PROBABILITY AND STATISTICS - HOMEWORK 6

DUE DATE: 26/OCT/2016 (IN TUTORIALS)

Homework problems are marked blue (but trying all problems is strongly recommended).

1. Find the expectation and variance for a random variable with the following distributions. (a) Bin(n, p), (b) Geo(p), (c) Pois(λ), (d) Hypergeo(N_1, N_2, m). [**Note:** Although the computations are easy, the answers you get are worth remembering as they occur in various situations.]

2. Find the expectation and variance for a random variable with the following distributions. (a) $N(\mu, \sigma^2)$, (b) Gamma(ν, λ), (c) Beta(p, q). [**Note:** Although the computations are easy, the answers you get are worth remembering as they occur in various situations.]

3. Place r balls in n bins uniformly at random. Let X_k be the number of balls in the k th bin. Find $\mathbf{E}[X_k]$, Var(X_k) and Cov(X_k, X_ℓ) for $1 \le k, \ell \le n$. [**Hint:** First do the case when $r = 1$. Then think how to use that to get the general case].

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 $\mathbf{E}[X^p] = \int_0^\infty pt^{p-1}(1-F(t))dt$. [**Hint:** In showing this, you may assume that X has a density if you like, but it is not necessary for the above formulas to hold true]

(2) If X is non-negative integer valued, then $\mathbf{E}[X] = \sum_{k=1}^\infty \mathbf{P}\{X \ge k\}$.

5. A deck consists of cards labelled $1, 2, \ldots, N$. The deck is shuffled well. Let X be the label on the first card and let Y be the label on the second card. Find the means and variances of X and Y and the covariance of X and Y.

6. A box contains N coupons labelled $1, 2, \ldots, N$. A sample of size m is drawn from the population and the sample average $\overline{X}_m$ is computed. Find the mean and standard deviation of $\overline{X}_m$ in both the following cases.

(1) The m coupons are drawn with replacement.

(2) The m coupons are drawn without replacement (in this case, assume $m \le N$).

Date: November 1, 2016.

7. Let $X \sim N(0,1)$. Although it is not possible to get an exact expression for the CDF of X, show that for any $t > 0$,

$$\mathbf{P}\{X \geq t\} \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}$$

which shows that the tail of the CDF decays rapidly. **[Hint:** Use the idea used in the proof of Markov's inequality]

The following problem is for the more mathematically minded students. You may safely skip this.

8. *The coupon collector problem.* A box contains n coupons labelled $1, 2, \ldots, n$. Coupons are drawn at random from the box, repeatedly and with replacement. Let T_n be the number of draws needed till each of the coupons has appeared at least once.

(1) Show that $\mathbf{E}[T_n] \sim n \log n$ (this just means $\frac{1}{n \log n}\mathbf{E}[T_n] \to 1$).

(2) Show that $\text{Var}(T_n) \leq 2n^2$.

(3) Show that $\mathbf{P}\left(\left| \frac{T_n}{n \log n} - 1 \right| > \delta \right) \to 0$ for any $\delta > 0$.

[*Hint:* Consider the number of draws needed to get the first new coupon, the further number of draws needed to get the next coupon and so on].