Problem set 2

1. Show that the expectation, $E[X]$, of a random variable X has the following optimality property: For any real number a, $E[(X-a)^2] \geq E[(X-EX)^2]$, with equality only if $a = EX$. What does this say, in words?

2. (a) For any two random variables X and Y defined on the same probability space, show that

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

(b) If X and Y are independent of each other, then show that $\text{Cov}(X,Y) = 0$ and, consequently, that $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$. (c) If $\text{Cov}(X,Y) = 0$, does it follow that X and Y are independent?

3. (a) Let U_1 and U_2 be independent random variables, uniformly distributed on $[0,1]$. Compute $P(U_1 + U_2 < 1)$ by integrating their joint density over an appropriate region. (b) Alice and Bob agree to meet in front of the Wills Building at 12.30pm. Neither is very punctual, and is equally likely to turn up anytime between 12 noon and 1pm. Assuming that whoever comes first waits for the other, what is the probability that neither of them will have to wait longer than 10 minutes? (Hint: Draw a picture representing their respective arrival times on a graph. What is the region corresponding to the event of interest?)

4. Coupon collector problem: In order to increase sales of boxes of cereal, they are sold containing pictures of sports stars, and purchasers try to collect a complete set of these pictures. Suppose there are n pictures, and each box is equally likely to contain any of them. How many boxes (approximately) do you need to buy in order to collect all n pictures? Ignore trading of pictures.

Let N be the number of boxes you need to buy to collect all n coupons, and N_i the number of boxes you need to buy to get the i^{th} distinct picture after

you already have $i-1$ distinct pictures.
(a) It is clear that $N = N_1 + N_2 + \ldots + N_n$. Use this to compute $E[N]$ and $\mathrm{Var}(N)$. (b) If n is large, show using Chebyshev's inequality that the random variable N lies close to $E[N]$ with high probability. (Coming up with a "mathematical" statement of the above is part of what you are being asked to do.)

5. Given a set of positive real numbers $x_1, \ldots, x_n$, their arithmetic, geometric and harmonic means are defined as follows:

$$AM = \frac{1}{n}\sum_{i=1}^{n} x_i,\ GM = \exp\Big(\frac{1}{n}\sum_{i=1}^{n}\log(x_i)\Big),\ HM = \Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\Big)^{-1}.$$

Show using Jensen's inequality that $HM \le GM \le AM$.

6. (a) Suppose X is Uniform($[0,1]$) and $Y = -\lambda \log X$ for some given $\lambda > 0$. Compute the cdf of Y, and thereby obtain the pdf.
(b) Suppose R is an Exp(1) random variable and Θ is uniform on $[0, 2\pi]$. Define $X = R\cos(\Theta)$, $Y = R\sin(\Theta)$. What is the joint pdf of (X, Y)? What are the marginal pdfs, and are X and Y independent? Do you recognise this joint distribution? (c) Suppose X is Normal$(0,1)$, and we want to obtain a random variable Y which is Normal(μ, σ^2). Can you guess how to get Y as a function of X? Can you prove that your guess is correct?

7. Simulation of random variables. Method of inversion:
Computers are programmed to generate iid random variables uniformly distributed on $[0,1]$. (That statement needs qualification, but let's accept it.) But in many applications, we need random variables with other distributions. How can we obtain (samples of) them from iid samples of a uniform random variable?

Suppose we are given a sample X of a Uniform($[0,1]$) random variable, but want a sample of a random variable Y with a specified cdf F. We could ask whether there is some transformation $g(X)$ that will yield the cdf F. It turns out that the answer is yes, always. (But it is only in special cases that g will have a nice expression that makes it easy to compute.)
(a) Can you find the relationship between g and F?
(b) The Cauchy distribution has density

$$f(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

How would you generate a sample of a Cauchy random variable? (Hint: $\int_{-\infty}^{y} \frac{1}{1+x^2}dx = \tan^{-1}(y) + \frac{\pi}{2}$.)

(c) The Pareto distribution with parameter $\alpha > 1$ has density

$$f(x) = (\alpha - 1)(1 + x)^{-\alpha}, \; x \geq 0.$$

How would you generate a Pareto(α) random variable?

8. Simulation of random variables: Rejection sampling
The inversion method isn't always practical because F may not be easy to invert. Suppose we want to simulate (samples of) a random variable X with density f_X, and we know how to simulate a random variable Y with density f_Y. Suppose moreover that there is a constant $c > 0$ such that $f_X(x) \leq c f_Y(x)$ for all x. We then use the following procedure.

Generate a random sample of Y. Say it has the value y. Accept this sample with probability $f_X(y)/(c f_Y(y))$ (e.g., by generating an auxiliary random sample U uniformly distributed on $[0, 1]$ and accepting y only if $U < f_X(y)/(c f_Y(y))$). Repeat this procedure until a sample is accepted.

(a) Show that this procedure yields samples with the desired density f_X.
(b) Explain how to use it to obtain a sample from the density xe^{-x} on the positive real line, given samples of an Exp(1) random variable, which has density e^{-x} on the positive real line.

9. (a) Suppose X is Binomial(n, p), Y is Binomial(m, p), and X and Y are independent of each other. What is the distribution of $X + Y$? (If you can guess the answer without calculating it, that's fine, but explain your reasons in words.)
(b) Suppose X is Poisson(λ) and independent of Y which is Poisson(μ). Calculate the distribution of $Z = X + Y$ using a method of your choice. Do you recognise this distribution? Can you provide an intuitive explanation? (Hint: Recall that the Poisson distribution is a limit of Binomial distributions.)

10. (a) Suppose N is Binomial(n, p) and $X_1, X_2, \ldots$ are iid Bernoulli(q). Let $Y = X_1 + \ldots + X_N$ (with an empty set being defined to be zero). Use generating functions to find the distribution of Y. Explain the answer.
(b) Repeat if N is Poisson(λ).