Week 3: Discrete Distributions

At the end of this week, you should be able to:

1) Distinguish between a *continuous* and *discrete* random variable.
2) Distinguish between a random variable and a *realization* of a random variable.
3) Define a probability mass function for a discrete random variable X.
4) Calculate probabilities using pmfs.
5) Identify situations for which a Bernoulli, binomial, geometric, or Poisson distribution works as a good model.
6) Calculate the probability that a Bernoulli, Binomial, Negative Binomial, Geometric, or Poisson rv takes on particular value or set of values.
7) Define the cumulative distribution function (cdf) for a rv. Calculate the cdf for given values of x.

Two Types of Random Variables

Discrete random variable:

- finite number of values (eg, pass/fail or 1/0)
- countably many values – can be infinitely many, eg {1,2,3,...}

Continuous random variable:

1. Its possible values = real numbers **R**, an interval of **R**, or a disjoint union of intervals from **R** (e.g., [0, 10] ∪ [20, 30])

2. No one single value of the variable has positive probability, that is, P(X = c) = 0 for any possible value c.

Only intervals have postitive prob: for example, P(X in [3,6]) = 0.5)

Examples of random variables

Discrete random variable:

- X = number of heads in 50 consecutive coin flips
- Y = number of times a cell phone goes off during any class

Continuous random variable:

- Z_1 = Length of your commuting time to class
- Z_2 = Baby birth weight

Examples of a realization of random variables

Discrete random variable:

- X = number of heads in 50 consecutive coin flips
  - X = 27 heads in a particular sequence of 50 coin flips
  - We call 27 a particular value (realization) of X
  - Oftentimes, we'll use X = x to denote a generic realization of X
- Y = number of times a cell phone goes off during any class
  - Eg, Y = 3 during today's class
  - Y = y in general

Continuous random variable:

- Z_1 = z = 15.2 min is the length of your commuting time to today's class
- Z_2 = z = 4123g is the birth weight of a baby born at noon today at BCH

Probability distribution of a discrete random variable

1. Probability density (or mass) function of X

2. Describes how probability is distributed among the various possible values of the random variable X

p(X=x), for each value x that X can take

3. Often, p(X=x) is simply written as p(x). Note p(X=x) is
P(all s $\in$ S : X(s) = x).

Example

A lab has 6 computers.

Let *X* denote the number of these computers that are in use during lunch hour -- {0, 1, 2... 6}.

Suppose that the probability mass function of *X* is as given in the following table:

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

Example, cont

cont'd

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

From here, we can find many things:

1) Probability that at most 2 computers are in use:

$P(X \leq 2) = P(X = 0 \text{ or } 1 \text{ or } 2)$

$= p(0) + p(1) + p(2)$

$= .05 + .10 + .15 \quad = .30$

2) Probability that half or more computers are in use:

$1- P(X \leq 2) = 1- 0.30 = 0.70$

3) Probability that there are 3 or 4 computers free:

$P(X = 3) + P(X=4) = 0.45$

The Cumulative Distribution Function

The **cumulative distribution function** (CDF):
F(*x*) of a discrete rv variable *X* with pmf *p*(*x*)
is defined for every real number *x* by

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

For any number *x*, *F*(*x*) is the probability that the observed value of *X* will be at most *x*.

Example

$$p(x) = \begin{cases} .500 & x = 0 \\ .167 & x = 1 \\ .333 & x = 2 \\ 0 & \text{otherwise} \end{cases}$$

$P(X \leq 0) = P(X = 0) = .5$
$P(X \leq 1) = p(0) + p(1) = .500 + .167 = .667$
$P(X \leq 2) = p(0) + p(1) + p(2) = .500 + .167 + .333 = .$

For any x satisfying $0 \leq x < 1$, $P(X \leq x) = .5$.
$P(X \leq 1.5) = P(X \leq 1) = .667$
$P(X \leq 20.5) = 1$

$F(y)$ will equal the value of F at the closest possible value of Y to the left of y.

Notice that $P(X < 1) < P(X \leq 1)$ since the latter includes the probability of the X value 1, whereas the former does not.

More generally, when X is discrete and x is a possible value of the variable, $P(X < x) < P(X \leq x)$.

If X is continuous, $P(X < x) = P(X \leq x)$.

Back to theory: Mean (Expected Value) of X

Let X be a discrete rv with set of possible values D and pmf $p(x)$. The **expected value** or **mean value** of X, denoted by $E(X)$ or μ_X or just μ, is

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

Example

Consider a university having 15,000 students and let X = of courses for which a randomly selected student is registered. The pmf of X is given to you as follows:

x	1	2	3	4	5	6	7
$p(x)$	.01	.03	.13	.25	.39	.17	.02
Number registered	150	450	1950	3750	5850	2550	300

$\mu = 1\ p(1) + 2\ p(2) + \ldots + 7\ p(7)$

$= (1)(.01) + 2(.03) + \ldots + (7)(.02)$

$= .01 + .06 + .39 + 1.00 + 1.95 + 1.02 + .14$

$= 4.57$

The Expected Value of a Function

Sometimes interest will focus on the expected value of some function $h(X)$ rather than on just $E(X)$.

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

$$E[h(X)] = \sum_D h(x) \cdot p(x)$$

That is, $E[h(X)]$ is computed in the same way that $E(X)$ itself is, except that $h(x)$ is substituted in place of x.

Example

A computer store has purchased 3 computers of a certain type at \$500 apiece. It will sell them for \$1000 apiece. The manufacturer has agreed to repurchase any computers still unsold after a specified period at \$200 apiece.

Let X denote the number of computers sold, and suppose that

$$p(0) = .1, \quad p(1) = .2, \quad p(2) = .3 \quad \text{and} \quad p(3) = .4.$$

With $h(X)$ denoting the profit associated with selling X units, the given information implies that

$$h(X) = \text{revenue} - \text{cost} =$$
$$= 1000X + 200(3 - X) - 1500 = 800X - 900$$

The expected profit is then

$$E[h(X)] = h(0)\ p(0) + h(1)\ p(1) + h(2)\ p(2) + h(3)\ p(3)$$

$$= (-900)(.1) + (-100)(.2) + (700)(.3) + (1500)(.4) = \$700$$

Rules of Averages (Expected Values)

The $h(X)$ function of interest is often a linear function $aX + b$. In this case, $E[h(X)]$ is easily computed from $E(X)$.

Proposition

$$E(aX + b) = a\ E(X) + b$$

(Or, using alternative notation, $\mu_{aX+b} = a\ \mu_x + b$)

To paraphrase, the expected value of a linear function equals the linear function evaluated at the expected value $E(X)$.

In the previous example, $h(X)$ is linear – so:

$E(X) = 2$, $E[h(x)] = 800(2) - 900 = \$700$, as before.

The Variance of X

Definition

Let X have pmf $p(x)$ and expected value μ. Then the **variance** of X, denoted by $V(X)$ or σ^2, is

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

The **standard deviation** (SD) of X is

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

Note these are population (theoretical) values, not sample values as before.

Example

Let X denote the number of books checked out to a randomly selected individual (max is 6). The pmf of X is as follows:

x	1	2	3	4	5	6
$p(x)$	.30	.25	.15	.05	.10	.15

The expected value of X is easily seen to be $\mu = 2.85$.

The variance of X is

$$V(X) = \sigma^2 = \sum_{x=1}^{6} (x - 2.85)^2 \cdot p(x)$$

$$= (1 - 2.85)^2(.30) + (2 - 2.85)^2(.25) + \ldots + (6 - 2.85)^2(.15) = 3.2275$$

The standard deviation of X is $\sigma = \sqrt{3.2275} = 1.800$.

A Shortcut Formula for σ^2

The number of arithmetic operations necessary to compute σ^2 can be reduced by using an alternative formula.

$V(X) = \sigma^2 = E(X^2) - [E(X)]^2$

In using this formula, $E(X^2)$ is computed first without any subtraction; then $E(X)$ is computed, squared, and subtracted (once) from $E(X^2)$.

Rules of Variance

The variance of $h(X)$ is the expected value of the squared difference between $h(X)$ and its expected value:

$$V[h(X)] = \sigma^2_{h(X)} = \sum_D \{h(x) - E[h(X)]\}^2 \cdot p(x)$$

When $h(X) = aX + b$, a linear function,

$$h(x) - E[h(X)] = ax + b - (a\mu + b) = a(x - \mu)$$

then

Rules of Variance

$V(aX + b) = \sigma^2_{aX+b} = a^2 \; \sigma^2_{X \; a}$

$\sigma_{aX+b} = |a| \cdot \sigma_X$

The absolute value is necessary because a might be negative, yet a standard deviation cannot be.

Usually multiplication by "*a*" corresponds to a change of scale, or of measurement units (e.g., kg to lb or dollars to euros).

Families of random variables

Discrete random variables can be categorized into different distribution families (Bernoulli, Geometric, Poisson…).

Each family corresponds to a model for many different real-world situations.

Each family has many members

Each specific member has its own particular set of parameters.

Bernoulli random variable

Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable.**

This distribution is specified with a single parameter:

$$\pi_1 = p(X=1)$$

Which corresponds to the proportion of 1's.

From here, p(X=0) = 1- p(X=1)

PMF shorthand: $P(X=x) = \pi_1^{\,x}(1-\pi_1)^{(1-x)}$

Example: fair coin-tossing $\pi_1 = 0.5$

Binomial experiments

Binomial experiments conform to the following:

1. The experiment consists of a sequence of *n* identical and independent Bernoulli experiments called *trials,* where *n* is fixed in advance:
2. Each trial outcome is a Bernoulli variable – ie, each trial can result in only one of 2 possible outcomes. We generically denote one oucome by "success" (*S,* or 1) and "failure" (*F,* or 0).
3. The probability of success *P*(*S*) (or *P(1)*) is identical across trials; we denote this probability by *p.*
4. The trials are independent, so that the outcome on any particular trial does not influence the outcome on any other trial.

Binomial random variable

Binomial random variable counts the **<u>total number of *1's*</u>**:

Definition

The **binomial random variable *X*** associated with a binomial experiment consisting of *n* trials is defined as

X = the number of *1*'s among the *n* trials

This is an identical definition as X = sum of *n* independent and identically distributed Bernoulli random variables

X ~ Bin(n,p)

Suppose, for example, that *n* = 3. Then the sample space elements are: *SSS SSF SFS SFF FSS FSF FFS FFF*

From the definition of *X, which simply counts the number of S for each member of the sample space, X*(*SSF*) = 2, *X*(*SFF*) = 1, and so on.

Possible values for *X* in an *n*-trial experiment are *x* = 0, 1, 2, . . . , *n*.

We will often write *X* ~ Bin(*n*, *p*) to indicate that *X* is a binomial rv based on *n* Bernoulli trials with success probability *p.*

For *n* = 1, the binomial r.v. reverts to the Bernoulli r.v.

Example – Binomial r.v.

A coin is tossed 6 times.

From the knowledge about fair coin-tossing probabilities,

$p = P(H) = P(S) = 0.5.$

Thus, if X = the number of heads among six tosses, then

$X \sim \text{Bin}(6, 0.5).$

Then, $P(X = 3) = \binom{6}{3} (.5)^3(.5)^3 = 20(.5)^6 = .313$

In general, $P(X = x) = (\text{ n choose x }) \, p^x (1-p)^{(n-x)}$

Example

cont'd

The probability that at least three come up heads is

$$P(3 \le X) = \sum_{x=3}^{6} \binom{6}{x} (.5)^x(.5)^{6-x}$$

$$= .656$$

and the probability that at most one come up heads is

$$P(X \le 1) = .109$$

Mean and Variance of a Binomial R.V.

The mean value of a Bernoulli variable is $\mu = p$
(= 0 x (1-p) + 1 x p)

So, the expected number of S's on any single trial is p.

Since a binomial experiment consists of n trials, intuition suggests that for $X \sim \text{Bin}(n, p)$ we have

- $E(X) = np$
  the product of the number of trials and the probability of success on a single trial.

Mean and Variance of Binomial r.v.

If $X \sim \text{Bin}(n, p)$, then

$E(X) = np$,

$V(X) = np(1 - p) = npq$, and

$\sigma_X = \sqrt{npq}$

(where $q = 1 - p$).

Example

A biased coin is tossed 10 times, so that the odds of "heads" are 3:1. Then, the number of heads follows

$$X \sim \text{Bin}(10, .75)$$

Then, $E(X) = np = (10)(.75) = 7.5$,

$V(X) = npq = 10(.75)(.25) = 1.875$,

and $\sigma = \sqrt{1.875}$

$= 1.37$.

Example, cont.

cont'd

Again, even though X can take on only integer values, $E(X)$ need not be an integer.

If we perform a large number of independent binomial experiments, each with $n = 10$ trials and $p = .75$, then the average number of *1*'s per experiment will be close to 7.5.

The probability that X is within 1 standard deviation of its mean value is

$$P(7.5 - 1.37 \le X \le 7.5 + 1.37) = P(6.13 \le X \le 8.87)$$

$$= P(X = 7 \text{ or } 8)$$

$$= .532.$$

Sidenote: simulating Bernoulli variables in R

R function for simulating binomial random variable realizations is:

`rbinom(n, size, prob)`

<u>Where</u>:

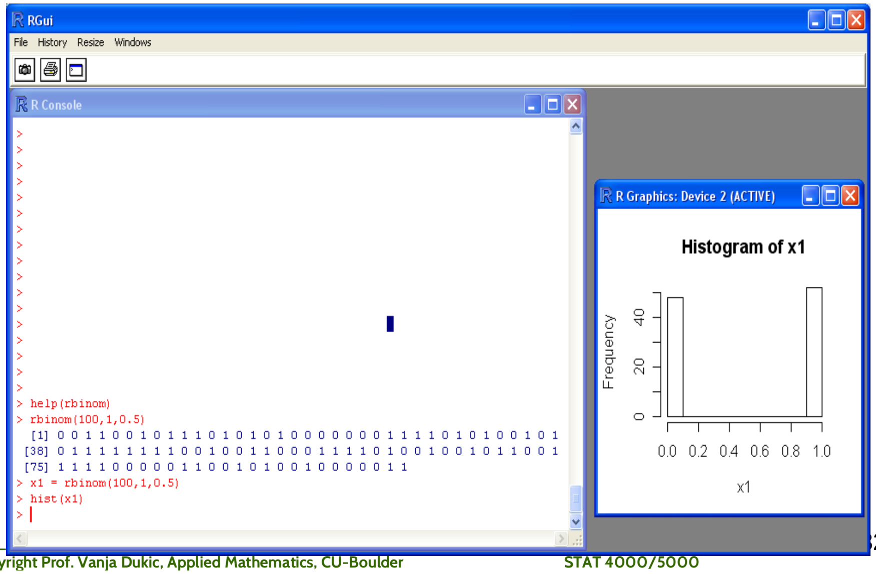
`n` is the number of simulations,

`size` is the number of Bernoulli trials (1 or more)

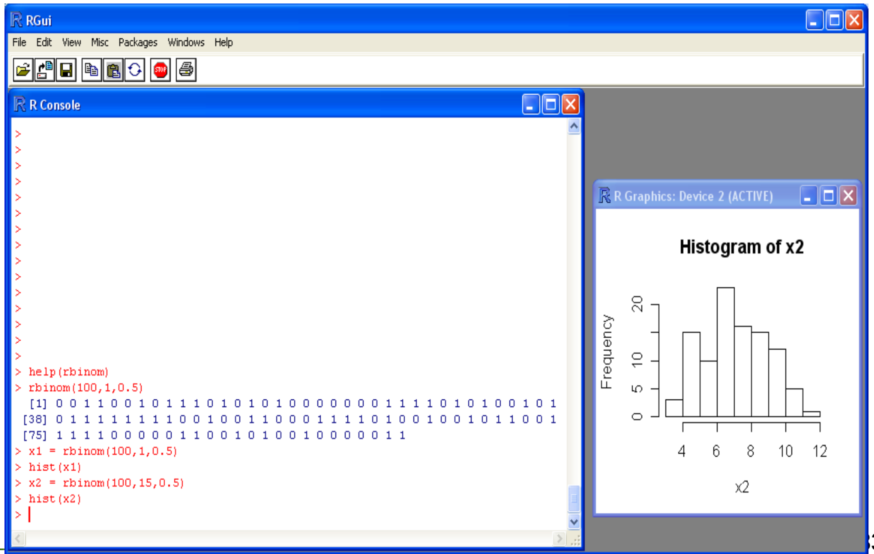
`prob` is the probability of success on each trial.

`rbinom(n, 1, prob)` generates `n` Bernoulli random variable realizations.

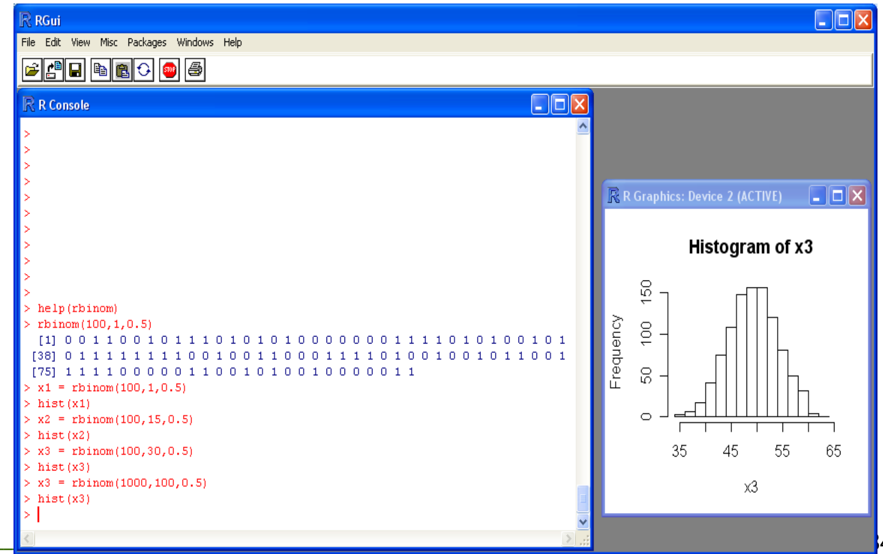
Sidenote: simulating Bernoulli and Binomial variables in R



Sidenote: simulating Bernoulli and Binomial variables in R



Sidenote: simulating Bernoulli and Binomial variables in R



Geometric random variable -- Example

Starting at a fixed time, we observe the gender of each newborn child at a certain hospital until a boy (B) is born.

Let $p = P(B)$, assume that successive births are independent, and let X be the number of births observed.

Then

$p(1) = P(X = 1)$

$= P(B)$

$= p$

Example, cont.

cont'd

$p(2) = P(X = 2)$

$= P(GB)$

$= P(G)\ \ P(B)$

$= (1 - p)\, p$

and

$p(3) = P(X = 3)$

$= P(GGB)$

$= P(G)\ \ P(G)\ \ P(B)$

$= (1 - p)^2 p$

Example, cont.

cont'd

Continuing in this way, a general formula emerges:

$$p(x) = \begin{cases} (1-p)^{x-1}p & x = 1, 2, 3, \ldots \\ 0 & \text{otherwise} \end{cases}$$

The parameter p can assume any value between 0 and 1.

Depending on what parameter p is, we get different members of the *geometric* distribution.

Sidenote: simulating Geometric variables in R

R function for simulating geometric random variables is:

```
X = rgeom(n, prob)
```

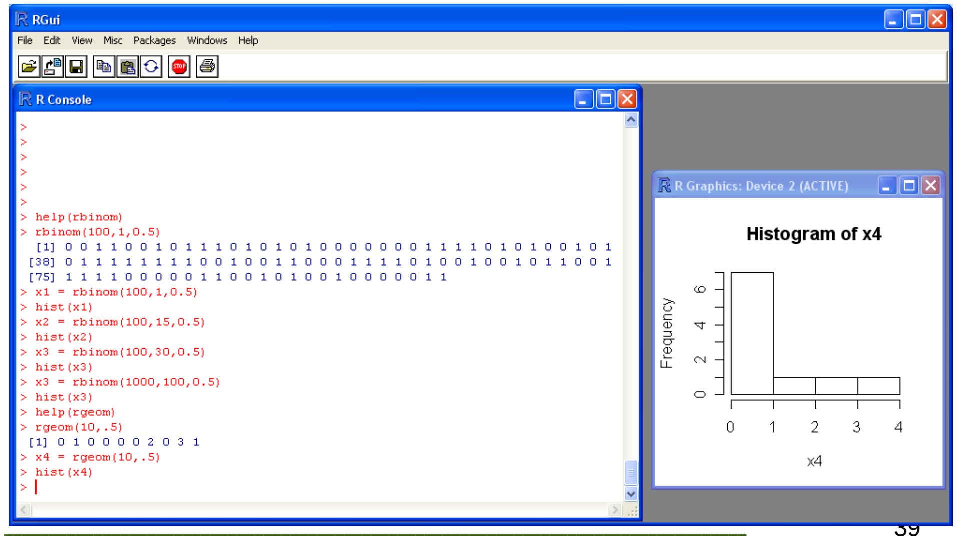
```
NOTE: In R, X represents the number of
failures in a sequence of Bernoulli trials
before a success occurs.
```

Where:

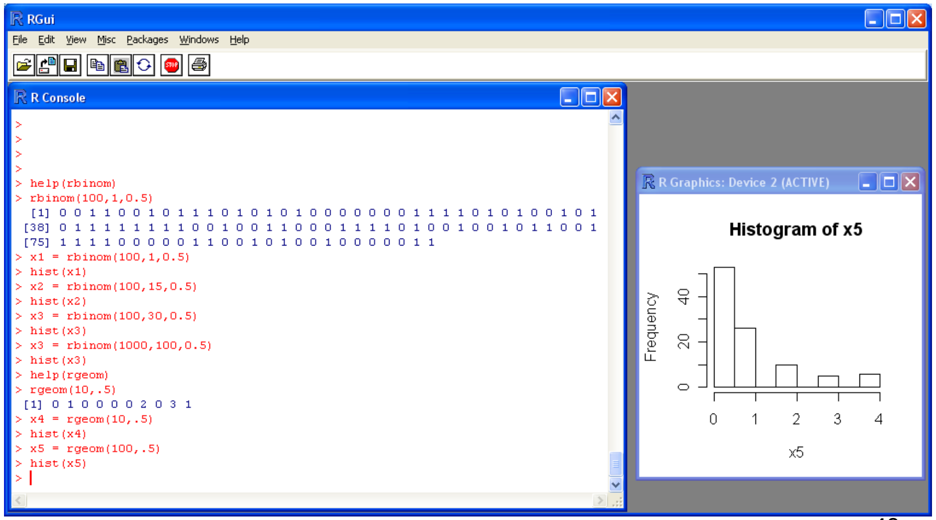
`n` is the number of simulations,

`prob` is the probability of success on each trial.

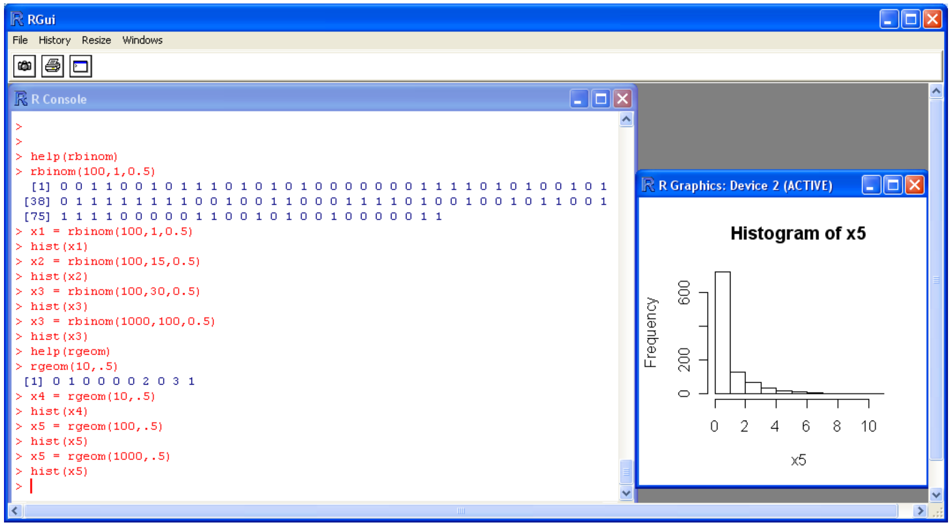
Sidenote: simulating Geometric variables in R



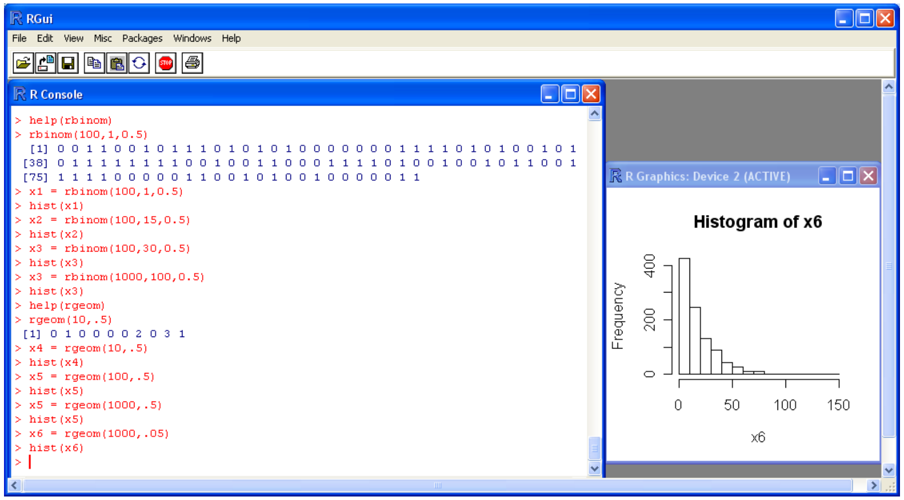
Sidenote: simulating Geometric variables in R



Sidenote: simulating Geometric variables in R



Sidenote: simulating Geometric variables in R



The Negative Binomial Distribution

The Negative Binomial Distribution

1. The experiment is a sequence of independent trials where each trial can result in a success (*S*) or a failure (*F*)

3. The probability of success is constant from trial to trial

4. The experiment continues (trials are performed) until a total of *r* successes have been observed

5. The random variable of interest is

   X = the number of failures that precede the *r*th success

6. In contrast to the binomial rv, the number of successes is fixed and the number of trials is random.

The Negative Binomial Distribution

Possible values of X are 0, 1, 2, ...

Let $nb(x; r, p)$ denote the pmf of X. Consider

$$nb(7; 3, p) = P(X = 7)$$

the probability that exactly 7 F's occur before the 3rd S.

In order for this to happen, the 10th trial must be an S and there must be exactly 2 S's among the first 9 trials. Thus

$$nb(7; 3, p) = \left\{ \binom{9}{2} \cdot p^2(1-p)^7 \right\} \cdot p = \binom{9}{2} \cdot p^3(1-p)^7$$

Generalizing this line of reasoning gives the following formula for the negative binomial pmf.

The Negative Binomial Distribution

The pmf of the negative binomial rv X with parameters r = number of S's and $p = P(S)$ is

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

Then,

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

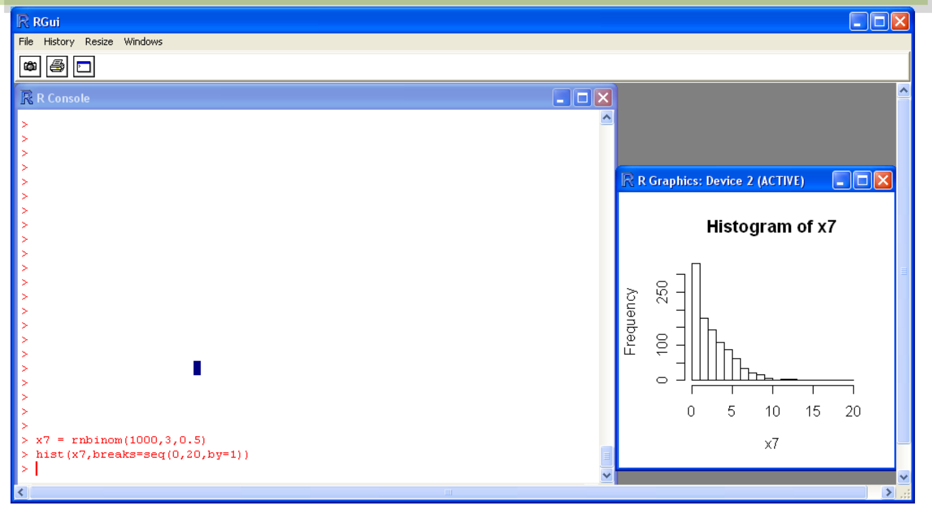
Simulating negative binomial random variables in R

```
help(rbinom)
```

```
rnbinom(n, size, prob)
```

```
Where
n = number of simulations
size = number of successful trials desired
prob = probability of success in each trial
```


Simulating negative binomial random variables in R



The Hypergeometric Distribution

The Hypergeometric Distribution

1. The population consists of *N* elements (a *finite* population)

2. Each element can be characterized as a success (*S*) or failure (*F*)

3. There are *M* successes in the population, and *N-M* failures

4. A sample of *n* elements is selected without replacement, in such a way that each sample of *n* elements is equally likely to be selected

The random variable of interest is

X = the number of S's in the sample of size n

Example

Last week the IT office received **20 service orders** for problems with printers: **8 were laser printers and 12 were inkjets**

A **sample of 5** of these orders is to be sent out for a customer satisfaction survey.

What is the probability that exactly *x* (where *x* can be any of these numbers: 0, 1, 2, 3, 4, or 5) of the 5 selected service orders were for inkjet printers?

Example

cont'd

- Here, the population size is $N = 20$,
- the sample size is $n = 5$
- the number of S's (inkjet = S) is 12
- The number of F's is 8

Consider the value $x = 2$. Because all outcomes (each consisting of 5 particular orders) are equally likely,

$$P(X = 2) = h(2; 5. 12. 20) = \frac{\text{number of outcomes having } X = 2}{\text{number of possible outcomes}}$$

$$h(2; 5, 12, 20) = \frac{\binom{12}{2}\binom{8}{3}}{\binom{20}{5}}$$

The Hypergeometric Distribution

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 the probability distribution of X, called the **hypergeometric distribution,** is given by

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

for x, an integer, satisfying
$\max(0, n - N + M) \le x \le \min(n, M)$.

The Hypergeometric Distribution

Proposition

For hypergeometric rv X having pmf $h(x; n, M, N)$:

$$E(X) = n \cdot \frac{M}{N} \qquad V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

The ratio M/N is the proportion of S's in the population. If we replace M/N by p in $E(X)$ and $V(X)$, we get

$$E(X) = np$$

$$V(X) = \left(\frac{N-n}{N-1}\right) \cdot np(1 - p)$$

Example

Five individuals from an animal population thought to be near extinction in a certain region have been caught, tagged, and released to mix into the population.

After they have had an opportunity to mix, a random sample of 10 of these animals is selected. Let x = the number of tagged animals in the second sample.

If there are actually 25 animals of this type in the region, what is the $E(X)$ and $V(X)$?

Example

cont'd

In the animal-tagging example,

$n = 10$, $M = 5$, and $N = 25$, so $p = \frac{5}{25} = .2$

and $$E(X) = 10(.2) = 2$$

$$V(X) = \frac{15}{24}(10)(.2)(.8) = (.625)(1.6) = 1$$

Example

cont'd

Suppose the population size *N* is not actually known, so the value *x* is observed and we wish to estimate *N*.

It is reasonable to equate the observed sample proportion of *S*'s, *x*/*n*, with the population proportion, *M*/*N*, giving the estimate

$$\hat{N} = \frac{M \cdot n}{x}$$

If *M* = 5, *n* = 10, and *x* = 2, then

$\hat{N}$ = 25.

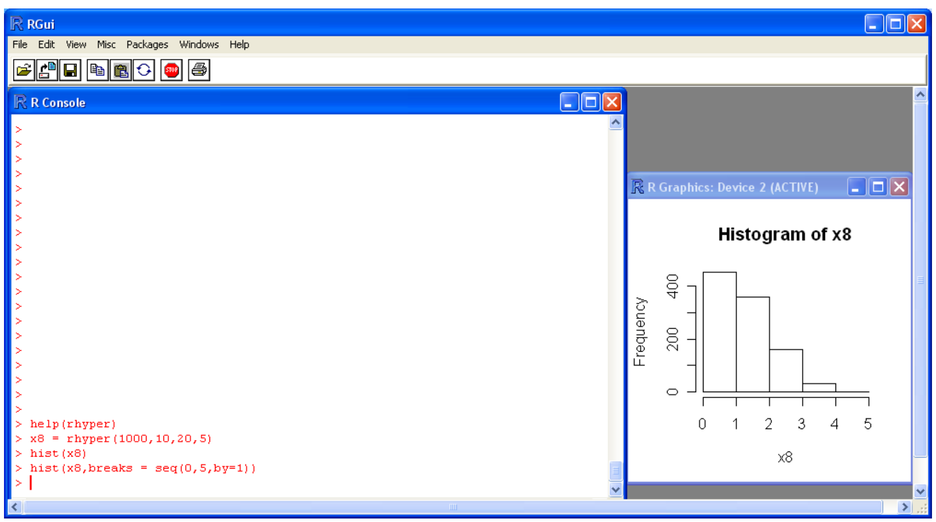
Hypergeometric in R

```
rhyper(nn, m, n, k)
```

Where

```
nn  -- number of simulations
m   -- number of successes in the population
n   -- number of failures in the population
k   -- size of the sample
```


Hypergeometric in R



The Poisson Distribution

The Poisson Probability Distribution

Poisson r.v. describes the total number of events that happen in a certain time period.

Eg:

- arrival of vehicles at a parking lot in one week
- number of gamma rays hitting a satellite per hour
- number of neurons firing per minute

A discrete random variable *X* is said to have a **Poisson distribution** with parameter μ ($\mu > 0$) if the pmf of *X* is

$$p(x;\mu) = \frac{e^{-\mu} \cdot \mu^x}{x!} \qquad x = 0, 1, 2, 3, \ldots$$

The Poisson Probability Distribution

It is no accident that we are using the symbol μ for the Poisson parameter; we shall see shortly that μ is in fact the expected value of *X*.

The letter *e* in the pmf represents the base of the natural logarithm; its numerical value is approximately 2.71828.

The Poisson Probability Distribution

It is not obvious by inspection that $p(x;\mu)$ specifies a legitimate pmf, let alone that this distribution is useful.

First of all, $p(x;\mu) > 0$ for every possible *x* value because of the requirement that $\mu > 0$.

The fact that $\Sigma\, p(x;\mu) = 1$ is a consequence of the Maclaurin series expansion of e^{μ} (check your calculus book for this result):

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\mu^x}{x!} \qquad \textbf{(3.18)}$$

The Mean and Variance of Poisson

Proposition

If *X* has a Poisson distribution with parameter μ, then $E(X) = V(X) = \mu$.

These results can be derived directly from the definitions of mean and variance.

Example

Let X denote the number of mosquitoes captured in a trap during a given time period.

Suppose that X has a Poisson distribution with μ = 4.5, so on average traps will contain 4.5 mosquitoes.

The probability that a trap contains exactly five mosquitoes is

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = .1708$$

Example

cont'd

The probability that a trap has at most five is

$$P(X \leq 5) = \sum_{x=0}^{5} \frac{e^{-4.5}(4.5)^x}{x!}$$
$$= e^{-4.5}\left[1 + 4.5 + \frac{(4.5)^2}{2!} + \cdots + \frac{(4.5)^5}{5!}\right]$$
$$= .7029$$

Example

Example continued…

Both the expected number of mosquitos trapped and the variance of the number trapped equal 4.5, and

$$\sigma_X = \sqrt{\mu}$$
$$= \sqrt{4.5}$$
$$= 2.12.$$

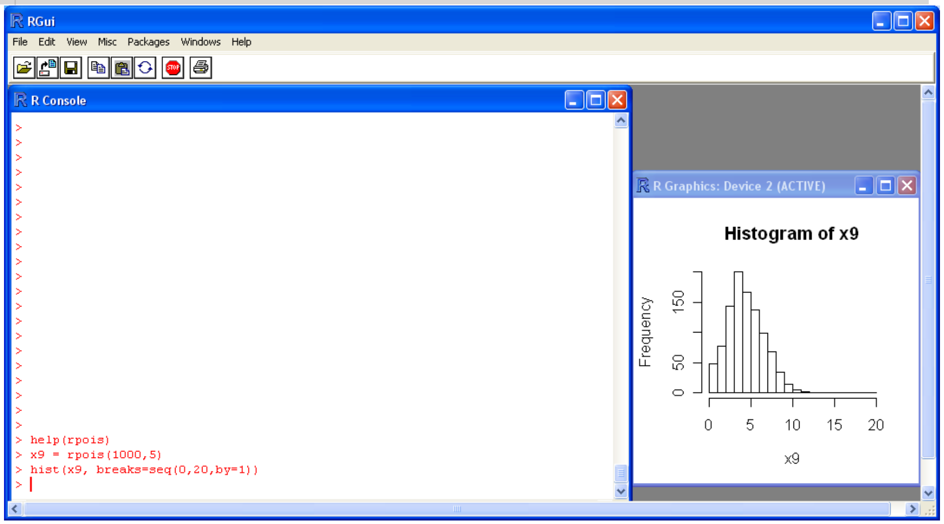
Poisson in R

```
rpois(n,lambda)

Where

n      -- the number of simulations
lambda  -- the mean number
```

Poisson in R



The Poisson Distribution as a Limit

The rationale for using the Poisson distribution in many situations is provided by the following proposition.

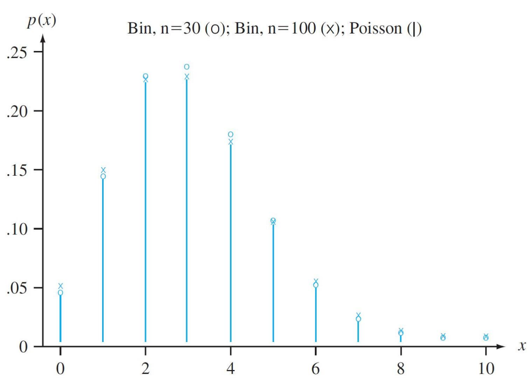
Proposition

Suppose that in the binomial pmf $b(x; n, p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then $b(x; n, p) \to p(x; \mu)$.

According to this proposition, *in any binomial experiment in which n is large and p is small*, $b(x; n, p) \approx p(x; \mu)$, *where* $\mu = np$. As a rule of thumb, this approximation can safely be applied if $n > 50$ and $np < 5$.

The Poisson Distribution as a Limit

The approximation is of limited use for $n = 30$, but the accuracy is better for $n = 100$ and much better for $n = 300$.



Comparing a Poisson and two binomial distributions

Example

A publisher takes great pains to ensure that its books are free of typographical errors: the probability of any given page containing at least 1 such error is .005.

If the errors are independent from page to page, what is the probability that one of the 400-page novels will contain exactly one page with errors? At most three pages with errors?

With S denoting a page containing at least one error and F an error-free page, the number X of pages containing at least one error is a binomial rv with $n = 400$ and $p = .005$, so $np = 2$.

Example

cont'd

We need to find out

$P(X = 1) = b(1; 400, .005) \approx p(1; 2) = \frac{e^{-2}(2)^1}{1!} = .270671$

The binomial value is $b(1; 400, .005) = .270669$, so the approximation is very good.

Similarly,

$$P(X \leq 3) \approx \sum_{x=0}^{3} p(x, 2) = \sum_{x=0}^{3} e^{-2} \frac{2^x}{x!}$$

The Poisson Process

The Poisson Process

A very important application of the Poisson distribution arises in connection with the occurrence of events of some type over time.

Events of interest might be visits to a particular website, pulses of some sort recorded by a counter, email messages sent to a particular address, accidents in an industrial facility, or cosmic ray showers observed by astronomers at a particular observatory.

Example

Suppose photons arrive at a plate at an average rate of six per minute, ie. $\alpha = 6$.

To find the probability that in a 0.5-min interval at least one photon is received, note that the number of photons in such an interval has a Poisson distribution with parameter $\alpha t = 6(0.5) = 3$ (0.5 min is used because α is expressed as a rate per minute).

Then with X = the number of pulses received in the 30-sec interval,

$$P(1 \leq X) = 1 - P(X = 0) = 1 - \frac{e^{-3}(3)^0}{0!} = .950$$

The Poisson Process

$P_k(t) = e^{-\alpha t}\ (\alpha t)^k/k!$ so that the number of events during a time interval of length t is a Poisson rv with parameter $\mu = \alpha t$.

The expected number of events during any such time interval is then αt, so the expected number during a unit interval of time is α.

The occurrence of events over time as described is called a *Poisson process*; the parameter α specifies the *rate* for the process.