Hongyi Nie hn327

Kaitlin Tubbs kt392

ORIE 5530: Homework 2

1. Simulation Optimization.

   A taxi driver in the city of New York wants to decide how much time he should work per day to maximize his income. Assume that the average money per ride is p dollars. The number of trips he could fulfill depends, of course, on how much time he works. If he works for t hours, the number of dispatches D he gets is a Poisson random variable with mean $\mu(t) = \log(1+t)$. In other words (as expected), the longer he works the more (on average) rides he gets.

   Also, the cost (fuel, car depreciation, etc) of working increases as the working time t increases. Let us assume here that there is a cost of \$$c$ for working per hour (half an our costs then $c/2$ etc.). If he chooses to work 3 hours, the total operating cost is $3c$.

   (a) Let $p = 12$, $c = 3$. Assume the driver only works for an integer number of hours, and can not work more than 10 hours a day. Design and code a simulation based algorithm to find the amount of time he should work in order to maximize his expected net income.

   Of course, once you have such a code you could have much more general specifications (cost need not be linear, dispatches could follow an arbitrary structure etc.). The reason for this specific stucture we used is that now we can compare your simualtion result to

   (b) Analytically derive the optimal amount of time a driver should work if he wishes to maximize his/her *expected* income? How does this compare to your answer to part (a)?

2. Conditional probability. An urn contains three white, six red, and five black balls. Six of these balls are randomly selected from the urn. Let W, B and R, denote respectively the number of white, black and red balls selected.

   (a) Compute the probability that $P[W = 2, B = 3]$? Hint: we are returning here to *counting*. Since all outcomes are equally likely, this is about couting outcomes with the required property and dividing by the total number of outcomes.

   (b) Compute the conditional probability mass function of W given that $B = 3$.

   (c) Assume the balls are not drawn simutaneously, but drawn one by one, and each time a ball is picked, another ball with the same color is placed into the urn to replace it. Compute the conditional probability mass function of W given that $B = 3$ again.

   Hint: What has changed here relative to (a) is now that we are returning a ball after we see it. The answer will require you to generalize a bit the idea of the binomial distribution but this should not be difficult. Just remember how we motivated the expression for the binomial.

3. Assume a retailer chain has 2 branches. Let X_1 and X_2 be the number of customers arriving 10:00-11:00am on Mondays to each of the branches. Assume $X_i \sim Poisson(\lambda_i)$ for $i = 1, 2$ and that the demand to one is indepndent of demand to the other.

   (a) Derive the proability density function of the total number of customers of the 2 branches. What distribution does this random variable have? Show your work. What is the mean and variance of this random variable?

   (b) Starting the other way.

   Suppose that you do not know the distribution of demand to each store but you do know that the total demand, Y, (the sum of demand to both stores) is a Poisson random variable with rate $\lambda_1 + \lambda_2$. Suppose that when a customer has to choose which store to go to (for example, the wholefoods in Downtown Evanston or the one in the South) the customer chooses with probabilty $\lambda_1/(\lambda_1 + \lambda_2)$ to go store 1 and with the remaining probability to store 2.

   This way, the demand to store 1 is the number of customers that chose store 1 over store 2 (among the total number of customers).

What is $P[X_1 = i|Y = n]$?

What is now the distribution of the demand to store 1? That is, what is $P[X_1 = i]$? To store 2?

(c) Now suppose, more realistically, that demand is positively correlated and the correlation coefficient is 0.8. What is the mean and variance of the total demand $X_1 + X_2$?

4. Application of Central Limit Theorem.

You are trying to determine the melting point of a new material, of which you have a large number of measurements. For each measurement you make, you find a value close to the actual melting point c but there is also a measurement error associated with it. We model this with random variables.:

$$M_i = c + U_i$$

where M_i is yhr measured value in degree Kelvin, and U_i is the random error. It is known that $E[U_i] = 0$ and $Var[U_i] = 3$, for each i, and we may assume the measurements are independent among each other.

(a) Using one of the inequalities we learned in class, find an upper bound on the number of measurements you have to make in order to be 90% sure that the average of your measurements is within half a degree of c.

(b) Your sample size constructed in part (a) is probably too conservative. How can Central Limit Theorem help you to find a less conservative value of the number of measurements n that has to be taken? Find the new necessary sample size. Show your work.

1. Simulation Optimization.

A taxi driver in the city of New York wants to decide how much time he should work per day to maximize his income. Assume that the average money per ride is p dollars. The number of trips he could fulfill depends, of course, on how much time he works. If he works for t hours, the number of dispatches D he gets is a Poisson random variable with mean $\mu(t) = \log(1 + t)$. In other words (as expected), the longer he works the more (on average) rides he gets.

Also, the cost (fuel, car depreciation, etc) of working increases as the working time t increases. Let us assume here that there is a cost of \$$c$ for working per hour (half an our costs then $c/2$ etc.). If he chooses to work 3 hours, the total operating cost is $3c$.

(a) Let $p = 12$, $c = 3$. Assume the driver only works for an integer number of hours, and can not work more than 10 hours a day. Design and code a simulation based algorithm to find the amount of time he should work in order to maximize his expected net income.

Of course, once you have such a code you could have much more general specifications (cost need not be linear, dispatches could follow an arbitrary structure etc.). The reason for this specific stucture we used is that now we can compare your simualtion result to

(b) Analytically derive the optimal amount of time a driver should work if he wishes to maximize his/her *expected* income? How does this compare to your answer to part (a)?

a) Stimulation Result $t=3$

Code is on next page.

b) Exp(Net income)

$= \text{Exp}(\text{income} - \text{cost})$

$= \text{Exp}(\text{income}) - \text{Exp}(\text{cost})$

$= \text{Exp}(p * D) - \text{Exp}(c * t)$

$= p\,\text{Exp}(D) - c * t.$

$= d\,(12 * \log(1+t) - 3t)$

$$\frac{d\,(12 * \log(1+t) - 3t)}{dt} = \frac{12}{1+t} - 3 = 0 \Rightarrow t = 3$$

$\Rightarrow$ Reach extreme point at $t=3$.

Since the graph is concave, the extreme point is maximum.

The result is same to part a) stimulation.

Q1 a)

In [1]:
```
import numpy as np
from statistics import mode
```

In [6]:
```
c = 3
p = 12
t = list(range(0,11))
optimal_hour_list = []
for i in t:
    profit_list = []
    for j in range(100000):
        D = np.random.poisson(float(np.log(1+t[i])))
        profit = D*p - c*t[i]
        profit_list.append(profit)
    expected_value = np.mean(profit_list)
    optimal_hour_list.append(expected_value)
    print('working ' + str(i) + ' hours per day gives expected net inc
ome = ' + str(expected_value))

print('The optimal working hour that maximize expected net income is '
+ str(optimal_hour_list.index(max(optimal_hour_list))))
```

```
working 0 hours per day gives expected net income = 0.0
working 1 hours per day gives expected net income = 5.28948
working 2 hours per day gives expected net income = 7.21548
working 3 hours per day gives expected net income = 7.58832
working 4 hours per day gives expected net income = 7.32012
working 5 hours per day gives expected net income = 6.52416
working 6 hours per day gives expected net income = 5.39772
working 7 hours per day gives expected net income = 3.89916
working 8 hours per day gives expected net income = 2.25108
working 9 hours per day gives expected net income = 0.66216
working 10 hours per day gives expected net income = -1.22124
The optimal working hour that maximize expected net income is 3
```

In [ ]:

2. Conditional Probability

(a) $P(W=2, B=3)$

- # Total Outcomes: $\binom{14}{6}$
- $W=2$: $\binom{3}{2}$
- $B=3$: $\binom{5}{3}$
- $R=1$: $\binom{6}{1}$

$\longrightarrow P(W=2, B=3) = \dfrac{\binom{3}{2}\cdot\binom{5}{3}\cdot\binom{6}{1}}{\binom{14}{6}}$

$= \dfrac{3\cdot 10\cdot 6}{3003} = \boxed{0.05994}$

(b) PMF of W given $B=3$:

$$P(W \mid B=3) = \frac{P(W, B=3)}{P(B=3)}$$

$$= \frac{\binom{5}{3}\binom{3}{w}\binom{6}{3-w}}{\binom{14}{6}} \cdot \frac{\binom{14}{6}}{\binom{5}{3}\binom{9}{3}}$$

$$= \boxed{\frac{\binom{3}{w}\binom{6}{3-w}}{\binom{9}{3}}}$$

, where w = # White Selected
$w = \{0, 1, 2, 3\}$

$$P(B=3) = \frac{\binom{5}{3}\binom{9}{3}}{\binom{14}{6}}$$

$$P(W, B=3) = \frac{\binom{5}{3}\binom{3}{w}\binom{6}{3-w}}{\binom{14}{6}}$$

(c) PMF of W, with Replacement:

$$P(B=3) = \left(\frac{5}{14}\right)^3 \left(\frac{9}{14}\right)^3 \binom{6}{3}$$

$$P(W, B=3) = \left(\frac{5}{14}\right)^3 \left(\frac{3}{14}\right)^w \left(\frac{6}{14}\right)^{3-w} \binom{6}{3}\binom{3}{w}$$

$$\hookrightarrow P(W \mid B=3) = \boxed{\frac{\left(\frac{3}{14}\right)^w \left(\frac{6}{14}\right)^{3-w} \binom{3}{w}}{\left(\frac{9}{14}\right)^3}}$$

, $w = \{0, 1, 2, 3\}$

3. Assume a retailer chain has 2 branches. Let X_1 and X_2 be the number of customers arriving 10:00-11:00am on Mondays to each of the branches. Assume $X_i \sim Poisson(\lambda_i)$ for $i = 1, 2$ and that the demand to one is indepndent of demand to the other.

(a) Derive the proability density function of the total number of customers of the 2 branches. What distribution does this random variable have? Show your work. What is the mean and variance of this random variable?

(b) Starting the other way.
Suppose that you do not know the distribution of demand to each store but you do know that the total demand, Y, (the sum of demand to both stores) is a Poisson random variable with rate $\lambda_1 + \lambda_2$. Suppose that when a customer has to choose which store to go to (for example, the wholefoods ~~in Downtown~~ Evanston or the one in the South) the customer chooses with probabilty $\lambda_1/(\lambda_1 + \lambda_2)$ to go store 1 and with the remaining probability to store 2.
This way, the demand to store 1 is the number of customers that chose store 1 over store 2 (among the total number of customers).
What is $P[X_1 = i|Y = n]$?
What is now the distribution of the demand to store 1? That is, what is $P[X_1 = i]$? To store 2?

(c) Now suppose, more realistically, that demand is positively correlated and the correlation coefficient is 0.8. What is the mean and variance of the total demand $X_1 + X_2$?

a). Let $Z = X_1 + X_2$, $X_2 = Z - X_1$

$$P(Z=z) = \sum_{k=0}^{z} P(X_1=k, X_2=z-k)$$

$$= \sum_{k=0}^{z} P(X_1=k)\,P(X_2=z-k)$$

$$= \sum_{k=0}^{z} \frac{e^{-\lambda_1}\lambda_1^{k}}{k!}\cdot\frac{e^{-\lambda_2}\lambda_2^{z-k}}{(z-k)!}$$

$$= \sum_{k=0}^{z} \frac{z!}{k!(z-k)!}\cdot\frac{e^{-(\lambda_1+\lambda_2)}\lambda_1^{k}\lambda_2^{z-k}}{z!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}\underbrace{\sum_{k=0}^{z}\binom{z}{k}\lambda_1^{k}\lambda_2^{z-k}}_{\text{binomial } = (\lambda_1+\lambda_2)^{z}}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^{z}}{z!}$$

$\sim$ poisson ($\lambda = \lambda_1 + \lambda_2$)

Thus total number follows poisson distribution with mean $= \lambda_1+\lambda_2$, Var $= \lambda_1 + \lambda_2$

b) Since given $Y=n$, X_1, X_2 are discrete probability distribution of # of success in a sequence of n independent events with $p = \frac{\lambda_1}{\lambda_1+\lambda_2}, \frac{\lambda_2}{\lambda_1+\lambda_2}$ Respectively. According to the definition of Binomial distribution, we conduct:

$$X_1 \mid Y=n \sim \text{Binominal}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$$

$$X_2 \mid Y=n \sim \text{Binominal}\left(n, \frac{\lambda_2}{\lambda_1+\lambda_2}\right),$$

Thus $P(X_1=i \mid Y=n) = \binom{n}{i}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{i}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-i}$

$$P(x_1=i)=\sum_{n=0}^{\infty} P(x_1=i \mid Y=n)\, P(Y=n)$$

$$=\sum_{n=i}^{\infty}\binom{n}{i}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{i}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-i}\cdot\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}$$

$$=\sum_{n=i}^{\infty}\frac{\lambda_1^{i}\lambda_2^{n-i}e^{-\lambda_1}\cdot e^{-\lambda_2}}{(n-i)!\,i!}$$

$$=\frac{e^{-\lambda_1}e^{-\lambda_2}\lambda_1^{i}}{i!}\underbrace{\sum_{n=0}^{\infty}\frac{\lambda_2^{n}}{n!}}\to e^{\lambda_2} \text{ by factorial series formula}$$

$$=\boxed{\frac{e^{-\lambda_1}\lambda_1^{i}}{i!}\Rightarrow x_1\sim \text{possion}(\lambda=\lambda_1)}$$

$$P(X_2=i)=\sum_{n=0}^{\infty} P(X_2=i \mid Y=n)\, P(Y=n)$$

$$=\sum_{n=i}^{\infty}\binom{n}{i}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{i}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-i}\cdot\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}$$

$$=\sum_{n=i}^{\infty}\frac{\lambda_2^{i}\lambda_1^{n-i}e^{-\lambda_1}e^{-\lambda_2}}{i!\,(n-i)!}$$

$$=\frac{e^{-\lambda_1}e^{-\lambda_2}\lambda_2^{i}}{i!}\underbrace{\sum_{n=0}^{\infty}\frac{\lambda_1^{n}}{n!}}=e^{\lambda_1}$$

$$=\boxed{\frac{e^{-\lambda_2}\lambda_2^{i}}{i!}\Rightarrow X_2\sim \text{possion}(\lambda=\lambda_2)}$$

Therefore, x_1 & x_2 are poisson distribution with mean λ_1 and λ_2 respectively.

c). X_1 hasdemand λ_1, X_2 has demand λ_2,

$$\boxed{Exp(X_1+X_2)=\lambda_1+\lambda_2}$$

$$Var(X_1+X_2)=Var(X)+Var(Y)+2Cov(X,Y)$$

$$cov(X,Y)=\rho_{X_1}\cdot\sigma_{X_1}\sigma_{X_2}$$

$$=0.8\times\sqrt{X_1}\cdot\sqrt{X_2}$$

$$\boxed{Var(X_1+X_2)=X_1+X_2+1.6\sqrt{X_1X_2}}$$

4. $M_i = c + U_i$

$E[M_i] = c$

$Var[M_i] = Var(c + U_i) = 3$

$E[\bar{M}] = E\left[\frac{1}{n}\sum_{i=1}^{n} M_i\right]$

$= \frac{1}{n}\sum_{i=1}^{n} E[M_i] = c$

$Var[\bar{M}] = Var\left(\frac{1}{n}\sum_{i=1}^{n} M_i\right) = \frac{n \cdot 3}{n^2} = \frac{3}{n}$

(a) Chebyshev's Inequality:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} M_i - E[M_1]\right| \geq 0.5\right) \leq \frac{Var(M_1)}{n(0.5^2)}$$

$$1 - \frac{Var(M_1)}{n(0.5^2)} = 0.90$$

$$n = \frac{3}{0.10\,(0.5^2)} = \boxed{120}$$

(b) Central Limit Theorem:

$P(c - 0.5 \leq \bar{M} \leq c + 0.5)$

since $\bar{M} \sim N\left(c, \sqrt{\frac{3}{n}}\right)$:

$$P\left(\frac{\bar{X} - c}{\sqrt{3/n}} \leq Z \leq \frac{c + .05 - c}{\sqrt{3}/\sqrt{n}}\right) = 0.95$$

$$\rightarrow Z \leq \frac{.5}{\sqrt{3}/\sqrt{n}} = 0.95$$

$$1.65 = \frac{.5\sqrt{n}}{\sqrt{3}}$$

$$n = 32.67 \approx \boxed{33}$$

The CLT finds a less-conservative sample size by normalizing the distribution of M_i and estimating the sample size according to the Normal Distribution.

Alternative Method:

$E = 0.5$, $\sigma = \sqrt{3}$, $1 - \alpha = 0.90$

$\hookrightarrow \alpha = 0.1$

$Z_{\alpha/2} = 1.6449$

$$\rightarrow n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

$$= \left(\frac{1.6449 \cdot \sqrt{3}}{0.5}\right)^2$$

$$= 32.46$$

$$\approx \boxed{33}$$