Chapter 3: Discrete Random Variables and Probability Distributions

Curtis Miller

2018-05-14

Introduction

AFTER WE DEFINE PROBABILITY measures and sample spaces, we can talk about random variables. The next two chapters focus on random variables, which translate random outcomes into mathematical objects, such as numbers.[1] This first chapter introduces random variables and a theory for discrete random variables. The second chapter focuses on continuous random variables.

Section 1: Random Variables

A **random variable** (sometimes abbreviated with **rv**) is a function taking values from the sample space $\mathcal{S}$ and associating numbers with them.[2] Conventional notation for random variables uses capital letters from the end of the English alphabet, while lower-case letters are used to denote a non-random value or outcome. If $\omega \in \mathcal{S}$, the notation $X(\omega) = x$ can be used to say that the value of the random variable X when the outcome ω occurs is x. The set $\{\omega : X(\omega) = x\}$ is the event that an element of $\mathcal{S}$ is drawn that causes the random variable X to equal x, and the set $\{\omega : X(\omega) \in A\}$ is the event that an element of $\mathcal{S}$ is drawn that causes the random variable X to assume a value that is in A.[3] Instead of writing $\mathbb{P}\left(\{\omega : X(\omega) \in A\}\right)$ we often write $\mathbb{P}\left(X \in A\right)$.

Random variables are commonly classified as being either discrete or continuous.[4] **Discrete** (real-valued) random variables take values in a finite or countably infinite (or enumerable, if you prefer) set with positive probability; these are effectively the only possible values. **Continuous** (real-valued) random variables satisfy the following two properties:

1. The random variable takes values in intervals (possibly infinite in length) or disjoint unions of intervals of the real line $\mathbb{R}$ with positive probability.
2. For any $c \in \mathbb{R}$, $\mathbb{P}\left(X = c\right) = 0$.

Perhaps the simplest non-trivial random variable is the **Bernoulli random variable**. If X is a Bernoulli random variable, then $\mathbb{P}\left(X = 1\right) = 1 - \mathbb{P}\left(X = 0\right) = p$, and we say $X \sim \text{Ber}(p)$ to mean this. The set $\mathcal{S}$

[1] In general random variables can produce any mathematical object. In this class random variables almost always produce real numbers, but in general random variables can also produce vectors or even functions, but this is well beyond the scope of the course.

[2] From this definition it's clear that random variables are neither random nor variables; they are functions mapping values from $\mathcal{S}$ to some other space, commonly the real numbers $\mathbb{R}$. They can be written $X : \mathcal{S} \to \mathbb{R}$ to emphasize this fact.

[3] The latter set is known as the **preimage** of A under X.

[4] There are random variables considered neither discrete nor continuous. One obvious example is a random variable that is a mixture of discrete and continuous random variables. For example, if a random variable X quantifies the number of hours slept per day, you may have $\mathbb{P}\left(X = 0\right) > 0$ and $\mathbb{P}\left(X = 24\right) > 0$ but all other outcomes are treated like the continuous case. Yet even then it's possible to define random variables that are not discrete, not continuous, and not a mixture of the two, those these are not seen in practice. In measure-theoretic probability theory, all of these cases are effectively indistinguishable; there isn't a separate theory for each type of random variable. But without this theory we handle discrete and continuous random variables separately.

on which $X(\omega)$ is defined could be just about anything. An inter-
esting result of probability theory is that if all I gave you was the
values of $X(\omega)$ without saying anything about $\mathcal{S}$ or how specifically
X assumes values given $\omega \in \mathcal{S}$, it is impossible statistically to deter-
mine what $\mathcal{S}$ is. The sample space is effectively forgotten. (In other
words, you wouldn't be able to tell the difference between a fair coin
and a Bernoulli random variable taking a value of 1 when the coin
lands heads-up after being flipped, or a fair die being rolled and
a Bernoulli random variable taking a value of 1 when the number
rolled is even.)

———————————————————

Example 1

Which of the following random variables are likely to be considered
discrete and which continuous? Describe the space of outcomes the
random variable takes with positive probability.

1. Flip a coin, record 1 for H, and 0 for T.

2. Roll a die, record the number of pips showing.

3. Roll a die, record 1 for an even number of pips and -1 for an odd
    number of pips.

4. The time (in minutes) needed to complete a race.

5. The length (in cm) of a hair plucked from a person's head.

6. Roll two dice and record the sum of the number of pips showing.

7. Flip a coin until H is seen and count the number of flips.

Example 2

Consider an experiment of rolling two six-sided die. Define two random variables for this experiment. Are they continuous or discrete?

Section 2: Probability Distributions for Discrete Random Variables

A **probability distribution** for a random variable is a function that describes the probability that a random variable takes on certain values. Discrete rv's are determined completely by the **probability mass function** (abbreviated **pmf**):

The pmf can be visualized using a **line graph**, where a line is placed on each point x of $\mathbb{R}$ that X takes with positive probability and extends to a height representing $p(x)$.

A **probability histogram** functions similarly to a line graph, but is a histogram, with bins centered on x of length 1 (usually) and with height $p(x)$.

Example 3

A fair coin is flipped; $X(H) = 1$ and $X(T) = 0$. Find the pmf of X, $p(x)$. Visualize $p(x)$ with a line graph.

The R package **discreteRV** [Buja et al., 2015] allows for defining
and working with discrete random variables in R. (It's pedagogically
useful but R's supplied discrete random variable functions are more
practical.)

```r
suppressPackageStartupMessages(  # Startup messages are annoying
  library(discreteRV)  # An R package for working with discrete random variables
)
```

```r
# A statement enclosed in parenthesis prints the variable that was assigned
(X <- RV(0:1, probs = c(1/2, 1/2)))
```
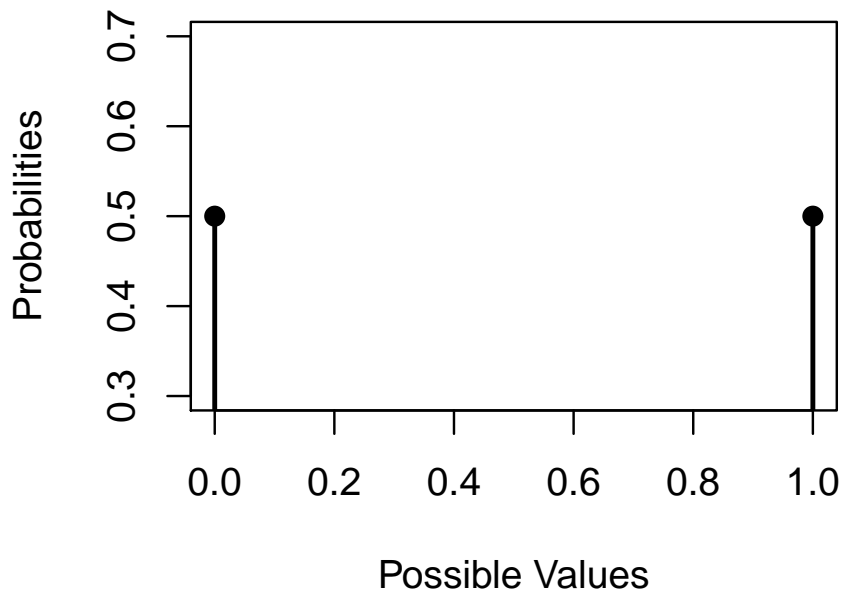
```
## Random variable with 2 outcomes
##
## Outcomes   0   1
## Probs    1/2 1/2
```

```r
plot(X)
```


Example 4

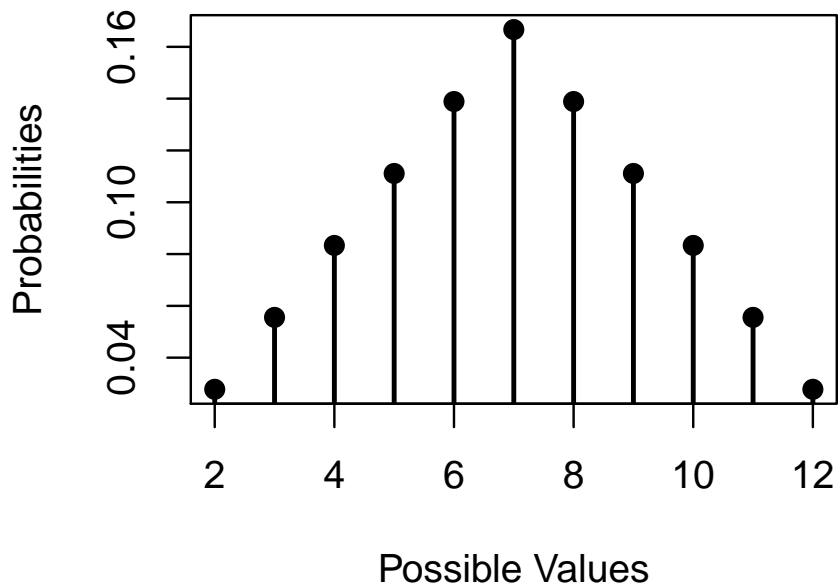Let *S* be the sum of the number of pips rolled on two dice. Find $p(s)$
and plot it.

```
(D <- RV(1:6, probs = rep(1/6, times = 6)))
```

```
## Random variable with 6 outcomes
##
## Outcomes   1   2   3   4   5   6
## Probs    1/6 1/6 1/6 1/6 1/6 1/6
```

```
(S <- SofIID(D))
```

```
## Random variable with 11 outcomes
##
## Outcomes    2    3    4    5    6    7    8    9   10   11   12
## Probs    1/36 1/18 1/12  1/9 5/36  1/6 5/36  1/9 1/12 1/18 1/36
```
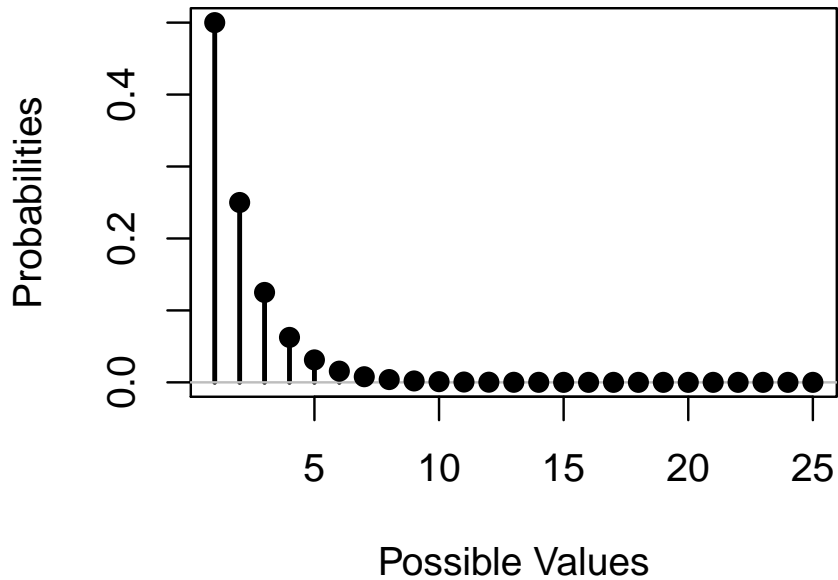
```
plot(S)
```


Example 5

Consider flipping a fair coin until H is seen. Let N be the number of flips. Find a pmf describing the distribution of N and plot the first few values of the pmf.

```
(N <- RV("geometric"))
```

```
## Random variable with outcomes from 1 to Inf
##
## Outcomes     1     2     3     4     5     6     7     8     9    10    11    12
## Probs    0.500 0.250 0.125 0.063 0.031 0.016 0.008 0.004 0.002 0.001 0.000 0.000
##
## Displaying first 12 outcomes
```

```
plot(N)
```

Sometimes we describe a distribution in terms of a **parameter**, which is a value that can be set to different possible values to generate a pmf. Probability distributions that differ only in the choice of parameters are called a **family** of distributions.

Example families with parameters include:

Example 6

Confirm that $p(x; N) = \frac{1}{N}$ for $x \in \{1, 2, ..., N\} = [N]$ is a valid pmf. This is the pmf of the discrete uniform distribution, $X \sim$ DUNIF$(1, N)$.

Example 7

Confirm that $f(n; p) = p(1 - p)^{n-1}$ is a valid pmf. This is the pmf of the geometric distribution, $X \sim$ GEOM(p).

The **cumulative distribution function** (abbreviated **cdf**) is defined below:

Notice the following relation between the cdf and the pdf:

In general, a function $F(x)$ is a cdf if it satisfies the following three properties:[5]

[5] If a function $F(x)$ satisfies these properties than there exists a random variable X with a cdf identical to $F(x)$.

For discrete rv's, cdf's are jump functions, resembling the following plot:

Like a pmf, a cdf completely characterizes a random variable. We can use it for computing probabilities of regions using the following rule:

Example 8

Compute and plot the cdf for a random variable $X \sim \text{Ber}(p)$.

```
# For p = 1/2

bercdf <- function(q) {pbinom(q, size = 1, prob = 1/2)}
plot(stepfun(0:1, bercdf((-1):1), right = TRUE), verticals = FALSE,
     main = "cdf")
```
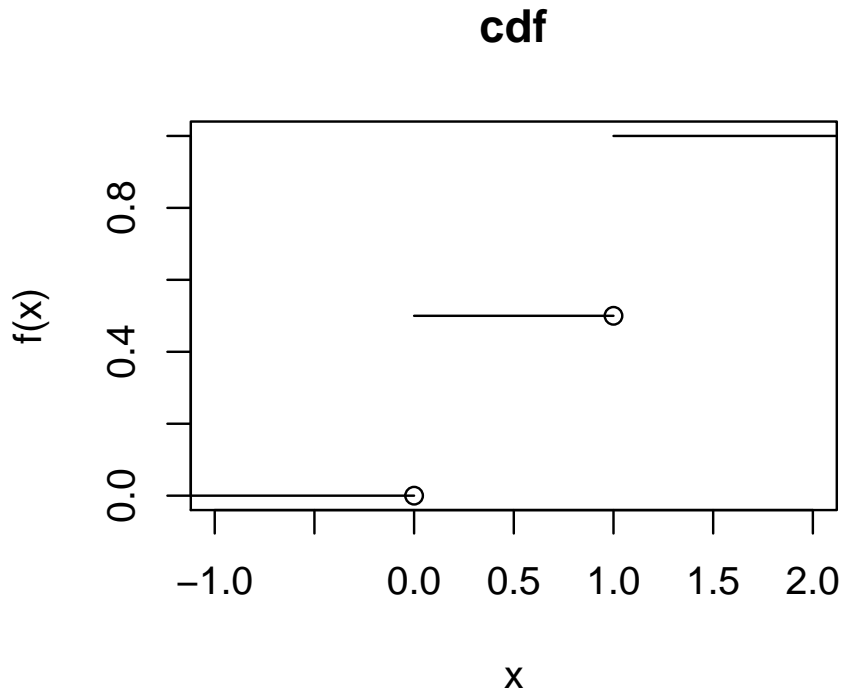

Example 9

Consider rolling a four-sided dice that produces numbers from 1 to 4
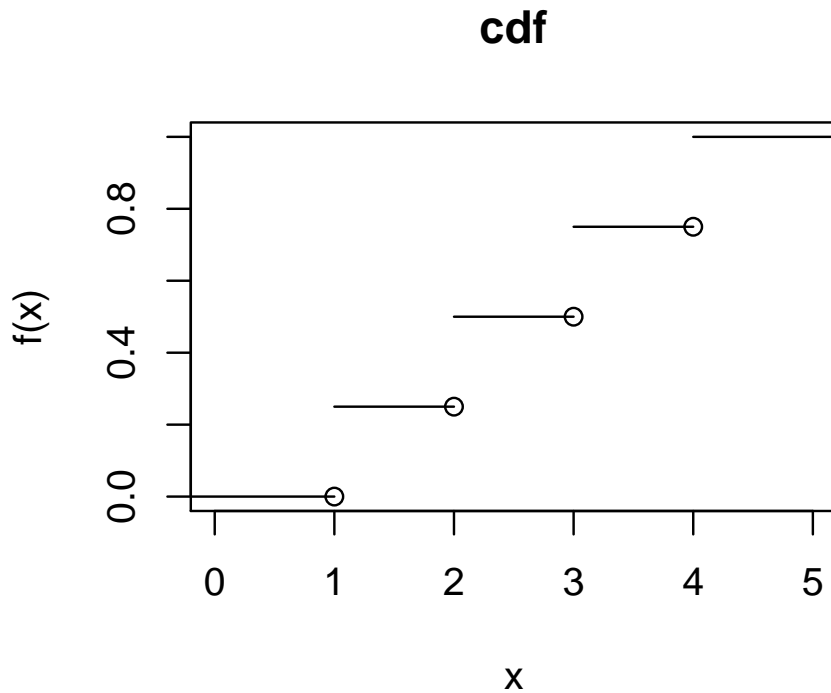(so $X \sim \text{DUNIF}(1, 4)$). Compute the cdf of X and plot it.

```
(U <- RV(1:4, probs = rep(1/4, 4)))

## Random variable with 4 outcomes
##
## Outcomes   1   2   3   4
## Probs    1/4 1/4 1/4 1/4

discunifcdf <- function(u) {P(U <= u)}
discunifcdf <- Vectorize(discunifcdf)
plot(stepfun(1:4, discunifcdf(0:4), right = TRUE), verticals = FALSE,
     main = "cdf")
```
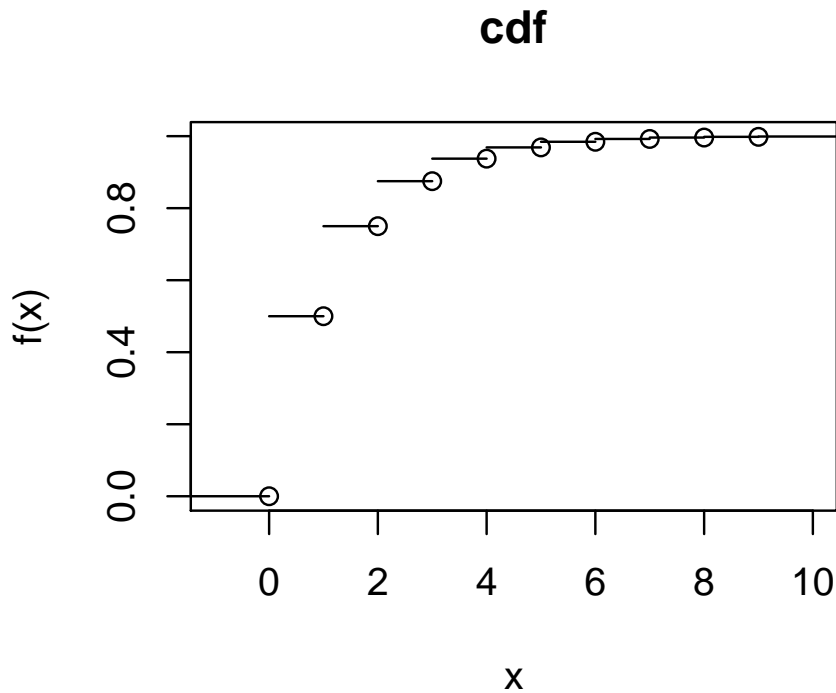

Example 10

Find the cdf of a geometric random variable with parameter p. What would a plot of the cdf look like?

```
geomcdf <- function(q) {pgeom(q, 1/2)}
plot(stepfun(0:9, geomcdf((-1):9), right = TRUE), vertical = FALSE,
     main = "cdf")
```


Section 3: Expected Values

The **expected value** for a discrete random variable, $\mathbb{E}[X]$, is given below:

$\mathbb{E}[X]$ is viewed as the population mean, μ, described in previous chapters. We can also compute the expected value of functions of X, $\mathbb{E}[h(X)]$, in a natural way:

The expected value is, in some sense, a "best prediction"[6] for the value of X.

[6] Specifically, let $\mathbb{E}\left[(X - \hat{\mu})^2\right]$ represent the expected squared error, and $\hat{\mu}$ represents a prediction for X; when this quantity exists (sometimes it doesn't), the value of $\hat{\mu}$ that minizes the expected squared error is $\hat{\mu} = \mathbb{E}[X]$.

Example 11

Compute the expected value for $X \sim \text{Ber}(p)$, $S \sim \text{DUNIF}(1,6)$, and $N \sim \text{GEOM}(p)$ (as seen in examples 3, 4, and 5).

```
E(X)
```

```
## [1] 0.5
```

```
E(S)
```

```
## [1] 7
```

```
E(N)  # Approximate
```

```
## [1] 1.999999
```

Expectations are linear functions acting on random variables.[7]

Proposition 1.

$$\mathbb{E}\left[aX + b\right] = a\mathbb{E}\left[X\right] + b$$

[7] This is true for all random variables though we work with the discrete case only for now.

The **variance** of a random variable is given by:

There is a handy formula for computing the variance that is often easier than computing it directly:

Proposition 2.

$$\mathrm{Var}\left(X\right) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}\left[X\right]\right)^2$$

Var (X) is thought of as the population variance and is often denoted Var $(X) = \sigma^2$. From this we get the population standard deviation, $\sigma = \sqrt{\sigma^2}$.

Example 12

Compute the variance and standard deviation of the random variables listed in Example 11.

```
V(X)
```

```
## [1] 0.25
```

```
SD(X)
```

```
## [1] 0.5
```

```
V(S)
```

```
## [1] 5.833333
```

```
SD(S)
```

```
## [1] 2.415229
```

```
V(N)    # Approximate
```

```
## [1] 1.999981
```

```
SD(N)
```

```
## [1] 1.414207
```

Proposition 3.

$$\text{Var}\,(aX + b) = a^2 \,\text{Var}\,(X)$$
$$\sigma_{aX+b} = |a|\,\sigma_X$$

where σ_X is the standard deviation of X.

Expectations need not be finite or even exist, as demonstrated in the following example:[8]

[8] This example is known as the St. Petersburg paradox, and is famous for how unintuitive its solution is and how strongly humans underestimate the game's value. The game gets its name due to its resolution by Daniel Bernoulli in 1738 [Bernoulli, 1954], who lived in St. Petersburg at the time.

Example 13

Consider a game where a fair coin is flipped until it lands heads-up. A player would earn $1 if the game ends with 1 flip, $2 if it ends with two flips, $4 if it ends with three flips, $8 if it ends with four flips, and so on. The "fair price" of a game corresponds to the game's expected payout. What, then, is the fair price to play this game?

Section 4: The Binomial Probability Distribution

A **binomial experiment** is an experiment that satisfies the following requirements:

1. The experiment consists of n Bernoulli trials that end in either in "success", S, or "failure", F.

2. The trials are independent.

3. For each trial, $\mathbb{P}(S) = 1 - \mathbb{P}(F) = p \in (0, 1)$.

   We can think of the outcome of an experiment as a sequence of S and F, such as $SSFSF$ (here, $n = 5$).
   The **binomial random variable** associated with a binomial experiment counts the number of "successes" in the experiment: $X(\omega) = \{\# \text{ of } S \text{ in } \omega\}$; we write $X \sim \text{BIN}(n, p)$. For example, $X(SSFSF) = 3$.
   We denote the pmf of X with $b(x; n, p)$. This is 0 for x that is not an integer from 0 to n. For $x \in \{0, 1, \ldots, n\}$, it can be computed:

The cdf of X is given below:

$\mathbb{E}[X]$, Var (X), and σ_X are given below:[9]

9 We can compute these algebraically or with a probabilistic argument. The former is algebraically tedious while the latter is illuminating and easy. We revisit this in Chapter 5.

Select values of $B(x; n, p)$ are given in Table A.1 of the textbook.

Example 14

You flip a fair coin ten times.

1. What is the probability you see exactly 4 heads? (Do so without using a table.)

2. If $X \sim \text{BIN}(10, 0.5)$, compute $\mathbb{P}(4 < X \leq 6)$.

3. Compute $\mathbb{P}\left(2 \leq X \leq 4\right)$.

4. What is the probability you see more than 7 heads?

5. Compute $\mathbb{E}\left[X\right]$, $\text{Var}\left(X\right)$, and σ_X.

```r
dbinom(4, size = 10, prob = 0.5)  # 1
```

```
## [1] 0.2050781
```

```r
pbinom(6, size = 10, prob = 0.5) - pbinom(4, size = 10, prob = 0.5)  # 2
```

```
## [1] 0.4511719
```

```r
pbinom(4, size = 10, prob = 0.5) - pbinom(2 - 1, size = 10, prob = 0.5)  # 3
```

```
## [1] 0.3662109
```

```r
1 - pbinom(7, size = 10, prob = 0.5)  # 4
```

```
## [1] 0.0546875
```

```r
pbinom(7, size = 10, prob = 0.5, lower.tail = FALSE)  # Alternative to 4
```
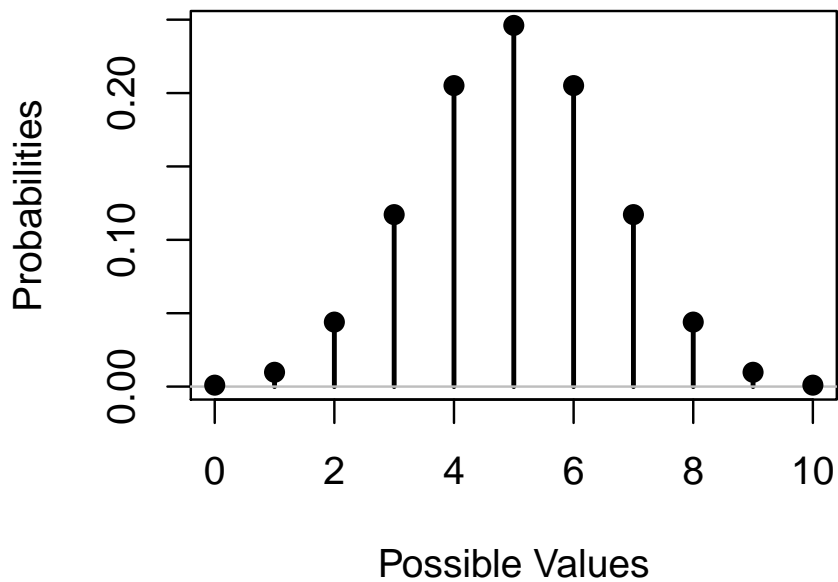
```
## [1] 0.0546875
```

```r
X <- RV(0:10, probs = dbinom(0:10, size = 10, prob = 0.5))
E(X)  # 5
```

```
## [1] 5
```

```r
V(X)
```

```
## [1] 2.5
```

```r
SD(X)
```

```
## [1] 1.581139
```

```r
plot(X)
```

Example 15

A manufacturer of widgets send batches of widgets in giant bins. Your company will accept a shipment of widgets if no more than 7% of widgets are defective. The procedure for deciding whether a shipment is defective is to choose four widgets from the batch at random, without replacement. If more than one widget is defective, the batch is rejected. What is the probability of rejecting the batch if 7% of the widgets are defective? Model the process using a binomial random variable.[10]

[10] We can view the batch of widgets as the entire population and we are choosing a subsample of that population without replacement. Binomial random variables draw "successes" and "failures" from an infinite population, not a finite one, and thus a different probability distribution should describe this experiment (it is the subject of the next section). It is safe, though, to treat a finite population like an infinite one if your sample size does not exceed 5% of the population size.

```
pbinom(1, 4, 0.07, lower.tail = FALSE)
```

```
## [1] 0.02672803
```

Example 16

I claim that I can make 80% of my free-throw shots when playing basketball. You plan to test me by having me shoot 20 baskets; if I make fewer baskets than a specified amount, you will call me a liar. The threshold amount of baskets is chosen so that the probability I make less than this amount given that I am, in fact, an 80% free-throw shooter does not exceed 5%. What is the threshold amount?

Additionally, compute the mean and standard deviation of the number of shots I would make if my claim is true.

```r
qbinom(0.05, 20, 0.8) - 1
```

```
## [1] 12
```

```r
# The subtraction is due to how qbinom defines quantiles; see documentation
X <- RV(0:20, probs = dbinom(0:20, size = 20, prob = 0.8))
plot(X)
```



```r
E(X)
```

```
## [1] 16
```

```r
V(X)
```

```
## [1] 3.199998
```
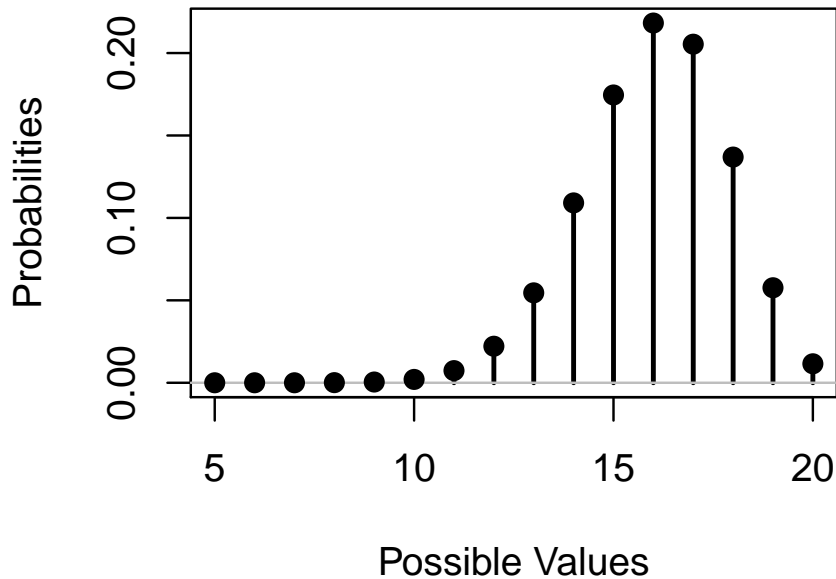
```r
SD(X)
```

```
## [1] 1.788854
```

Section 5: Hypergeometric and Negative Binomial Distributions

The **hypergeometric distribution** is the finite population analogue to the binomial distribution. The population has N elements labeled S or F (for "success" or "failure"). There are M S's in the population (and thus $N - M$ F's). A sample of size n is chosen from the sample without replacement in such a way that each subset of n elements is equally likely.[11] $X \sim \text{HYPERGEOM}(n, M, N)$ denotes a random

[11] The hypergeometric distribution should resemble the binomial distribution as N becomes large. In fact it can be shown that if M is replaced with M_N and $\frac{M_N}{N} \to p$ as $N \to \infty$, the hypergeometric distribution becomes the binomial distribution in the limit.

variable following the hypergeometric distribution.

For an integer x satisfying $\max(0, n - N + M) \leq x \leq \min(n, M)$, the pmf of the hypergeometric distribution is given below:

Below are $\mathbb{E}[X]$ and $\mathrm{Var}(X)$:

Example 16

A manufacturer of widgets send batches of widgets in giant bins. Your company will accept a shipment of widgets if no more than 6% of widgets are defective. The procedure for deciding whether a shipment is defective is to choose four widgets from the batch at random, without replacement. If more than one widget is defective, the batch is rejected. The batch sent contains 50 widgets. What is the probability of rejecting the batch if 6% of the widgets are defective?

Also, compute the mean and variance of X, the number of defective widgets in the sample, under the assumption that 6% of widgets are defective.
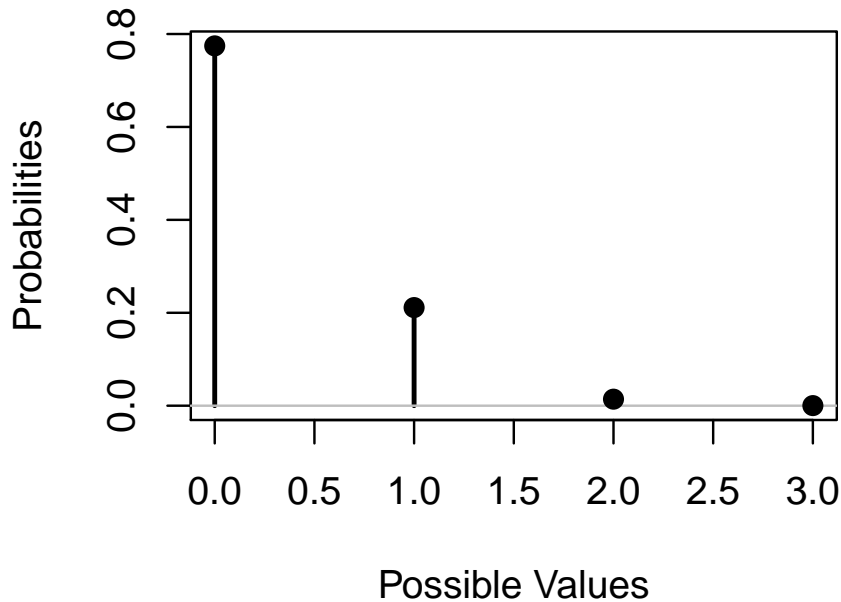
```r
phyper(1, 50 * .06, 50 * (1 - .06), 4, lower.tail = FALSE)
```

```
## [1] 0.01428571
```

```r
pbinom(1, 4, .06, lower.tail = FALSE)  # For comparison
```

```
## [1] 0.01991088
```

```r
X <- RV(0:4, probs = dhyper(0:4, 50 * .06, 50 * (1 - .06), 4))
plot(X)
```



```r
E(X)
```

```
## [1] 0.24
```

```r
V(X)
```

```
## [1] 0.2117878
```

Example 17

It is election night in the small town of Studentsville, and Jack Johnson is running for mayor against bitter rival, John Jackson. Votes have been cast and are being counted. There are 1024 ballots cast and among the 200 ballots counted, 116 were cast for Jack Johnson. If the election were actually a tie, what would be the probability of observing 116 ballots or more cast for Jack Johnson? What does this say about who is likely winning the election?

```r
phyper(116 - 1, 512, 512, 200, lower.tail = FALSE)
```

```
## [1] 0.007203238
```

Consider flipping a coin with probability p of landing heads-up (all flips independent). Flip a coin until r heads have been seen, and count the number of tails seen until the experiment ended; let the random variable X represent this count. Then X follows the **negative binomial distribution**, or $X \sim \text{NB}(r, p)$.[12] The pmf of X is given below:

[12] Let $r = 1$. Then $Y = 1 + X$ follows the geometric distribution, or $Y \sim \text{GEOM}(p)$.

Additionally, below are $\mathbb{E}[X]$ and $\text{Var}(X)$:

Example 18

A husband and wife plan to have children until they have exactly two boys; after this, they will stop attempting to have children. Assume that the probability of giving birth to a boy is 51%.

1. What is the probability they will have two girls before stopping attempting to have more children?

2. What is the probability they will need at least four children?

3. What is the expected number of children they will have? What is the variance of this random variable?

```r
dnbinom(2, 2, 0.51)  # 1
```

```
## [1] 0.18735
```

```r
pnbinom(4 - 2 - 1, 2, 0.51, lower.tail = FALSE)  # 2
```

```
## [1] 0.485002
```

```r
nbinom_func <- function(x) {dnbinom(x, 2, 0.51)}
(X <- RV(c(0, Inf), nbinom_func))
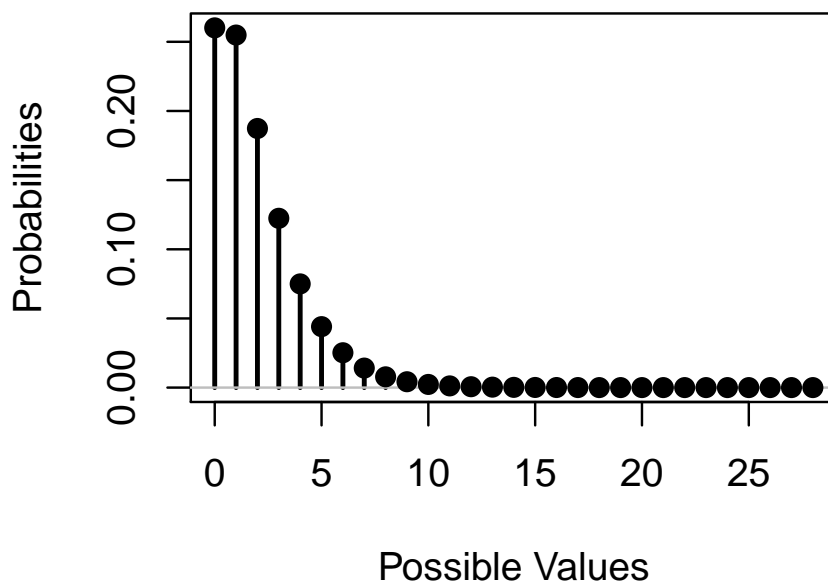```

```
## Random variable with outcomes from 0 to Inf
##
## Outcomes     0     1     2     3     4     5     6     7     8     9    10    11
## Probs    0.260 0.255 0.187 0.122 0.075 0.044 0.025 0.014 0.008 0.004 0.002 0.001
##
## Displaying first 12 outcomes
```

```r
plot(X)
```



```r
E(X)  # 3
```

```
## [1] 1.921568
```

```r
V(X)
```

```
## [1] 3.767769
```

Section 6: The Poisson Probability Distribution

$X \sim \text{POI}(\mu)$, or X follows the Poisson distribution with parameter μ, if the pmf of X is given by:

Is this a valid pmf? Yes.

If $X \sim POI(\mu)$, $\mathbb{E}[X] = \text{Var}(X) = \mu$.

Table A.2 of your textbook gives the cdf of select Poisson distributions.

The Poisson distribution describes random variables that follow the Poisson process. This process describes the number of times an event occurs over an interval of time.[13] So the probability an event occurs k times during an interval of time of length t is given by $P_k(t) = \frac{e^{-\alpha t}(\alpha t)^k}{k!}$.

[13] This interpretation comes from a relationship between Poisson random variables and binomial random variables; if $p_n \to 0$ but $np_n \to \mu$ as $n \to \infty$, $b(x; n, p_n) \to p(x; \mu)$ as $n \to \infty$, where $p(\cdot; \mu)$ is the pmf of a Poisson random variable.

Below is a plot of a simulated Poisson process.

```r
jumps <- rexp(4)
plot(stepfun(cumsum(jumps), 0:4, right = TRUE), vertical = FALSE,
     xlab = "Time", ylab = "Value", main = "Simulated Poisson Process",
     xlim = c(0, ceiling(sum(jumps))))
```
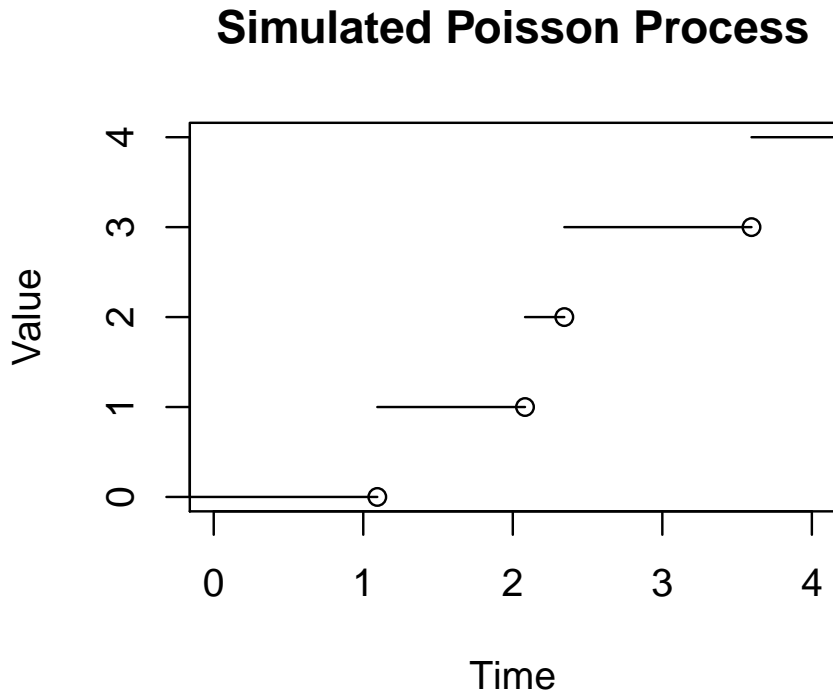

Example 19

On average, your daughter's soccer team scores 10 points per game. Assume that the number of points scored per game by her team follows a Poisson process.

1. What is the probability her team will score 7 points during the game?

2. What is the probability that by half time your daughter's team will have scored two points?

3. What is the probability your daughter's team will score between 3 and 6 points (inclusive) during the game?

4. What is the probability your daughter's team will score more than three points in the last half of the game?

5. What is the probability that in two games your daughter's team will score between 15 and 18 points (inclusive)?

```
poiproc <- function(x, t) {dpois(x, 10 * t)}
(X1 <- RV(c(0, Inf), poiproc, t = 1))

## Random variable with outcomes from 0 to Inf
##
## Outcomes     0     1     2     3     4     5     6     7     8     9    10    11
## Probs    0.000 0.000 0.002 0.008 0.019 0.038 0.063 0.090 0.113 0.125 0.125 0.114
##
## Displaying first 12 outcomes

(Xhalf <- RV(c(0, Inf), poiproc, t = 1/2))

## Random variable with outcomes from 0 to Inf
##
## Outcomes     0     1     2     3     4     5     6     7     8     9    10    11
## Probs    0.007 0.034 0.084 0.140 0.175 0.175 0.146 0.104 0.065 0.036 0.018 0.008
##
## Displaying first 12 outcomes

(X2 <- RV(c(0, Inf), poiproc, t = 2))

## Random variable with outcomes from 0 to Inf
##
## Outcomes     1     2     3     4     5     6     7     8     9    10    11    12
## Probs    0.000 0.000 0.000 0.000 0.000 0.000 0.001 0.001 0.003 0.006 0.011 0.018
##
## Displaying first 12 outcomes

plot(X1)
```
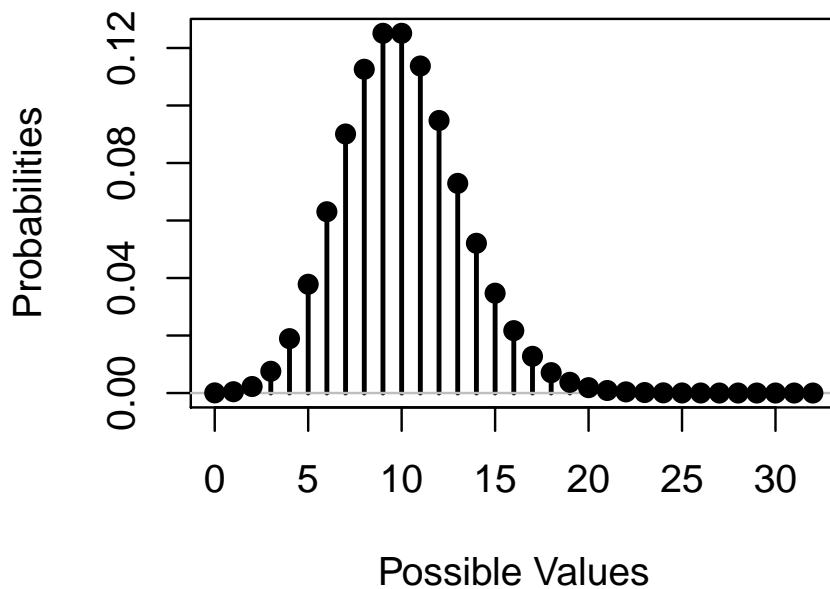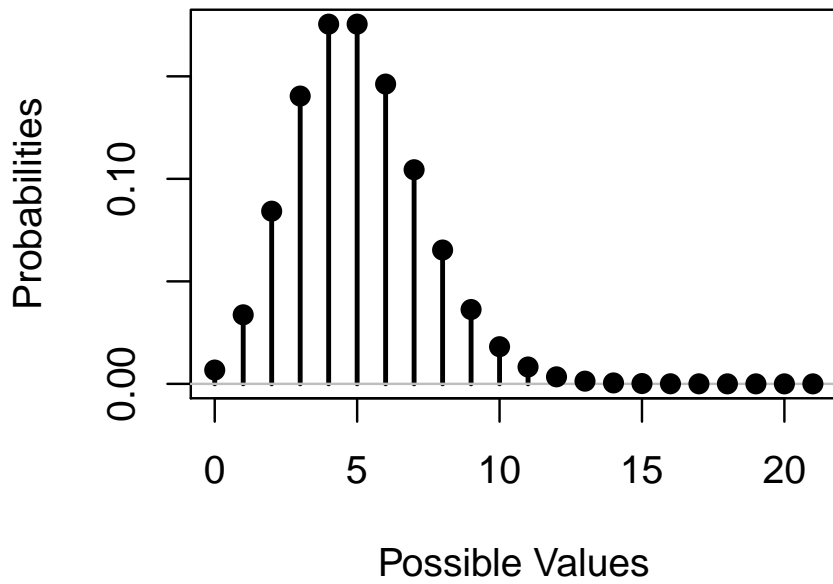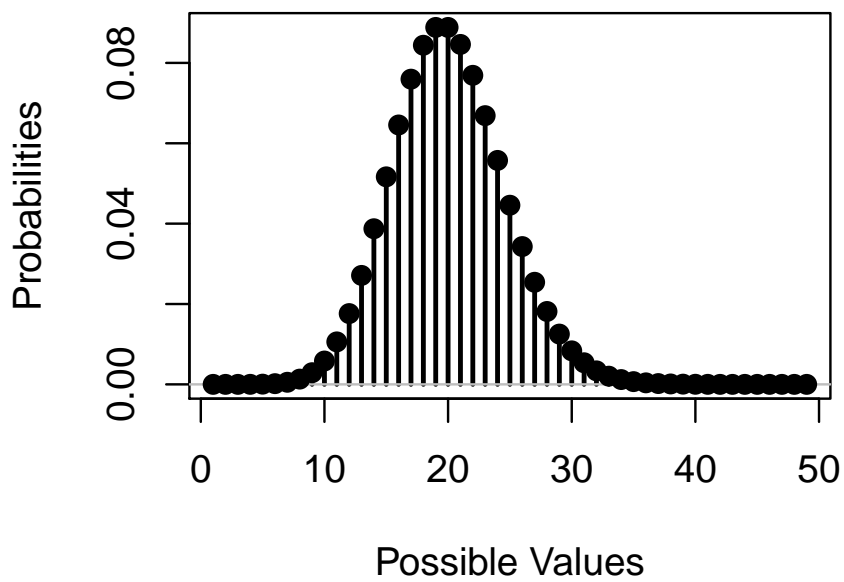
```
plot(Xhalf)
```



```
plot(X2)
```



```
P(X1 == 7)   # 1
```

```
## [1] 0.09007923
```

```
P(Xhalf == 2)   # 2
```

```
## [1] 0.08422434
```

```
P((X1 >= 3) %AND% (X1 <= 6))   # 3
```

```
## [1] 0.127372
```

```
P(Xhalf > 3)  # 4
```

```
## [1] 0.7349741
```

```
P((X2 >= 15) %AND% (X2 <= 18))  # 5
```

```
## [1] 0.2765577
```

Tables for the Poisson distribution can be used for approximating binomial distribution probabilities when n is large and p is small. Then $b(x; n, p) \approx p(x; np)$.

Example 20

Use the Poisson approximation to estimate $B(4; 200, .01)$.

```r
pbinom(4, 200, .01)
```

```
## [1] 0.9482537
```

```r
ppois(4, 200 * .01)
```

```
## [1] 0.947347
```

References

Daniel Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):23–36, 1954. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/1909829`.

Andreas Buja, Eric Hare, and Heike Hofmann. *discreteRV: Create and Manipulate Discrete Random Variables*, 2015. URL `https://CRAN.R-project.org/package=discreteRV`. R package version 1.2.2.