Probability & Statistics for Engineers & Scientists

N I N T H E D I T I O N

By Ronald E. Walpole
Roanoke College
Raymond H. Myers
Virginia Tech
Sharon L. Myers
Radford University
Keying Ye
University of Texas at San Antonio

BINOMIAL DISTRIBUTION

Example 5.2: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

Solution: Here n=15, p=0.40

R-Code:

```
    rm(list=ls())
    n=15; p=0.40
# (a)
    1-pbinom(q=9,size=15,prob=0.40)    # or
    prob=dbinom(x=10:15,size=15,prob=0.40)
    ans= sum (prob)
    ans            # or
    sum(dbinom(10:15,15,0.40))
# (b)
    prob=dbinom(3:8,15,0.40)
    ans= sum (prob)
    ans            # or
    b=pbinom(8,15,0.40)- pbinom(2,15,0.40)
    b
# (c)
    dbinom(5,15,0.40)
```

Example 5.3: A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

(a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

(b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

Answer:

```
rm(list=ls())
n=20; p=0.03
# (a)
1-pbinom(0,n,p)    # or
```

```
prob=dbinom(1:n,n,p)
ans= sum (prob)
ans
# (b)
ship=10
prob=dbinom(3,ship,ans)
ans= sum (prob)
ans
```

Example 5.4 , 5.5 Same as above.

Exercise Question no 5.4 to 5.16

HYPER GEOMETRIC DISTRIBUTION

```
dhyper(x, m, n, k, log = FALSE)
phyper(q, m, n, k, lower.tail = TRUE, log.p = FALSE)
qhyper(p, m, n, k, lower.tail = TRUE, log.p = FALSE)
rhyper(nn, m, n, k)
```

Arguments

`x, q` vector of quantiles representing the number of white balls drawn without replacement from an urn which contains both black and white balls.

`m` the number of white balls in the urn.

`n` the number of black balls in the urn.

`k` the number of balls drawn from the urn, hence must be in $0, 1, \ldots, m+n$.

`p` probability, it must be between 0 and 1.

`nn` number of observations. If `length(nn) > 1`, the length is taken to be the number required.

`log, log.p` logical; if TRUE, probabilities p are given as log(p).

Example 5.9: Lots of 40 components each are deemed unacceptable if they contain 3 or more defectives. The procedure for sampling a lot is to select 5 components at random and to reject the lot if a defective is found. What is the probability that exactly 1 defective is found in the sample if there are 3 defectives in the entire lot?

***Solution*:** Using the hypergeometric distribution with $n = 5$, $N = 40$, $k = 3$, and $x = 1$, we find the probability of obtaining 1 defective to be

$$h(1; 40, 5, 3) = \frac{\binom{3}{1}\binom{37}{4}}{\binom{40}{5}} = 0.3011.$$

Using R-code:

Solution

```
rm(list=ls())
m=37; n=3; k=5 ; x=1
prob=dhyper(x,3,37,5)
prob
```

-

5.29 A homeowner plants 6 bulbs selected at random from a box containing 5 tulip bulbs and 4 daffodil bulbs. What is the probability that he planted 2 daffodil bulbs and 4 tulip bulbs?

R-code:

```
dhyper(2,4,5,6)    #   or
dhyper(x=2, m=4,n=5,k=6)
 # or   by using specified parameter we can change order of syntax, for example
dhyper(k=6, x=2, m=4, n=5)
```

5.30 To avoid detection at customs, a traveler places 6 narcotic tablets in a bottle containing 9 vitamin tablets that are similar in appearance. If the customs official selects 3 of the tablets at random for analysis, what is the probability that the traveler will be arrested for illegal possession of narcotics?

R-code:

```
p1=dhyper(1:3,6,9,3)
sum(p1)  #   or
p2=1-phyper(0,6,9,3)
p2
```

5.31 A random committee of size 3 is selected from 4 doctors and 2 nurses. Write a formula for the probability distribution of the random variable X representing the number of doctors on the committee. Find $P(2 \leq X \leq 3)$.

```
rm(list=ls())
# solution
p1=dhyper(2:3,4,2,3)
sum(p1)  #   or
```

p2=phyper(3,4,2,3)-phyper(1,4,2,3)
p2

Practice Questions

Exercise 5.33 to 5.41

Description of Negative Binomial Distribution

Density, distribution function, quantile function and random generation for the negative binomial distribution with parameters `size` and `prob`.

Usage

```
dnbinom(x, size, prob, mu, log = FALSE)
pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
rnbinom(n, size, prob, mu)
```

Arguments

`x`	vector of (non-negative integer) quantiles.
`q`	vector of quantiles.
`p`	vector of probabilities.
`n`	number of observations. If `length(n) > 1`, the length is taken to be the number required.
`size`	target for number of successful trials, or dispersion parameter (the shape parameter of the gamma mixing distribution). Must be strictly positive, need not be integer.

x=Total attempts-No. of successes, size=No of successes, prob= probability of success

Example 5.14: In an NBA (National Basketball Association) championship series, the team that wins four games out of seven is the winner. Suppose that teams A and B face each other in the championship games and that team A has probability 0.55 of winning a game over team B.

(a) What is the probability that team A will win the series in 6 games?

(b) What is the probability that team A will win the series?

(c) If teams A and B were facing each other in a regional playoff series, which is decided by winning three out of five games, what is the probability that team A would win the series?

Solution: (a) $b^*(6;4,0.55) = \binom{5}{3}0.55^4(1-0.55)^{6-4} = 0.1853$

(b) P(team A wins the championship series) is

$$b^*(4;4,0.55) + b^*(5;4,0.55) + b^*(6;4,0.55) + b^*(7;4,0.55)$$
$$= 0.0915 + 0.1647 + 0.1853 + 0.1668 = 0.6083.$$

(c) P(team A wins the playoff) is

$$b^*(3;3,0.55) + b^*(4;3,0.55) + b^*(5;3,0.55)$$
$$= 0.1664 + 0.2246 + 0.2021 = 0.5931.$$

R-code:

```
rm(list=ls())
# solution (a)
p1=dnbinom(2,4,0.55)
p1
  #   or
p2=dnbinom(x=6-4,size=4,prob=0.55)
p2
# Solution (b)
p1=dnbinom(x=0:3,size=4,prob=0.55)
sum(p1)
  #   or
p2=pnbinom(q=3,size=4,prob=0.55)
p2
# solution(c)
n=3:5
k=3
p1=dnbinom(x=n-k,size=k,prob=0.55)
sum(p1)
  #   or
p2=pnbinom(q=5-k,size=k,prob=0.55)
p2
```

GEOMETRIC DISTRIBUITON

Usage

```
dgeom(x, prob, log = FALSE)
pgeom(q, prob, lower.tail = TRUE, log.p = FALSE)
qgeom(p, prob, lower.tail = TRUE, log.p = FALSE)
rgeom(n, prob)
```

Arguments

x, q	vector of quantiles representing the number of failures in a sequence of Bernoulli trials before success occurs.
p	vector of probabilities.
n	number of observations. If `length(n) > 1`, the length is taken to be the number required.
prob	probability of success in each trial. `0 < prob <= 1`.
log, log.p	logical; if TRUE, probabilities p are given as log(p).

Example 5.15: For a certain manufacturing process, it is known that, on the average, 1 in every 100 items is defective. What is the probability that the fifth item inspected is the first defective item found?

***Solution*:** Using the geometric distribution with $x = 5$ and $p = 0.01$, we have

$$g(5; 0.01) = (0.01)(0.99)^4 = 0.0096.$$

Example 5.16: At a "busy time," a telephone exchange is very near capacity, so callers have difficulty placing their calls. It may be of interest to know the number of attempts necessary in order to make a connection. Suppose that we let $p = 0.05$ be the probability of a connection during a busy time. We are interested in knowing the probability that 5 attempts are necessary for a successful call.

***Solution*:** Using the geometric distribution with $x = 5$ and $p = 0.05$ yields

$$P(X = x) = g(5; 0.05) = (0.05)(0.95)^4 = 0.041.$$

Solution 5.15:

```
rm(list=ls())
n=5
k=1
p1=dnbinom(x=n-k,size=k,prob=0.01)
p1
 #   or
p2=dgeom(x=4,prob=0.01)   # here x= no of failours
p2
```

Solution 5.16

```
rm(list=ls())
n=5
k=1
p1=dnbinom(x=n-k,size=k,prob=0.05)
p1
 #   or
p2=dgeom(x=4,prob=0.05)   # here x= no of failours
p2
```

Practice questions for Negative binomial and Geometric distributions: 5.49 to 5.55, 5.59, 5.74-5.76

POISSON DISTRIBUTION

Usage

```
dpois(x, lambda, log = FALSE)
ppois(q, lambda, lower.tail = TRUE, log.p = FALSE)
qpois(p, lambda, lower.tail = TRUE, log.p = FALSE)
rpois(n, lambda)
```

Arguments

`x`	vector of (non-negative integer) quantiles.
`q`	vector of quantiles.
`p`	vector of probabilities.
`n`	number of random values to return.
`lambda`	vector of (non-negative) means.

Example 5.17: During a laboratory experiment, the average number of radioactive particles passing through a counter in 1 millisecond is 4. What is the probability that 6 particles enter the counter in a given millisecond?

***Solution*:** Using the Poisson distribution with $x = 6$ and $\lambda t = 4$ and referring to Table A.2, we have

$$p(6; 4) = \frac{e^{-4}4^6}{6!} = \sum_{x=0}^{6} p(x; 4) - \sum_{x=0}^{5} p(x; 4) = 0.8893 - 0.7851 = 0.1042.$$

Example 5.18: Ten is the average number of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers per day. What is the probability that on a given day tankers have to be turned away?

***Solution*:** Let X be the number of tankers arriving each day. Then, using Table A.2, we have

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x; 10) = 1 - 0.9513 = 0.0487.$$

Solution with R-code:

```
rm(list=ls())
# Solution of Example 5.17
p1=dpois(x=6,lambda=4)
p1
```

```
# Solution of Example 5.18
p2=1-ppois(q=15,lambda=10)
p2   # or
p3=1-sum(dpois(x=0:15,lambda=10))
p3
```

Practice questions Poisson distribution 5.56 -5.58, 5.60-5.73, 5.87

NORMAL DISTRIBUTION

Usage

```
dnorm(x, mean = 0, sd = 1, log = FALSE)
pnorm(q, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE, log.p = FALSE)
rnorm(n, mean = 0, sd = 1)
```

Example 6.5: Given that X has a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

Solution: The normal probability distribution with the desired area shaded is shown in Figure 6.12. To find $P(X > 362)$, we need to evaluate the area under the normal curve to the right of $x = 362$. This can be done by transforming $x = 362$ to the corresponding z value, obtaining the area to the left of z from Table A.3, and then subtracting this area from 1. We find that

$$z = \frac{362 - 300}{50} = 1.24.$$

Hence,

$$P(X > 362) = P(Z > 1.24) = 1 - P(Z < 1.24) = 1 - 0.8925 = 0.1075.$$

Solution using R

```
rm(list=ls())
p1=1-pnorm(q=362,mean=300, sd=50)
p1
```

Example 6.6: Given a normal distribution with $\mu = 40$ and $\sigma = 6$, find the value of x that has

(a) 45% of the area to the left and

(b) 14% of the area to the right.

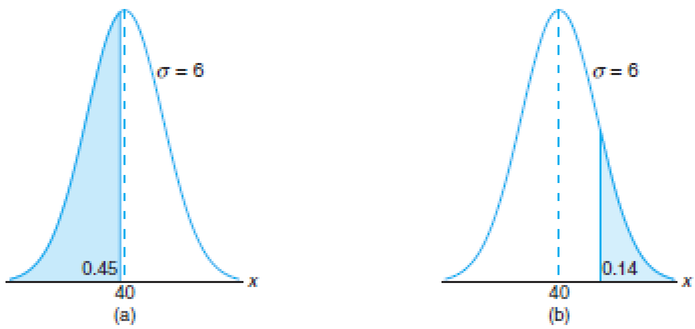


Figure 6.13: Areas for Example 6.6.

Solution: (a) An area of 0.45 to the left of the desired x value is shaded in Figure 6.13(a). We require a z value that leaves an area of 0.45 to the left. From Table A.3 we find $P(Z < -0.13) = 0.45$, so the desired z value is -0.13. Hence,

$$x = (6)(-0.13) + 40 = 39.22.$$

(b) In Figure 6.13(b), we shade an area equal to 0.14 to the right of the desired x value. This time we require a z value that leaves 0.14 of the area to the right and hence an area of 0.86 to the left. Again, from Table A.3, we find $P(Z < 1.08) = 0.86$, so the desired z value is 1.08 and

$$x = (6)(1.08) + 40 = 46.48.$$

Solution:

```
rm(list=ls())
#(a)
qnorm(.45,40,6)

#(b)
qnorm(1-.14,40,6)
```

Practice questions for normal distribution 6.5 to 6.22

LAW OF LARGE NUMBER

Statement: As the Sample size say n increases the average value of iid data get closer and closer to the expected value.

Example: If a die is rolled M times and a roll of “6” is recorded as a success.

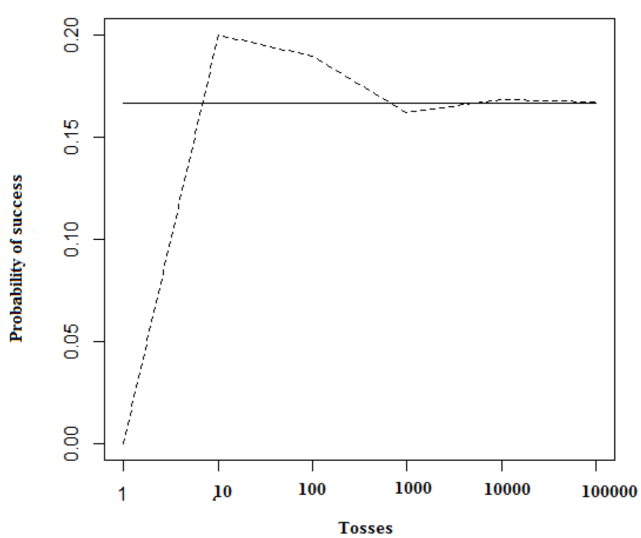
- Compute the empirical probability of success for M= 1, 10, 100, 1000, 10000, 100000.
- Calculate the theoretical probability of success
- Plot the theoretical probability of success against the empirical.
- Discuss the effect of M on empirical probability.

Note: Same experiment can be done with coin or other changes are highlighted in R-code after # sign.

R-Code for the above example is presented as:

```
rm(list=ls())
x=1:6     # for coin use x=1:2
N=c(1,10,100,1000,10000,100000)
px=c()
for(i in 1:length(N)){
sam=sample(x,N[i],replace=T)
nx=length(which(sam==6))    # for coin use 1=head, 2=tail
px[i]=nx/N[i]
}
plot(1:length(N),px,type="l",lty=2)
lines(1:length(N),rep(1/6,length(N)),type="l")    # for coin use ½ instead of 1/6
```

Graph of the output



Example statement for coin:

Write an R program to simulate the following experiment
A balanced coin is tossed M times, and a toss of "head" was recorded as a success. Compute the empirical probability of success for M=1, 10, 100, 1000, 10000, 100000. Also, calculate the theoretical probability of success. Plot the theoretical probability of success against the empirical. Discuss the effect of M on empirical probability.

CENTRAL LIMIT THEOREM

Statement:

Let the random variable X1, X2,…,Xn be sequence of independent and identically distributed, each with finite population mean μ and finite variance σ^2 , then the random variable

$$z = \frac{Statistic - E(Statistic)}{S.D.(Statistic)}$$

For large sample size follows Normal distribution with mean 0 and variance 1. i.e. Z~N(0,1).

Example 1:

Generate the N= 1000 random numbers from Poisson distribution with parameter 0.90.

i) A random sample of size n= 2, 10, 30 and 100 is selected by with replacement.

ii) Find the sample mean of each sample

the steps i) and ii) are repeated M=10,000 times.

iii) Plot the histogram for given sample sizes and draw the conclusion about the shape of the sample means.

R-Code for above algorithm

- rm (list=ls())
- n=5
- M=10000
- xbar=c(); sdx=c()
- pop=rpois(1000, 0.9)
- for(i in 1:M){
- sam=sample(pop,n,replace=T)
- xbar[i]=mean(sam)
- }
- hist(xbar)

Example 2:

The N=1000 random numbers from the binomial distribution with parameters n=10 and p=0.2 have been generated.

(i) A random sample of size n=2, 10, 30 and 100 is selected by with replacement
(ii) the sample proportion of the number divisible by 3 is calculated.

Steps (i) and (ii) are repeated M=10,000 times.

the histogram is drawn for each sample size.

Comment on the shape of the distribution of the sample proportion. This conclusion about the shape of the distribution is based on what Law? Also, provide the R-code of the above algorithm.

R-Code for above example:

```
rm(list=ls())
N=1000
pop=rbinom(N,10,0.2)
P=length(which(pop%%3==0))/N
n= 2      # similary n= 10,30,100
m=length(n);M=10000
par(mfrow=c(2,2))
pr=c()
for(i in 1:M){
sam=sample(pop,n,replace=T)  #
pr[i]=length(which(sam%%3==0))/n
}
hist(pr,main=paste("Sample prop for n=",n),col=2, xlab="sample prop")
```

Graphical representation of above code: