

Ch 3: Discrete RV

Contents

1	PMF and CDF	2
1.1	Random Variable	3
1.2	PMF and CDF	4
2	Expected Value and Variance	14
2.1	The Expected value (Theoretical Mean)	15
2.2	Expectation is a long-run average	17
2.3	How to calculate $E(g(X))$	21
2.4	(Theoretical) Variance	23
3	Popular Discrete Distributions	27
3.1	Binomial Random Variable	29
3.2	Negative Binomial Random Variable	41
3.3	Hypergeometric Random Variable	46
3.4	Poisson Random Variable	56
3.5	R code for the Four Distributions	65

PMF and CDF

[\[ToC\]](#)

1.1 Random Variable

[\[ToC\]](#)

is a function whose domain is a sample space, and whose range is a real numbers.

Discrete Random Variable is a r.v. whose range is a finite or countably infinite set.

Continuous Variable is a r.v. whose range is a interval on a real line or a disjoint union of such intervals. It also must satisfy that for any constant c , $P(X = c) = 0$.

Example

1. Throw a die: $\{1\} \rightarrow 1$
2. Throw two dice at once and add: $\{2, 5\} \rightarrow 7$

1.2 PMF and CDF

[\[ToC\]](#)

- **Probability Mass Function:** (pmf) of a discrete RV is defined as

$$p(x) = P(X = x)$$

- **Cumulative Distribution Function:** (cdf) of a discrete random variable X is defined as

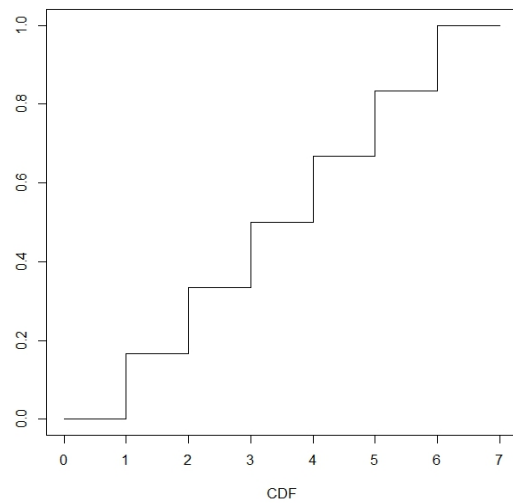
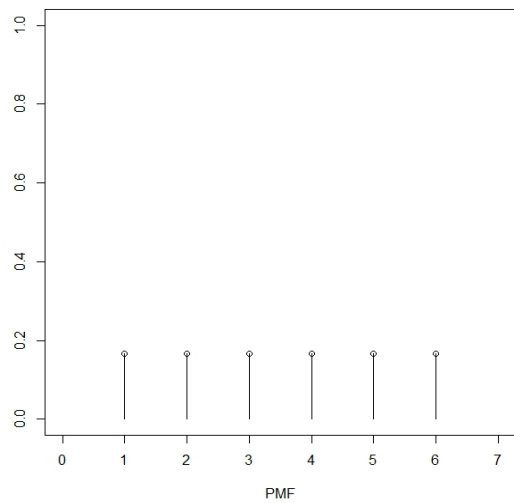
$$F(x) = P(X \leq x)$$

R example: Roll a Die Once

```
layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side

#- PMF plot -
x <- 1:6
plot(x, rep(1/6,6), type="h", xlim=c(0,7), xlab="PMF", ylab="", ylim=c(0,1))
lines(x, rep(1/6,6), type="p")

#- CDF plot -
plot(0:7, c(0,cumsum(rep(1/6, 6)),1), type="s", xlim=c(0,7), xlab="CDF", ylab="", ylim=c(0,1))
```

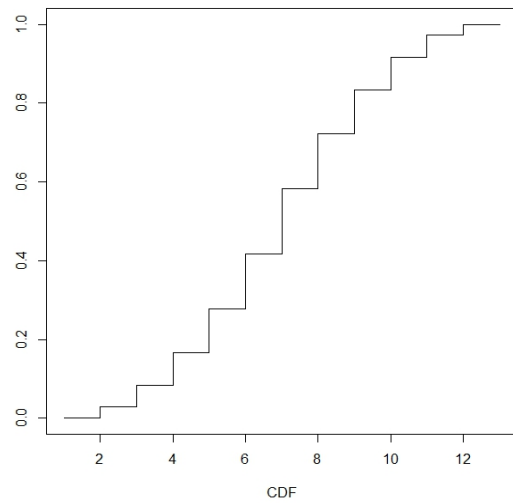
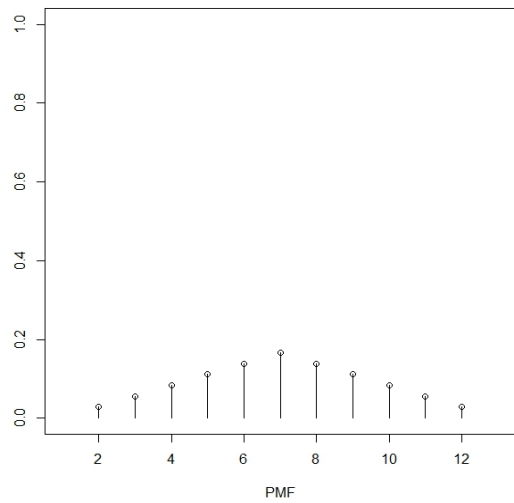


R example: Roll a Die Twice

```
layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side

#- PMF plot -
x <- 2:12
pmf1 <- c(1,2,3,4,5,6,5,4,3,2,1)/36
plot(x, pmf1, type="h", xlim=c(1,13), xlab="PMF", ylab="", ylim=c(0,1))
lines(x, pmf1, type="p")

#- CDF plot -
plot(c(1,x,13), c(0,cumsum(pmf1),1), type="s", xlim=c(1,13), xlab="CDF", ylab="", ylim=c(0,1))
```



PMF and CDF

- You can calculate CDF from PMF, or vice versa.
- CDF always start at 0, and end at 1.
- pmf at any point must be ≥ 0 .
- If you add all values of pmf, it must add up to 1.

Calculating probability using CDF

If X is a discrete random variable, then

$$\begin{aligned}P(a \leq X \leq b) &= P(X \leq b) - P(X \leq a) \\&= F(b) - F(a)\end{aligned}$$

$$\begin{aligned}P(a \geq X) &= 1 - P(X < a) \\&= 1 - P(X \leq a - 1) \\&= 1 - F(a - 1)\end{aligned}$$

$$\begin{aligned}P(X = a) &= P(X \leq a) - P(X \leq a - 1) \\&= F(a) - F(a - 1) \\&= p(a)\end{aligned}$$

Example: Roll a Die Once

pmf:

$$p(x) = 1/6 \quad \text{for } x = 1, \dots, 6.$$

Calculate:

- $F(4) = P(X \leq 4)$
- $P(2 \leq X \leq 4)$
- $P(X > 4)$

Exercises:

Given the CDF:

$$F(1) = .2, \quad F(2) = .35, \quad F(3) = .60, \quad F(4) = .90, \quad F(5) = .95, \quad F(6) = 1$$

Calculate the following:

$$P(X > 3)$$

$$P(2 \leq X \leq 5)$$

$$P(2 < X < 5)$$

$$p(2)$$

Exercises:

Given the CDF:

$$F(1) = .2, \quad F(2) = .35, \quad F(3) = .60, \quad F(4) = .90, \quad F(5) = .95, \quad F(6) = 1$$

Calculate the following:

$$\begin{aligned} P(X > 3) \\ &= 1 - P(X \leq 3) = 1 - F(3) \end{aligned}$$

$$\begin{aligned} P(2 \leq X \leq 5) \\ &= P(X \leq 5) - P(X < 2) = F(5) - F(1) \end{aligned}$$

$$\begin{aligned} P(2 < X < 5) \\ &= P(X < 5) - P(X \leq 2) = F(4) - F(2) \end{aligned}$$

$$\begin{aligned} p(2) \\ &= P(X = 2) = F(2) - F(1) \end{aligned}$$

Expected Value and Variance

[\[ToC\]](#)

2.1 The Expected value (Theoretical Mean)

[\[ToC\]](#)

- Expected Value of a random variable X , whose range is $x_1, x_2, x_3, \dots, x_n$ is defined as

$$E(X) = \mu = \sum_{i=1}^n x_i \cdot p(x_i)$$

Example: Throw a Die Once

Suppose you are given pmf:

#	1	2	3	4	5	6
p(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Then the expectation is

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

2.2 Expectation is a long-run average

[ToC]

Theoretical Mean (Expectation)

$$E(X) = 1 \cdot \left(\frac{1}{6}\right) + 2 \cdot \left(\frac{1}{6}\right) + 3 \cdot \left(\frac{1}{6}\right) + 4 \cdot \left(\frac{1}{6}\right) + 5 \cdot \left(\frac{1}{6}\right) + 6 \cdot \left(\frac{1}{6}\right) = 3.5$$

Sample Mean (Average)

$$\begin{aligned}\bar{X} &= 1 \cdot \left(\text{Rel. Freq. of \#1}\right) + 2 \cdot \left(\text{Rel. Freq. of \#2}\right) + 3 \cdot \left(\text{Rel. Freq. of \#3}\right) \\ &\quad + 4 \cdot \left(\text{Rel. Freq. of \#4}\right) + 5 \cdot \left(\text{Rel. Freq. of \#5}\right) + 6 \cdot \left(\text{Rel. Freq. of \#6}\right) \\ &= \dots\end{aligned}$$

- Sample Mean converges to The Expectation as $n \rightarrow \infty$.

Example: Pooled Blood Testing

- Each blood sample has .1 chance of testing positive.
- New procedure called "pooled testing" combines 10 blood samples before testing.
- If comes back negative, no further test is done. If comes back positive, then 10 more tests must be done using individual samples.
- What is the long-run average of the test number in the new scheme?

Example: Casino Simplified

- You charge each person \$1 to play this game.
- Each player has 1% chance of winning \$100.
- Does this make sense as business?
- What if you change 1% chance to .5% chance?

Example: Casino Simplified

- You charge each person \$1 to play this game.
- Each player has 1% chance of winning \$100.
- Does this make sense as business?
- What if you change 1% chance to .5% chance?

$$1(.99) + (-99)(.01) = 0$$

$$1(.995) + (-99)(.005) = .5$$

2.3 How to calculate $E(g(X))$

- Expected Value of a function of random variable X , say $g(X)$ is defined as

$$E(g(X)) = \sum_{i=1}^n g(x_i) \cdot p(x_i)$$

- If a and b are constants, then

$$E(aX + b) = aE(X) + b$$

Example: Throw a Die Once

Recall Let a random variable X to be a number of the rolled die. Then

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = 15.16667$$

Note that they are not equal to $(E(X))^2$.

$$\begin{aligned} E(3X + 5) &= (3 \cdot 1 + 5) \cdot \frac{1}{6} + (3 \cdot 2 + 5) \cdot \frac{1}{6} + (3 \cdot 3 + 5) \cdot \frac{1}{6} \\ &\quad + (3 \cdot 4 + 5) \cdot \frac{1}{6} + (3 \cdot 5 + 5) \cdot \frac{1}{6} + (3 \cdot 6 + 5) \cdot \frac{1}{6} \\ &= 15.5 \end{aligned}$$

Note that it is equal to $3E(X) + 5$.

2.4 (Theoretical) Variance

[\[ToC\]](#)

- Variance of a random variable X is defined as

$$V(X) = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 \cdot p(x_i) = E[(X - \mu)^2] = E(X^2) - (E(X))^2$$

Note that $\mu = E(X)$. Standard Deviation of X is defined as $\sigma = \sqrt{\sigma^2}$.

- Compare to Sample Variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Alternative formula for variance

$$\begin{aligned}V(X) &= E[(x - \mu)^2] \\&= E[x^2 - 2\mu x + \mu^2] \\&= E[x^2] + E[-2\mu x] + E[\mu^2] \quad E \text{ can be distributed over any linear equation.}\end{aligned}$$

But $\mu = E(X)$, and they are just a number. Taking E again doesn't do anything. So,

$$\begin{aligned}V(x) &= E[x^2] - 2\mu E[x] + \mu^2 \\&= E[x^2] - 2\mu^2 + \mu^2 \\&= E[x^2] - \mu^2\end{aligned}$$

Similarly to the discrete case, we have

$$V(aX + b) = a^2V(X)$$

Example: Throw a Die Once

Suppose you are given pmf:

#	1	2	3	4	5	6
p(x)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Expectation was $E(X) = 3.5$.
- Now calculate $E(X^2)$,

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} = 15.167$$

- So the variance is

$$V(X) = E(X^2) - [E(X)]^2 = 15.167 - 3.5^2 = 2.917.$$

Example: Casino Simplified

- You charge each person \$1 to play this game.
- Each player has 1% chance of winning \$100.
- Calculate Variance of game profit.
- What if you change 1% chance to .5% chance?

$$1^2(.99) + (-99)^2(.01) =$$

Popular Discrete Distributions

[\[ToC\]](#)

Four Distributions

If you have a coin which has probability p of landing head up when tossed, then,

Binomial(p, n)	Number of heads in n tosses
Negative Binomial(p, r)	Number of tails until you get r heads
Hypergeometric(n, N, m)	Number of red balls if n balls selected from an urn with N balls which includes m red.
Poisson(λ)	Number of events with rate λ per unit time.

3.1 Binomial Random Variable

[\[ToC\]](#)

-
- **Analogy:** Number of heads in n tosses. **Possible Values:** $0, 1, 2, \dots, n$
 - pmf of Binomial(n, p) is

$$P(X = x) = \text{dbinom}(x, n, p) = \binom{n}{x} (1-p)^{n-x} p^x$$

- CDF of Binomial(n, p) is

$$P(X \leq x) = \text{pbinom}(x, n, p) = \sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k},$$

Some of the values are listed in Table A.1.

- Expectation and Variance:

$$E(X) = np \quad \text{and} \quad V(X) = np(1-p)$$

Derivation of Binomial pmf

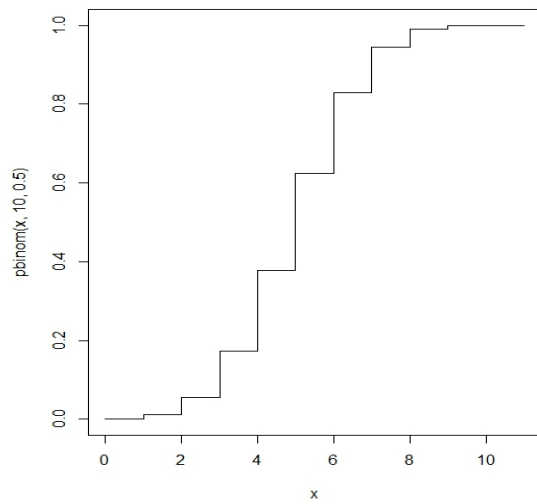
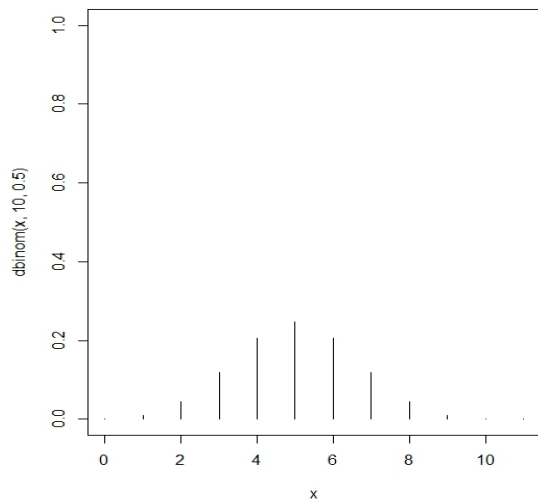
R code for Binomial(n, p)

x =[number of heads] n =[number of flip] p =[prob. of head in 1 flip]

```
pbinom(3,10,.5)      #- F(3):  CDF of Bin(n=10, p=.5) at x=3
dbinom(3,10,.5)      #- p(3):  pmf of Bin(n=10, p=.5) at x=3

layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side

x <- 0:11
plot(x, dbinom(x, 10, .5), type="h", ylim=c(0,1))  #- PMF plot -
plot(x, pbinom(x, 10, .5), type="s", ylim=c(0,1))  #- CDF plot -
```



Example: Crystal Picking

A company that produces fine crystal knows from experience that 10% of its goblets have cosmetic flaws and must be classified as “seconds”.

1. Among six randomly selected goblets, how likely is it that only one is a second?
2. Among six randomly selected goblets, what is the probability that at least two are seconds?

Example: Crystal Picking

A company that produces fine crystal knows from experience that 10% of its goblets have cosmetic flaws and must be classified as “seconds”.

1. Among six randomly selected goblets, how likely is it that only one is a second?

If we let X =(number of second within six selected), it has Binomial(6, .1) distribution. So answer is

$$\text{dbinom}(1,6,.1) = \binom{6}{1}(.9)^5(.1) = 0.354.$$

2. Among six randomly selected goblets, what is the probability that at least two are seconds?

$$\begin{aligned}
P(\text{at least two are seconds}) &= 1 - P(\text{at most one is second}) \\
&= 1 - P(X = 0) - P(X = 1) \\
&= 1 - \text{dbinom}(0, 6, .1) - \text{dbinom}(1, 6, .1) \\
\\
&= 1 - P(X \leq 1) \\
&= 1 - \text{pbinom}(0, 6, .1)
\end{aligned}$$

Example: Multiple Choice Exam

Multiple Choice Exam has 30 questions, each with 5 choices. What is the probability that you get above 80% (≥ 24 questions) of you guesses all the questions?

```
pbinom(3,10,.5)      #- F(3):  CDF of Bin(n=10, p=.5) at x=3
dbinom(3,10,.5)      #- p(3):  pmf of Bin(n=10, p=.5) at x=3

layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side

x <- 0:11
plot(x, dbinom(x, 10, .5), type="h", ylim=c(0,1))  #- PMF plot -
plot(x, pbinom(x, 10, .5), type="s", ylim=c(0,1))  #- CDF plot -
```

Example: Psychic or not

One person claims to have a psychic power. He claims that when one card is chosen out of 4 cards having different symbols, (star, triangle, circle, cross), he can tell the symbol while the chosen card is inside an envelope.

If he is allowed to try 10 times, over what % of correct answer is impressive?

Example: Minority Represented in a Committee

Suppose among large number of students. 20% of them are minority. To form a student council, 10 students are randomly selected. What is the probability that the minority is over-represented or under-represented?

Example: Best of 7

Example: Catch a cheater?

Suppose you flip a coin n times. When a coin is not fair, and $P(\textit{Head})$, how many times do you have to flip to see the difference in behavior?

3.2 Negative Binomial Random Variable

[\[ToC\]](#)

- **Analogy:** Number of **tails** until you get r th head. **Possible Values:** 0, 1, 2, 3,...

- pmf of NBin(r, p)

$$\begin{aligned}P(X = x) = \text{dnbinom}(\mathbf{x}, \mathbf{r}, \mathbf{p}) &= \binom{x + r - 1}{r - 1} (1 - p)^x p^r \quad \text{for } x = 0, 1, 2, \dots \\&= \binom{x + r - 1}{x} (1 - p)^x p^r \quad \text{for } x = 0, 1, 2, \dots\end{aligned}$$

- CDF of NBin(r, p)

$$P(X \leq x) = \text{pnbinom}(\mathbf{x}, \mathbf{r}, \mathbf{p})$$

- Expectation and Variance

$$E(X) = \frac{r(1-p)}{p} \quad \text{and} \quad V(X) = \frac{r(1-p)}{p^2}$$

Derivation of Negative Binomial pmf

R code for Negative Binomial(r, p)

x = [number of tails before r th head]

r = [number of heads to flip to]

p = [prob. of head in 1 flip]

```
pnbinom(3,5,.3)      #- F(3):  CDF of NegBin(n=5, p=.4) at x=3
```

```
dnbinom(3,5,.3)      #- p(3):  pmf of NegBin(n=5, p=.4) at x=3
```

```
layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side
```

```
x <- 0:30
```

```
plot(x, dnbinom(x, 5, .3), type="h", ylim=c(0,.5))  #- PMF plot -
```

```
plot(x, pnbinom(x, 5, .3), type="s", ylim=c(0,1))  #- CDF plot -
```

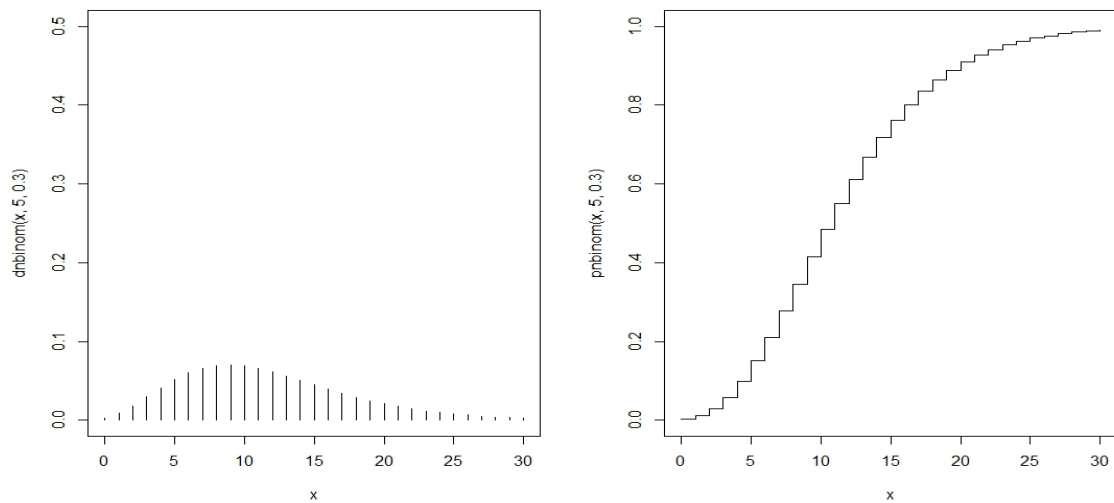


Figure 1: $r=5$, $p=.3$

Example: Shoot free throw until you make 10 shots

Suppose your free-throw percentage is 90%. Assume independence between each shots. You can't go home until you make 10 baskets, how many shots do you need to take before you go home?

3.3 Hypergeometric Random Variable

[\[ToC\]](#)

-
- **Analogy:** There are N balls with m reds. Pick n balls at once. $X = [\# \text{ of red picked}]$.
 - Pmf of Hypergeometric(n, m, N) is

$$P(X = x) = \text{dhyper}(x, n, N-m, n) = \frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$$

for $\max(0, n - N + m) \leq x \leq \min(n, m)$, and 0 otherwise.

- Expectation and Mean

$$E(X) = \frac{nm}{N} \quad \text{and} \quad V(X) = \frac{N-n}{N-1} n \frac{m}{N} \left(1 - \frac{m}{N}\right)$$

Derivation of Hypergeometric pmf

R code for Hypergeometric(n, m, N)

x =[number of heads]

n =[number of balls picked]

m =[number of Red balls]

N =[number of total balls]

```
dhyper(3,4,6,5)      #- p(3):  pmf of Hypergeoretric(m=4, N-m=6 ,n=5) at x=3    (That means N=10)
phyper(3,4,6,5)      #- F(3):  CDF of Hypergeoretric(m=4, N-m=6 ,n=5) at x=3
```

```
layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side
```

```
x <- 0:10
plot(x, dhyper(x, 4,6,5), type="h", ylim=c(0,1))  #- PMF plot -
plot(x, phyper(x, 4,6,5), type="s", ylim=c(0,1))  #- CDF plot -
```

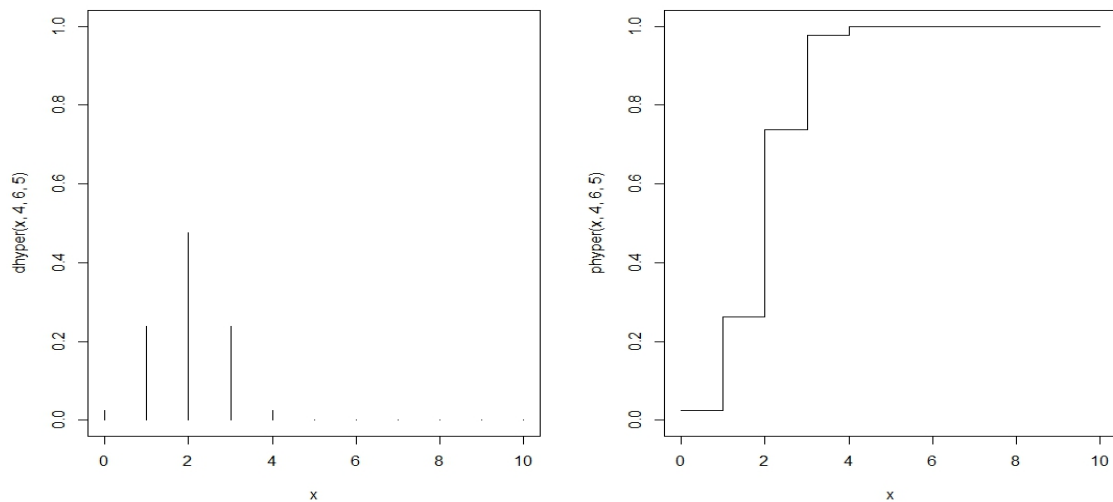



Figure 2: 10 balls, 4 are red. Pick 5 at once. X =number of red picked.

Example: Rock sampling

A geologist has collected 10 specimens of basaltic rock and 10 specimens of granite. The geologist instructs a laboratory assistant to randomly select 15 of the specimen for analysis.

- a What is the pmf of the number of granite specimen selected for analysis.
- b What is the probability hat all specimens of one of the two types of rock are selected for analysis.
- c What is the probability that the number of granite specimens selected for analysis is within 1 standard deviation of its mean value?

So if you let X to be the number of granite specimen selected, then X is Hypergeometric(15, 10, 20). Therefore,

b

$$\begin{aligned}
 & P\left((10 \text{ basaltic } 5 \text{ granite}) \cup (10 \text{ granite } 5 \text{ basaltic})\right) \\
 &= P((10 \text{ b } 5 \text{ g}) + P(10 \text{ g } 5 \text{ b}) \\
 &\quad - P((10b5g) \cap (10g5b)) \\
 &= P(X = 5) + P(X = 10) \\
 &= h(5; 15, 10, 20) + h(10; 15, 10, 20)
 \end{aligned}$$

c SD for Hypergeometric(15, 20, 10) is

$$\begin{aligned}
 \sigma &= \sqrt{\frac{N-n}{N-1} n \frac{m}{N} \left(1 - \frac{m}{N}\right)} = \sqrt{\frac{20-15}{20-1} 15 \frac{10}{20} \left(1 - \frac{10}{20}\right)} \\
 &= \sqrt{\frac{5}{19} \frac{15}{4}} \\
 &= 0.993
 \end{aligned}$$

Mean of X is

$$E(X) = \frac{mn}{N} = \frac{(10)15}{20} = 7.5.$$

Therefore we are looking for

$$\begin{aligned} P(7.5 - 0.993 \leq X \leq 7.5 + 0.993) &= P(7 \leq X \leq 8) \\ &= h(7; 15, 10, 20) + h(8; 15, 10, 20) \end{aligned}$$

Example: Capture-Recapture

One popular method of estimating population size of wild animal is called Capture-Recapture method. First, you capture m subjects, tag and release them. Then sometime later, you come back and capture n subjects. Within this n subjects, we count how many of them has a tag. Let X be the number of tagged subjects. Our logic is that since

$$E(X) = \frac{nm}{N},$$

It must be that

$$X \approx E(X) = \frac{nm}{N}.$$

If we go with this logic, then our estimator for N will be $\hat{N} = nm/X$.

If we use this estimator in the case $(N = 500, n = 100, m = 100)$, what will be the accuracy of this estimator \hat{N} ?

Capture-Recapture cont'd

Let's define our 'accuracy' as estimator \hat{N} being within 10% of the true value N .

If N is 500, then $P(\hat{N} \text{ will be within true value } \pm 10\%) = P(450 < \hat{N} < 550) = P(a < X < b)$

$N = 500$

n	m	a	b	$P(a < X < b)$ $= P(450 < \hat{N} < 550)$
10	10	0.18	0.22	0
20	20	0.72	0.88	0
50	50	4.5	5.5	0.19
100	100	18.1	22.2	0.42
200	200	72.7	88.8	0.86
300	300	163.6	200	0.999

If N is 5000, then $P(\hat{N} \text{ will be within true value } \pm 10\%) = P(4500 < \hat{N} < 5500) = P(a < X < b)$

$N = 5000$

n	m	a	b	$P(a < X < b)$ $= P(4500 < \hat{N} < 5500)$
100	100	1.8	2.2	.28
100	200	3.6	4.4	.20
200	200	7.3	8.9	.15
500	500	45.5	55.6	.57
1000	1000	181.8	222.2	.93

n = number of second round capture

m = number of first round capture-tag-release

3.4 Poisson Random Variable

[\[ToC\]](#)

- **Analogy:** events with rate λ per unit time. **Possible Values:** $0, 1, 2, 3, \dots$
- pmf of $\text{Poisson}(\lambda)$ is

$$P(X = x) = \text{dpois}(x, \text{lambda}) = \frac{e^{-\lambda} \lambda^x}{x!}$$

- Expectation and Variance:

$$E(X) = \lambda \quad \text{and} \quad V(X) = \lambda.$$

- **Poisson as a limit** If we let $n \rightarrow \infty$, $p \rightarrow 0$, in such a way that $np \rightarrow \lambda$, then the pmf

$$\text{dbinom}(x, n, p) \rightarrow \text{dpois}(x, \text{lambda}).$$

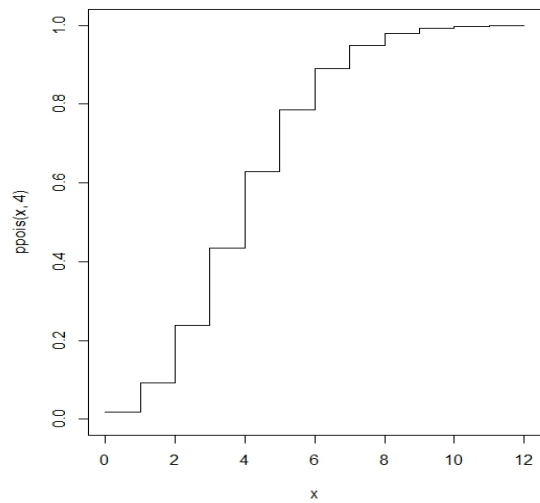
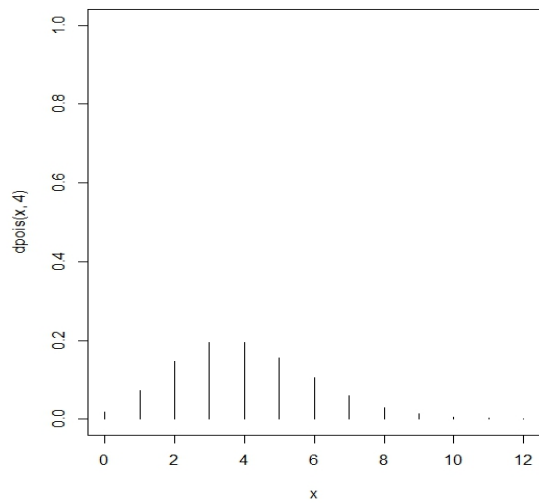
R code for Poisson(λ)

x =[number of events in a unit time] λ =[average number of events per unit time]

```
dpois(3,4)      #- p(3):  pmf of Poi(lambda=.5) at x=3
ppois(3,4)      #- F(3):  CDF of Poi(lambda=.5) at x=3

layout( matrix(1:2, 1, 2) )  #- Make plot layout side by side

x <- 0:12
plot(x, dpois(x, 4), type="h", ylim=c(0,1))  #- PMF plot -
plot(x, ppois(x, 4), type="s", ylim=c(0,1))  #- CDF plot -
```



When Time Units are Changed

Example: Number of TORNADOS

Suppose the number X of tornadoes observed in a particular region during a 1-year period has a Poisson distribution with $\lambda = 8$.

1. What is the probability we get fewer than 4 tornados next year?
2. What is the probability we get fewer than 6 tornados in next two years?

Example: Number of Tornados

Suppose the number X of tornadoes observed in a particular region during a 1-year period has a Poisson distribution with $\lambda = 8$.

1. What is the probability we get fewer than 4 tornados next year?

$$\begin{aligned} &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= \text{dpois}(0,8) + \text{dpois}(1,8) + \text{dpois}(2,8) + \text{dpois}(3,8) \\ \\ &= P(X \leq 3) = \text{ppois}(3, 8) \end{aligned}$$

2. What is the probability we get fewer than 6 tornados in next two years?

For two-year period, number of tornados has $\text{Pois}(16)$ distribution.

$$P(X \leq 5) = \text{ppois}(5, 16)$$

□

Example: Aircraft arrivals

Suppose small aircraft arrive at a certain airport according to a Poisson process with rate $\alpha = 8$ per hour, so that the number of arrivals during a time period of t hours is a Poisson r.v. with $\lambda = 8t$.

1. What is the probability that exactly 6 small aircraft arrive during 1-hour period?
2. What are the expected value and standard deviation of the number of small aircraft that arrive during a 90-min period?
3. What is the probability that at least 20 small aircraft arrive during 3 hour period?

Example: circuit boards

If proof testing of circuit boards, the probability that an particular diode will fail is .01. Suppose a circuit board contains 200 diodes.

- a How many diodes would you expect to fail, and what is the standard deviation of the number that are expected to fail?

Here we have $\text{Binomial}(200, .01)$, which can be approximated by $\text{Poisson}(200 \times .01)$. Therefore, expected value is 2, and standard deviation is $\sqrt{2}$.

- b What is the approximate probability that at least four diodes will fail on a randomly selected board?

Since we are looking at $\text{Poisson}(2)$,

$$P(\text{ at least 4}) = 1 - P(X \leq 3) = 1 - .857 = .143$$

from the table A.2.

- c If five boards are shipped to a particular customer, how likely is it that at least four of them will work properly?

Board will work only if all of 200 diodes works. Then

$$P(\text{ Board will work}) = P(X = 0) = .135$$

Now each board will work with probability of .135, Choosing five board and see how many of them work is like Binomial(5, .135). Therefore,

$$P(\text{ at least four boards work}) = b(4; 5, .135) + b(5; 5, .135) = 0.00418$$

3.5 R code for the Four Distributions

[\[ToC\]](#)

Binomial(n, p)

x = [number of heads] n = [number of flip] p = [prob. of head in 1 flip]

```
pbinom(3,10,.5)      #- F(3):  CDF of Bin(n=10, p=.5) at x=3
dbinom(3,10,.5)      #- p(3):  pmf of Bin(n=10, p=.5) at x=3
```

Negative Binomial(r, p)

x = [number of tails before r th head] r = [number of heads to flip to] p = [prob. of head in 1 flip]

```
pnbinom(3,5,.4)      #- F(3):  CDF of NegBin(r=5, p=.4) at x=3
dnbinom(3,5,.4)      #- p(3):  pmf of NegBin(r=5, p=.4) at x=3
```

Hypergeometric(n, m, N)

x =[number of heads]

n =[number of balls picked]

m =[number of Red balls]

N =[number of total balls]

```
phyper(3,4,6,5)      #- F(3):  CDF of Hypergeoretric(m=4, N-m=6 ,n=5) at x=3  (That means N=10)
dhyper(3,4,6,5)      #- p(3):  pmf of Hypergeoretric(m=4, N-m=6 ,n=5) at x=3
```

Poisson(λ)

x =[number of events in a unit time]

λ =[average number of events per unit time]

```
ppois(3,5)          #- F(3):  CDF of Poi(lambda=5) at x=3
dpois(3,5)          #- p(3):  pmf of Poi(lambda=5) at x=3
```