STAT 403 Spring 2018
HW02

Nan Tang
1662478

April 17, 2019

Q1

Q1-1

$$p(x) = \begin{cases} 6x(1-x), & \text{if } x \in [0,1] \\ 1, & \text{otherwise} \end{cases}$$

For $x \in [1,0]$, the cdf of X is

$$\begin{aligned} F_X(x) &= \int_0^x p(t)dt \\ &= \int_0^x 6t(1-t)dt \\ &= \int_0^x 6t - 6t^2 dt \\ &= 3x^2 - 2x^3 \end{aligned}$$

Otherwise, if $x > 1, F(x) = 1$; if $x < 0, F(x) = 0$, since $p(x) = 0$ if $x \notin [0,1]$.

$$F_X(x) = \begin{cases} 3x^2 - 2x^3, & \text{if } x \in [0,1] \\ 1, & \text{if } x > 1 \\ 0, & \text{if } x < 0 \end{cases}$$

Q1-2

For $x \in [0, 1]$, the edf of X is

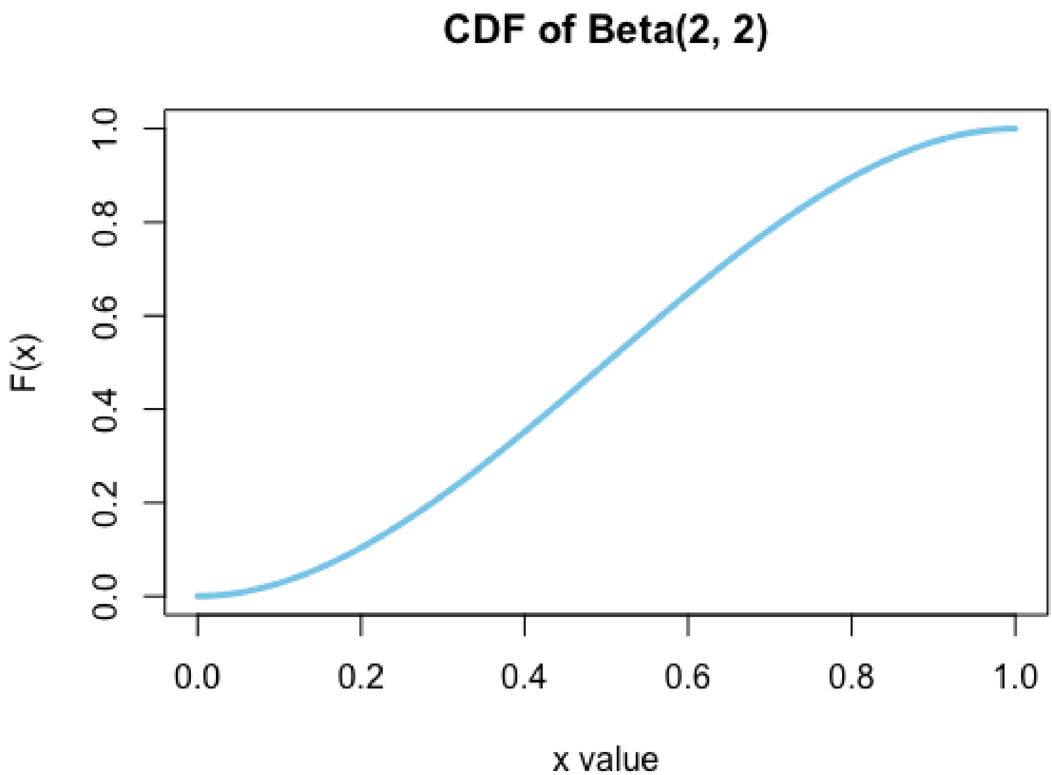
$$\widehat{F_n}(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)$$

$$\begin{aligned}
\mathbb{E}(\widehat{F_n}) &= \mathbb{E}(\frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(x_i \leq x) \\
&= \mathbb{E}(x_i \leq x) \\
&= P(x_i < x) \\
&= F_X(x), \text{ since } (x_i < x) \text{ follows Bernoulli distribution} \\
&= 3x^2 - 2x^3
\end{aligned}$$

$$\begin{aligned}
Var(\widehat{F_n}) &= Var(\frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} Var(I(x_i \leq x)) \\
&= \frac{1}{n} Var(P(x_i < x)) \\
&= \frac{1}{n} F(x)(1 - F(x)), \text{ since } Var(Bern(p)) = p(1-p) \\
&= \frac{(3x^2 - 2x^3)(1 - 3x^2 + 2x^3)}{n}
\end{aligned}$$

Q1-3

```
x_base <- seq(0, 1, 0.01)
beta_cdf <- function(x) {return(3 * x^2 - 2 * x^3)}
plot(x_base, beta_cdf(x_base), type='l', col='skyblue', lwd=3,
     xlim=c(0, 1), main='CDF of Beta(2, 2)', xlab='x value', ylab='F(x)')
```



Q2

Q2-1

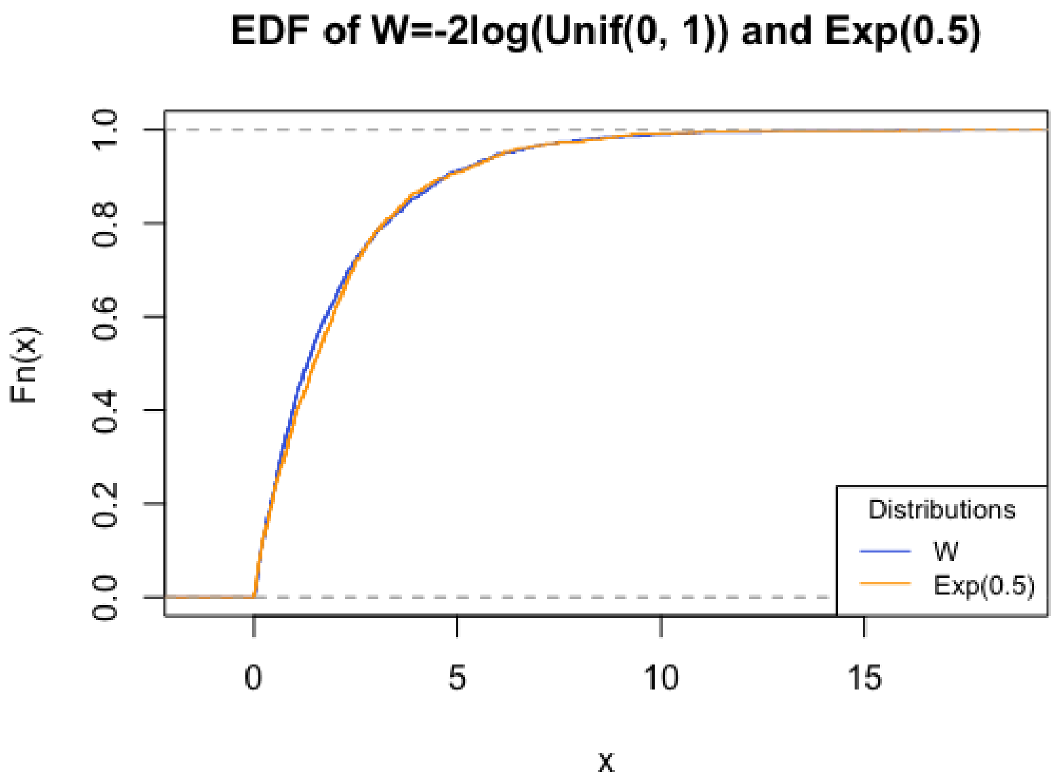
Given $U \sim Uniform(0,1)$ and $W = -2log(U)$.
The cdf of U $F_U(x) = x, x \in [0,1]$.
First find the cdf of W.

$$\begin{aligned}
F_W(x) &= P(W < x) \\
P(W < x) &= P(-2log(U) < x) \\
&= P(e^{-2log(U)} < e^x) \\
&= P(U^{-2} < e^x) \\
&= P(U \geq e^{-\frac{x}{2}}) \\
&= 1 - P(U < e^{-\frac{x}{2}}) \\
&= 1 - e^{-\frac{x}{2}}
\end{aligned}$$

Note that the cdf of $Exp(0.5)$, $F = 1 - e^{-\frac{x}{2}}$, which is equal to the cdf of W, therefore, we can conclude that W and $Exp(0.5)$ have the same distribution.

Q2-2

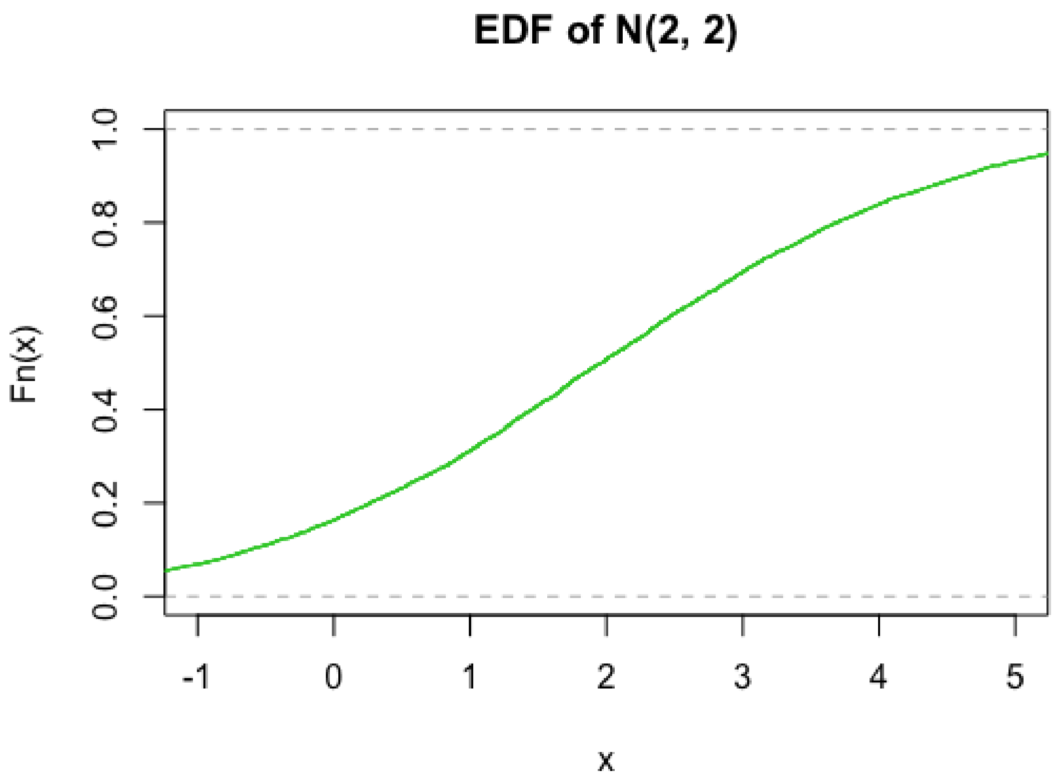
```
unif_value <- runif(1000, 0, 1)
w_value <- -2*log(unif_value)
exp_value <- rexp(1000, 0.5)
w_edf <- ecdf(w_value)
exp_edf <- ecdf(exp_value)
plot(w_edf, col='royalblue', lwd=1,
     main='EDF of W=-2log(Unif(0, 1)) and Exp(0.5)')
lines(exp_edf, col='orange', lwd=1)
legend('bottomright', legend=c('W', 'Exp(0.5)'), lty=1:1,
       col=c('royalblue', 'orange'), title='Distributions', cex=0.8)
```



Q3

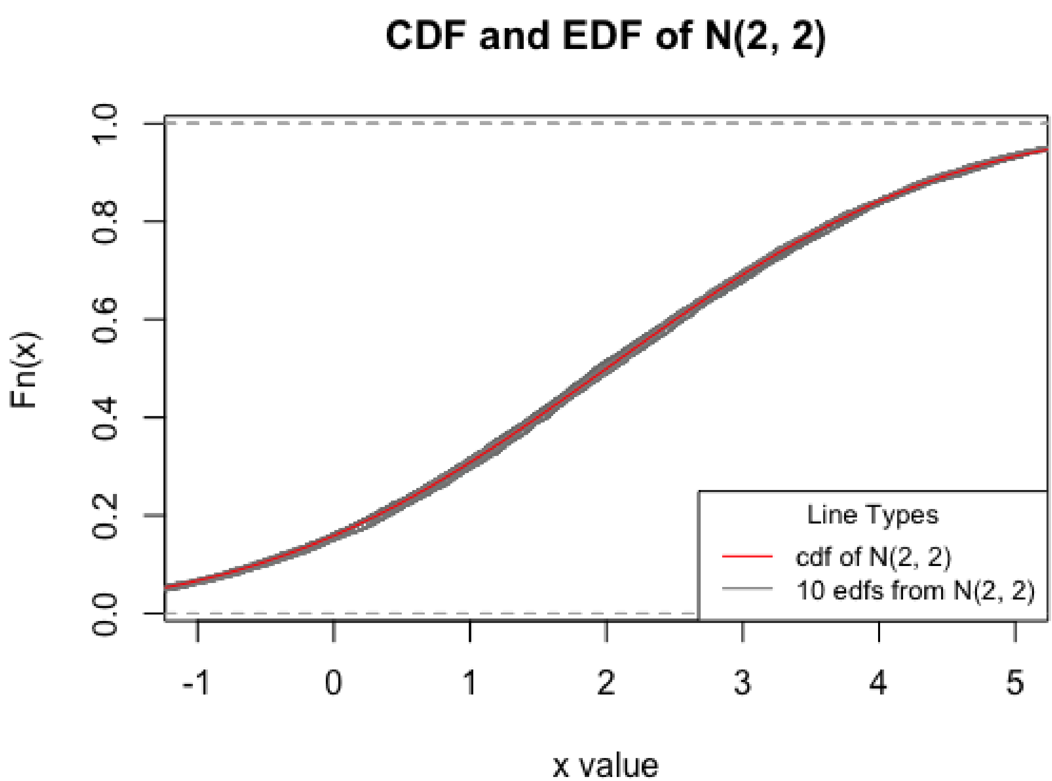
Q3-1

```
norm_value <- rnorm(5000, 2, 2)
norm_edf <- ecdf(norm_value)
plot(norm_edf, xlim=c(-1, 5), col='limegreen', lwd=1,
     main='EDF of N(2, 2)')
```



Q3-2

```
x_base <- seq(-2, 6, 0.01)
norm_cdf <- pnorm(x_base, 2, 2)
plot(x_base, norm_cdf, xlim=c(-1, 5), type='l', lwd=1, col='red', xlab='x value',
     ylab='Fn(x)', main='CDF and EDF of N(2, 2)')
for(i in 1:10) {
  edf_value <- ecdf(rnorm(5000, 2, 2))
  lines(edf_value, col=paste('gray', i+45), lwd=1)
}
lines(x_base, norm_cdf, xlim=c(-1, 5), col='red', lwd=1)
legend('bottomright', legend=c('cdf of N(2, 2)', '10 edfs from N(2, 2)'), lty=1:1,
       col=c('red', 'gray50'), title='Line Types', cex=0.8)
```



The edf from samples of 5000 data points seems to be a good estimator of cdf.

Q4

We have proved that $\mathbb{E}(\widehat{F_n}) = F_X$, $\widehat{F_n}$ is an unbiased estimator for F_X.

$$
\begin{aligned}
SE(\widehat{F_n}) =& \sqrt{Var(\widehat{F_n})} \\
Var(\widehat{F_n}) =& Var(\frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)) \\
Var(\widehat{F_n}) =& \frac{\widehat{F_n}(1-\widehat{F_n})}{n} \text{ since } I(x_i \leq x) \text{ follows } Bern(\widehat{F_n}) \\
SE(\widehat{F_n}) =& \sqrt{\frac{\widehat{F_n}(1-\widehat{F_n})}{n}}
\end{aligned}
$$

By Central Limit Theorem, $\widehat{F_n}$ follows normal distribution, therefore $\frac{\widehat{F_n}-\mathbb{E}(\widehat{F_n})}{SE(\widehat{F_n}(x_0))}$ follows standard normal distribution

The confidence interval of $\widehat{F_n}$ under α can be written as

$$
\begin{aligned}
\widehat{F_n} &\pm MOE \\
MOE &= Z_{1-\frac{\alpha}{2}} SE(\widehat{F_n}) \\
&= Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{F_n}(1-\widehat{F_n})}{n}}
\end{aligned}
$$

Therefore, for given point x_0, the $1-\alpha$ confidence interval of $F_X(x_0)$ is

$$
\widehat{F_n}(x_0) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\widehat{F_n}(x_0)(1-\widehat{F_n}(x_0))}{n}}
$$