Chapter 3 - Discrete Random Variables and Probability Distributions
Outline

1. Random variables
2. Discrete random variables and distributions
3. Expected values of discrete random variables
4. Binomial probability distribution
5. Hypergeometric and negative binomial distributions
6. Poisson probability distribution
Outline

1. Random variables
2. Discrete random variables and distributions
3. Expected values of discrete random variables
4. Binomial probability distribution
5. Hypergeometric and negative binomial distributions
6. Poisson probability distribution
Outline

1. Random variables
2. Discrete random variables and distributions
3. Expected values of discrete random variables
4. Binomial probability distribution
5. Hypergeometric and negative binomial distributions
6. Poisson probability distribution
Outline

1. Random variables
2. Discrete random variables and distributions
3. Expected values of discrete random variables
4. Binomial probability distribution
5. Hypergeometric and negative binomial distributions
6. Poisson probability distribution
Outline

1. Random variables
2. Discrete random variables and distributions
3. Expected values of discrete random variables
4. Binomial probability distribution
5. Hypergeometric and negative binomial distributions
6. Poisson probability distribution
Outline

1. Random variables
2. Discrete random variables and distributions
3. Expected values of discrete random variables
4. Binomial probability distribution
5. Hypergeometric and negative binomial distributions
6. Poisson probability distribution
Random Variables

Sample spaces with general types of outcomes are not easy to model mathematically. For mathematical modeling we define

A random variable is any rule that associates a number with each outcome in the sample space $S$.

Bernoulli experiment - in an experiment with a binary outcome, such as success/failure, we will define the random variable success $\rightarrow 1$; failure $\rightarrow 0$.

Counts - frequently the outcome of an experiment is the total number of times that a particular event occurred. This is already a numeric outcome.

Measurements - many of the examples in chapter 1 (temperature at space shuttle launches, material strength, yardage of golf courses, power consumption) are already numeric outcomes on a continuous scale.
Discrete random variables

**Definition** A random variable that can only assume distinct values is said to be *discrete*. Usually these represent a count.

- **Bernoulli** A Bernoulli experiment provides a 0/1 response.
- **Binomial** A binomial rv gives the number of successes in $n$ independent, identical trials. Possible values are $0, 1, \ldots, n$.
- **Geometric** Number of objects tested until a success. Possible values are $1, 2, 3, \ldots$.

Note that discrete random variables can have a finite range or an infinite range.
Continuous random variables

Definition  A random variable that can (theoretically) assume any value in a finite or infinite interval is said to be continuous.

Measurements  Selection of a random location in the continental United States and measure the height above sea level. The result could be any value in the approximate range $[-290, 14500]$.

Time to failure  The result is potentially any positive number.

Round-off error  Round-off error in calculations is generally modeled as a uniform continuous distribution.
A discrete distribution is described by giving its *probability mass function*, or *pmf*, either as a table or as a function.

Properties:

- For any $x$, $p(x) = P(X = x) = P(\text{all } x \in \Omega : X(s) = x)$. In theory $p(x)$ is defined for any $x$ but in practice it is zero except at selected distinct values.
- In $R$ the pmf's for various distribution classes are functions whose names start with *d*, such as `dbinom`, `dgeom`, `dpois`.
- $p(x) \geq 0$, $-\infty < x < \infty$
- $\sum_x p(x) = 1$
Parameters of a distribution

Frequently we describe an entire family of distributions that depend upon one or more variable quantities, called *parameters*.

**Bernoulli** A Bernoulli probability distribution depends upon the success of the trial. If we call this $\alpha$ then the probability function is

$$p(x; \alpha) = \begin{cases} 
1 - \alpha & \text{if } x = 0 \\
\alpha & \text{if } x = 1 \\
0 & \text{otherwise}
\end{cases}$$

**Geometric** A related distribution is the total number of trials until the first “success” occurs. The parameter is $p$, the probability of success on a given trial.

$$p(x; p) = \begin{cases} 
(1 - p)^{x-1}p & x = 1, 2, \ldots \\
0 & \text{otherwise}
\end{cases}$$
Cumulative Distribution Function

Definition  The cumulative distribution function (cdf) $F(x)$ of a discrete random variable $X$ with pmf $p(x)$ is

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

Interpretation  The cdf accumulates the probability to the left of $x$. For a discrete random variable its graph is a step function.

$R$ functions  In $R$ the cdf's for various distribution are functions whose names start with $p$, such as `pbinom`, `pgeom`, `ppois`.

Limits  If $X$ has a finite range then $F(x) = 0$ to the left of the range and $F(x) = 1$ to the right. In all cases

$$\lim_{x \downarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \uparrow \infty} F(x) = 1$$
Use One use of the cdf is in calculating probabilities. When $a$ and $b$ are any two numbers such that $a \leq b$

$$P(a \leq X \leq b) = F(b) - F(a^-)$$

The notation $F(a^-)$ indicates evaluation of $F$ at the point immediately to the left of $a$. For most discrete random variables, $F$ only changes at integer values so $F(a^-) = F(a - 1)$. In this case

$$p(a) = P(X = a) = F(a) - F(a - 1).$$

General R calculation The function `cumin` can be used with a vector of probabilities to calculate the cumulative distribution function values. A plot with `type="s"` is used to get the “stair-step” pattern of a CDF.
Plot of CDF

> xyplot(cumsum(c(0.4, 0.3, 0.2, 0.1)) ~ 1:4, ylab = "F(x)",
+       type = "s")

Cumulative distribution function for Example 3.11, p. 105
Enhanced Plot of CDF

Adding a couple of zeros extends the line to define its extent better.

```r
> xyplot(cumsum(c(0, 0.4, 0.3, 0.2, 0.1, 0)) ~ c(0.9, 1:4, + 4.1), ylab = "F(x)", type = "s")
```

Cumulative distribution function for Example 3.11, p. 105
Plots from CDF’s defined in R

The pmf and cdf defined in Example 3.12, page 106 is known as the geometric probability distribution. In the text the value $x$ is the total number of trials until a “success”. In $R$, $x$ is defined as the number of failures before a success.

```r
> xyplot(pgeom(0:40, prob = 0.51) ~ 0:40, type = c("g", + "s"))
```

Cumulative distribution function for geometric distribution with $p = 0.51
Expected value

Definition The *expected value* of a discrete random variable is the weighted sum of the outcomes using the probabilities as weights. If $X$ is a discrete r.v. with possible values in $D$ then

$$E(X) = \mu_x = \sum_{x \in D} x \cdot p(x)$$

The expected values for the distributions in examples 3.16 and 3.18 are

```r
> prob <- c(0.002, 0.001, 0.002, 0.005, 0.02, 0.04, 0.18,
>           0.37, 0.25, 0.12, 0.01)
> sum(prob)
[1] 1
> sum(0:10 * prob)
[1] 7.15
> sum(0:40 * dgeom(0:40, prob = 0.5))
[1] 1
```

Note that the definition of $\mu_x - x$ is off by 1 from the definition in the text.
Expected value of a function of $x$

Definition: If the rv $X$ has set of possible values $D$ and pmf $p(x)$, then the expected value of any function $h(X)$, denoted $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

Special case: $E(aX + b) = a \cdot E(X) + b$ or $\mu_{aX+b} = a \cdot \mu_X + b$
Definition Let $X$ have pmf $p(x)$ and expected value $\mu$. Then the variance of $X$, denoted $V(X)$ or $\sigma_X^2$ or just $\sigma^2$, is

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E[(X - \mu)^2]$$

The standard deviation (SD) of $X$ is $\sigma_X = \sqrt{\sigma_X^2}$

Shortcut $V(X) = \sigma^2 = \left[ \sum_D x^2 \cdot p(x) \right] - \mu^2 = E(X^2) - [E(X)]^2$

Linear combinations $V(aX + b) = \sigma_{aX+b}^2 = a^2 \cdot \sigma_X^2$ and $\sigma_{aX+b} = |a| \sigma_X$.

These imply that $\sigma_{aX}^2 = a^2 \cdot \sigma_X^2$ and $\sigma_{X+b}^2 = \sigma_X^2$.
Definition  A binomial experiment satisfies the conditions

1. The experiment consists of $n$ trials ($n$ is fixed in advance)
2. The trials are identical, and each trial can result in one of two possible outcomes called success ($S$) or failure ($F$).
3. The trials are independent, so the outcome on any particular trial does not affect the other outcomes.
4. The probability of success, called $p$, does not vary from trial to trial. (This is implied by condition 2 — identical trials.)

A count of the total number of successes in $n$ trials of a binomial experiment is a binomial random variable with pmf

$$b(x : n, p) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$
Binomial distribution in R

Description: Density, distribution function, quantile function and random generation for the binomial distribution with parameters ‘size’ and ‘prob’.

Usage:
- `dbinom(x, size, prob, log = FALSE)`
- `pbinom(q, size, prob, lower.tail = TRUE, log.p = FALSE)`
- `qbinom(p, size, prob, lower.tail = TRUE, log.p = FALSE)`
- `rbinom(n, size, prob)`

Arguments:
- `x, 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`: number of trials.
- `prob`: probability of success on each trial.
- `...`
Example 3.31, page 123 is a binomial with \( n = 15 \) and \( p = 0.2 \)

\[
\text{> xyplot(dbinom(0:15, size = 15, prob = 0.2) \sim 0:15, type = "h")}
\]
\[
\text{> xyplot(pbinom(0:15, size = 15, prob = 0.2) \sim 0:15, type = "s")}
\]
\[
\text{> sum(dbinom(0:8, size = 15, prob = 0.2))}
\]

[1] 0.999215

\[
\text{> pbinom(8, 15, 0.2)}
\]

[1] 0.999215

\[
\text{> dbinom(8, 15, 0.2)}
\]

[1] 0.003454764

\[
\text{> pbinom(8, 15, 0.2) - pbinom(7, 15, 0.2)}
\]

[1] 0.003454764

\[
\text{> sum(dbinom(8:15, 15, 0.2))}
\]

[1] 0.00423975
Calculations with binomial distributions (cont'd)

```r
c > pbinom(7.5, 15, 0.2, lower = FALSE)
[1] 0.00423975
c > 1 - pbinom(7, 15, 0.2)
[1] 0.00423975
c > sum(dbinom(4:7, 15, 0.2))
[1] 0.3475981
c > sum(0:15 * dbinom(0:15, 15, 0.2))
[1] 3
c > sum((0:15 - 15 * 0.2)^2 * dbinom(0:15, 15, 0.2))
[1] 2.4
c > c(expected = 15 * 0.2, variance = 15 * 0.2 * 0.8)
expected variance
    3.0 2.4```
pmf of binomial

pmf of a binomial distribution with $n = 15$ and $p = 0.2$
cdf of binomial

cdf of a binomial distribution with $n = 15$ and $p = 0.2$
Sampling with or without replacement

**Binomial**  Technically, a binomial experiment requires independent trials.

**Sampling with replacement**  If we sample $k$ objects from a population of size $N$ the samples will be independent only when we replace the object in the current sample before taking the next sample.

**Sampling without replacement**  If we do not replace the object after every sample we change the probabilities on the second and subsequent samples. For a small sampling fraction ($k/N \ll 1$) we can act as if we were sampling with replacement.

**Hypergeometric**  If $X$ is the number of $S$s in a completely random sample of size $n$ drawn from a population consisting of $M$ $S$s and $N-M$ $F$s then $X$ has a hypergeometric distribution.
Hypergeometric distribution

**pmf** The pmf $h(x; n, M, N)$ of a hypergeometric distribution is non-zero only for integer values $x$ that satisfy $\max(0, n - N + M) \leq x \leq \min(n, M)$

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$$

**R functions**

- `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)`

Note that the R functions define the parameters differently than the text. In R $m$ is the number of Ss, $n$ is the number of F's, and $k$ is the sample size.
Properties and Examples

Expected value and variance

\[ E(X) = n \cdot \frac{M}{N} \quad V(X) = \left( \frac{N - n}{N - 1} \right) \cdot n \cdot \frac{M}{N} \cdot \left( 1 - \frac{M}{N} \right) \]

Example 3.34 Hypergeometric with \( N = 20, M = 12, \) and \( n = 5 \)
> `round(dhyper(0:5, 12, 8, 5), 4)`

> [1] 0.0036 0.0542 0.2384 0.3973 0.2554 0.0511

Example 3.35 Hypergeometric with \( N = 25, M = 5, \) and \( n = 10 \)
> `dhyper(2, 5, 20, 10)`

> [1] 0.3853755

> `sum(dhyper(0:2, 5, 20, 10))`

> [1] 0.6988142

Binomial approximation: When the sampling fraction is small, the hypergeometric can be approximated by the binomial with \( p = \frac{M}{N} \). This is unnecessary in \( R \), which has stable computational methods for the probabilities.
Approximation of hypergeometric by binomial

Example 3.29 500,000 licensed drivers in a state of whom 400,000 are insured. Obtain a sample of size 10. The following table gives the computed hypergeometric and the approximate binomial \((n = 10, p = 0.8)\) probabilities.

<table>
<thead>
<tr>
<th>Hypergeometric</th>
<th>Binomial</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0000001024</td>
</tr>
<tr>
<td>1</td>
<td>0.0000040949</td>
</tr>
<tr>
<td>2</td>
<td>0.0000737138</td>
</tr>
<tr>
<td>3</td>
<td>0.0007863317</td>
</tr>
<tr>
<td>4</td>
<td>0.0055046111</td>
</tr>
<tr>
<td>5</td>
<td>0.0264231903</td>
</tr>
<tr>
<td>6</td>
<td>0.0880797234</td>
</tr>
<tr>
<td>7</td>
<td>0.2013281020</td>
</tr>
<tr>
<td>8</td>
<td>0.3019929079</td>
</tr>
<tr>
<td>9</td>
<td>0.2684354560</td>
</tr>
<tr>
<td>10</td>
<td>0.1073717665</td>
</tr>
</tbody>
</table>

The binomial approximation is quite accurate but unnecessary because the exact hypergeometric probability can be calculated easily.
Definition In an experiment of repeated independent trials where each trial produces “success” ($S$) with probability $p$ or “failure” ($F$) with probability $1 - p$, the distribution of the number of failures that precede the $r$th success is called a negative binomial distribution.

pmf The probability mass function is

$$ nb(x; r, p) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x \quad x = 0, 1, 2, \ldots $$
Negative Binomial (cont’d)

\[ R \text{ functions} \]
\[ \text{dnbinom}(x, \text{size}, \text{prob}, \mu, \log = \text{FALSE}) \]
\[ \text{pnbinom}(q, \text{size}, \text{prob}, \mu, \text{lower.tail} = \text{TRUE}, \log.p = \text{FALSE}) \]
\[ \text{qnbinom}(p, \text{size}, \text{prob}, \mu, \text{lower.tail} = \text{TRUE}, \log.p = \text{FALSE}) \]
\[ \text{rnbinom}(n, \text{size}, \text{prob}, \mu) \]

* \( R \) allows specification of any two of \( r, p \) and \( \mu = \frac{r(1-p)}{p} \)

**Geometric**  
The \( R \) definition of a geometric distribution is a special case of the negative binomial when \( r = 1 \). The text’s definition of the geometric distribution is off by 1 because it counts the last trial (which must be a success).
Poisson distribution

pmf

\[ p(x; \lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, \ldots \quad \lambda > 0 \]

R functions

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

Properties: \( E(X) = V(X) = \lambda \)

Example 3.38 Number of creatures caught in a trap is a Poisson distribution with \( \lambda = 4.5 \).

> dpois(5, 4.5)

[1] 0.1708269

> sum(dpois(0:5, lambda = 4.5))

[1] 0.7029304
Approximating binomials

Formal statement  In the binomial probability mass function if \( n \to \infty \) and \( p \downarrow 0 \) in such a way that \( np \) approaches a finite value \( \lambda \) then

\[
b(x; n, p) \to p(x; \lambda)
\]

Practical application  For \( n \) large and \( p \) small you can approximate the binomial with the Poisson by setting \( \lambda = np \). (This is not necessary in \( \mathbb{R} \).)

Example 3.39  Binomial with \( n = 400 \) and \( p = 0.005 \)

\[
\text{> dbinom(1, 400, 0.005)}
\]

\[
[1] \ 0.2706694
\]

\[
\text{> dpois(1, lambda = 2)}
\]

\[
[1] \ 0.2706706
\]

\[
\text{> sum(dbinom(0:3, 400, 0.005))}
\]

\[
[1] \ 0.8575767
\]

\[
\text{> sum(dpois(0:3, lambda = 2))}
\]

\[
[1] \ 0.8575767
\]
Comparison plot of binomial and Poisson approximation
A *Poisson process* is generated by discrete events that occur over time (or some other continuum such as distance) subject to the conditions:

1. There exists a parameter $\alpha > 0$ such that, for short time intervals of length $\Delta t$, the probability of exactly one event is approximately $\alpha \cdot \Delta t$.

2. The probability of more than one event in very short intervals is approximately zero.

3. The number of events in an interval is independent of the number of events prior to this interval. That is, the process has "no memory". We also say that the process is restartable.

The parameter $\alpha$ is called the *rate* of the process. The number of events in an interval of length $t$ has a Poisson distribution with parameter $\lambda = \alpha t$. 