Homework 6 (due 04/Nov/2013)

Try all the exercises. Submit only those marked with an asterisk (*).

1. Find the means and variances of X in each of the following cases.

$$(a)\, X \sim \text{Bin}(n,p). \qquad (b)\, X \sim \text{Pois}(\lambda). \qquad (c) X \sim \text{Geo}(p).$$

2. Find the means and variances of X in each of the following cases.

$$(a)\, X \sim \text{N}(\mu,\sigma^2). \qquad (b)\, X \sim \text{Gamma}(\nu,\lambda). \qquad (c)\, X \sim \text{Beta}(\nu_1,\nu_2). \qquad (d)\, X \sim \text{Unif}[a,b].$$

3.
1. Let $\xi \sim \text{Exp}(\lambda)$. For any $t,s \geq 0$, show that $\mathbf{P}\{\xi > t+s \,|\, \xi > t\} = \mathbf{P}\{\xi > s\}$. (This is called the *memoryless property* of the exponential distribution).
2. Show that if a non-negative random variable ξ has memoryless property (i.e., $\mathbf{P}\{\xi > t+s \,|\, \xi > t\} = \mathbf{P}\{\xi > s\}$), then ξ must have exponential distribution.

4. Let X be a non-negative random variable with CDF $F(t)$.
1. Show that $\mathbf{E}[X] = \int_0^\infty (1-F(t))dt$ and more generally $\mathbf{E}[X^p] = \int_0^\infty pt^{p-1}(1-F(t))dt$.
2. If X is non-negative integer valued, then $\mathbf{E}[X] = \sum_{k=1}^\infty \mathbf{P}\{X \geq k\}$.

5. (*) Find all possible joint distributions of (X,Y) such that $X \sim \text{Ber}(1/2)$ and $Y \sim \text{Ber}(1/2)$. Find the correlation for each such joint distribution.

6. Let (X,Y) have the bivariate normal distribution with density

$$f(x,y) = \frac{\sqrt{ab-c^2}}{2\pi} e^{-\frac{1}{2}\left[a(x-\mu)^2+b(y-\nu)^2+2c(x-\mu)(y-\nu)\right]}.$$

1. Find the marginal distributions of X and of Y.
2. Find means and variances of X and Y and the covariance and correlation of X with Y. Under what conditions on the parameters are X and Y independent?

[**Note:** It is very useful to introduce the matrix $\Sigma = \begin{bmatrix} a & c \\ c & b \end{bmatrix}^{-1} = \begin{bmatrix} \frac{b}{\Delta} & -\frac{c}{\Delta} \\ -\frac{c}{\Delta} & \frac{a}{\Delta} \end{bmatrix}$ which is called the *covariance matrix* of (X,Y). The answers can be written in terms of the entries of Σ.]

7. Let r balls be placed in m bins at random. Let X_k be the number of balls in the k^{th} bin. Recall that $(X_1,\ldots,X_m)$ has a multinomial distribution.

1. Find the joint distribution of (X_1, X_2) and the marginal distribution of X_1 and of X_2.
2. Find the means, variances, covariance and correlation of X_1 and X_2.
3. Let Y be the number of empty bins. Find the mean and variance of Y. [**Hint:** Write Y as $\mathbf{1}_{A_1} + \ldots + \mathbf{1}_{A_m}$ where A_k is the event that the k^{th} bin is empty].

8. (*) A box contains N coupons where the number w_k is written on the k^{th} coupon. Let $\mu = \frac{1}{N}\sum_{k=1}^{N} w_k$ be the "population mean" and let $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(w_i - \mu)^2$ be the "population variance". A sample of size m is drawn from the population, the values seen are $X_1, \ldots, X_m$. The sample mean $\overline{X}_m = (X_1 + \ldots + X_m)/m$ is formed. Find the mean and variance of $\overline{X}_m$ in the following two cases.

1. The samples are drawn with replacement (i.e., draw a coupon, note the number, put the coupon back in the box, and draw again…).
2. The samples are drawn without replacement.

More challenging problems - optional!

9. Recall the coupon collector problem where coupons are drawn repeatedly (with replacement) from a box containing coupons labelled $1, 2, \ldots, N$. Let T_N be the number of draws made till all the coupons are seen.

1. Find $\mathbf{E}[T_N]$ and $\text{Var}(T_N)$.
2. Use Chebyshev's inequality to show that for any $\delta > 0$, as $N \to \infty$ we have

$$\mathbf{P}\{(1-\delta)N\log N \le T_N \le (1+\delta)N\log N\} \to 1.$$

10. Let X be a non-negative random variable. Read the discussion following the problem to understand the significance of this problem.

1. Suppose X_n takes the values n^2 and 0 with probabilities $1/n$ and $1-(1/n)$, respectively. Compare $\mathbf{P}\{X_n > 0\}$ and $\mathbf{E}[X_n]$ for large n.
2. Show the *second moment inequality* (aka *Paley-Zygmund inequality*): $\mathbf{P}\{X > 0\} \ge (\mathbf{E}[X])^2/\mathbf{E}[X^2]$.

[**Discussion:** Markov's inequality tells us that that the tail probability $\mathbf{P}\{X \ge t\}$ can be bounded from above using $\mathbf{E}[X]$. In particular, $\mathbf{P}\{X \ge r\mathbf{E}[X]\} \le \frac{1}{r}$. A natural question is whether there is a lower bound for the tail probability in terms of the expected value. In other words, if the mean is large, must the random variable be large with significant probability?

The first part shows that the answer is 'No' in general. The second part shows that the answer is 'Yes', provided we have control on the second moment $\mathbf{E}[X^2]$ from above. Notice why the inequality does not give any useful bound in the first part of the problem (what happens to the second moment of X_n?)]

11. Let X be a random variable. Let $f(a) = \mathbf{E}[|X-a|]$ (makes sense if the first moment exists) and $g(a) = \mathbf{E}[(X-a)^2]$ (makes sense if the second moment exists).

1. Show that f is minimized uniquely at $a = \mathbf{E}[X]$.
2. Show that the minimizers of g are precisely the medians of X (recall that a number b is a median of X if $\mathbf{P}\{X \geq t\} \geq \frac{1}{2}$ and $\mathbf{P}\{X \leq t\} \geq \frac{1}{2}$).