Ben Juarez
UID: 2134561

ACM 116
Fall 2021

Problem Set 2

1 Problem 1

We are given two random variables X and Y with an auxiliary random variable Z. Let $Z \sim U[0,1]$ and $0 < \alpha < \beta < 1$. Let

$$X = \begin{cases} 1, & \text{if } 0 < Z < \beta \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y = \begin{cases} 1, & \text{if } \alpha < Z < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let us find the covariance between X and Y as follows using the definition of covariance:

$$Cov(X,Y) = E[XY] - E[X]E[Y] = \mathbb{P}(X=1, Y=1) - \mathbb{P}(X=1)\mathbb{P}(Y=1)$$

$$\implies (\beta - \alpha) - (\beta)(1-\alpha) = \beta - \alpha - \beta + \beta\alpha$$

$$\implies \beta\alpha - \alpha = \alpha(\beta - 1)$$

2 Problem 2

We have that X and Y are independent normal random variables, each having a mean μ and variance σ^2. Let us show that $X+Y$ and $X-Y$ are independent by showing that they are jointly normally distributed and uncorrelated. First, we clearly have that $X+Y \sim N(2\mu, 2\sigma^2)$ and $X-Y \sim N(0, 2\sigma^2)$. With this, we can determine the following:

$$M_{X+Y}(t) = exp(2\mu t + \sigma^2 t^2) \qquad M_{X-Y}(t) = exp(\sigma^2 t^2)$$

Let us now show that $(X+Y) \perp (X-Y)$, or equivalently, $Cov(X+Y, X-Y) = 0$:

$$\implies Cov(X+Y, X-Y) = E[X+Y, X-Y] - E[X+Y]E[X-Y]$$
$$= E[X^2 - Y^2] - E[X+Y]E[X-Y]$$

We know that $E[X] = \mu$ for $X \sim N(\mu, \sigma^2)$ and $E[Y] = \mu$ for $Y \sim N(\mu, \sigma^2)$ for the given $X \perp\!\!\!\perp Y$. This allows us to arrive at the following:

$$\implies Cov(X+Y, X-Y) = E[X^2] - E[Y^2] - E[X+Y]E[X-Y]$$
$$= E[X]^2 - E[Y]^2 - E[X+Y]E[X-Y]$$
$$= \mu^2 - \mu^2 - (2\mu) \cdot 0 = 0$$

Thus, we have $Cov(X+Y, X-Y) = 0 \implies (X+Y) \perp (X-Y)$. Let us now examine the joint MGF of $X+Y$ and $X-Y$:

$$M_{(X+Y, X-Y)}(t,s) = exp(2\mu t + \frac{2\sigma^2 t^2}{2} + 0 + \frac{2\sigma^2 s^2}{2}) = exp(2\mu t + \sigma^2 t^2 + \sigma^2 s^2)$$

$$= exp(2\mu t + \sigma^2 t^2) + exp(\sigma^2 s^2)$$

Following our establishments regarding $M_{X+Y}(t), M_{X-Y}(t)$, we arrive at the joint MGF of our two independent normal random variables, as desired:

$$\implies M_{(X+Y,X-Y)}(t,s) = M_{X+Y}(t) \cdot M_{X-Y}(t) \implies (X+Y) \perp\!\!\!\perp (X_Y)$$

Thus, we conclude that $X+Y$ and $X-Y$ are independent.

3 Problem 3

Let $X \sim N(0,1)$ and let Z be a discrete random variable such that $Z = 1$ or -1 each with probability $\frac{1}{2}$. Also, let $Y = ZX$.

3.1 Part a

Let us determine the distribution of Y. Let us examine the CDF of Y:

$$F_Y(x) = \mathbb{P}(Y \leq x) = \mathbb{P}(ZX \leq x)$$

Given what we know about Z, we can rewrite $F_Y(x)$ as follows:

$$\implies F_Y(x) = \mathbb{P}(-X \leq x, Z = -1) + \mathbb{P}(X \leq x, Z = 1) = \frac{1}{2} \cdot \mathbb{P}(-X \leq x) + \frac{1}{2} \cdot \mathbb{P}(X \leq x)$$

$$\implies F_Y(x) = \frac{1}{2}[\mathbb{P}(X \geq -x) + \mathbb{P}(X \leq x)]$$

We can see how $\mathbb{P}(X \geq -x) = \mathbb{P}(X \leq x)$ considering the normal distribution, which allows us to arrive at:

$$\implies F_Y(x) = \frac{1}{2}[\mathbb{P}(X \leq x) + \mathbb{P}(X \leq x)] = \frac{1}{2}[2 \cdot \mathbb{P}(X \leq x)] = \mathbb{P}(X \leq x)$$

Since we know that the CDF of $X \sim N(0,1)$ is $\mathbb{P}(X \leq x)$, we have determined that the CDF of Y is equivalent to the CDF of X, meaning that $Y \sim N(0,1)$.

$$\implies F_Y(x) = F_X(x) \implies Y \sim N(0,1)$$

3.2 Part b

Let us determine if X and Y are correlated by examining $Cov(X,Y)$. We know that $Cov(X,Y) = 0 \implies X \perp Y$, meaning X and Y are uncorrelated:

$$Cov(X,Y) = E[XY] - E[X]E[Y] = E[X^2 Z] - E[X]E[Y] = E[X^2]E[Z] - E[X]E[Y]$$

Since $X \sim N(0,1)$ and $Y \sim N(0,1)$ we know $E[X] = E[Y] = \mu = 0$. We can calculate $E[Z]$ as follows:

$$E[Z] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0.$$

Now, we can clearly see that $Cov(X,Y) = 0$, implying that $X \perp Y$.

3.3 Part c

Let us determine if X and Y are independent. We can see that if $\mathbb{P}(X > 0, Y > 0) = \mathbb{P}(X > 0)\mathbb{P}(Y > 0)$, then $X \perp\!\!\!\perp Y$.

For $X \sim N(0,1)$ and $Y \sim N(0,1)$, we know that $\mathbb{P}(X > 0) = \mathbb{P}(Y > 0) = \frac{1}{2}$. On the other hand, let us rewrite $\mathbb{P}(X > 0, Y > 0)$ as follows considering that $Y = ZX$:

$$\mathbb{P}(X > 0, Y > 0) = \mathbb{P}(X = \frac{Y}{Z} < 0, Z = -1) + \mathbb{P}(X = \frac{Y}{Z} > 0, Z = 1)$$

$$\implies \mathbb{P}(X > 0, Y > 0) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

So, we can now properly compute $\mathbb{P}(X > 0, Y > 0) = \mathbb{P}(X > 0)\mathbb{P}(Y > 0)$ and arrive at our conclusion:

$$\implies \mathbb{P}(X > 0, Y > 0) = \frac{1}{2} \neq \mathbb{P}(X > 0)\mathbb{P}(Y > 0) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

Therefore, since $\mathbb{P}(X > 0, Y > 0) \neq \mathbb{P}(X > 0)\mathbb{P}(Y > 0)$, we conclude that X and Y are not independent.

4 Problem 4

Let $X_1, ..., X_n$ be independent Poisson random variables with mean λ and let $S_n = \sum_{i=1}^n X_i$.

4.1 Part a

Let us use the Markov inequality $\mathbb{P}(X \geq a)$ to get a bound on $\mathbb{P}(S_n > m)$ such that $a = m$ and $X = S_n$:

$$\implies \mathbb{P}(S_n > m) = \frac{E[S_n]}{m} = \frac{E[\sum_{i=1}^n X_i]}{m} = \frac{\sum_{i=1}^n E[X_i]}{m} = \frac{\sum_{i=1}^n \lambda}{m} = \frac{n \cdot \lambda}{m}$$

4.2 Part b

Let us use the CLT to estimate $\mathbb{P}(S_n \geq m)$, assuming n is large. Given the definition of S_n, we can see that $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{S_n}{n}$. Let us estimate $\mathbb{P}(S_n \geq m)$ as follows:

$$\mathbb{P}(S_n \geq m) \implies \mathbb{P}(\frac{S_n}{n} \geq \frac{m}{n}) = \mathbb{P}(\bar{X_n} \geq \frac{m}{n}) = \mathbb{P}(\frac{\bar{X_n} - \lambda}{\sqrt{\frac{\lambda}{n}}} \geq \frac{\frac{m}{n} - \lambda}{\sqrt{\frac{\lambda}{n}}})$$

$$\implies \mathbb{P}(S_n \geq m) = 1 - \Phi(\frac{\frac{m}{n} - \lambda}{\sqrt{\frac{\lambda}{n}}})$$

4.3 Part c

See corresponding MATLAB section at end of file.

5 Problem 5

See corresponding MATLAB section at end of file.

6 Problem 6

See corresponding MATLAB section at end of file.

7 MATLAB

Problem Set 2


```
% PS2Q4c
clear;
lambda = 1;
n = 100;
m = 120;
N = 10^4;
est = 0;

for i = 1:N
    s = sum(poissrnd(lambda, [1 n]));
    if (s >= m)
        est = est + 1;
    end
end

clt = 1 - normcdf(((m./n)-(lambda)) ./ (sqrt(lambda./n)) )
markov = (lambda .* n) / m
prob_estimate = est ./ N
% Clearly, it looks like Markov bound more loose while the CLT estimate is
 quite
% accurate.
```

```
clt =

    0.0228


markov =

    0.8333


prob_estimate =

    0.0277
```

Published with MATLAB® R2021b

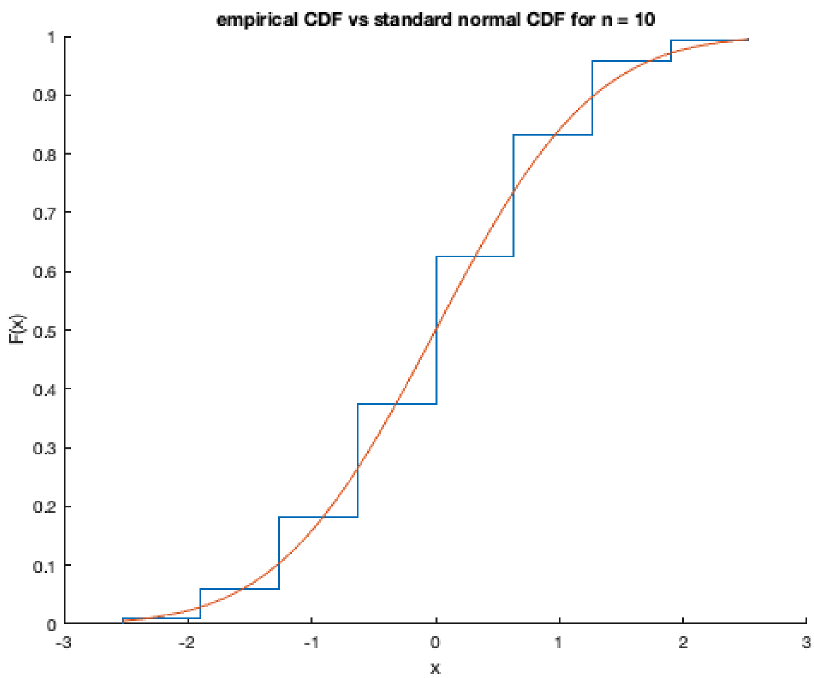
```
% PS2Q5a
b = 1;
a = 0;
N = 100;
sum = 0;
for i = 1:N
    sum = sum + rand()^3;
end
I_a = (b - a) ./ N .* sum

% PS2Q5b
alpha = 4;
beta = 1;
x = betarnd(alpha, beta);
% We know g(x) = x^3.
% For f(x | alpha, beta), we have the following:
%   (alpha + beta - 1)! / [(alpha - 1)!  (beta - 1)!] = 4! / 3! = 4
%   x^(alpha - 1) = x^3
% So, we have that f(x) must be 4x^3.
% Clearly, for one sample, we have g(x)/f(x) = 1/4 = 0.25.
I_b = x^3 / betapdf(x, 4, 1)
```

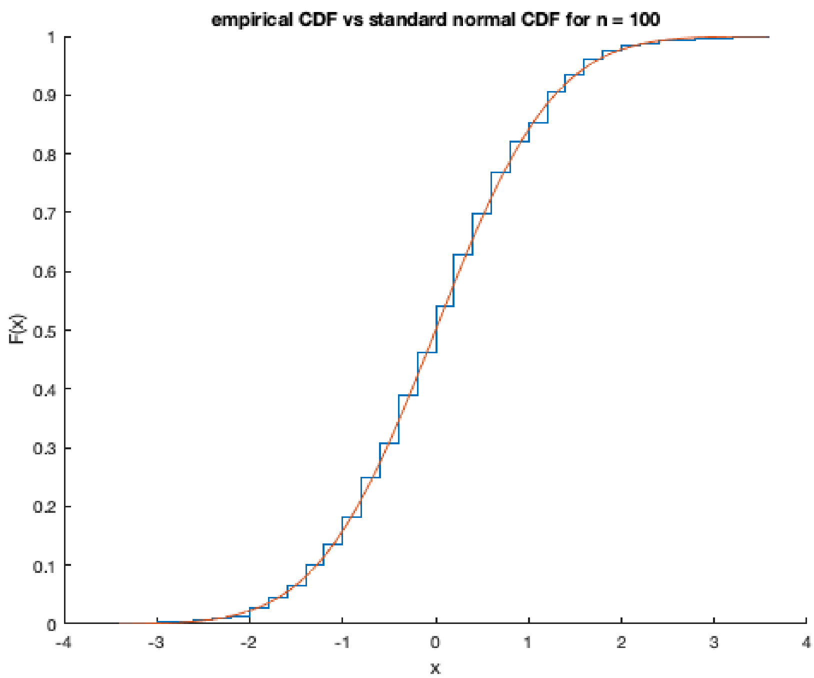
```
I_a =

    0.2306


I_b =

    0.2500
```

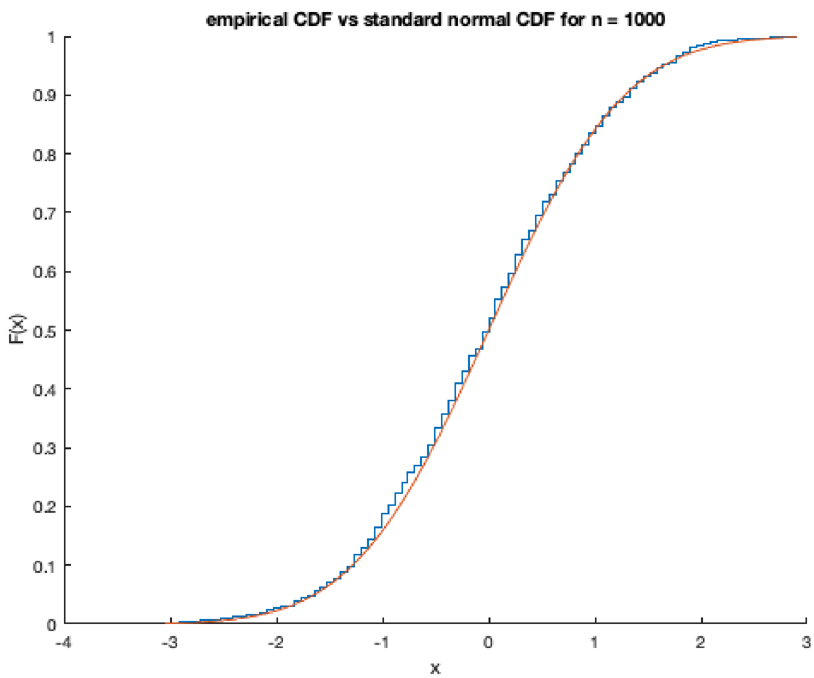
Published with MATLAB® R2021b

```
% PS2Q6a
clear;
n = 10;
N = 10^3;
y = zeros(N, 1);
for i = 1:N
    y(i) = sqrt(n) .* mean(datasample([-1 1], n));
end

% PS2Q6b
figure
hold on
ecdf(y)
low = min(y);
high = max(y);
fplot(@(x) normcdf(x), [low high])
title("empirical CDF vs standard normal CDF for n = 10")
hold off

% PS2Q6c

% n = 100
n = 100;
for i = 1:N
    y(i) = sqrt(n) .* mean(datasample([-1 1], n));
end
figure
hold on
ecdf(y)
low = min(y);
high = max(y);
fplot(@(x) normcdf(x), [low high])
title("empirical CDF vs standard normal CDF for n = 100")
hold off

% n = 1000
n = 1000;
for i = 1:N
    y(i) = sqrt(n) .* mean(datasample([-1 1], n));
end
figure
hold on
ecdf(y)
low = min(y);
high = max(y);
fplot(@(x) normcdf(x), [low high])
title("empirical CDF vs standard normal CDF for n = 1000")
hold off

% n = 10000
n = 10000;
for i = 1:N
```

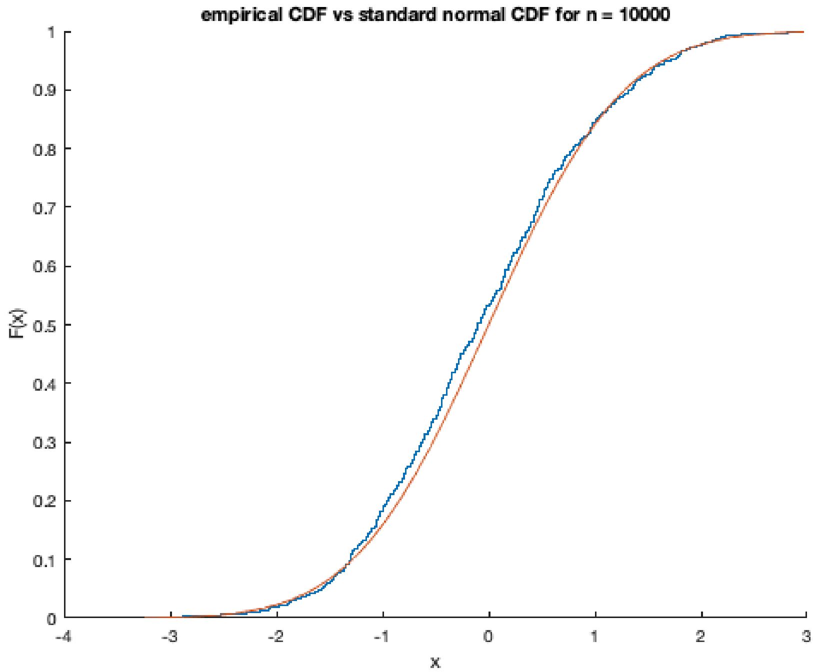
```
    y(i) = sqrt(n) .* mean(datasample([-1 1], n));
end
figure
hold on
ecdf(y)
low = min(y);
high = max(y);
fplot(@(x) normcdf(x), [low high])
title("empirical CDF vs standard normal CDF for n = 10000")
hold off
```




empirical CDF vs standard normal CDF for n = 100
F(x)
x
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3
4


empirical CDF vs standard normal CDF for n = 1000
F(x)
x
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-4
-3
-2
-1
0
1
2
3

4



Published with MATLAB® R2021b