Chapter 5

Random Variable and Probability Distribution

1 Random Variable (r.v.)

Definition

A **random variable** is a function that assigns a real number to each outcome of an experiment. A **discrete random variable** is a random variable whose range is either finite or countable infinite. If the range of value is (uncountably) infinite, the random variable is called **continuous random variable.**

Example 1

Consider the experiment of tossing a coin twice, and let X denote the random variable that measures the number of times head comes up.

a. List all of the possible outcomes of this experiment and the value that the random variable X assigns to each outcome.

b. List the outcomes that make up the event X=1.

Solution

a. The possible outcomes are: HH, HT, TH, and TT. Since X assigns to each output the number of times heads comes up, we obtain

Outcome	HH	HT	TH	TT
X	2	1	1	0

c. The event [X=1] is the set of outcomes for which [X=1]. It is the event {HT, TH}.

Example 2

In a certain experiment, one mouse is selected at random from a collection of 5 mice whose weights (in grams) are summarized in the following table:

Mouse	A	B	C	D	E
Weight	65	47	59	31	47

Let X be the weight of the selected mouse. List the outcomes that make up the following events:

a. [X=65]; **b.** [X=47]; **c.** $[X \geq 30]$

Solution

a. Only mouse A weights 65 grams, so [X=65] ={A}

b. Mice B and E weight 47, so [X=47]={B, E}

c. All mice weight 30 grams or more, so $[X \geq 30]$={A,B,C,D,E}

2 Discrete Probability Distribution

A probability distribution for random variable X is the set of probabilities corresponding to all possible values of X such that

(i) $0 \leq P(X) \leq 1$ and

(ii) $\sum P(X) = 1$

Example 1

A box contains 3 red balls and 2 green balls. Two balls are selected at random, with replacement. Find the probability distribution of the random variable X that measures the number of red balls selected.

Solution

R stands for red ball and *G* for green one. If we choose two balls with replacement, the sample space is {*RR*, *RG*, *GR*, *GG*}.

The probability of getting no red balls, $P(GG)=\frac{2}{5}\times\frac{2}{5}=\frac{4}{25}$

The probability of getting one red balls, $P(GR \text{ or } RG)=\frac{2}{5}\times\frac{3}{5}+\frac{2}{5}\times\frac{3}{5}=\frac{12}{25}$

The probability of getting two red balls, $P(RR)=\frac{3}{5}\times\frac{3}{5}=\frac{9}{25}$

Hence, we can get a table of probability distribution

X	0	1	2
$P(X)$	$\frac{4}{25}$	$\frac{12}{25}$	$\frac{9}{25}$

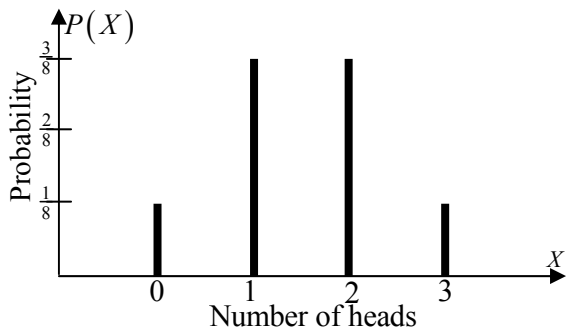
This probability distribution can be represented by a graph.

Example 2

Represent graphically the probability distribution for the sample space for tossing tree coins.

Number of heads X	0	1	2	3
Probability $P(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Solution



Example 3

During the summer months, a rental agency keeps track of the number of chain saws it rents each day during a period of 90 days. The number of saws rented per day is represented by the variable *X*. The results are shown here. Compute the probability $P(X)$ for each *X*, and construct a probability distribution and graph for the data.

X	**Number of days**
0	45
1	30
2	15

Solution

The probability $P(X)$ can be computed for each *X* by dividing the number of days that *X* saws were rented by total days.

For 0 saws: $\frac{45}{90}=0.50$

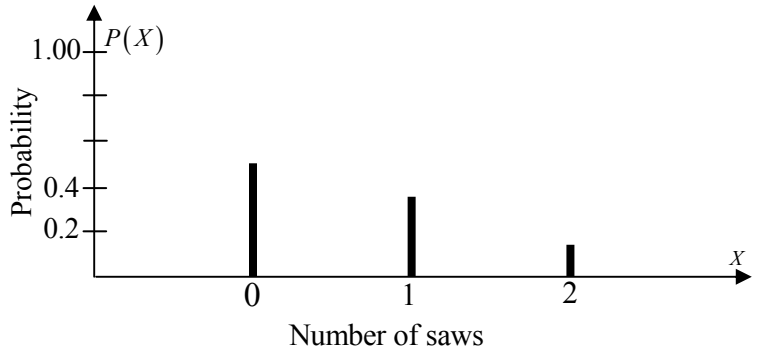
For 1 saws: $\frac{30}{90} = 0.33$

For 2 saws: $\frac{15}{90} = 0.17$

The distribution is shown here.

Number of saws rented X	0	1	2
Probability $P(X)$	0.50	0.35	0.17

Solution



3 Mean, Variance, and Expectation

3.1 Mean

The mean of a random variable with a discrete probability distribution is

$$\mu = x_1P(x_1) + x_2P(x_2) + x_3P(x_3) + \cdots + x_nP(x_n) = \sum xP(x)$$

Example 4

Find the mean of the number of spots that appear when a die is tossed.

Solution

Outcome	1	2	3	4	5	6
Probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mu = \sum xP(x) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + 3\cdot\frac{1}{6} + 4\cdot\frac{1}{6} + 5\cdot\frac{1}{6} + 6\cdot\frac{1}{6} = \frac{21}{6} = 3.5$$

Example 5

In a family with 2 children, find the mean of the number of children who will be girl.

Solution

Number of girl X	0	1	2
Probability $P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Hence, the mean is $\mu = \sum x\cdot P(x) = 0\cdot\frac{1}{4} + 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4} = 1$

Example 6

The probability distribution shown represents the number of trips of 5 nights or more that American adults take per year. (That is, 6% do not take the trip lasting 5 nights or more, 70% take one trip lasting 5 nights or more per year, etc.) Find the mean.

Number of trip X	0	1	2	3	4
Probability $P(X)$	0.06	0.70	0.20	0.03	0.01

Solution

$$\begin{aligned}\mu &= \sum x \cdot P(x) \\ &= 0\times 0.06 + 1\times 0.70 + 2\times 0.20 + 3\times 0.03 + 4\times 0.01 \\ &= 1.23 \approx 1.2\end{aligned}$$

3.2 Variance

To find the variance for the random variable of a probability distribution, subtract the theoretical mean of the random variable from each outcome and square the difference. Then multiply each difference by its corresponding probability and add the products. The formula is

$$\sigma^2 = \sum\left[(x-\mu)^2 \cdot P(x)\right]$$

Formula for the computation of variance of a probability distribution

$$\sigma^2 = \sum\left[x^2 \cdot P(x)\right] - \mu^2$$

The standard deviation of a probability distribution is

$$\sigma = \sqrt{\sigma^2}$$

Example 7

Find the variance and standard deviation for the probability distribution of the number of spots that appear when a die is tossed.

Solution

From the previous example, we have $\mu = 3.5$, then

$$\sigma^2 = \left(1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}\right) - (3.5)^2 = 2.9$$

To get the standard deviation, find the square root of the variance

$$\sigma = \sqrt{2.9} = 1.7$$

Example 8

Five balls numbered 0, 2, 4, 6, and 8 are placed in a bag. After the balls are mixed, one is selected, its number is noted, and then it is replaced. If this experiment is repeated many times, find the variance and standard deviation of the numbers on the balls.

Solution

The probability of getting each number is $\frac{1}{5}$ or 0.2.

x	$P(x)$	$x \cdot P(x)$	$x^2 \cdot P(x)$
0	0.2	0	0
2	0.2	0.4	0.8
4	0.2	0.8	3.2
6	0.2	1.2	7.2
8	0.2	1.6	12.8
		$\sum x \cdot P(x) = 4.0$	$\sum x^2 \cdot P(x) = 24.0$

Hence, $\sigma^2 = 24 - 4^2 = 8$

3.3 Expectation

3.3.1 Expectation Value

The **expectation value** of a discrete random variable of a probability distribution is the theoretical average of the variable. The formula is

$$\mu = E(X) = \sum x \cdot P(x)$$

The symbol $E(X)$ is used for the expectation value.

Comment: Expected value is a 'long-run average' in the sense that if repeated trials are made of an experiment with a random variable *X*, the sample average of the

observed values of X will get closer and closer to the expected value $E(X)$ as the number of repetitions gets large.

Example 9

A bag contains three \$1 bills, two \$5 bills, and one \$10 bill. One bill is selected at random. If X denotes the denomination of the selected bill, find the expected value $E(X)$.

Solution

X	1	5	10
$P(X)$	$\frac{3}{6}=\frac{1}{2}$	$\frac{2}{6}=\frac{1}{3}$	$\frac{1}{6}$

Hence, $E(X)=1\left(\frac{1}{2}\right)+5\left(\frac{1}{3}\right)+10\left(\frac{1}{6}\right)=\frac{23}{6}=3.83$

In other words, if this procedure is repeated many times, the selector would be expected to win $\$\frac{23}{6}$ or \$3.83

Example 10

One thousand tickets are sold at \$1 each for a color television valued at \$350. What is the expected value of the gain if a person purchases one ticket?

Solution

The problem can be set up as follow

	Win	**Lose**
Gain X	\$349	-\$1
Probability $P(X)$	$\frac{1}{1000}$	$\frac{999}{1000}$

For a win, the net gain is \$349, since the person does not get the cost of the ticket (\$1) back. For a loss, the gain is represented by a negative number, in this case -\$1. The solution, then, is

$$E(X)=\$349\times\frac{1}{1000}+(-\$1)\times\frac{999}{1000}=-\$0.65$$

This expectation means that the average of the losses is \$0.65 for each of the 1000 ticket holder. Another way, it means that if a person purchased one ticket each week over a long period of time, the average loss would be \$0.65 per ticket.

Example 11

One thousand tickets are sold at \$1 each for four prizes of \$100, \$50, \$25, and \$10. What is the expect value if a person purchases 2 tickets.

Gain (\$ X)	\$98	\$48	\$23	\$8	-\$2
Probability $P(X)$	$\frac{2}{1000}$	$\frac{2}{1000}$	$\frac{2}{1000}$	$\frac{2}{1000}$	$\frac{992}{1000}$

Solution:

$$E(X)=98\times\frac{2}{1000}+48\times\frac{2}{1000}+23\times\frac{2}{1000}+8\times\frac{2}{1000}+(-2)\times\frac{992}{1000}=-1.63$$

An alternative solution is found by

$$E(X)=100\cdot\frac{2}{1000}+50\cdot\frac{2}{1000}+25\cdot\frac{2}{1000}+10\cdot\frac{2}{1000}-2=-1.63.$$

3.3.2 Properties of Expectation Value

Sometimes it is necessary to compute the expected value of a combination of random variables, and the following result may be quite useful.

For any random variables $X, X_1, X_2, X_3, ..., X_k$ and number c

$E(X_1 + X_2 + \cdots + X_k) = E(X_1) + E(X_2) + \cdots + E(X_k)$

$E(cX) = cE(X)$ and

$E(c) = c$

and also, for all constants a, b $E(aX + b) = aE(X) + b$

4 Discrete Distributions

4.1 Binomial

Binomial experiment (distribution) arises in the following situation:

(i) the underlying experiment consists of n independent and identical trials.

(ii) each trial results in one of two possible outcomes, a successes of a failure.

(iii) the probability of success in a single trial is equal to p and remains the same through the experiment.

(iv) the experimenter is interested in the r.v. X that counts the number of successes in n trials.

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, x = 1, 2, \ldots, n$$

where $q = 1 - p$.

Mean: $\mu = np$

Variance: $\sigma^2 = npq$, $\sigma = \sqrt{npq}$

Example 1

A coin is tossed 3 times. Find the probability of getting exactly 2 heads.

Solution:

This problem can be solved by looking at the sample space.

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

The answer is $\frac{3}{8}$ or 0.375

In this case, $n = 3, X = 2, p = \frac{1}{2}$, and $q = \frac{1}{2}$. Hence, substituting in the formula gives

$$P(2 \text{ heads}) = \frac{3!}{(3-2)!2!}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^1 = \frac{3}{8} = 0.375$$

Example 2

A survey found that one out of five Americans says he or she has visited a doctor in any given month. If ten people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

Solution

In this case, $n = 10, X = 3, p = \frac{1}{5}$, and $q = \frac{4}{5}$. Hence,

$$P(3) = \frac{10!}{(10-3)!3!}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7 = 0.201$$

Example 3

Suppose that X has a binomial distribution with $n = 10, p = .4$ .Find (i) $P(X \le 4)$, (ii) $P(X < 6)$, (iii) $P(X > 4)$.

Example 4

A survey from Teenage Research Unlimited found that 30% of teenage consumers receive their spending money from part-time jobs. If 5 teenagers are selected at random, find the probability that at least 3 of them will have part-time jobs.

Solution

Let E be the event that at least 3 of teenagers will have part-time jobs.

To find the probability that at least 3 have part-time jobs, it is necessary to find the individual probabilities for either 3, or 4, or 5, and then add them to get the total probability.

$$P(3)=\frac{5!}{(5-3)!3!}(0.3)^3(0.7)^2=0.132$$

$$P(4)=\frac{5!}{(5-4)!4!}(0.3)^4(0.7)^1=0.028$$

$$P(5)=\frac{5!}{(5-5)!5!}(0.3)^5(0.7)^0=0.002$$

Hence

$$P(E)=0.132+0.028+0.002=0.162$$

4.2 Poisson[1]

The Poisson random variable arises when counting the number of event that occurs in an interval of time when the events are occurring at a constant rate, for instance, number of calls per minute in a switchboard, number of arrivals at an emergency room, number of items demanded from an inventory, number of error in a manuscript.

If X is a Poisson r.v. with a parameter λ, then

$$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!};\ x=0,1,2,\ldots$$

Note: $e\approx 2.71828$

Mean: $\mu=\lambda$

Variance: $\sigma^2=\lambda$, $\sigma=\sqrt{\lambda}$

Example 5

If there are 200 typographical errors randomly distributed in a 500-page manuscript, find the probability that a given page contains exactly three errors.

Solution

First, find the mean number λ of errors. Since there are 200 errors distributed over 500 pages, each page has an average of

$$\lambda=\frac{200}{500}=0.4$$

or 0.4 error per page. Since $X=3$, substituting into the formula yields

$$P(X=3)=\frac{e^{-\lambda}\lambda^3}{3!}=\frac{(2.71828)^{-0.4}(0.4)^3}{3!}=0.0072$$

Example 6

A sales firm receives, on the average, three calls per hour on its toll-free number for any given hour, find the probability that it will receive the following

[1] Siméon Denis Poisson (1781-1840), a French Mathematician

a. At most three calls, **b.** At least three calls, **c.** Five or more calls.

Solution

a. "At most three calls" means 0, 1, 2, or 3 calls. Hence,

$$\begin{aligned} P(\text{at most three calls}) &= P(X=0)+P(X=1)+P(X=2)+P(X=3) \\ &= 0.0498+0.1494+0.2240+0.2240 \\ &= 0.6472 \end{aligned}$$

b. "At least three calls" means 3 or more calls" means 3 or more calls. Let E be the event of receiving 0, 1, or 2 calls. Then

$$\begin{aligned} P(E) &= P(X=0)+P(X=1)+P(X=2) \\ &= 0.0498+0.1494+0.2240 \\ &= 0.4232 \end{aligned}$$

$$P(\bar{E}) = 1-P(E) = 1-0.4232 = 0.5768$$

c.

$$\begin{aligned} P(\text{Five or more calls}) &= 1-[P(X=0)+P(X=1) \\ &\quad +P(X=2)+P(X=3)+P(X=4)] \\ &= 1-(0.0498+0.1494+0.2240+0.2240+0.1680) \\ &= 1-0.8152 \\ &= 0.1848 \end{aligned}$$

4.3 Hyper geometric

The hyper-geometric distribution arises when one selects a random sample of size n, without replacement, from a finite population of size N divided into two classes consisting of D elements of the first kind and $N-D$ of the second kind. Such a scheme is called sampling without replacement from a finite dichotomous population. This situation can be summarized by the contingency table as follow:

	Drawn	Not Drawn	Total
One kind	x	D-x	D
Other kind	n-x	N-D-n+x	N-D
Total	n	N-n	N

Formula: $$P(X=x) = \frac{\binom{D}{x}\times\binom{N-D}{n-x}}{\binom{N}{n}}$$

Example 7

Suppose a committee of four people is to be selected from seven women and five men. What is the probability that the committee will consist of three women and one man?

Solution

$$P(X=3) = \frac{\binom{7}{3}\times\binom{5}{1}}{\binom{12}{4}} = \frac{\frac{7!}{4!3!}\cdot\frac{5!}{4!1!}}{\frac{12!}{8!4!}} = \frac{35}{99}$$

Example 8

Ten people apply for a job as assistant manager of a restaurant. Five have completed college and five have not. If the manager selects three applicants at random find the probability that all three are college graduates. (Answer: $1/12$)

Example 9

A lot of 12 compressor tanks are checked to see whether there are any defective tanks. Three tanks are checked for leaks. If one or more of the three is defective, the lot is rejected. Find the probability that the lot will be rejected if there are actually three defective tanks in the lot. (Answer: 0.62)

Example 10

Assume that there are 5 white and 45 black marbles in the urn. Standing next to the urn, you close your eyes and draw 10 marbles without replacement. What is the probability that exactly 4 of the 10 are white? (Answer: 0.00396)

5 Normal Distribution or Gaussian[2] distribution

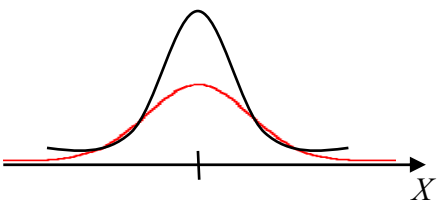
5.1 Normal and Standard Normal Distribution

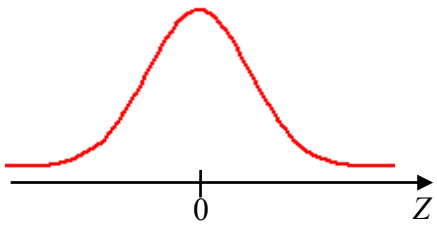
The **normal distribution** or **Gaussian distribution** is a **continuous probability distribution** that often gives a good description of data that cluster around the mean. The graph of the associated probability density function is bell-shaped, with a peak at the mean, and is known as the *Gaussian function* or *bell curve*. This bell-shaped curve is characterize by the function

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

Standard Normal Distribution is the normal distribution with the mean $\mu = 0$ and variance $\sigma^2 = 1$. Hence the corresponding function is

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





Standard Normal Curve

To transform a nonstandard normally distributed r.v X to a standard one, we use the formula

$$Z = \frac{X - \mu}{\sigma}$$

The probability of Z between a and b (a and b are any two values on horizontal axis) is the area under the standard normal curve between a and b.

Example 1

Suppose that the r.v Z has the standard normal distribution. Find the following probabilities.

(a) $P(Z < 1)$ (b) $P(Z > 0.03)$ (c) $P(Z < 1.87)$

[2] Carl Friedrich Gauss (1777–1855), a German Mathematician and Physicist.

(d) $P(-1.87 < Z < 1)$ (e) $P(Z > 4.7)$

Answer: (a) 0.8413 (b) 0.4880 (c) 0.0307 (d) 0.8106 (e) ~1

Example 2

Find the value of z so that (a) $P(Z < z) = 0.9$ (b) $P(Z > z) = 0.01$ (c) $P(Z < z) = 0.1$ (c) $P(-z < Z < z) = 0.95$.

Example 3

Suppose X is normally distributed with $\mu = 20$ and $\sigma = 4$ ($X \sim N(\mu, \sigma^2)$). Find the following probabilities: (a) $P(X < 26)$ (b) $P(X > 18)$ (c) $P(15 < X < 21)$

Solution

We apply the formula $Z = \dfrac{X - \mu}{\sigma}$, then

(a) $P(X < 26) = P\left(Z < \dfrac{26 - 20}{4}\right) = P(Z < 1.5) = 0.9332$

(b) and (c) are left for the reader.

5.2 Applications of Normal Distribution

Example 4

On a certain scale, measurement errors are made. They are normally distributed with mean equal to 0 and standard deviation equal to 0.1 ounce. If you weight an object on this scale, what is the probability that the measurement error is no greater than 0.15 ounce?

Solution

Let X denote the measurement error made when an object is weighted on the scale. Our goal is to find $P(-0.15 < X < 0.15)$. Since $\mu = 0$ and $\sigma = 0.1$ we use the transformation formula to obtain

$$\begin{aligned} P(-0.15 < X < 0.15) &= P\left(\frac{-0.15 - 0}{0.1} < Z < \frac{0.15 - 0}{0.1}\right) \\ &= P(-1.5 < Z < 1.5) \\ &= 0.8664 \end{aligned}$$

Hence, there is about 87% chance that the error in measurement is no greater than 0.15 ounce.

Example 5

The number of loaves of bread that can be sold during a day by a certain supermarket is normally distributed with $\mu = 1000$ loaves and $\sigma = 100$ loaves. If the market stocks 1200 loaves on a given day, what is the probability that the loaves will be sold out before the day is over?

Solution

Let X denote the # of loaves that can be sold during a day. So, we wish to find $P(X > 1200)$.

$$P(X > 1200) = P\left(Z > \frac{1200 - 1000}{100}\right) = P(Z > 2) = 0.0228$$

Hence, there's about 2% chance that the loaves will be sold out before the day is over.

Example 6

To qualify for a police academy, candidates must score in the top 10% on a general abilities test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume the test scores are normally distributed.

(Answer: Anybody scoring 226 or higher qualifies.)

Example 7

The lengths of trout in a certain lake are normally distributed with a mean of 7 inches and a standard deviation of 2 inches. If the fish and game department would like fishermen to keep only the largest 20% of the trout, what should the minimum size for 'keepers' be? (Answer: 8.68 inches)

5.3 Normal Approximation to Binomial Distribution

Suppose X is a binomial r.v. arising from an experiment with n trials and success probability p. If n is large, then the distribution of X is approximately normal with mean μ and standard deviation σ, where

$$\mu = np \text{ and } \sigma = \sqrt{npq}; q = 1 - p$$

In order that this approximation can be suitable n should satisfy $np>5$ and $nq>5$.

To apply the approximation and use normal distribution we need the correction for continuity. It is summarized in the table below

Binomial	**Normal**
When finding	Use
(i). $P(X = a)$	$P(a - 0.5 < X < a + 0.5)$
(ii). $P(X \geq a)$	$P(X > a - 0.5)$
(iii). $P(X > a)$	$P(X > a + 0.5)$
(iv). $P(X \leq a)$	$P(X < a + 0.5)$
(v). $P(X < a)$	$P(X < a - 0.5)$

Remark For simplicity of the problem the correction for continuity is neglected if n is large enough.

Example 8

The probability that a certain surgical operation is successful is 0.8. If the operation is performed on 120 people, find the probability that 90 or more operations are successful.

Solution

The suitability for the approximation:

p=0.8, n=120, then q=1-0.8=0.2 and np=120×0.8=96>5 and nq=120×0.2=24>5. So, the approximation is suitable. $\mu = np = 120 \times 0.8 = 96$ and $\sigma = \sqrt{120 \times 0.8 \times 0.2} = 4.38$. For this problem we wish to seek for $P(X \geq 90)$. Using the correction for continuity, we will have to find $P(X > 89.5)$

$$P(X > 89.5) = P\left(Z > \frac{89.5 - 96}{4.38}\right) = P(Z > -1.48) = 0.9147$$

That is, it is about 91% likely that at least 90 of the 120 operations are successful.

Example 9

If a baseball player is battering average is 0.320 (32%), find the probability that the player will get at most 26 hits in 100 times at bat.

(Hint: $p = 0.32, q = 0.68, n = 100,$ answer: 11.9%)