

# Homework 4 DS-GA 1002

Yuhao Zhao  
N17578783

November 19, 2015

## Problem 1.

(a) The Markov transition matrix for this problem is :

$$P = \begin{array}{cc} & \begin{array}{cc} \text{Made the next} & \text{didn't make the next} \end{array} \\ \begin{array}{c} \text{Made the shot} \\ \text{Didn't make the shot} \end{array} & \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} \end{array}$$

The probability that he makes the 14th shot given he made 2nd and 12th is only relevant to the 13th shoot. The distribution of 13th shoot is only relevant to the 12th shoot. Let the distribution of ith shoot be  $\pi_i$   $\pi_{13} = (0.6, 0.4)$

$$\pi_{14} = (0.6, 0.4) \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix} = (0.48, 0.52)$$

The probability that he makes the 14th shoot is 0.48

(b) We know that  $\pi_1 = (0.4, 0.6)$ . Let A be that he made the first, B be that he didn't made the 3rd and C be that he didn't made 12th. We know that  $P(A|B, C) = P(A|B)$  from the Markov property.

$$P(A|B) \times P(B) = P(B|A) \times P(A), P(B|A) = 0.52, P(A) = 0.4$$

$$\pi_3 = \pi_1 \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}^2 = (0.4, 0.6) \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}^2 = (0.426, 0.574), P(B) = 0.574$$

Therefore, the probability is  $P(A|B) = \frac{0.52 \times 0.4}{0.574} \approx 0.3623693$

(c) Let K be the number of shots he made in a row.

$$P(K = k | \text{made the first}) = 0.6^k * 0.4, P(K = k | \text{didn't make the first}) = 0.3 * 0.6^{k-1} * 0.4$$

$$\begin{aligned} E(k) &= \sum_{k=1}^{\infty} k(0.4 \times 0.6^k \times 0.4 + 0.6 \times 0.3 \times 0.6^{k-1} \times 0.4) = \sum_{k=1}^{\infty} 0.16k \times 0.6^k + 0.072k \times 0.6^{k-1} \\ &= \sum_{k=1}^{\infty} 0.28k \times 0.6^k = 0.28 \times \frac{0.6}{(1-0.6)^2} = 1.05 \end{aligned}$$

## Problem 2.

$$(a) P_N(N = n) = \sum_{k=0}^n \left( \frac{1}{30} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} + \frac{1}{60} \binom{n}{k} \left(\frac{4}{5}\right)^k \left(\frac{1}{5}\right)^{n-k} \right) = \frac{1}{30} + \frac{1}{60} = \frac{1}{20}$$

$$P_C(C = \frac{1}{4}) = \sum_{n=0}^{20} \sum_{k=0}^n \frac{1}{30} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} = \frac{1}{30} \times 20 = \frac{2}{3},$$

$$\text{Since } C \text{ can only be } \frac{1}{4} \text{ or } \frac{4}{5}, P_C(C = \frac{4}{5}) = 1 - \frac{2}{3} = \frac{1}{3}$$

N and C are independent, since  $P(N, C) = P(N)P(C)$ , in particular:

$$P(N, C = \frac{1}{4}) = \sum_{k=0}^n \frac{1}{30} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k} = \frac{1}{30} = \frac{1}{20} \times \frac{2}{3} = P(n)P(C = \frac{1}{4})$$

$$P(N, C = \frac{4}{5}) = \sum_{k=0}^n \frac{1}{60} \binom{n}{k} \left(\frac{4}{5}\right)^k \left(\frac{1}{5}\right)^{n-k} = \frac{1}{60} = \frac{1}{20} \times \frac{1}{3} = P(n)P(C = \frac{4}{5})$$

(b) C is the probability of a person taking taxi to work in a particular day,  $C = \frac{1}{4}$  if he/she has a car,  $\frac{4}{5}$  if she/he doesn't have a car.

N is the number of days in a month the person has to work ( $N < 20$ ).

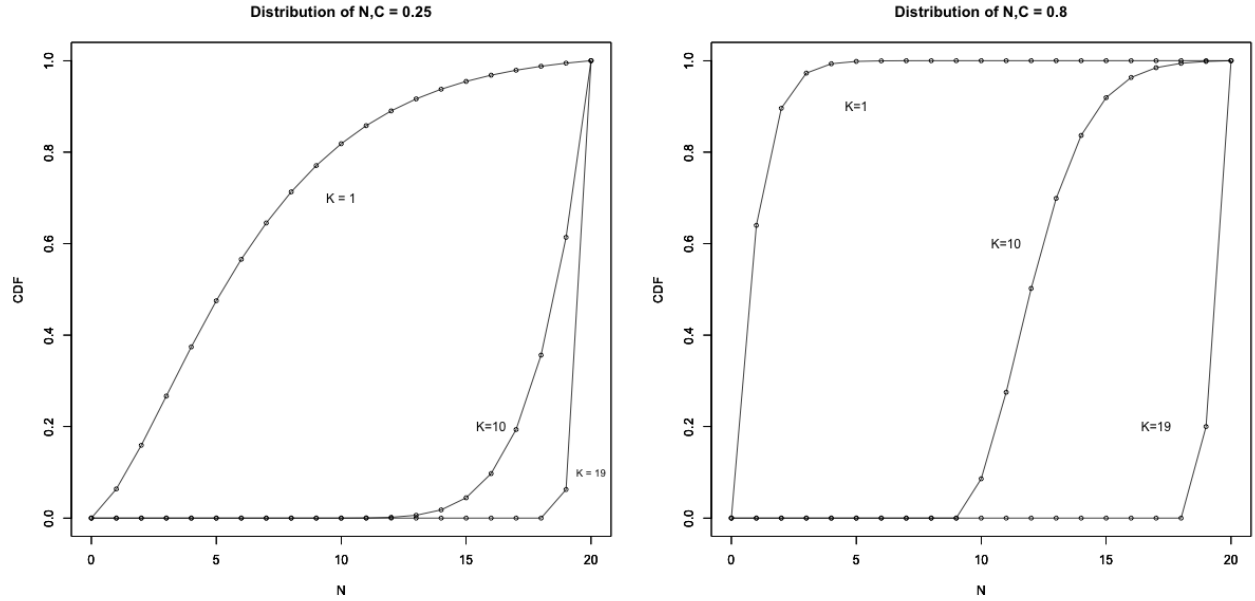
K is the number of days that the person takes taxi to work.

$$(c) P(N|K = k, C = \frac{1}{4}) = \frac{P(N, K, C = \frac{1}{4})}{P(K, C = \frac{1}{4})} = \frac{\frac{1}{30} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k}}{\sum_{n=k}^{20} \frac{1}{30} \binom{n}{k} \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k}}$$

$$P(N|K = k, C = \frac{4}{5}) = \frac{P(N, K, C = \frac{4}{5})}{P(K, C = \frac{4}{5})} = \frac{\frac{1}{60} \binom{n}{k} \left(\frac{4}{5}\right)^k \left(\frac{1}{5}\right)^{n-k}}{\sum_{n=k}^{20} \frac{1}{60} \binom{n}{k} \left(\frac{4}{5}\right)^k \left(\frac{1}{5}\right)^{n-k}}$$

$P(N|K, C)$  is the probability of the number of days that the person need to work given he has car or not and number of days that he go to work by taxi.

(d)



For my example, on the one hand, for the same person, the more days that he/she takes taxi to work, the more days she need to work per month. Although this is common sense, but it proves the model is valid.

On the other hand, given the same number of days that two person taking taxi to work, the person who has a car has higher probability to work more days than the person who doesn't have a car.

**Problem 3.**

(a) Let the number of packets arriving at the router be  $X$ , and the number of packets routed through connection 1 be  $Y$ . The model to this problem is that:  $X \sim \text{Poisson}(\lambda), Y|X \sim \text{Binomial}(X, p)$   
 $E(Y) = E(E(Y|X)) = \sum_{x=0}^{\infty} (px) \times p_X(x) = p \sum_{x=0}^{\infty} x \times p_X(x) = p \times E_X(x) = \lambda p$   
 $\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X)) = \sum_{x=0}^{\infty} p(1-p)x \times p_X(x) + \text{Var}_x(px) = p(1-p)E(x) + p^2\text{Var}(x) = p(1-p)\lambda + p^2\lambda = \lambda p$

(b)  $P(Y = y) = \sum_{k=y}^{\infty} P(Y = y|k) \times p_X(k) = \sum_y^{\infty} \binom{k}{y} p^y (1-p)^{k-y} \times \frac{\lambda^k e^{-\lambda}}{k!}$   
 $= \sum_y^{\infty} \frac{1}{y!(k-y)!} \frac{(\lambda p)^y}{\lambda^y} (1-p)^{k-y} \lambda^x e^{-\lambda} = \frac{e^{-\lambda}}{y!} (\lambda p)^y \sum_y^{\infty} \frac{(\lambda(1-p))^{k-y}}{(k-y)!} = \frac{e^{-\lambda}}{y!} (\lambda p)^y \sum_{k=0}^{\infty} \frac{(\lambda(1-p))^k}{k!}$   
 $\sum_{k=0}^{\infty} \frac{(\lambda(1-p))^k}{k!}$  is the Taylor expansion of  $e^{\lambda(1-p)}$   
Therefore,  $P(Y = y) = \frac{e^{-\lambda}}{y!} (\lambda p)^y e^{\lambda(1-p)} = e^{-\lambda p} \frac{(\lambda p)^y}{y!} \sim \text{poisson}(\lambda p)$

**Problem 4.**

(a) We want to find  $\Delta$  such that,  $P(|X_i - d_i| > \Delta) < 1\%$ .

We know that by chebyshev inequality,  $P(|Z_i - 0| \geq 10\sigma) \leq \frac{1}{10^2} = 0.01$ ,  $\Delta = 10\sigma = 10$

(b)  $|Y_i - d_i| = |\frac{1}{m} \sum_{j=i-m+1}^i X_j - d_i| = |\frac{1}{m} \sum_{j=i-m+1}^i X_j + \frac{1}{m}(d_i - d_{i-1}) + \frac{1}{m}(d_i - d_{i-2}) + \dots + \frac{1}{m}(d_i - d_{i-m+1})|$

Since the max speed is 2 m/s,  $d_i - d_j \leq 2(i - j)$

Therefore, by triangular inequality ,

$$\text{RHS} \leq |\frac{1}{m} \sum_{j=i-m+1}^i X_j| + |\frac{1}{m} \sum_{j=i-m+1}^{i-1} d_i - d_j| \leq |\frac{1}{m} \sum_{j=i-m+1}^i X_j| + |\frac{2}{m}(1 + 2 + \dots + m - 1)|$$

$$\text{RHS} \leq |\frac{1}{m} \sum_{j=i-m+1}^i X_j| + \frac{m(m-1)}{m}$$

Since  $X_j$  are iid normal(0,1),  $\frac{1}{m} \sum_{j=i-m+1}^i X_j \sim N(0, \frac{1}{\sqrt{m}})$

By Chebyshev inequality:  $P(|\frac{1}{m} \sum_{j=i-m+1}^i X_j - 0| \geq 10\frac{1}{\sqrt{m}}) \leq 0.01$

$$\rightarrow P(|\frac{1}{m} \sum_{j=i-m+1}^i X_j - 0| + (m-1) \geq 10\frac{1}{\sqrt{m}} + (m-1)) \leq 0.01$$

Since  $0 < |Y_i - d_i| \leq |\frac{1}{m} \sum_{j=i-m+1}^i X_j - 0| + (m-1)$ ,  $P(|Y_i - d_i| \geq \Delta_2) \leq P(|\frac{1}{m} \sum_{j=i-m+1}^i X_j - 0| + (m-1) \geq \Delta_2)$

Therefore, let  $\Delta_2 = 10\frac{1}{\sqrt{m}} + (m-1)$ ,  $P(|Y_i - d_i| > \Delta_2) < 0.01$

(c) Let  $m = 4$ ,  $\Delta_2 = 10 \times 0.5 + 3 = 8$ . She use  $m = 4$  because  $\sqrt{4}$  is an integer.

(d) We want to minimize  $\Delta_2$ .  $\Delta_2(m) = \frac{10}{\sqrt{m}} + m - 1$

$$\frac{d\Delta}{dm} = 1 - \frac{5}{m^{3/2}} = 0, m = 5^{2/3} \approx 2.92.$$

Since  $m$  can only be integer,  $\Delta_2(2) \approx 8.07, \Delta_2(3) \approx 7.77$ . The best precision at be attained at  $m = 3$ , which is  $\frac{10}{\sqrt{3}} + 2$

**Problem 5.**

(a) We can define  $E(X_I|X_J)$  as  $E(X_I|X_J) = \int_{X_I} X_I f_{X_I|X_J}(X_I|X_J) dX_I$ . It's a random vector.

$$E(E(X_I|X_J)) = \int_{X_J} E(X_I|X_J) f(X_J) dX_J = \int_{X_J} \int_{X_I} X_I f_{X_I|X_J}(X_I|X_J) dX_I f(X_J) dX_J$$

$$\text{Since } f_{X_I|X_J} = \frac{f_{X_I, X_J}}{f_{X_J}}$$

$$\begin{aligned} \text{Therefore, } E(E(X_I|X_J)) &= \int_{X_J} E(X_I|X_J) f(X_J) dX_J = \int_{X_J} \int_{X_I} X_I \frac{f_{X_I, X_J}}{f_{X_J}} f(X_J) dX_I dX_J \\ &= \int_{X_J} \int_{X_I} X_I f_{X_I, X_J} dX_I dX_J = \int_{X_I} X_I \int_{X_J} f_{X_I, X_J} dX_J dX_I = \int_{X_I} X_I f_{X_I} dX_I = E(X_I) \end{aligned}$$

(b)  $E(K) = E(E(K|N, C))$

$K|N, C$  is just a binomial distribution with  $N = n$  and  $p = c$ .

$$E(K|N, C) = \frac{n}{4}, c = \frac{1}{4}, \text{ and } \frac{4n}{5}, c = \frac{4}{5}$$

$$P_{N, C}(n, \frac{1}{4}) = \frac{1}{30}, P_{N, C}(n, \frac{4}{5}) = \frac{1}{60}$$

$$\text{Therefore, } E(K) = \sum_0^{20} \frac{n}{4} \times \frac{1}{30} + \frac{4n}{5} \times \frac{1}{60} = 4.55$$

$$E(K^2|N, C) = \text{Var}(K|N, C) + E(E(K|N, C))^2$$

$$\text{for } c = \frac{1}{4}, E(K^2|N, C) = \frac{n}{4}(\frac{3}{4} + \frac{n}{4})$$

$$\text{for } c = \frac{4}{5}, E(K^2|N, C) = \frac{4n}{5}(\frac{1}{5} + \frac{4n}{5})$$

$$\text{Therefore, } E(K^2) = E(E(K^2|C, N)) = \sum_0^{20} \frac{n}{4}(\frac{3}{4} + \frac{n}{4}) \times \frac{1}{30} + \frac{4n}{5}(\frac{1}{5} + \frac{4n}{5}) \times \frac{1}{60} = 38.465$$

$$\text{Var}(K) = E(K^2) - E(K)^2 = 17.7625$$