Eric Kong

Data 102
October 24, 2019

Homework 3

Problem 2

1. If $i < k$, the probability that i is the max serial number in the sample set is 0, because the minimum possible serial number is at least k.
   If $k \leq i \leq N$,

2. $E[Y_{(k)}] = Y_{(k)} * Pr(Y_{(k} = i)$
   $\Rightarrow \sum_{i=k}^{N} \binom{i}{k} * Pr(Y_{(k} = i)$

3. Part (D)
   $Pr(N|Y_{(k)}) = \pi(N) * Pr(Y_{(k)}|N)$
   $\Rightarrow \prod_{i=k}^{N} Pr(Y_{(k)} = i) * \sum_{n \in N} \pi(N)$

Problem 3

1. Let $X = I_1 + ... + I_n$ be the sum of independent indicator variables
   $\mu = E[X] = E[I_1 + ... + I_n] = E[I_1] + ... + E[I_n]$
   $\mu = Pr(I_1) + ... + Pr(I_n) = p_1 + ... + p_n$
   $\mu = \sum_{j=1}^{n} p_j$

   $\sigma^2 = Var(X) = Var(I_1 + ... + I_n) = Var(I_1) + ... + Var(I_n)$
   $\sigma^2 = Pr(I_1) * (1 - Pr(I_1)) + ... + Pr(I_n) * (1 - Pr(I_n))$
   $\sigma^2 = p_1 * (1 - p_1) + ... + p_n * (1 - p_n)$
   $\sigma^2 = \sum_{j=1}^{n} p_j * (1 - p_j)$

2. Markov's Bound: $Pr(X \geq \mu(1 + c)) \leq \frac{1}{1+c}$

   $Pr(X \geq c) \leq \frac{\mu}{c} \Rightarrow$ Let $c = c * \mu \Rightarrow Pr(X \geq c * \mu) \leq \frac{1}{c}$
   $\Rightarrow \hat{c} = \mu * (1 + c) \Rightarrow Pr(X \geq \mu * (1 + c)) \leq \frac{1}{1+c}$

3. Chebyshev's Bound:

   $P(X \geq \mu(1 + c)) = Pr(X \geq \mu + \mu * c) = Pr(X - \mu \geq \mu * c)$
   Definition of Chebyshev: $Pr(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$
   $\Rightarrow$ Let $k = \mu * c \Rightarrow$
   $Pr(|X - \mu| \geq \mu * c) \leq \frac{\sigma^2}{\mu^2 c^2}$

4. From part (c): $Pr(|X - \mu| \geq \mu * c) \leq \frac{\sigma^2}{\mu^2 c^2}$
   $\Rightarrow Pr(|X - np| \geq npc) \leq \frac{n^2 p^2 (1-p)^2}{n^2 p^2 c^2} = \frac{(1-p)^2}{c^2}$

   From part (b): $Pr(X \geq \mu(1 + c)) \leq \frac{1}{1+c}$
   $Pr(X \geq np(1 + c)) \leq \frac{1}{1+c}$

5. $M_{I_j}(t) = 1 + p_j(e^t - 1)$
   $M_{I_j}(t) = E[e^{t * I_j}] = e^{t*0} * Pr(I_j = 0) + e^{t*1} * Pr(I_j = 1)$
   $M_{I_j}(t) = (1 - p_j) + p_j * e^t = 1 + p_j(e^t - 1)$

6. $M_X(t) \leq e^{\mu(e^t - 1)}$
   $\Rightarrow M_X(t) = M_{I_1}(t) * ... * M_{I_n}(t) = (1 + p_j(e^t - 1))^n$
   $\Rightarrow (1 + x) \leq e^x$ and $\mu = p_j$ because I_j is an indicator random variable
   $(1 + \mu(e^t - 1))^n \leq e^{\mu(e^t - 1)}$ where $x = \mu(e^t - 1)$

7.