1. Let $X_1, \cdots, X_n$ be i.i.d. $\mathsf{Ber}(p)$, a Bernoulli random variable with parameter p. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average.

   (a) What is the expected value $\mathbb{E}\left(\bar{X}_n\right)$? What is the variance $\mathsf{Var}(\bar{X}_n)$?

   (b) Explain what is meant by a *consistent* estimator of p, and show that $\bar{X}_n$ a consistent estimator of p?

   (c) Explain why $\sqrt{n}\left(\bar{X}_n - p\right)$ can be approximated by a $N(0, p(1-p))$ distribution when n is large.

(a) $E(\bar{X}_n) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n} E(\sum_{i=1}^{n} X_i) = \frac{1}{n}[1 \times p \times n + 0 \times (1-p) \times n] = p$

$\mathrm{var}[\bar{X}_n] = \mathrm{var}(\frac{1}{n}\sum_{i=1}^{n} X_i)$

$= \frac{1}{n^2}\mathrm{var}(\sum_{i=1}^{n} X_i)$

$= \frac{1}{n^2} np(1-p)$

$= \frac{1}{n} p(1-p)$

(b) A consistent estimator means that as the sample size gets large the estimate gets closer and closer to the true value of the parameter.

$\lim_{n\to\infty} p(|\bar{X}_n - p| > a) = 0, \quad a > 0$

(c) $Y_n = \sqrt{n}(\bar{X}_n - p)$

$E(Y_n) = 0$

$\mathrm{Var}(Y_n) = \mathrm{var}(\sqrt{n}(\bar{X}_n - p)) = (\sqrt{n})^2 \frac{1}{n} p(1-p) = p(1-p)$

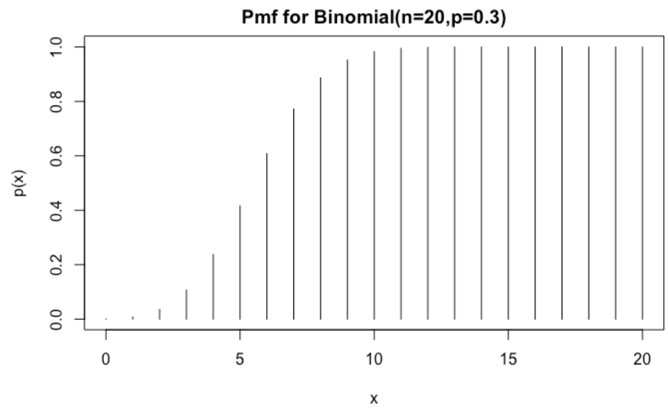
According to the central limit theorem, when n gets larger, $Y_n \sim N(0, p(1-p))$.

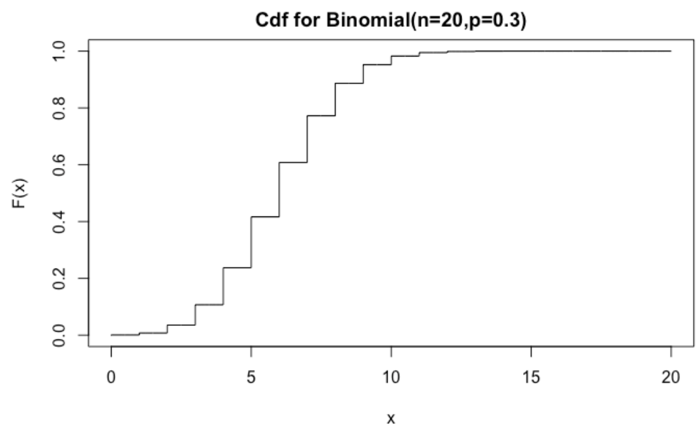
2. Use the R-functions **dbinom()**, **dnorm()** and/or **pbinom()**, **pnorm()** to complete this question.

   (a) Plot the probability mass function and cumulative distribution function (cdf) of a Binomial random variable $Y \sim Bin(n, p)$ when $n = 20$ and $p = 0.3$.

   (b) What Normal distribution might approximate the distribution of Y?

   (c) Superimpose on your plots the probability density function and cdf of this Normal distribution (in a contrasting color such as blue or red)

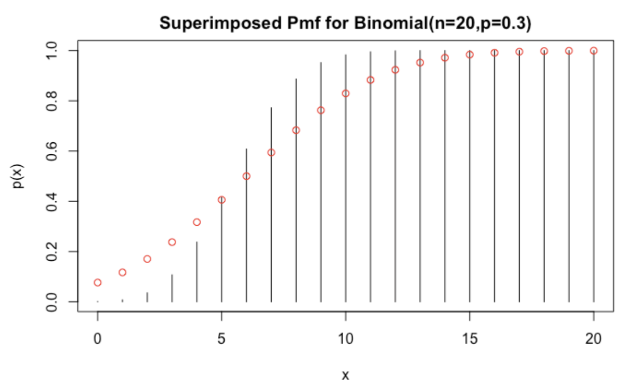


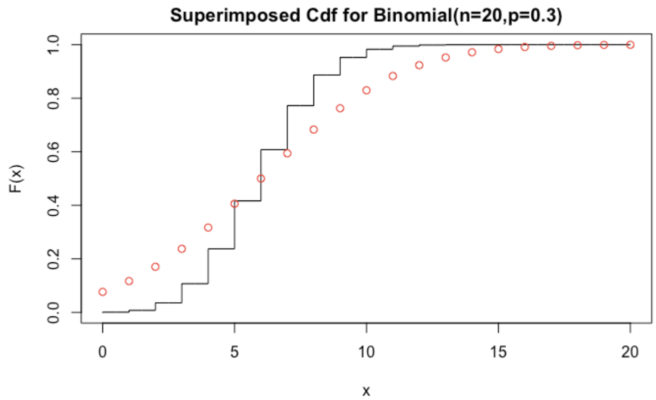


(b) We can use $N(np, npq)$ to approximate the distribution.

$n=20$, $p=0.3$, $q=1-0.3=0.7$

$N(6, 4.2)$





3. Let $X_1, \cdots, X_n$ be .i.d from $Exp(1/\beta)$, with pdf $f(x) = \frac{1}{\beta}e^{-x/\beta}$ on $x \geq 0$.
Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample average.

(a) What are the bias and variance of $\bar{X}_n$ as an estimator of β?

(b) What is the mean square error of using $\bar{X}_n$ as the estimator of β? Does $\bar{X}_n$ converges to β? Why?

(c) Now consider a new estimator $\widetilde{\beta} = a\,\bar{X}_n$, where $a \in \mathbb{R}$ is a real number. Show that the mean square error of $\widetilde{\beta}$ is $[n(a-1)^2 + a^2]\beta^2/n$.

(d) Find the value of a that minimizes this MSE and show that the resulting MSE is smaller than the MSE of $\bar{X}_n$. (Note: this new estimator is related to the *Stein's Shrinkage estimator*.)

3.(a) $E(x) = \int_0^\infty x \frac{1}{\beta} e^{-\frac{x}{\beta}} = \beta$

$\text{Bias}(x) = E(x) - \bar{x}_n = \beta - \frac{1}{n}\sum_{i=1}^{n} x_i$

$\text{var}(x) = E(x^2) - (E(x))^2 = \beta^2$

(b) $MSE = \frac{1}{n}\sum_{i=1}^{n} (\bar{x}_n - x_i)^2$

$= \frac{1}{n}\sum_{i=1}^{n} (\bar{x}_n^2 + x_i^2 - 2\bar{x}_n x_i)$

$= \bar{x}_n^2 - 2\bar{x}_n^2 + \frac{1}{n}\sum_{i=1}^{n} x_i^2$

$= \text{var}(\bar{x}_n) + (E(\bar{x}_n) - \beta)^2$

$E(\bar{x}_n) = E(\frac{1}{n}\sum_{i=1}^{n} x_i)$

$= \frac{1}{n}\sum_{i=1}^{n} E(x_i)$

$= \frac{1}{n} \cdot n \cdot E(x)$

$= \beta$ So it does converge to β

(c) $E(a\bar{x}_n) = a\beta$

$\text{var}(a\bar{x}_n) = a^2 \frac{\beta^2}{n}$

$MSE(a\bar{x}_n)$

$= E[(a\bar{X}_n - E(a\bar{X}_n)] + E(a\bar{X}_n) - a\beta$

$= var(a\bar{X}_n) + (E(a\bar{X}_n) - \beta)^2$

$= [n(a-1)^2 + a^2] \frac{\beta^2}{n}$

(d) $MSE' = [2n(a-1) + 2a] \frac{\beta^2}{n} = 0 \quad \Rightarrow a = \frac{n}{n+1}$

$a \in (-\infty, \frac{n}{n+1})$, $MSE' < 0$, MSE decreases

$a \in (\frac{n}{n+1}, \infty)$, $MSE' > 0$, MSE increases.

So when $a = \frac{n}{n+1}$, MSE has the smallest value.

4. Let $X_1, X_2 \sim \text{Uni}[2,4]$. Let $Y = (X_1 + X_2)/2$.

(a) Use the R-function **runif**() to generate at least 10000 realizations of Y and plot the corresponding histogram.

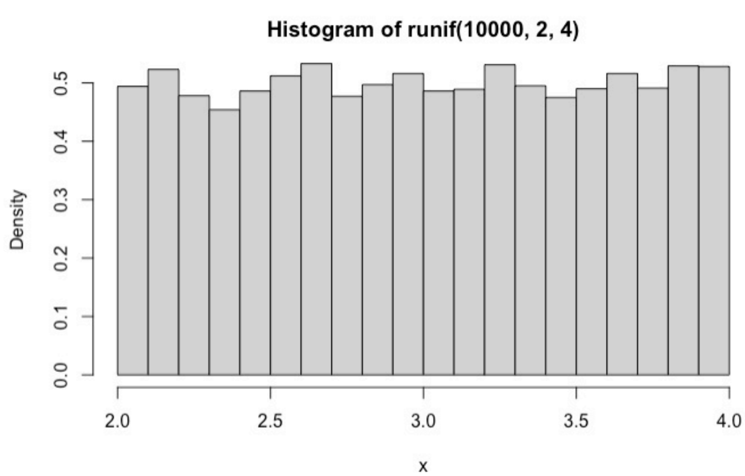
(b) The density curve of Y is

$$f_Y(x) = \begin{cases} x-2, & \text{when } 2 \le x \le 3 \\ 4-x, & \text{when } 3 < x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

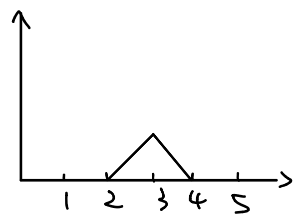
Verify that $f_Y(x)$ is a density function (pdf).

(c) Superimpose the $f_Y(x)$ density curve on your histogram (in a contrasting color).

(a)



(b) $f_Y(x) = \begin{cases} x-2 & 2 \le x \le 3 \\ 4-x & 3 < x \le 4 \\ 0 & \text{otherwise} \end{cases}$



$$\int_2^4 f_Y(x)\,dx$$
$$= \int_2^3 (x-2)\,dx + \int_3^4 (4-x)\,dx$$
$$= \left(\frac{x^2}{2} - 2x\right)\Big]_2^3 + \left(4x - \frac{x^2}{2}\right)\Big]_3^4$$
$$= \frac{9}{2} - 6 - \frac{4}{2} + 4 + 16 - \frac{16}{2} - 12 + \frac{9}{2}$$
$$= \frac{1}{2} + \frac{1}{2}$$
$$= 1$$

It integrates to 1, so it's a density function.

(c)