that independent trials, each of which is equally likely to have any of m possible outcomes, are performed repeatedly until the same outcome occurs k consecutive times. Let N denote the number of trials needed. Show that

$$E[N] = \frac{m^k - 1}{m - 1}$$

22. Let Y be a continuous random variable with density

$$f_Y(y) = \frac{1}{\Gamma(0.5)}\sqrt{\frac{0.5}{y}}e^{-0.5y},\ y > 0.$$

Let the conditional density of X given Y be

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}}\sqrt{y}\,e^{-0.5yx^2},\ -\infty < x < \infty,\ y > 0$$

Show that $E[X|Y] = 0$. Does this mean $EX = 0$?

Show that marginal of X is given by

$$f_X(x) = \frac{1}{\pi(1+x^2)}$$

(Note that $\Gamma(0.5) = \sqrt{\pi}$). Does EX exist?
(Notice that $f_{X|Y}$ is Gaussian with mean zero and variance $1/\sqrt{y}$, Y is Gamma with parameters $0.5, 0.5$ and X has Cauchy distribution).

23. Consider Bernoulli trials where p, the probability of success, itself may be random. Suppose p is distributed uniformly over (0, 1). That is, we first choose a random number from $(0, 1)$ and then using this as p (the probability of success) now perform a sequence of n Bernoulli trials. Let X denote the number of successes. Using the identity $\int_0^1 p^k(1-p)^{n-k}dp = \frac{k!(n-k)!}{(n+1)!}$ (which comes from our knowledge of beta-density), show that

$$P[X = k] = \int_0^1 P[X = k|p]f(p)dp = \frac{1}{(n+1)},\ k = 0, 1, \cdots, n.$$

Also show that the probability that $(r+1)st$ is a success given that there were k successes in the first r trials is $\frac{k+1}{r+2}$.

24. Let X be a non-negative integer-valued random variable. Then the function $\Phi_X(t) = Et^X$ is called its probability generating function. Suppose $\Phi_X(t_0) < \infty$ for some $t_0 > 1$. Then, by properties of power series, $\Phi_X(t)$ is defined and is differentiable on the interval $(-t_0, t_0)$. Now show that $\Phi'_X(1) = EX$ and $\Phi''_X(1) = EX(X-1)$. (Here, $\Phi'_X(\cdot)$ is its first derivative, $\Phi''_X(\cdot)$ is the second derivative and so on). Thus we can use Φ_X to obtain moments.

Note that

$$\Phi_X(t) = f_X(1)t + f_X(2)t^2 + f_X(3)t^3 + \cdots$$

Show that $f_X(1) = \Phi'_X(0)/1!$, $f_X(2) = \Phi''_X(0)/2!$ and so on.

Calculate the probability generating function when X is (i). Binomial, (ii). Poisson.

25. Let X be a nonnegative integer valued random variable. Let $\Phi_X(t) = Et^X$ be its probability generating function and assume that $\Phi_X(t)$ is finite for all t. By arguing as in the proof of Chebeshev inequality, show that for any positive integer, y,

    a. $P[X \le y] \le \frac{\Phi_X(t)}{t^y}$, $0 \le t \le 1$;

    b. $P[X \ge y] \le \frac{\Phi_X(t)}{t^y}$, $t \ge 1$.

    Now suppose X is a Poisson random variable with parameter λ. Use the above to show that

    $$P[X \le \frac{\lambda}{2}] \le \left(\frac{2}{e}\right)^{\lambda/2}.$$

    (Hint: Note that since the earlier inequalities are true for all values of t in a range, we can get tight inequality by choosing a t to minimise the righthandside of those inequalities). What will be the bound we get on the above probability if we use Chebeshev inequality?

26. 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?

27. 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 $E(X_k - \mu)(X_j - \mu) = 0$).

Let $S' = \sum_{k=1}^{n} (X_k - \bar{X})^2$. Suppose the first and third moments of X_i are zero. Find the covariance between $\bar{X}$ and S'.

28. Using the result in the previous problem, can you suggest an unbiased estimator for the variance of a random variable based on n iid samples?

29. Let $X_1, X_2, \cdots, X_n$ be iid random variables. 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 $\mathrm{cov}(\bar{X},\ X_i - \bar{X}) = 0,\ i = 1, 2, \cdots, n$. Now suppose that the iid random variables X_i have normal distribution. Show that $\bar{X}$ and S^2 are independent random variables.

30. Let $X_1, \cdots, X_n$ be iid random variables with a density (or mass) function having a parameter θ. Let $\mathbf{X} = (X_1, \cdots, X_n)$. Let $T(\mathbf{X})$ be a function of $\mathbf{X}$. If the conditional distribution of $\mathbf{X}$ given $T(\mathbf{X})$ does not depend on θ then $T(\mathbf{X})$ is called a *sufficient statistic* for θ. Show that $T(\mathbf{X}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic in the following cases (i). X_i are normal with mean θ and variance unity. (ii). X_i are poisson with mean θ.