Probability Distribution of Continuous R. V.

Question: **Consider the binomial experiment with $n = 10,000$ and the probability of success $p = 0.3$. Let X be the number of successes in the $10,000$ trials. Find the approximate probability $\mathbb{P}(X \leq 6000)$.**

Approaching Idea: **We know the distribution table of binomial:**

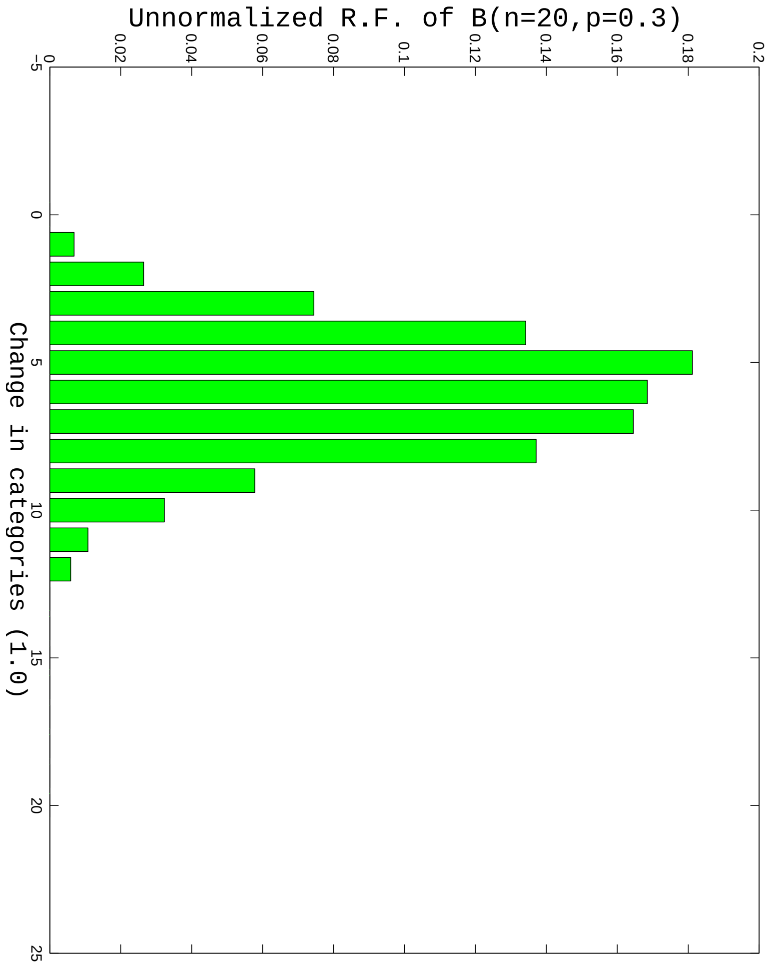
The distribution table of r.v X with $n = 10,000$

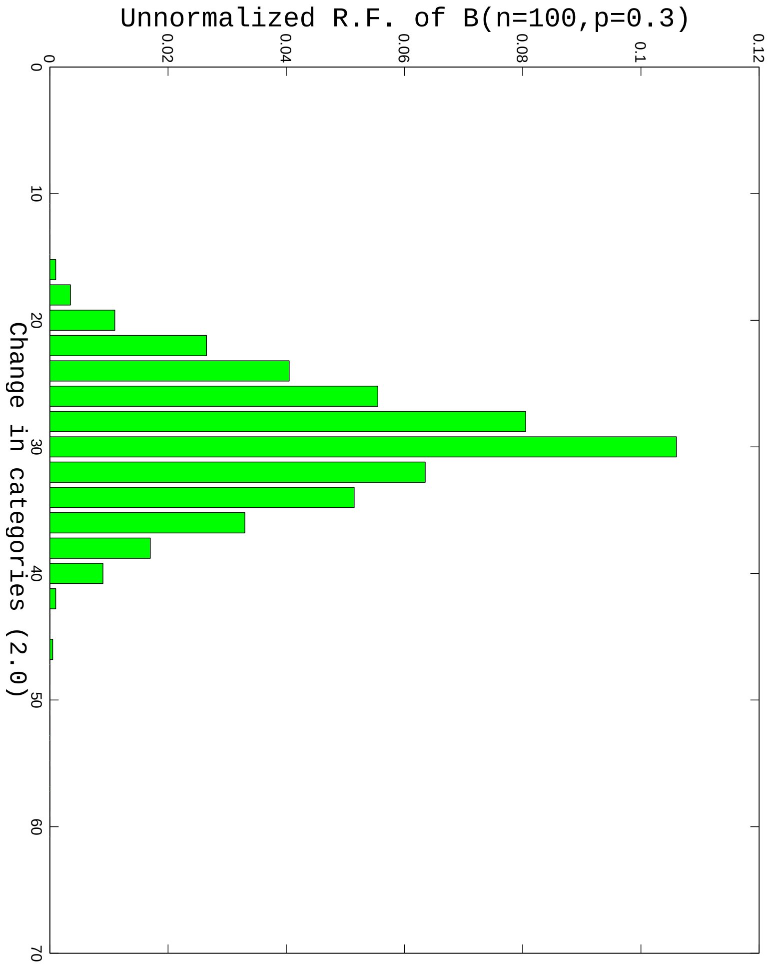
x	0	1	2	$\cdots$	n
$p(x)$	$C_0^n p^0 q^n$	$C_1^n p^1 q^{n-1}$	$C_2^n p^2 q^{n-2}$	$\cdots$	$C_n^n p^n q^0$

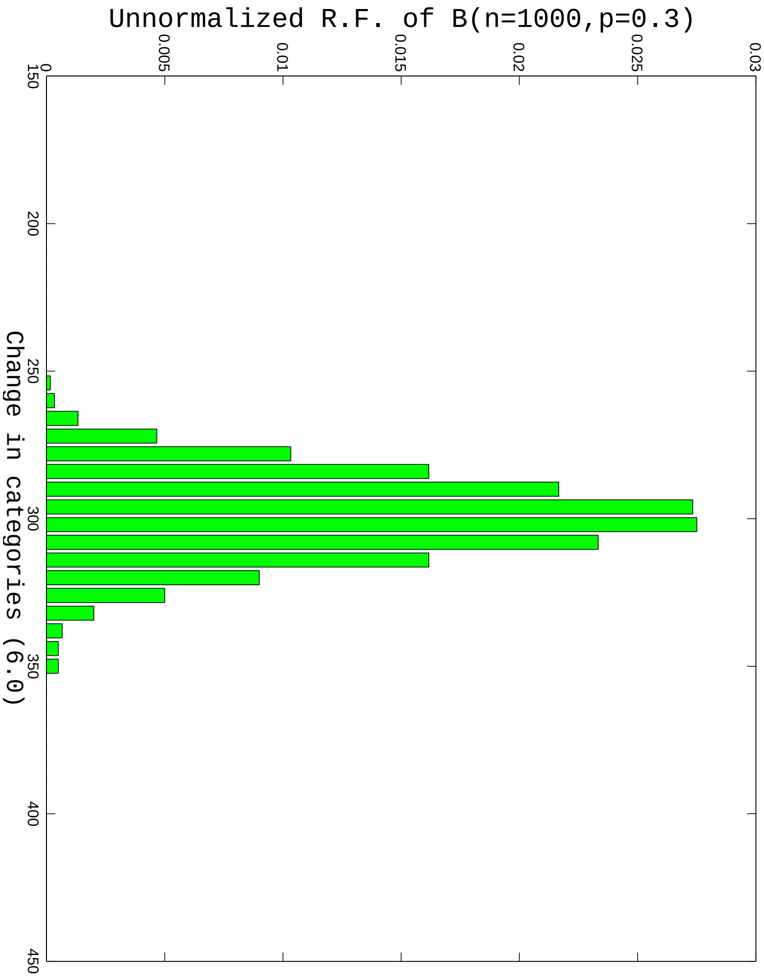
We have

$$\mathbb{P}(X \leq 6000) = \sum_{i=1}^{6000} C_i^n p^i q^{n-i}$$

This is equal to the area of the corresponding histogram between 0 and 6000 (See corresponding graph). Do the histogram top broken lines converge to a curve as $n \to \infty$? If the corresponding area under the curve is easy to be calculated, this would be a wonderful idea. See the following histograms:


Unnormalized R.F. of B(n=20,p=0.3)
Change in categories (1.0)
0 0.02 0.04 0.06 0.08 0.1 0.12 0.14 0.16 0.18 0.2
−5 0 5 10 15 20 25


Unnormalized R.F. of B(n=100,p=0.3)
Change in categories (2.0)
0
0.02
0.04
0.06
0.08
0.1
0.12
0
10
20
30
40
50
60
70

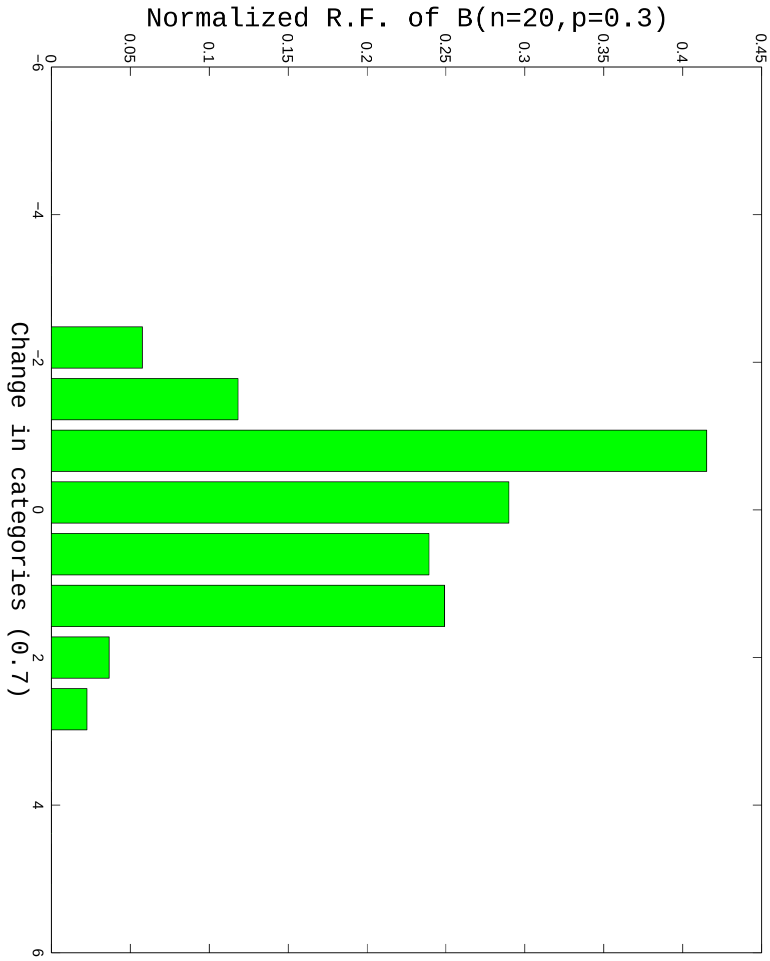


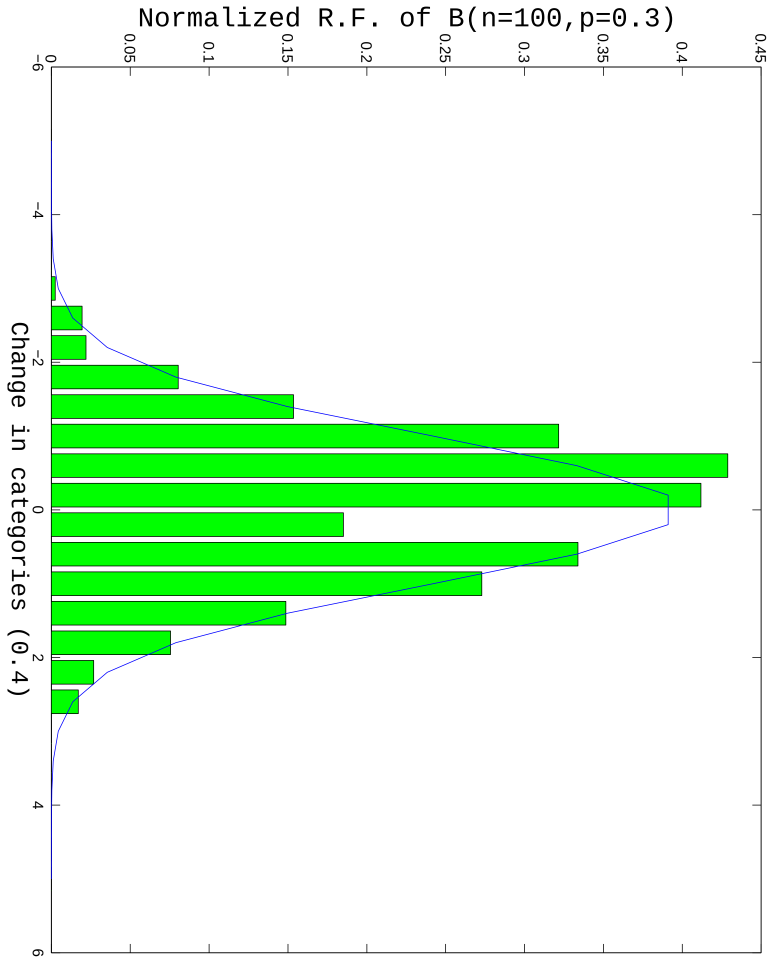
Unnormalized R.F. of B(n=1000,p=0.3)
0
0.005
0.01
0.015
0.02
0.025
0.03
150
200
250
300
350
400
450
Change in categories (6.0)

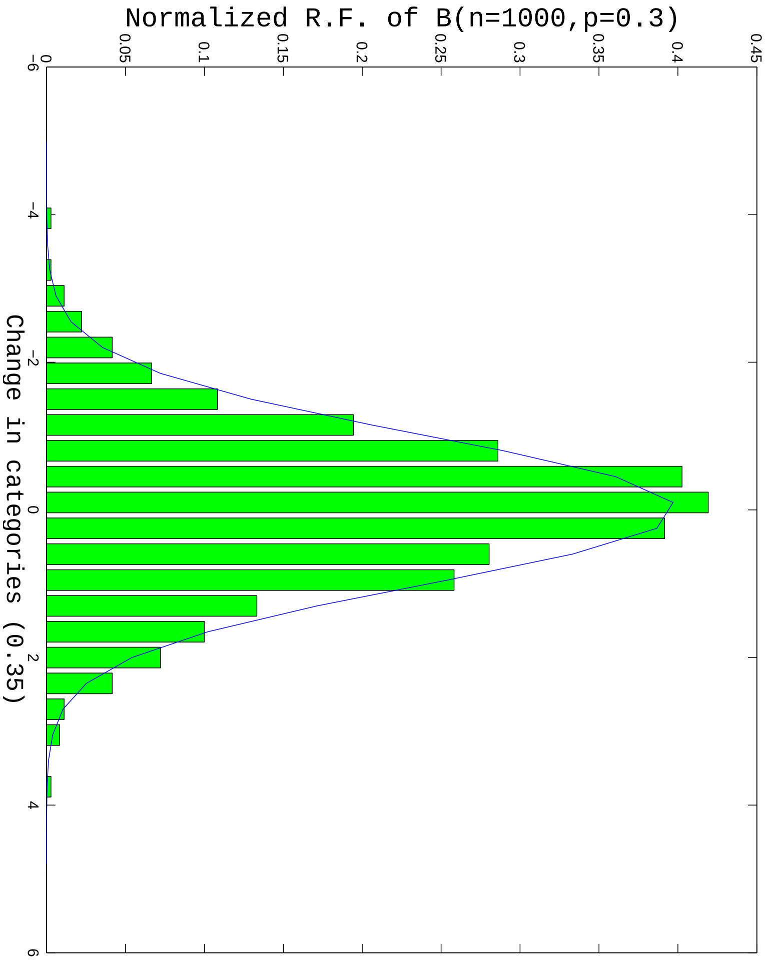
The means keep moving as n changes. This is not convenient. Therefore, instead of considering X itself, we consider a normalized random variable

$$Y = \frac{X - np}{\sqrt{npq}}$$

Then, we have following histograms of Y.


Normalized R.F. of B(n=20,p=0.3)
Change in categories (0.7)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
-6
-4
-2
0
2
4
6


Normalized R.F. of B(n=100,p=0.3)
Change in categories (0.4)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
−6
−4
−2
0
2
4
6


Normalized R.F. of B(n=1000,p=0.3)
Change in categories (0.35)
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
−6
−4
−2
0
2
4
6

In the case of r.v. Y, its mean is always equal to zero and its variance is always equal to 1. Its histograms converge to a curve with following function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Then it is easy to find any area under this curve, above 0, and between any two numbers $a < b$ from the normal distribution table (at the end of your textbook).

Definition 11.1 **If a r.v., say W, has a distribution density which is given by a continuous curve , then, W is called a continuous r.v..**

Remark:

(a) The range space of a continuous r.v. is an uncountable infinite set such as interval $[a, b]$, (c, d), and so on.

(b) For any continuous r.v. and any constant $a \in \mathbb{R}$, $\mathbb{P}(W = a) = 0$.

Definition 11.2 **If a r.v., say W, has a distribution density which is given by the following bell shaped continuous curve or precisely the function**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\{-(x-\mu)^2/(2\sigma^2)\}}$$

Then W is called a normal r.v., its distribution is called normal distribution, and its distribution density is called normal density. In particular, if $\mu = 0$ and $\sigma = 1$, then W is called a standard normal r.v., its distribution is called standard normal distribution, and its distribution density is called standard normal density.

Definition 11.2 If a r.v., say W, has a distribution density which is given by the following bell shaped continuous curve or precisely the function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\{-(x-\mu)^2/(2\sigma^2)\}}$$

Then W is called a normal r.v., its distribution is called normal distribution, and its distribution density is called normal density. In particular, if $\mu = 0$ and $\sigma = 1$, then W is called a standard normal r.v., its distribution is called standard normal distribution, and its distribution density is called standard normal density.

Example 11.1 Let X be a standard normal r.v.. Find following probabilities:

(a) $\mathbb{P}(X > 2)$;

(b) $\mathbb{P}(-1 \leq X < 3)$.

Solution: According to p672 Table , we get

$$\mathbb{P}(-\infty \leq X \leq 2) = 0.9772,$$

therefore,

$$\mathbb{P}(X > 2) = 1 - \mathbb{P}(-\infty \leq X \leq 2) = 1 - 0.9772 = 0.0228,$$

(b)

$$\mathbb{P}(-1 \leq X < 3)$$

$$= \mathbb{P}(-\infty \leq X < 3) - \mathbb{P}(-\infty \leq X \leq -1)$$

$$= 0.9987 - 0.1587 = 0.84$$

Example 12.2 :
Let X be a standard normal r.v.. Find the probability $\mathbb{P}(-2.25 < X \leq -1.16)$.
Solution : **According to the table, we have**

$$\begin{aligned}
&\mathbb{P}(-2.25 < X \leq -1.16) \\
&= \mathbb{P}(-\infty \leq X \leq -1.16) - \mathbb{P}(-\infty \leq X \leq -2.25) \\
&= 0.123 - 0.0122 \\
&= 0.1108
\end{aligned}$$

Non standard normal r.v. :
There is a transformation between standard normal r.v. and non-standard normal r.v..

Definition 13.1 **For given number μ and positive number σ^2, if X is a normal r.v. with mean equal to μ and variance equal to σ^2, then X is called a non-standard normal r.v.. Its distribution is denoted by $N(\mu, \sigma)$.**

Given a non-standard normal r.v. X with distribution $N(\mu, \sigma)$, define

$$Y = \frac{X - \mu}{\sigma}.$$

Then, Y is a standard normal r.v. with distribution $N(0, 1)$. Conversely, given a standard r.v. U with normal distribution $N(0, 1)$, then

$$V = \mu_1 + \sigma_1 U$$

is a non-standard normal r.v. with normal distribution $N(\mu_1, \sigma_1)$
The proof of above result is beyond the scope of this course. However, since continuous r.v. can be looked as a limit of a sequence of discrete r.v's, we can take a look at a discrete case.

Example 13.1 **Let** X **be discrete r.v. with distribution** $\mathbb{P}(X = 0) = 0.5$ **and** $\mathbb{P}(X = 1) = 0.5$**. Find the mean** μ **of** X **and the standard deviation** σ **of** X**. Define** $Y = (X - \mu)/\sigma$**. Find the mean and standard variance of** Y**.**
Solution:

$$\mu = 0(0.5) + 1(0.5) = 0.5$$

$$\sigma^2 = \mathbb{E}(X - \mu)^2 = (0 - 0.5)^2(0.5) + (1 - 0.5)^2(0.5)$$

$$= (0.5)[(0.5)^2 + (0.5)^2] = (0.5)(0.5) = 0.25$$

Thus, $\sigma = \sqrt{0.25} = 0.5$ **and** $Y = (X - 0.5)/(0.5)$**.**

$$\mathbb{E}(Y) = [(0 - 0.5)/(0.5)](0.5) + [(1 - 0.5)/(0.5)](0.5)$$

$$= [-1](0.5) + [1](0.5) = 0$$

$$\mathbb{E}(Y - 0)^2 = \mathbb{E}(Y)^2 = [(0 - 0.5)/(0.5)]^2(0.5) + [(1 - 0.5)/(0.5)]^2(0.5)$$

$$= [-1]^2(0.5) + [1]^2(0.5) = 1$$

Example 13.2 **Let** X **be a standard normal r.v. Find**

the probabilities of the following events:
(a) $\mathbb{P}((X-1)/2>0)$**;**
(b) $\mathbb{P}(X<1.2)$
(c) $\mathbb{P}(X^2<4)$

<u>Solution:</u> (a) We have

$$\mathbb{P}((X-1)/2>0)=\mathbb{P}(X-1>0)=\mathbb{P}(X>1)$$
$$=1-\mathbb{P}(-\infty\le X\le 1)$$
$$=1-0.8413=0.1587$$

(b)

$$\mathbb{P}(X<1.2)=0.8849$$

(c)

$$\mathbb{P}(X^2<4)=\mathbb{P}(-2<X<2)$$
$$=\mathbb{P}(-\infty\le X<2)-\mathbb{P}(-\infty\le X\le -2)$$
$$=0.9772-0.0228=0.9544$$

<u>Normal Approximation to Binomial Distribution :</u>
Suppose that X is a binomial r.v. with distribution $b(n,p)$. If the following condition (1) is satisfied, then the following formulae can be used to find the corresponding probabilities.

<u>Procedure:</u>
(1) Check conditions: $np>5$ and $nq>5$.
(2) Find $\mu=np$, p is the success probability. $\sigma=\sqrt{npq}$.
(3) Let

$$Z=\frac{X-np}{\sqrt{npq}}$$

Then Z is an approximate standard normal r.v.. For given constant integers a, b, use following formulae to find approximate probabilities:

(I)
$$\begin{aligned}\mathbb{P}(X \geq a) &= \mathbb{P}(X \geq a - 0.5) = \mathbb{P}(X - np \geq a - 0.5 - np) \\ &= \mathbb{P}(\tfrac{X-np}{\sqrt{npq}} \geq \tfrac{a-0.5-np}{\sqrt{npq}}) \\ &= \mathbb{P}(Z \geq \tfrac{a-0.5-np}{\sqrt{npq}})\end{aligned}$$

(II)
$$\begin{aligned}\mathbb{P}(X \leq b) &= \mathbb{P}(X \leq b + 0.5) = \mathbb{P}(X - np \leq b + 0.5 - np) \\ &= \mathbb{P}(\tfrac{X-np}{\sqrt{npq}} \leq \tfrac{b+0.5-np}{\sqrt{npq}}) \\ &= \mathbb{P}(Z \leq \tfrac{b+0.5-np}{\sqrt{npq}})\end{aligned}$$

(III)
$$\begin{aligned}\mathbb{P}(a \leq X \leq b) &= \mathbb{P}(a - 0.5 \leq X \leq b + 0.5) \\ &= \mathbb{P}(a - 0.5 - np \leq X - np \leq b + 0.5 - np) \\ &= \mathbb{P}(\tfrac{a-0.5-np}{\sqrt{npq}} \leq \tfrac{X-np}{\sqrt{npq}} \leq \tfrac{b+0.5-np}{\sqrt{npq}}) \\ &= \mathbb{P}(\tfrac{a-0.5-np}{\sqrt{npq}} \leq Z \leq \tfrac{b+0.5-np}{\sqrt{npq}})\end{aligned}$$

<u>Example 1</u> Consider the binomial experiment with $n = 10,000$ and the probability of success $p = 0.3$. Let X be the number of successes in the $10,000$ trials. Find the approximate probability $\mathbb{P}(X \leq 3100)$.

<u>Solution:</u> (I) $np = 10,000(0.3) = 3,000 > 5$ **and** $nq = 10,000(0.7) = 7,000 > 5$.

(II) $\mu = np = 3,000$**,** $\sigma = \sqrt{npq} = \sqrt{10,000(0.3)(0.7)} = 45.82$
(III) To find the probability $\mathbb{P}(X \leq 3,100)$**, according to (II), we have**

$$\begin{aligned}\mathbb{P}(X \leq b) &= \mathbb{P}(X \leq b + 0.5) = \mathbb{P}(X - np \leq b + 0.5 - np) \\ &= \mathbb{P}(\tfrac{X-np}{\sqrt{npq}} \leq \tfrac{b+0.5-np}{\sqrt{npq}}) \\ &= \mathbb{P}(Z \leq \tfrac{b+0.5-np}{\sqrt{npq}})\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X \leq 3,100) &= \mathbb{P}(Z \leq \tfrac{3,100+0.5-3,000}{45.82}) = \mathbb{P}(Z \leq 2.19) \\ &= 0.9857\end{aligned}$$

Example :
An airline finds that 5% **of the persons making reservations on a certain flight will not show up for the flight. If the airline sells** 160 **tickets for a flight that has only** 155 **seats, what is the probability that a seat will be available for every person holding a reservation and claiming the reservation later on?**
Solution:
Let X **be the number of persons making reservations and later on claiming their reservations. According to the information, the success probability** $p = 95\%$
(I) $np = 160(0.95) = 152 > 5$ **and** $nq = 160(0.05) = 8 > 5$**.**
(II) $\mu = np = 152$**,** $\sigma = \sqrt{npq} = \sqrt{160(0.95)(0.05)} = 2.756$
(III) The question is just to find the probability $\mathbb{P}(X \leq$

155). **According to (II), we have**

$$\begin{aligned}\mathbb{P}(X \le b) &= \mathbb{P}(X \le b + 0.5) = \mathbb{P}(X - np \le b + 0.5 - np)\\ &= \mathbb{P}(\tfrac{X-np}{\sqrt{npq}} \le \tfrac{b+0.5-np}{\sqrt{npq}})\\ &= \mathbb{P}(Z \le \tfrac{b+0.5-np}{\sqrt{npq}})\end{aligned}$$

$$\begin{aligned}\mathbb{P}(X \le 155) &= \mathbb{P}(Z \le \tfrac{155+0.5-152}{2.756}) = \mathbb{P}(Z \le 1.269)\\ &= \mathbb{P}(-\infty \le Z \le 1.269) = 0.898\end{aligned}$$

<u>**Example 3**</u> **Hotels often grant reservations in excess of capacity to minimize losses due to no-shows. Suppose the records of a motel show that, on the average,** 10% **of their prospective guests will not claim their reservations. If the motel accepts** 215 **reservations and there are only** 200 **rooms in the motel, what is the probability that all guests who arrive to claim a room will receive one?**

<u>**Solution:**</u> **Let** X **be the number of persons making reservations and later on not show up. According to the information, the not claiming probability** $p = 10\%$

(I) $np = 215(0.1) = 21.5 > 5$ **and** $nq = 215(0.9) = 193.5 > 5$**.**

(II) $\mu = np = 21.5$, $\sigma = \sqrt{npq} = \sqrt{215(0.1)(0.9)} = 4.398$

(III) The question is just to find the probability $\mathbb{P}(X \ge$

15). **According to (I), we have**

$$\mathbb{P}(X \geq a) = \mathbb{P}(X \geq a - 0.5) = \mathbb{P}(X - np \geq a - 0.5 - np)$$
$$= \mathbb{P}(\tfrac{X-np}{\sqrt{npq}} \geq \tfrac{a-0.5-np}{\sqrt{npq}})$$
$$= \mathbb{P}(Z \geq \tfrac{a-0.5-np}{\sqrt{npq}})$$

$$\mathbb{P}(X \geq 15) = \mathbb{P}(Z \geq \tfrac{15-0.5-21.5}{4.398}) = \mathbb{P}(Z \geq -1.591)$$
$$= \mathbb{P}(-\infty \leq Z \leq 1.591) = 0.9441$$

<u>Example :</u> Consider the binomial experiment with $n = 10,000$ **and the probability of success** $p = 0.3$**. Let** X **be the number of successes in the** $10,000$ **trials. Find the approximate probability** $\mathbb{P}(3000 \leq X \leq 3100)$**.**

<u>Solution:</u> (I) $np = 10,000(0.3) = 3,000 > 5$ **and** $nq = 10,000(0.7) = 7,000 > 5$**.**

(II) $\mu = np = 3,000$**,** $\sigma = \sqrt{npq} = \sqrt{10,000(0.3)(0.7)} = 45.82$

(III) To find the probability $\mathbb{P}(3000 \leq X \leq 3,100)$**, according to (III), we have**

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(\frac{a-0.5-np}{\sqrt{npq}} \leq Z \leq \frac{b+0.5-np}{\sqrt{npq}})$$

$$\mathbb{P}(3000 \leq X \leq 3,100) = \mathbb{P}(\tfrac{3,000-0.5-3,000}{45.82} \leq Z \leq \tfrac{3,100+0.5-3,000}{45.82})$$
$$= \mathbb{P}(-0.01 \leq Z \leq 2.19)$$
$$= \mathbb{P}(0 \leq Z \leq 0.01) + \mathbb{P}(0 \leq Z \leq 2.19)$$
$$= 0.004 + 0.4857 = 0.4861$$