Recitation Session 3 Solutions

1. *Expectation from cdf.*

a. Note that $F(0) = 0$, $F^{-1}(1) = \infty$ and $U = F(X)$, then,

$$E(X) = E(F^{-1}(U)) = \int_0^1 F^{-1}(u)du \tag{1}$$

$$= \int_0^\infty (1 - F(x))dx \tag{2}$$

$$= \int_0^\infty P(X \geq x)dx \tag{3}$$

b. For X a positive valued random variable with integer range we have:

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) \tag{4}$$

c. Note that X^2 is always positive so it satisfies the assumption. These formulae are useful for finding their means, for the second moment using this formula might not be practical. So let's change the problem to finding only means: For exponential with parameter λ:

$$E(X) = \int_0^\infty P(X \geq x)dx = \int_0^\infty e^{-\lambda x}dx = 1/\lambda \tag{5}$$

$$\tag{6}$$

Geometric, number of failures until success $X \in \{0, 1, \dots\}$:

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=1}^{\infty} (1-p)^k = (1-p)/p \tag{7}$$

$$\tag{8}$$

d. Assign negative area to the part of the negative x axis and below the cdf curve.

Note: One can obtain geometric distribution from exponential via taking its integer part.

Note 2: If you insist on finding the variation try to show $E(X(X+1)) = 2\sum_{k=1}^{\infty} kP(X \geq k)$.

2. *Coupon collecting.* First purchase is a guaranteed hit. In the subsequent purchase you have $(n-1)/n$ chance of getting a card that you don't already have. Hence the additional number of purchases you need to make in order to get two different cards is a geometric random variable with parameter $(n-1)/n$, with mean $\frac{n}{n-1}$ so the mean of the total number of purchases is $1 + \frac{n}{n-1}$. Once this is done, the number of additional purchases to obtain 3 different cards is geometric with success probability $\frac{n-2}{n}$. Continuing this way the number of purchases to finish the whole sequence is:

$$1 + \frac{n}{n-1} + \frac{n}{n-2} + \dots + \frac{n}{1} = n(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \dots + \frac{1}{1}) \tag{9}$$

$$= n(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}) \tag{10}$$

For $n = 6$ this number is 14.7 and $n = 10$ gives 29.29, what happens when n grows larger? Recall:

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \log(n) + \gamma + \frac{1}{2n} \tag{11}$$

where γ is the Euler's constant, so roughly you can take it to be $n\log(n)$.

3. *Coupon collecting continued.* From the previous discussion we know that $T_n = \sum_{k=1}^{n} X_k$ where X_k denotes the number of purchases one must make to get the k^{th} distinct card. We also know that X_k's are independent and distributed according to the geometric distribution with parameter $p_k = (n - (k-1))/n$ (so that X_1 is just constant).

   a. Since they are uncorrelated (remember that independence implies uncorrelatedness but not the other way around) we can write: $\mathrm{Var}(T_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k) = \sum_{k=1}^{n} (1 - p_k)/p_k^2$.

   b. $\mathrm{Var}(T_n) = n \sum_{k=1}^{n} \frac{k-1}{(n-k+1)^2} = n \sum_{k=1}^{n} \frac{k}{(n-k)^2} \leq n \sum_{k=1}^{n} \frac{k}{n-k} \leq n^2 \mathcal{O}(1) = c_1^2 n^2$.

   c. $\mathrm{P}(|T_n - n\log n| > c_2 n) \leq \mathrm{P}(|T_n - n\log n - \gamma n| > c_3 n) \leq \mathrm{Var}(T_n)/c_3^2 n^2 \leq (c_1/c_3)^2$ so that one can play with the constants to make this probability small.

   For a challenge find the asymptotic distribution as $n \to \infty$ of $(T_n - n\log n)/n$.

4. *Conditional expectation.*

   a. Given $X = x$ the distribution of Y is binomial$(x, 1/2)$ so that $E(Y|X = x) = x/2$ for $x = 1, 2, \ldots, 6$.

   b. From the previous part $E(Y|X) = X/2$ and $E(X) = 3.5$ so,

   $$E(Y) = E(E(Y|X)) = E(X/2) = E(X)/2 = 3.5/2 = 1.75 \tag{12}$$

   Note that in this case you can explicitly write the distribution of Y however this way of conditioning to another random variable is handy when it is difficult to find the distribution of Y.

   c. Write the two dimensional distribution of (X, Y) then find the conditional distribution of X given $Y = 2$, finally calculate the expected value from this table. The answer should be approximately 3.94.

5. *Conditional expectation continued.*

   a. First argue intuitively to get $\frac{m}{n} S_n$. Let $S_n = \sum X_i$,

$$E(S_m|S_n = k) = \sum_{i=1}^{n} E(X_i|S_n = k) \tag{13}$$

$$E(X_i|S_n = k) = \mathrm{P}(i^{th}\text{trial is success given there are } k \text{ successes}) \tag{14}$$

$$= \frac{\mathrm{P}(i\text{th is success and } k \text{ successes})}{\mathrm{P}(} \text{k} successes) \tag{15}$$

$$= \frac{p\binom{n-1}{k-1} p^{k-1} (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{k}{n} \tag{16}$$

$$E(S_m|S_n = k) = \frac{mk}{n} \tag{17}$$

b. First term of the RHS is x because X is constant, and the second term is $E(Y)$ because of independence.

$$\begin{aligned} E(X+Y|X=x) &= E(X|X=x) + E(Y|X=x) && (18)\\ &= x + E(Y) && (19) \end{aligned}$$

6. *Bounds.* X be a random number picked from the list. Then $E(X) = 10$ and $E(X^2) = 101$, so that $\mathrm{Var}(X) = 1$.

   a. By Markov/Chebychev type inequality we have:

   $$\mathrm{P}(X \geq 14) = \mathrm{P}((X-10)/1 \geq (14-10)/1) \leq \mathrm{P}(|X - E(X)| \geq 4std(X)) \leq 1/4^2 \quad (20)$$

   There are 10^6 entries in the list, so the total number of entries that are 14 or over is 62,500.

   b. This time by symmetry we would have $\mathrm{P}(X \geq 14) = \mathrm{P}(X \leq 6)$ and the Chebychev bound applies to both of them so we would half the previous result to get 31,250.

   c. $10^6(1 - \Phi(4)) \approx 32$ where $\Phi(x)$ is the cdf of the normal distribution whose value can be found on any standard normal table.