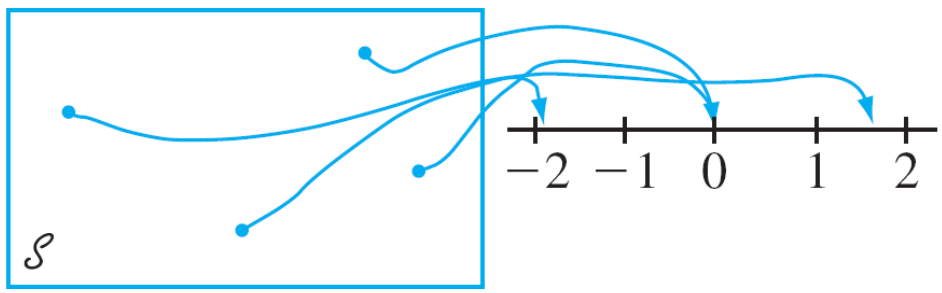
# 3 Discrete Random Variables and Probability Distributions

**Stat 4570/5570**

**Based on Devore's book (Ed 8)**

# Random Variables

We can associate each single outcome of an experiment with a real number:




We refer to the outcomes of such experiments as a **"random variable".**

**Why is it called a "random variable"?**

# Random Variables

**Definition**

For a given sample space ***S*** of some experiment, a **random variable (r.v.)** is a rule that associates a number with each outcome in the sample space ***S***.

In mathematical language, a random variable is a "function" whose domain is the sample space and whose range is the set of real numbers:

$$X : S \to \mathbb{R}$$

So, for any event *s*, we have *X(s)*=*x* is a real number.

# Random Variables

## Notation!

1. Random variables - usually denoted by uppercase letters near the end of our alphabet (e.g. $X$, $Y$).
2. Particular value - now use lowercase letters, such as $x$, which correspond to the r.v. $X$.

**Birth weight example**

# Two Types of Random Variables

A **discrete** random variable:

Values constitute a finite or countably infinite set

A **continuous** random variable:

1. Its set of possible values is the set of real numbers **R**, one interval, or a disjoint union of intervals on the real line (e.g., $[0, 10] \cup [20, 30]$).
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 positive probabilities.

## Probability Distributions for Discrete Random Variables

Probabilities assigned to various outcomes in the sample space ***S***, in turn, determine probabilities associated with the values of any particular random variable defined on ***S****.*

The *probability mass function (pmf) of X , p(X)* describes how the total probability is distributed among all the possible range values of the r.v. *X*:

*p(X=x),* for each value *x* in the range of *X*

Often, *p(X=x)* is simply written as *p(x)* and by definition

$$p(X = x) = P(\{s \in \mathcal{S} | X(s) = x\}) = P(X^{-1}(x))$$

Note that the domain and range of *p(x)* are **real numbers**.

# 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 distribution 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

From here, we can find many things:

1. Probability that at most 2 computers are in use
2. Probability that at least half of the computers are in use
3. Probability that there are 3 or 4 computers free

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

# Bernoulli r.v.

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

This is a discrete random variable – values?

This distribution is specified with a single parameter:

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

**Examples?**

# Geometric r.v. -- 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 until a first boy is born.

Then

$$p(1) = P(X = 1) = \mathrm{P(B)} = \mathrm{p}$$

And,

$$p(2)=?,\ p(3) = ?$$

# The Geometric r.v.

cont'd

Continuing in this way, a general formula for the pmf emerges:

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

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.

# The Cumulative Distribution Function

**Definition**

The **cumulative distribution function** (*cdf*) denoted *F*(*x*) of a discrete *r.v. X* with *pmf p*(*x*) is defined for every real number *x* by

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

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

# Example

Suppose we are given the following pmf:

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

Then, calculate:
F(0), F(1), F(2)

What about F(1.5)? F(20.5)?

Is P(X < 1) = P(X <= 1)?

# The Binomial Probability Distribution

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 r.v., *i.e.*, each trial can result in only one of 2 possible outcomes. We generically denote one outcome 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.

# The Binomial Random Variable and Distribution

The Binomial r.v. counts the **total number of successes**:

**Definition**

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

$X$ = the number of $S$'s among the $n$ trials

This is an identical definition as X = sum of $n$ independent and identically distributed Bernoulli random variables, where $S$ is coded as 1, and $F$ as 0.

# The Binomial Random Variable and Distribution

Suppose, for example, that $n$ = 3. What is the sample space?

Using the definition of $X$, $X(SSF)$ = ? $X(SFF)$ = ? What are the possible values for $X$ if there are $n$ trials?

NOTATION: We write $X \sim \text{Bin}(n, p)$ to indicate that $X$ is a binomial rv based on $n$ Bernoulli trials with success probability $p$.

What distribution do we have if n = 1?

# Example – Binomial r.v.

A coin is tossed 6 times.

From the knowledge about fair coin-tossing probabilities,

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

How do we express that *X* is a binomial r.v. in mathematical notation?

What is P(X = 3)? *P*(X >= 3)? P(X <= 5)?

Can we derive the binomial distribution?

# GEOMETRIC AND BINOMIAL RANDOM VARIABLES IN R.

# Back to theory: Mean (Expected Value) of *X*

Let *X* be a discrete r.v. with set of possible values *D* and pmf *p*(*x*). The **expected value** or mean value of *X*, denoted by *E*(*X*) or $\mu_X$ or just $\mu$, is

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

Note that if *p(x)=1/N* where *N* is the size of *D* then we get the arithmetic average.

# 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 |

Calculate $\mu$

# The Expected Value of a Function

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

**Proposition**
If the r.v. $X$ 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

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

What is the expected profit?

# 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**

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

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

How can this be applied to the previous example?

# The Variance of *X*

**Definition**

Let *X* have pmf *p*(*x*) and expected value $\mu$. Then the **variance** of *X*, denoted by *V*(*X*) or $\sigma^2_X$, or just $\sigma^2$, is

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

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

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

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 $\mu = 2.85$. What is Var(X)? Sd(X)?

# A Shortcut Formula for $\sigma^2$

The variance can also be calcualted using an alternative formula:

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

**Why would we use this equation instead?**

Can we show that the two equations for variance are equal?

# Rules of Variance

The variance of *h*(*X*) is calculated similarly:

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

**Proposition**

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

Why is the absolute value necessary? Examples of when this equation is useful?

Can we do a simple proof to show this is true?

# The Mean and Variance of a Binomial R.V.

The mean value of a Bernoulli variable is $\mu = p$.

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 $X \sim \text{Bin}(n, p)$, $E(X) = np$, the product of the number of trials and the probability of success on a single trial.

The expression for $V(X)$ is not so intuitive.

# Mean and Variance of *Binomial r.v.*

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

*Expectation:* $E(X) = np$ (let's prove this one)

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

*Standard Deviation:* $\sigma_X = \sqrt{npq}$ (where $q = 1 - p$)

# Example

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

What notation do we use to describe X?

What is the mean of X?  The variance?

# Example, cont.



NOTE: 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 $S$'s per experiment will be close to 7.5.

What is the probability that $X$ is within 1 standard deviation of its mean value?

# 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 (so the # of trials is not fixed)

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,

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

# 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 $n$**

# 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
$\max(0, n - N + M) \leq x \leq \min(n, M)$.

# Example

During a particular period a university's information technology office received **20 service orders** for problems with printers, of which **8 were laser printers and 12 were inkjet models**.

A **sample of 5** of these service orders is to be selected for inclusion in a customer satisfaction survey.

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

# The Hypergeometric Distribution

**Proposition**

The mean and variance of the hypergeometric rv *X* having pmf *h*(*x*; *n, M, N*) are

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

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 of a certain type of fox 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 foxes are selected. Let $x$ = the number of tagged foxes in the second sample.

If there are actually 25 foxes in the region, what is the $E(X)$ and $V(X)$?

# The Poisson Probability Distribution

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

Eg:

- # of vehicles arriving at a parking lot in one week
- # of gamma rays hitting a satellite per hour
- # of neurons firing per minute
- # of cookies chips in a length of cookie dough

A discrete random variable *X* is said to have a **Poisson distribution** with parameter $\mu$ ($\mu > 0$) 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 $\mu$ for the Poisson parameter; we shall see shortly that $\mu$ 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 $e^{\mu}$ (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 $\mu$, 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 $\mu = 4.5$, so on average traps will contain 4.5 mosquitoes.

What is the probability that the trap contains 5 mosquitoes? What is the probability that the trap has at most 5 mosquitoes? What is the standard deviation of the number of trapped mosquitoes?

# POISSON IN R

# WORKING WITH DATA FRAMES IN R