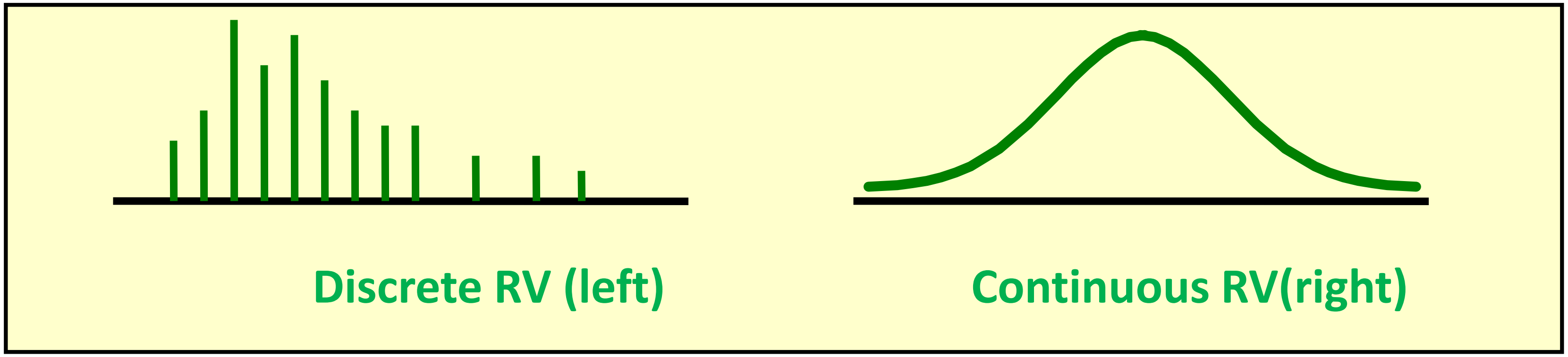
Stat 401
Fall 2017

Shyamala Nagaraj
Vijay Nair

Slides #2

Topics

- Characterizing Distributions: **Discrete RVs**
    - Probability Mass Function
    - Examples
    - Cumulative Distribution Functions
    - Means:
    - Variances
- Characterizing Distributions: **Continuous RVs**
    - Probability Density Functions
    - Cumulative Distribution Functions
    - Means and Variances
    - Medians and Quantiles



Random Variable

- Many different descriptions
- Informally: The “outcome” of some “random process”
- Can take on different values

  → sample space: set of all possible values

- Randomness → variable takes on different values with different probabilities

- There’s a correspondence between a RV and its probability distribution

Example – Binary Random Variable

- Flip a coin: possible values: head or tail
- Let X = **outcome** of the coin toss → random variable
- **Sample space** or all possible values – {H, T}

- If fair coin, **probabilities** are:
  - P(X = H) = ½
  - P(X = T) = ½
- Not fair coin → can write probabilities in general as
  - P(X = H) = p
  - P(X = T) = 1– p
  - p is a **parameter,** some constant, $0 < p < 1$

- X is a binary random variable – takes on two values → denote as {0, 1}

- Generic framework for general applications
  - Competition, election, exam, weather, manufacturing process, …
  - {win, lose}, {pass, fail}, {rain, no rain}, …, {defective, not defective}
  - Set up: Outcome → {0, 1}

Random Variables – Other Examples

For each example, what are:

possible values (discrete/ continuous; range)?

population?

source of randomness?

- X = GPA of randomly selected UM student
- N = number of typos in a randomly selected page of a given manuscript
- K = make (Ford, Toyota, …) of a randomly selected SUV on State Street on a Monday morning
- J = {poor, …, excellent} – ratings of a hotel by customers who've stayed there
- M = number of games that will be won by UM women's basketball team this year
- P = blood pressure measurement of a randomly selected patient during office visit
- D = diameter of a randomly selected piston from a manufacturing plant

Types of Random Variables

- Qualitative or Categorical
  - Nominal
  - Ordinal
- Quantitative
  - Discrete
  - Continuous

- Nominal → names, labels, etc.
  - Examples: race, gender, make of car, county of origin, disease type, …
  - Typically, they take on a finite set of labels
  - No ordering in the categories

- Ordinal → ranking, preference, etc. (poor, …, excellent)
  - Categorical data with ordering

- Quantitative → takes on numerical values – continuous or discrete
  - Continuous – can take on all possible numerical values in an interval (in principle)
  - Discrete – takes on finitely many values or infinitely many “separate” values
    (eg. number of defects in a circuit board: 0, 1, 2, …)

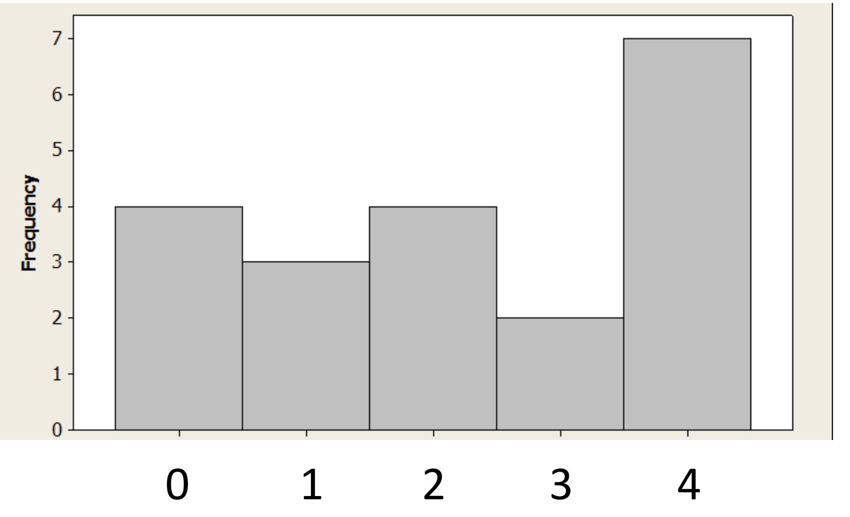
Empirical probabilities

- Sample of 20 active top-ranked women professional tennis players :
  - Data: y = number of tennis tournaments won
  - (hypothetical – assume no more than 4)
  - 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4, 4, 4, 4
  - Summarize the data?

(Relative) Frequency Distribution

x	Frequency	Proportion
0	4	0.20
1	3	0.15
2	4	0.20
3	2	0.10
4	7	0.35
Total	20	1.00

Histogram



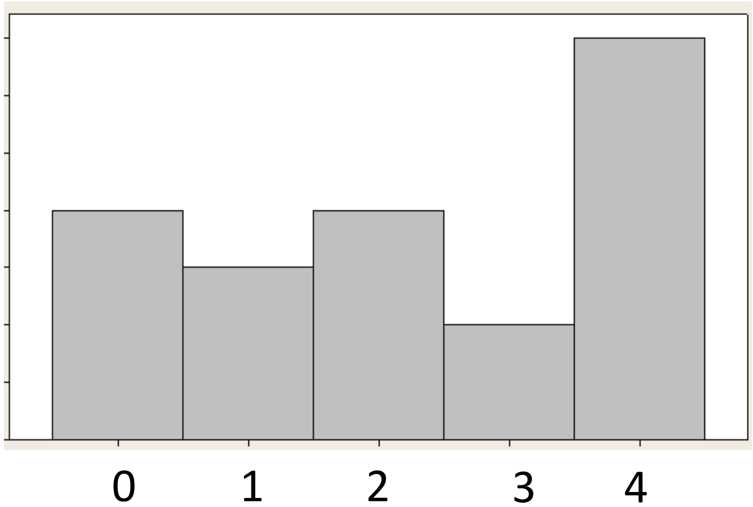
Use estimated proportions as estimates of true probabilities

Population (Discrete Data)

- Population of all top-ranked women players in the professional tennis circuit
- Variable of interest: Number of tennis tournaments won
- Hypothetical: assume no more than 4 wins
- Distribution →

Probability mass function, pmf

x	Probability
0	0.20
1	0.15
2	0.20
3	0.10
4	0.35
Total	1.00



Select a professional tennis player randomly from the population

Let M = number of tennis tournaments won by the player

M is a random variable with possible values (in this case): 0, 1, 2, 3, 4

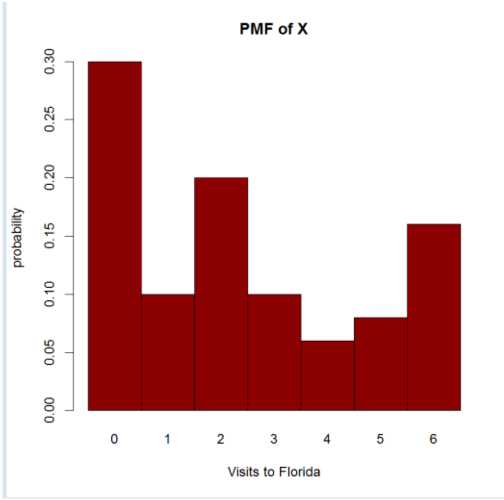
What is the distribution of M? or P(M = k), k = 0, 1, 2, 3, 4 ← **pmf**

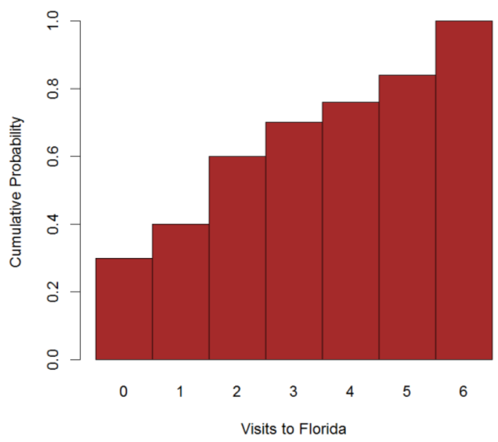
Correspondence between random variables and distributions

Visualizing Probability Distributions: CDF

Suppose we are interested in the number of times a Michigan resident has visited Florida, denoted X, as described by the following distribution:

x	0	1	2	3	4	5	6
$P(X = x)$	0.3	0.1	0.2	0.1	0.06	0.08	0.16
$P(X \leq x)$	0.3	0.4	0.6	0.7	0.76	0.84	1





Discrete RVs and Probability Distributions

- Let X be a RV with K possible values: $x_1, x_2, \ldots x_K$ ← notation for values in distribution

- Its distribution can be characterized by:
  - Probability mass function (pmf):
    - $P(X = x_j), j = 1, \ldots K$

      [What does the notation $P(X = x_j), j = 1, \ldots K$ mean?]
    - Denote $P(X = x_j)$ as $p(x_j)$
  - Cumulative distribution function (cdf):
    - $P(X \leq x_j) = P(X = x_1) + \ldots + P(X = x_j), j = 1, \ldots K$
    - Denote $P(X \leq x_j)$ as $F(x_j)$ – picture?

      - $p(x_j)$ is non-negative
      - $F(x_j)$ starts at zero and increases up to its maximum value of 1.

Selected examples of discrete distributions

- Bernoulli: Outcome: {S, F} → {1, 0}: p(1) = p, p(0) = 1 – p
- Binomial: X = Number of "successes" in n independent trials and probability of success is p (parameter n, p)
  - Sample space: {0, 1, …, n} and p(k) = $\binom{n}{k} p^k (1-p)^{n-k}$
- Geometric: K = number of independent "trials" until first "success" and probability of success is p (parameter p)
  - Sample space: {1, 2, …} and p(k) = $(1-p)^{k-1} p$
- Uniform distribution on the integers {1, …, M} (parameter M)
  - Sample space: {1, 2, …, M} and p(k) = $1/M$
- Poisson distribution (parameter λ, $\lambda > 0$)

  → useful in modeling number of events (defects, accidents)
  - Sample space: {0, 1, 2, …, M} and p(k) = $\frac{\lambda^k e^{-\lambda}}{k!}$

Shape of distribution depends **on parameter(s)**

Example: Binomial distribution

- Toss a fair coin 4 times and let

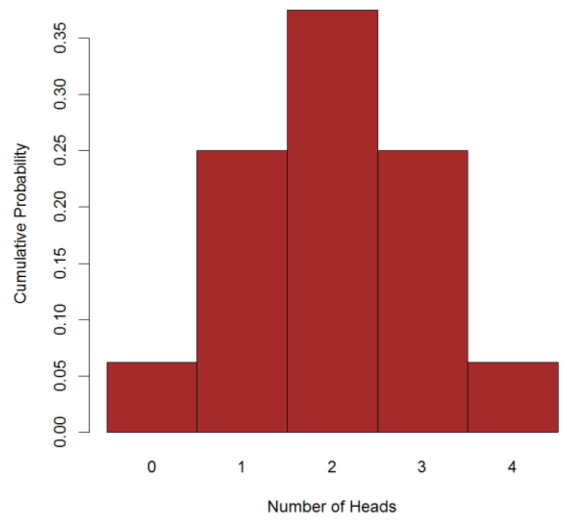
  X=the **number** of Heads in the 4 tosses

The list of the values X takes on along with the corresponding probabilities that X takes on those values is called a probability mass function.

		HTTH		
		HTHT		
	HTTT	THTH	HHHT	
	THTT	HHTT	HHTH	
	TTHT	THHT	HTHH	
TTTT	TTTH	TTHH	THHH	HHHH
$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$

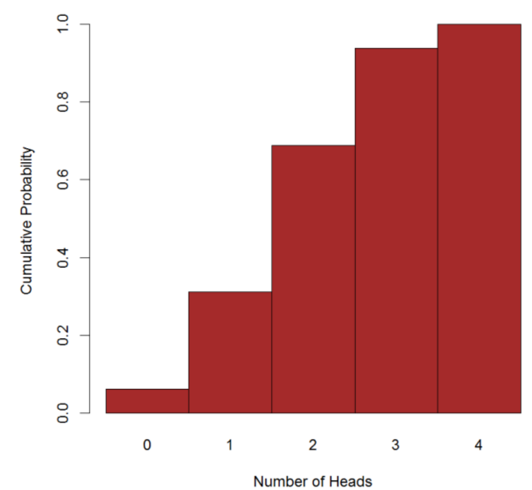
What is the probability mass function?

```
> pmf=dbinom(0:4,4,0.5); pmf
[1] 0.0625 0.2500 0.3750 0.2500 0.0625
> sum(pmf)
[1] 1
barplot(pmf, col="brown", space=0,
names.arg=c(0,1,2,3,4), xlab="Number of
Heads", ylab="Cumulative Probability")
```

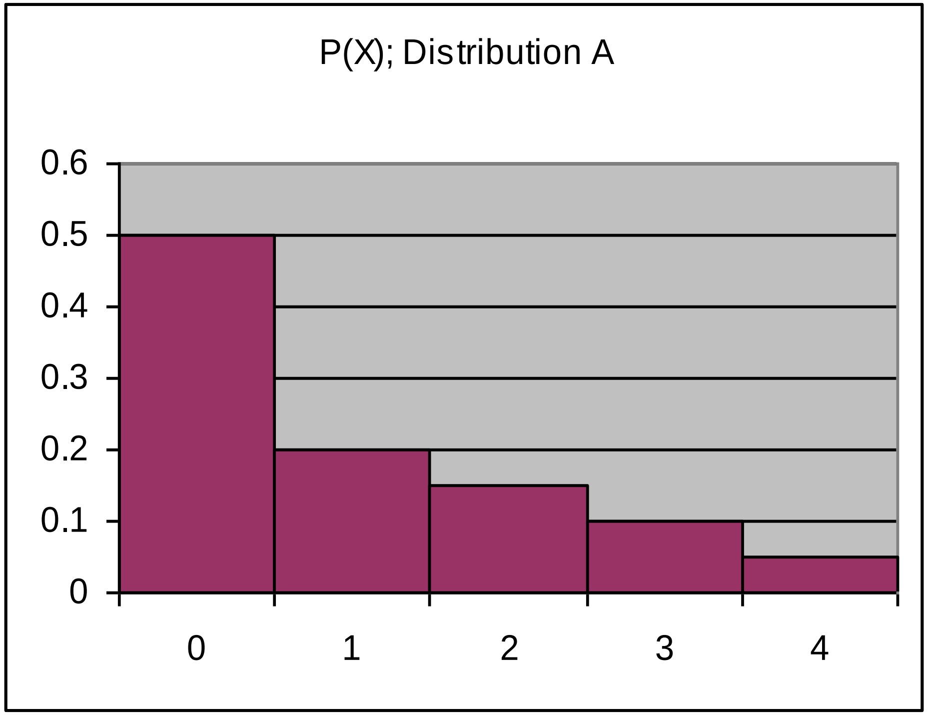


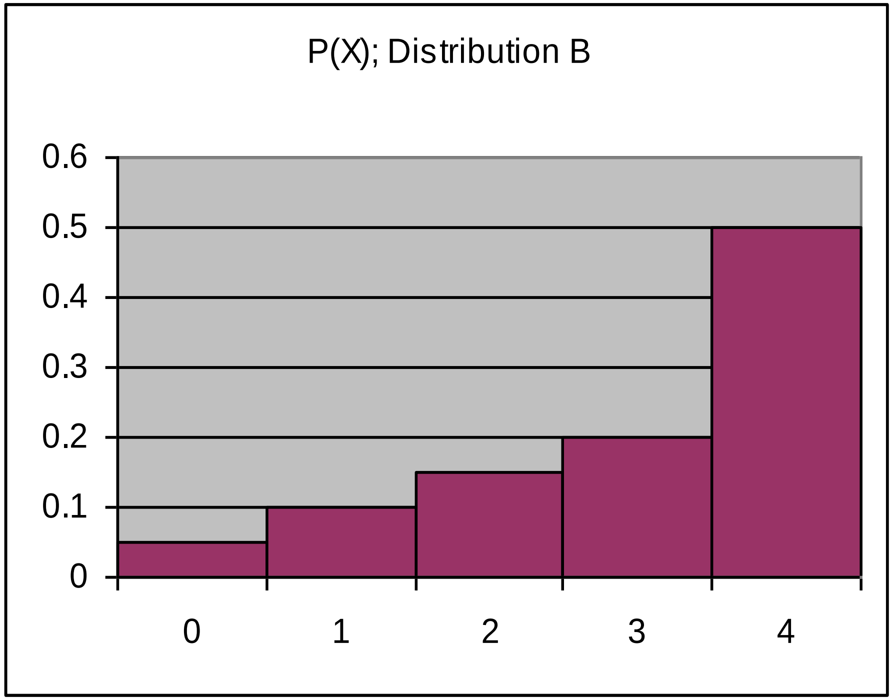
Cumulative distribution?

```
> cdf=cumsum(pmf); cdf
[1] 0.0625 0.3125 0.6875 0.9375 1.0000
barplot(cdf, col="brown", space=0,
names.arg=c(0,1,2,3,4), xlab="Number of
Heads", ylab="Cumulative Probability")
```



what does a skewed probability distribution mean (p is not ½)?





The Mean of a Distribution: Tennis example

RV: X = number of tennis tournaments won

EX=0*0.2+1*0.15+2*0.2+3*0.1+4*0.35=

Recall: data from 20 active highly-ranked women professional tennis players

- 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4, 4, 4, 4

Distinct values -- x	Frequency	Probability (Proportion)
0	4	0.20
1	3	0.15
2	4	0.20
3	2	0.10
4	7	0.35
Total	20	1.00

Expected value: Recall another example

Suppose we are interested in the number of times a Michigan resident has visited Florida, denoted X, as described by the following distribution:

	x	0	1	2	3	4	5	6
pmf p(x)	$P(X = x)$	0.3	0.1	0.2	0.1	0.06	0.08	0.16
cdf F(x)	$P(X \leq x)$	0.3	0.4	0.6	0.7	0.76	0.84	1

The expected value of X is

$$
\begin{aligned}
EX &= 0 \times 0.3 + 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.1 + 4 \times 0.06 + \\
&\quad 5 \times 0.08 + 6 \times 0.16 \\
&= 2.4
\end{aligned}
$$

Population Mean of a Discrete Distribution

- Summary of a distribution
  - A measure of "location"
  - Just like the sample mean for a set of data
- Consider a discrete RV X with K possible values $x_1, x_2, \ldots, x_K$
- Recall: its probability mass function (pmf) is

  $p(x_j)$, $j = 1, 2, \ldots , K$
- Then population mean (or expected value of X) is:

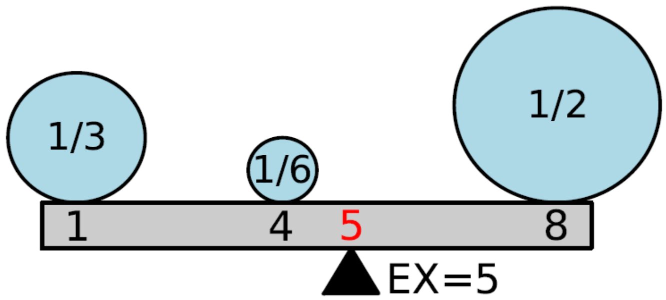
$$E(X) = \sum_{j=1}^{K} x_j P(X = x_j)$$

- For "countably" infinite number of

  possible values, K is infinity

Interpretation of Expected Value

The expected value is the balancing point if we weight each point in the sample space by its probability:



This distribution in tabular form:

x	1	4	8
P(X=x)	1/3	1/6	1/2

Suppose X is a constant –

e. g. X always takes on the value 2: P(X = 2) = 1. What is E(X)?

Suppose X takes two values, 0 with probability 999/1000 and 10,000 with probability 1/1000. What is E(X)?

If this was a lottery and the ticket costs$20, is it worthwhile to buy it?

What if the ticket were $5?

Expected Value of some distributions

- Bernoulli:
  - {1, 0}
  - p(1) = p
  - p(0) = 1 – p
  - E(X) = p

- Uniform distribution on {1, …, M}
  - {1, 2, … M}
  - p(i) = 1/M, i=1, 2, …, M
  - E(X) = (M+1)/2

- Binomial distribution on {0, …, n}
  - {0, 2, … n}
  - p(k) = $\binom{n}{k}p^k(1-p)^{n-k}$
  - E(X) = np

Expected value of linear function of RV: Example

RV X takes on two possible values: { - 1, 1}

- $P(X = -1) = 0.3$; $P(X = 1) = 0.7$
- $E(X) = (-1 \times 0.3) + (1 \times 0.7) = 0.4$

- $f(X) = 3 - 2X$

- Direct calculation
  - $E(f(X)) = f(-1) \times 0.3 + f(1) \times 0.7 = (5 \times 0.3) + (1 \times 0.7) = 2.2$

- Using linearity property:
  - $E(3 – 2X) = 3 – 2\,E(X) = 3 – 2 \times 0.4 = 2.2$

Expected value of a function of a RV: Example

$$Let\ f(x) = 2x$$

$$Then, E(2X) = 2E(X) = 2\sum_{j=1}^{K} x_j P(X = x_j)$$

Example

x	$P(X = x)$
-1	0.2
0	0.4
1	0.4

The expected values of some functions of X are:

$$EX = -1 \cdot 0.2 + 0 \cdot 0.4 + 1 \cdot 0.4 = 0.2$$

$$EX^2 = (-1)^2 \cdot 0.2 + 0^2 \cdot 0.4 + 1^2 \cdot 0.4 = 0.6$$

$$E1/(2+X) = 1 \cdot 0.2 + (1/2) \cdot 0.4 + (1/3) \cdot 0.4 \approx 0.53$$

$$In\ general, E(f(x)) = \sum_{j=1}^{K} f(x_j) P(X = x_j)$$

Expected value of linear function of RV

- Linear function of x:
  - $f(x) = a x + b$ for some constants a and b
  - Example: $f(x) = 3 x – 10$ → picture

- Then, what is $E(f(X))$?
- In this case, each value of X is shifted and scaled in the same way
- Then, the expected value is also shifted and scaled
- $E(aX + b) = a E(X) + b$
  - Why?
  - Interpretation

Implication of linearity

A consequence of linearity is that the expected value behaves naturally when we change the measurement units.

Example: Suppose we measure five peoples' heights in inches:

$$72, 68, 73, 65, 69$$

The mean height is 69.4 inches. One inch equals 2.54 centimeters, so the equivalent heights in centimeters are

$$182.88, 172.72, 185.42, 165.10, 175.26.$$

The mean height in centimeters is 176.28cm. If we simply convert the mean height in inches to centimeters we get $69.4 \times 2.54 = 176.28$cm, the same value.

This result does not hold if the function is not linear

In general, the expected value of $f(X)$ is not the same as applying f to the expected value of X (i.e. $f(EX) \neq Ef(X)$).

Example: Suppose we have a population in which $1/3$ of the people are 160cm tall, $1/3$ of the people are 170cm tall, and $1/3$ of the people are 180cm tall.

The expected height is $160/3 + 170/3 + 180/3 = 170$.

The expected squared height is $160^2/3 + 170^2/3 + 180^2/3 \approx 28967$.

But the square of the expected height is $170^2 = 28900$.

Exercises

For the following distribution, what are EX, EX^2, $\sqrt{EX^2}$ and $(EX)^2$?

x	-2	0	2
$P(X = x)$	0.2	0.5	0.3

How are the following distributions related?

	Distribution 1			Distribution 2			Distribution 3		
x	-2	0	2	-4	0	4	-1	1	3
$P(X = x)$	0.2	0.5	0.3	0.2	0.5	0.3	0.2	0.5	0.3

Variance: Tennis example

- Sample: 20 active highly-ranked women professional tennis players:
  - Data: y = number of tennis tournaments won (hypothetical – assume no more than 4)
  - 0, 0, 0, 0, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4, 4, 4, 4, 4, 4, 4

$$Variance: \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_{20} - \bar{y})^2}{20}$$

$$= \frac{1}{20}\sum_{i=1}^{20}(y_i - \bar{y})^2 = \frac{47.75}{20} = 2.3875$$

(Relative) Frequency Distribution

	Frequency	Proportion
0	4	0.20
1	3	0.15
2	4	0.20
3	2	0.10
4	7	0.35
Total	20	1.00

$$Var(X) = \sum_{j=1}^{5}(x_j - E(X))^2 P(X = x_j)$$

$$= (0 - 2.25)^2 + \cdots + (4 - 2.25)^2 = 2.3875$$

Variance of a RV or its Distribution

Variance of X: var(X) → a measure of spread of the distribution (in squared units)

$$Var(X) = \sum_{j=1}^{K} (x_j - E(X))^2 P(X = x_j)$$

Intuition:

Consider the new RV: D = X – E(X)

D is the deviation from the mean

What is E(D)?

Note that:

var(X) = E{ [X – E(X)] }^{2}

= E(D^2)

– expected value of the squared deviations from the mean

– so it is also an expected value, but of a different variable

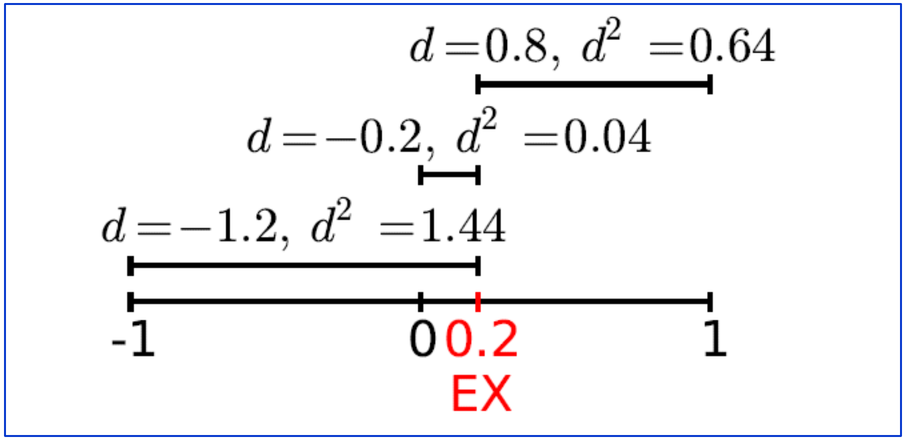
An Example

x	-1	0	1
$P(X = x)$	0.2	0.4	0.4

E(X) = ?

var(X) = E[X – E(X)]^2 = E(D^2)

The deviations from the mean and their squares are:



Therefore, the variance is

E(D^2) = $1.44 \times 0.2 + 0.04 \times 0.4 + 0.64 \times 0.4 = 0.56.$

Two pmfs: Same mean, different variances

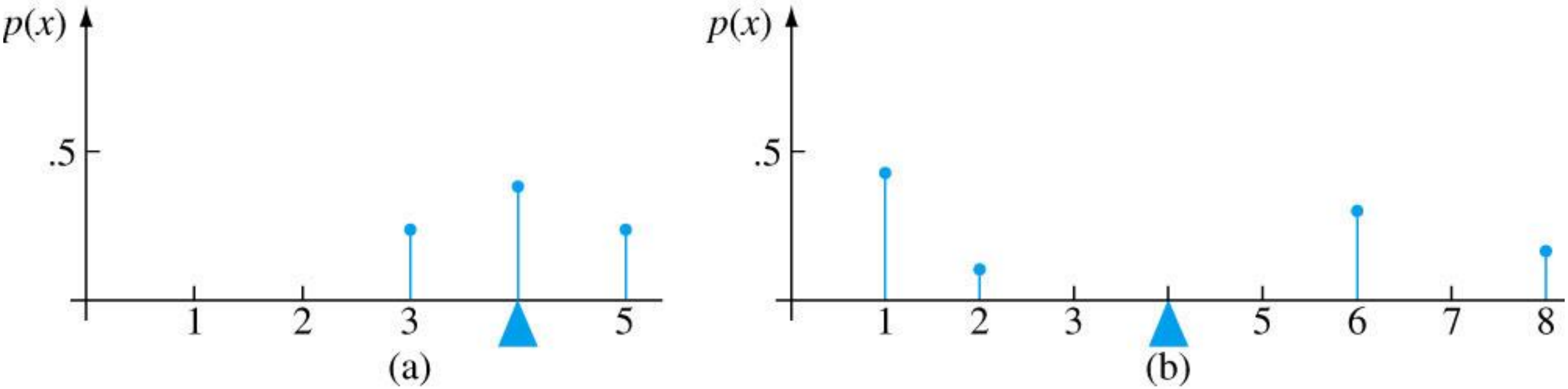


Fig. 3-7, p. 104

Some useful properties

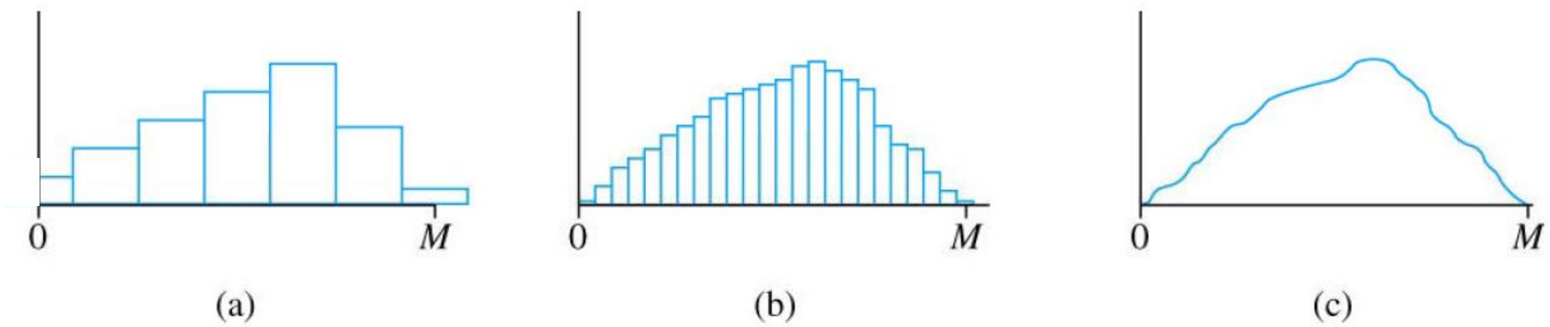
- $\text{var}(X) = E(X^2) - [E(X)]^2$ → show
  - Sample version
  - Useful for computation

- Constants have zero variance and zero standard deviation.
- Shifting a random variable X by a constant c does not change the variance or standard deviation: $\text{var}(X + c) = \text{var}(X)$, $\text{SD}(X + c) = \text{SD}(X)$.
- Scaling affects the variance and standard deviation as follows: $\text{var}(c \cdot X) = c^2\text{var}(X)$, $\text{SD}(c \cdot X) = |c|\text{SD}(X)$.
- If X is measured in some unit, say centimeters, the standard deviation of X also has centimeters as units, and the variance of X has centimeters^2 as units.

Notation: The conventional notations for expected value and standard deviation are the Greek letters μ and σ.

Continuous Random Variables

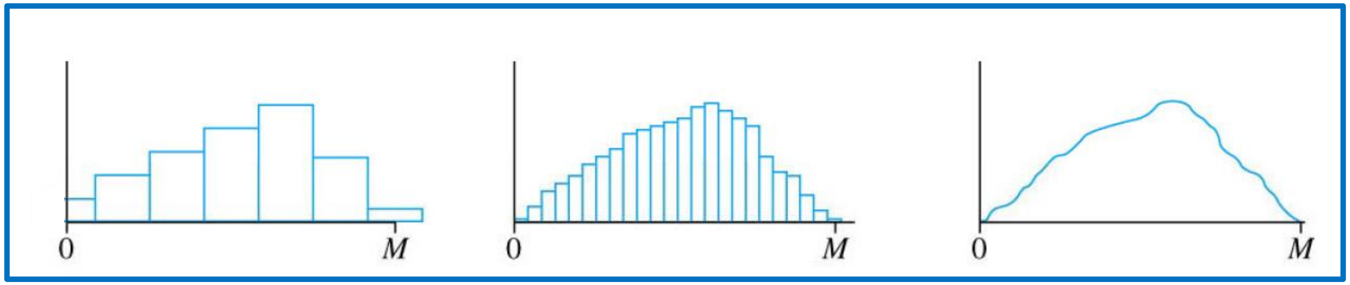
- Continuous – can take on all possible numerical values in an interval
- Most observations are discrete – due to limitations of measurements (distance, height, velocity, blood pressure)
- But if the measurement scale is “fine” enough, then can treat the data as continuous
- Example: measurement of distance in meters, centimeters, and so on.



- Continuous RV’s are useful as “idealizations” for the purpose of statistical analysis

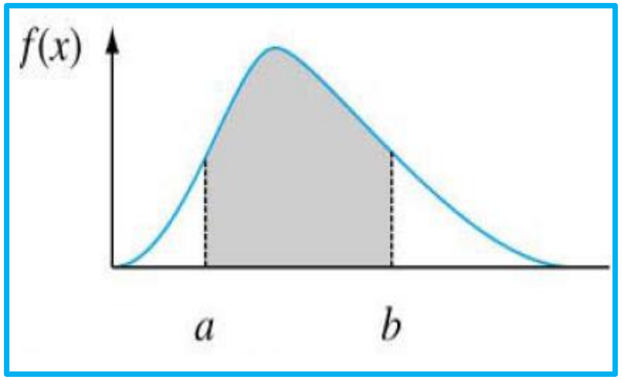
Continuous RVs and Distributions

- Notion of pmf does not work for continuous case: P(X = x) = 0 for any x



- f(x) → probability density function (pdf); f(x) is non-negative
- Intuitive interpretation: f(x) → how "likely" the value of x is (picture)
- Total area under the curve is 1
- We can use this to get the probability that X falls in an interval:
- P(X ≤ 3), P(X > - 1), P(2 < X ≤ 5) by computing the area under the curve
- In general for any two numbers a and b with a < b:

$$P(a < X < b) = \int_a^b f(x)dx$$



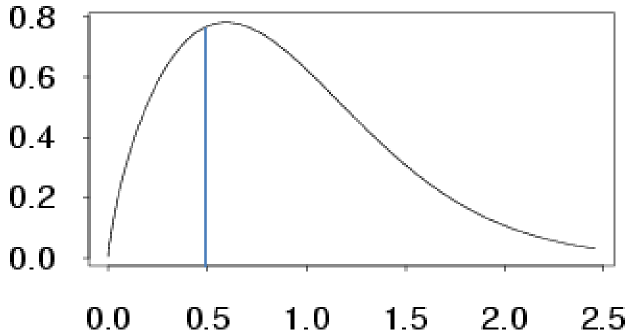
- Integral can be thought of as

  "computing the area under the curve" – like <u>*summing*</u> in the discrete case

Cumulative Distribution Function (cdf) for continuous RV

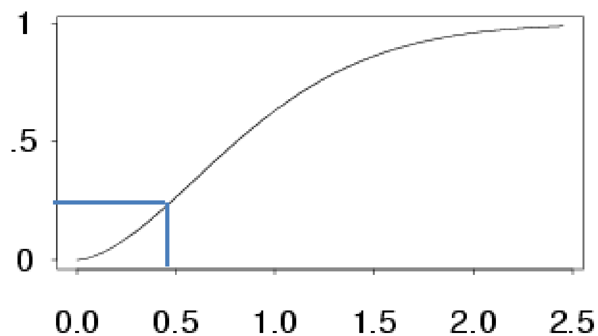
- cdf F(x) = P(X ≤ x) for any value of "little" x
- Notation: Big X and little x
- Example: F(0.5) = P(X ≤ 0.5)
  - area to the left of 0.5
- Shape of cdf F(x):
  - Starts at zero and increases to 1

Picture of pdf f(x)



- How to get P(0.5 < X ≤ 1.5)
- Area between 0.5 and 1.5

Picture of cdf F(x)



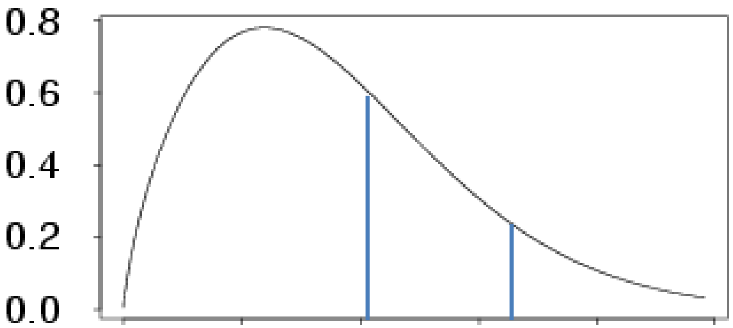
$$P(0.5 < X < 1.5)$$
$$= \int_{0.5}^{1.5} f(x)dx = F(1.5) - F(0.5)$$

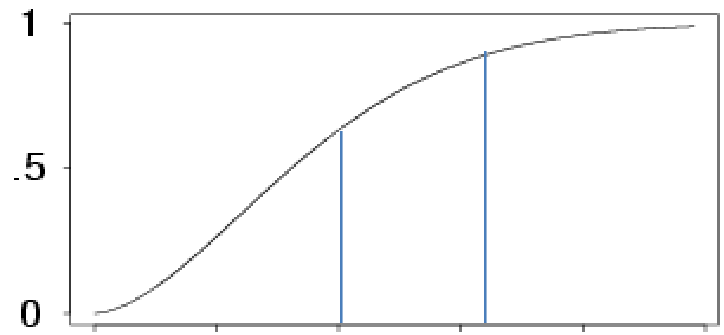
Example: Suppose we want to know the probability that a person in a certain population is between 170cm and 180cm tall. If we know that the probability of being less than 180cm tall is 0.8, and the probability of being less than 170cm tall is 0.6, then the probability of being between 170cm and 180cm tall is $0.8 - 0.6 = 0.2$.

To put this in mathematical notation, let X denote the height of a randomly selected person from the population. Then

$$P(170 < X \leq 180) = P(X \leq 180) - P(X < 170) = 0.8 - 0.6.$$

F(180) – F(170)



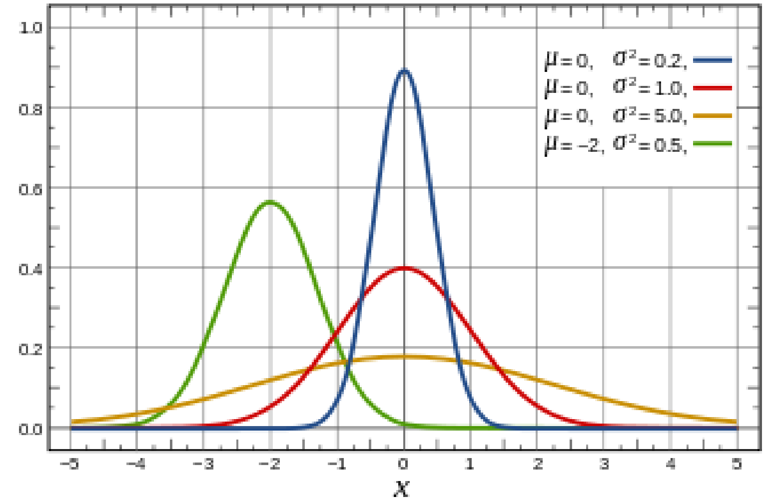


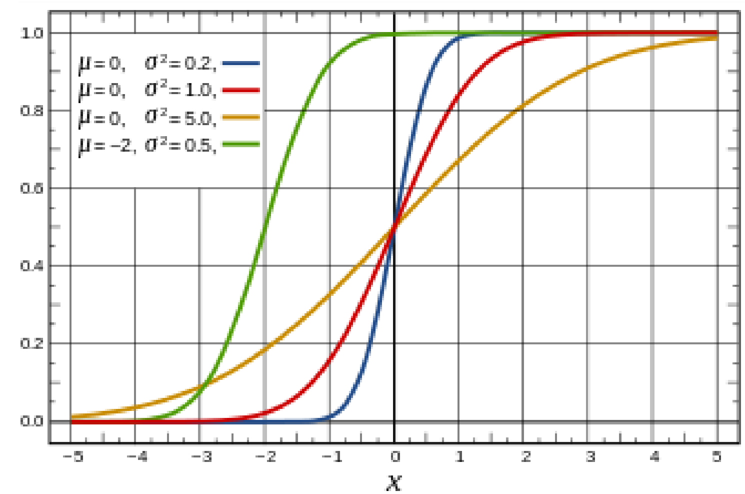
A well-known example of a continuous distribution

- Normal distribution
  - We'll write it as N(μ, σ)
  - μ and σ are "parameters"
  - σ is positive

- For normal distribution:
  - μ is the mean
  - σ is the SD

- Sometimes written as N(μ, σ^2)
  - N(5, 4) – 4 is SD or var?
  - use the context to figure out
- Standard normal distribution
  - μ = 0 and σ =1





Images from Wikipedia

Other examples of continuous distributions

- Continuous uniform distribution
- t distribution
- chi-squared distribution
- F distribution
- Will discuss the last three later
- Many others – but the above is enough for our purposes

Expected value and variance of continuous random variables

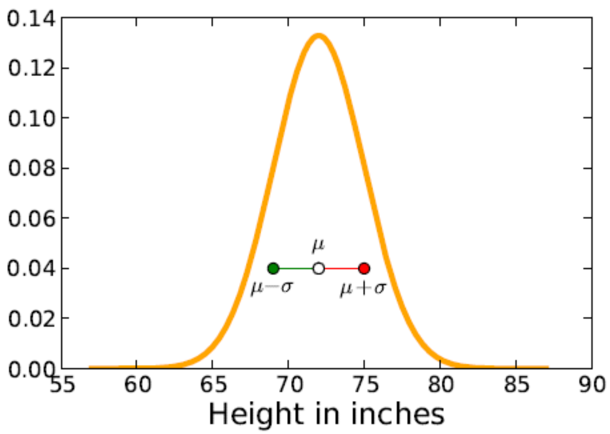
- $E(X) = \int x\, f(x)\, dx$
  - what does this mean?

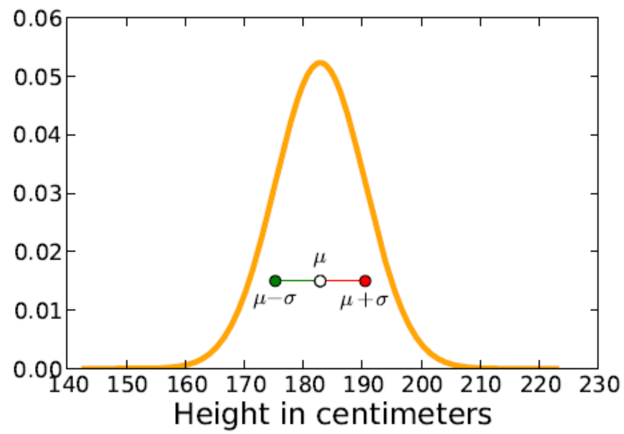
- Variance is defined the same way but with this notion of expectation
  - $\text{var}(X) = E[\ X - E(X)\ ]^2$

- All the properties of expectation and variance continue to hold

Shifting or Translation

- X → X + b
- Changing the "location" of the RV (or its distribution) by adding a constant b to each value
- Mean shifts by b
- No change to the variance
- **X → aX; mean and standard of new X now is multiplied by a:**





On the left $\mu = 72, \sigma = 3$; on the right $\mu = 72 \cdot 2.54 = 182.88$, $\sigma = 3 \cdot 2.54 = 7.62$.

Properties: review

- If X is a constant c:
  - its expected value is c and it has zero variance
- E(aX + b) = a E(X) + b
- In general, $f(EX) \neq Ef(X))$

- var(X + b) = var(X)
- var(cX) = c^2 var(X)
- SD(cX) = |c| SD(X)

Centering and Standardization

If X is a random variable with mean μ and variance σ^2, then

$$Y \equiv X - \mu$$

is a new random variable with mean 0 and variance σ^2 called the **centered** version of X, and

$$Z \equiv (X - \mu)/\sigma$$

is a new random variable with mean 0 and variance 1 called the **standardized** version of X.

Suppose the population of heights (in centimeters) is 168, 175, 163, 180, and 175, all with equal probabilities (1/5 each).

The mean is 172.2, so the centered values are -4.2, 2.8, -9.2, 7.8, and 2.8.

The standard deviation is 5.98, so the standardized values are -0.70, 0.47, -1.54, 1.30, and 0.47.