MA 677 Take-Home Assignment

Hsueh-Pin Liu

2023-03-24

Problem 1

Consider three boxes:

A box containing two gold coins.

A box containing two silver coins.

A box containing one gold coin and a silver coin.

You choose a box at random and withdraw one coin at random. It's a gold coin. What is the probability that the other coin in the same box is also a gold coin?

Simulation

Besides the basic calculation method that I did it correct in class, the answer should be $\frac{2}{3}$, and I decide to use simulate the experiment for 100000 times and see the results.

```
set.seed(1)
n <- 100000
x <- c("1G","1G","2S","2S","3G","3S")
y <- sample(x,size=n,replace=T)
y <- as.data.frame(y)
m <- count(y,y)
m
```

```
##    y     n
## 1 1G 33244
## 2 2S 33255
## 3 3G 16766
## 4 3S 16735
```

```
#To satisfy the problem, we should calculate the ratio of "1G" to "1G"+"3G".
print(33244/(33244+16766))
```

```
## [1] 0.6647471
```

As a conclusion the answer of this problem is approximately equal to $\frac{2}{3}$ by using simulation. And then I want to explore something about statistical paradoxes, after searching a lot of information, I find an interesting one.

Birthday Paradox

Imagine that there are 23 students in a class, what's the probability that two of them have the same birthday? Some people might think that there are 365 days in a year, so the probability is $\frac{23}{365}$ which is even smaller than 10%. But in fact, the correct formula should be $1 - \frac{364}{365} * \frac{363}{365} * ... * \frac{343}{365}$, and it is calculated as follows:

```
print(1-prod(343:364)/365^22)
```

```
## [1] 0.5072972
```

Surprisingly, we got a result that is over 0.5, which means that the probability two of them having the same birthday is higher than the probability that all of them have a different birthday!

Problem 3

Suppose that a random variable Y has a probability density function given by $f(y) = ky^3 e^{-(y/2)}$ for $y > 0$. Find the value of k and calculate $E[y^2]$.

Solution 1

Before I think about the gamma function, I use integration by parts to solve the problem. $\int udv = uv - \int vdu$, so the function can be calculated as follows.

$$\int_0^\infty f(y)dy = \int_0^\infty ky^3 e^{-\frac{y}{2}} dy = k(y^3(-2e^{-\frac{y}{2}}) + \int_0^\infty 6y^2 e^{-\frac{y}{2}} dy$$
$$\int_0^\infty 6y^2 e^{-\frac{y}{2}} dy = 6y^2(-2e^{-\frac{y}{2}}) + \int_0^\infty 24ye^{-\frac{y}{2}} dy$$
$$\int_0^\infty 24ye^{-\frac{y}{2}} dy = 24y(-2e^{-\frac{y}{2}}) + \int_0^\infty 48e^{-\frac{y}{2}} dy$$

So we can get the final result.

$$\int_0^\infty f(y)dy = k(y^3 + 6y^2 + 24y + 48)(-2e^{-\frac{y}{2}})|_0^\infty = 96k = 1$$
$$k = \frac{1}{96}$$

And we can calculate $E[y^2]$ using the same method.

$$E[y^2] = \int_0^\infty f(y) * y^2 dy = \frac{1}{96}(y^5 + 10y^4 + 80y^3 + 480y^2 + 1920y + 3840)(-2e^{-\frac{y}{2}})|_0^\infty = \frac{7680}{96} = 80$$
$$E[y^2] = 80$$

This is calculation without using the gamma function, so it's more complex and takes more time.

Solution 2

Let's consider $\Gamma(4,2)$.

$$f(y) = \frac{y^3 e^{-\frac{y}{2}}}{2^4\Gamma(4)} = \frac{y^3 e^{-\frac{y}{2}}}{96}$$

And we know that

$$\int_0^\infty \frac{y^3 e^{-\frac{y}{2}}}{96} = 1$$

So

$$k = \frac{1}{96}$$

In order to calculate $E[Y^2]$, we can consider $\Gamma(6,2)$

$$f(y) = \frac{y^5 e^{-\frac{y}{2}}}{2^6\Gamma(6)} = \frac{y^5 e^{-\frac{y}{2}}}{7680}$$
$$\int_0^\infty \frac{y^5 e^{-\frac{y}{2}}}{7680} = 1$$
$$E[y^2] = \int_0^\infty f(y) * y^2 dy = \int_0^\infty \frac{y^5 e^{-\frac{y}{2}}}{96} dy = 80 \int_0^\infty \frac{y^5 e^{-\frac{y}{2}}}{7680} dy = 80$$

Problem 5

Suppose that the random variables X and Y are such that $E[X] = 4$, $E[Y] = -1$, $V[X] = 2$, and $V[Y] = 8$.

(a) What is the largest possible value of $Cov(X,Y)$?

(b) Assuming $\rho_{X,Y}$= 0.3, find $V(X,Z)$ where $Z = X + bY$ and $b \in R$.

Solution

(a) Because

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}, \rho_{X,Y} \leq 1$$

So we can have

$$Cov(X,Y) \leq \sqrt{V(X)V(Y)} = \sqrt{2*8} = 4$$
$$\max Cov(X,Y) = 4$$

(b)

$$V(X,Z) = V(X,X+bY) = V(X,X) + bV(X,Y) = V(X,X) + 0.3b\sqrt{V(X)V(Y)} = 2 + 1.2b$$

Deep Thinking

From question(a), I have two problems.

1) When can $Cov(X, Y)$ get the largest value?
2) If $\rho_{X,Y} = 0$, does it mean that there's not any relation between them?

1)

When $Cov(X, Y)$ get the largest value, we can know that $\rho_{X,Y} = 1$, which also means that there are constants a, b such that $P\{Y = aX + b\} = 1$, and to say this in an easy way, there is a linear relationship between Y and X.

2)

In fact, we can consider the following example. Assume that $X_1 = \sin x$, $X_2 = \cos x$,we know that $\rho_{X1,X2} = 0$, but we can't say there are not any relation between them since $X_1^2 + X_2^2 = 1$, so when $\rho_{X,Y} = 0$ we can only get the conclusion that they don't have a linear relationship.

Problem 6

Suppose that the number of minutes you have to wait for a bus is uniformly distributed on the interval $[0, 15]$. You wait for the bus 3 times. Find the probability that your longest wait is less than 10 minutes.

Simulation

Besides the basic calculation method that I did it correct in class, the answer should be $\frac{8}{27}$, and I decide to use simulate the experiment for 100000 times and see the results.

```
n <- 100000
x <- matrix(runif(n*3,min=0,max=15),nrow=n,ncol=3)
max <- apply(x,1,max)
print(sum(max<10)/n)
```

```
## [1] 0.29732
```

As a conclusion the answer of this problem is approximately equal to $\frac{8}{27}$ by using simulation. And for the situation mentioned in the problem, I have a similar example. When shopping in a market, if you are buying some products that you haven't buy before, A has 7 left on the shelf and B has only 2 left, which one will you choose? Probably most of us will choose B because we think it's more popular. We can actually use this idea to sell products with high inventory.