UNIVERSITY AT ALBANY
State University of New York

Lecture 8: Probability and Applications

Dr. Chengjiang Long
Computer Vision Researcher at Kitware Inc.
Adjunct Professor at SUNY at Albany.
Email: **clong2@albany.edu**

Recap Previous Lecture

- Binomial Coefficients
- Direct Proof and Combinatorial Proof

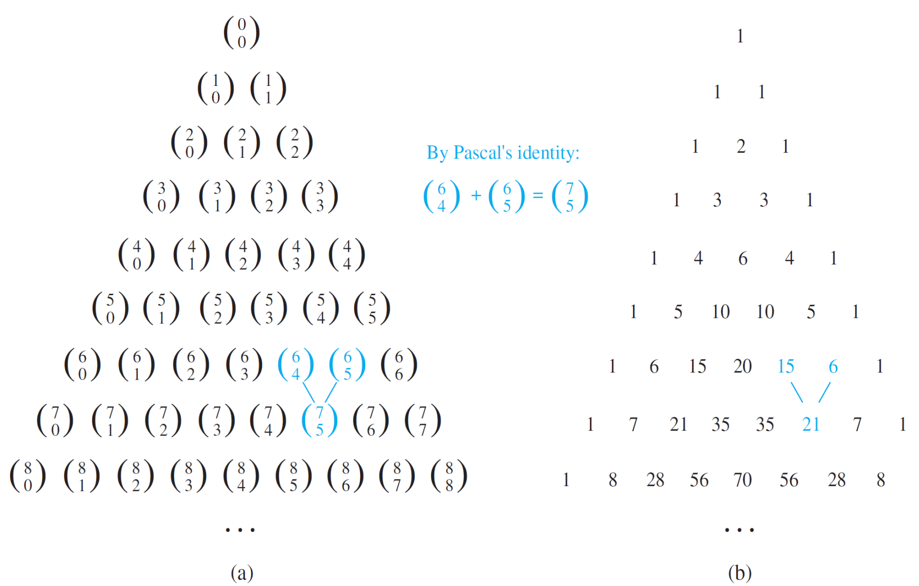
Pascal's Formula

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

$$\sum_{i \text{ odd}} \binom{n}{i} = \sum_{j \text{ even}} \binom{n}{j}$$

$$\binom{n}{k} = \binom{n}{n-k}$$



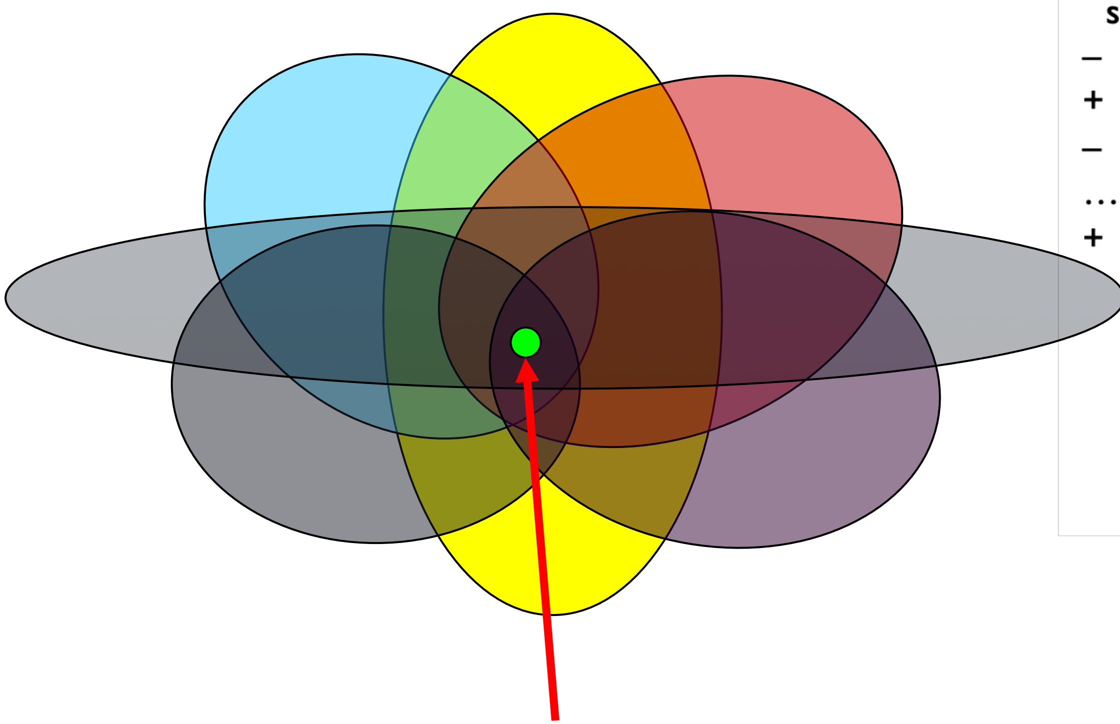
Recap Previous Lecture

- Inclusion-exclusion Principle

$$|A_1 \cup A_2 \cup \cdots \cup A_n| =$$

sum of sizes of all single sets
− sum of sizes of all 2-set intersections
+ sum of sizes of all 3-set intersections
− sum of sizes of all 4-set intersections
…
+ $(-1)^{n+1}$ × sum of sizes of intersections of all n sets

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{\substack{S \subseteq \{1,2,\ldots,n\} \\ |S|=k}} \left| \bigcap_{i \in S} A_i \right|$$



$$\binom{k}{1} - \binom{k}{2} + \binom{k}{3} - \binom{k}{4} + \ldots + (-1)^{k+1}\binom{k}{k} = 1$$

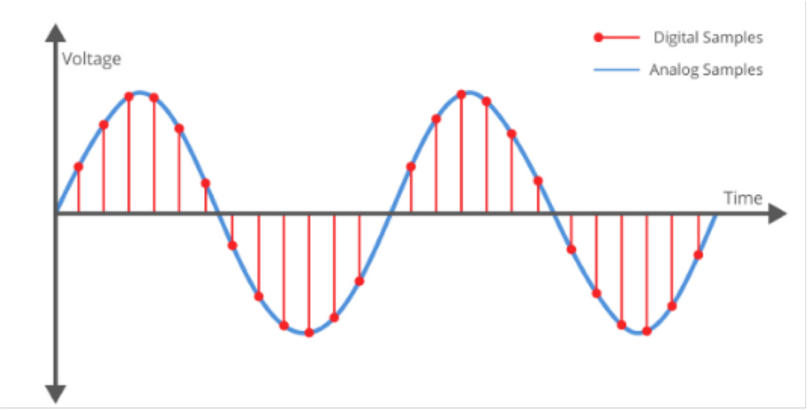
Outline

- Discrete Probability
- Conditional Probability
- Bayes Rules, Expected Value, Variances
- Binominal Distribution

Discrete Probability

Discrete Random Variables

- A Random Variable is a measurement on an outcome of a random experiment.
- Discrete versus Continuous random variable: a random variable *x* is discrete if it can assume a finite or countably infinite number of values. *x* is continuous if it can assume all values in an interval.



Example

- Which of the following random variables are discrete and which are continuous?
  - x = Number of houses sold by real estate developer per week?
  - x = Number of heads in ten tosses of a coin?
  - x = Weight of a child at birth?
  - x = Time required to run100 yards?

More examples

- If you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.
- If you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition 100" is a also a random variable (the percentage will be slightly differently every time you poll).

Roughly, probability is how frequently we expect different outcomes to occur if we repeat the experiment over and over ("frequentist" view)

Random variables can be discrete or continuous

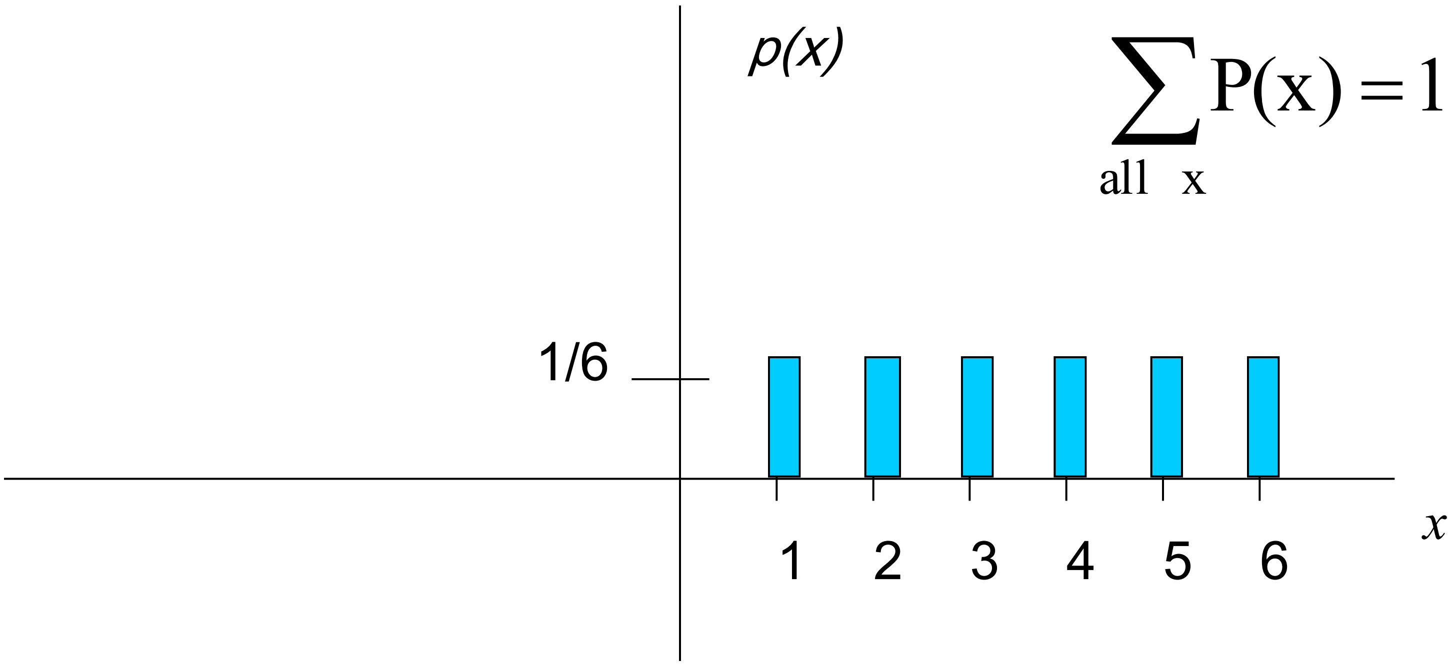
- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, treatment/placebo, dice, counts, etc.
- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

Probability functions

If S is a finite nonempty sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the *probability* of E is $p(E) = \frac{|E|}{|S|}$.

- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*
- *p(x)* is a number from 0 to 1.0.
- The area under a probability function is always 1.

Discrete example: roll of a die



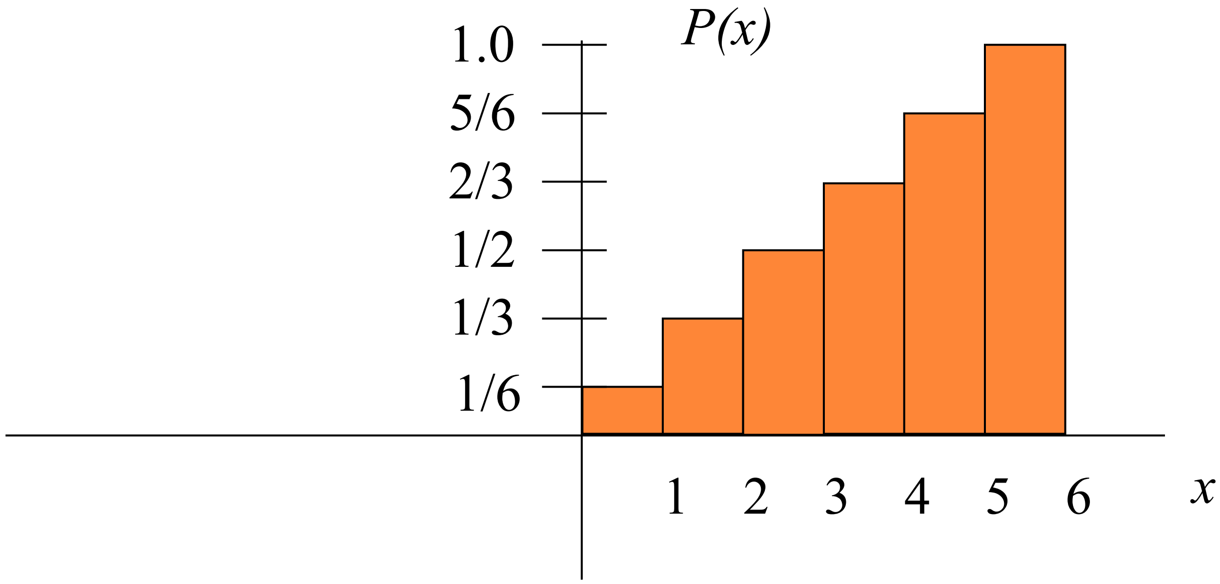
$$\sum_{\text{all } x} P(x) = 1$$

Probability mass function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$

1.0

Cumulative distribution function (CDF)



Cumulative distribution function

x	$P(x \leq A)$
1	$P(x \leq 1) = 1/6$
2	$P(x \leq 2) = 2/6$
3	$P(x \leq 3) = 3/6$
4	$P(x \leq 4) = 4/6$
5	$P(x \leq 5) = 5/6$
6	$P(x \leq 6) = 6/6$

Practice Problem:

- The number of patients seen in the ER in any given hour is a random variable represented by *x*. The probability distribution for *x* is:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

Find the probability that in a given hour:

a. exactly 14 patients arrive $p(x=14) = .1$

b. At least 12 patients arrive $p(x \geq 12) = (.2 + .1 + .1) = .4$

c. At most 11 patients arrive $p(x \leq 11) = (.4 + .2) = .6$

Example 1

If you toss a die, what's the probability that you roll a 3 or less?

The possible value of the die? **1, 2, 3**, 4, 5, 6

One of these three ≤ 3.

$\therefore 1/2$

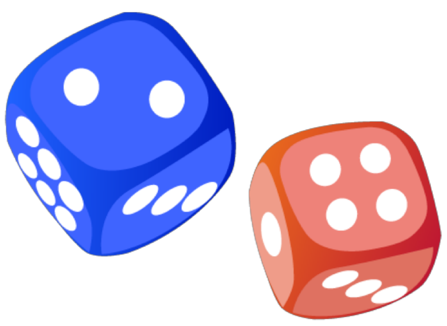
Example 2

Two dice are rolled and the sum of the face values is six? What is the probability that at least one of the dice came up a 3?

How can you get a 6 on two dice? 1-5, 5-1, 2-4, 4-2, 3-3

One of these five has a 3.

$\therefore$ 1/5



Property of probability

- We can use counting techniques to find the probability of events derived from other events.

Let E be an event in a sample space S. The probability of the event $\overline{E} = S - E$, the complementary event of E, is given by

$$p(\overline{E}) = 1 - p(E).$$

A sequence of 10 bits is randomly generated. What is the probability that at least one of these bits is 0?

Solution: Let E be the event that at least one of the 10 bits is 0. Then $\overline{E}$ is the event that all the bits are 1s. Because the sample space S is the set of all bit strings of length 10, it follows that

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{|\overline{E}|}{|S|} = 1 - \frac{1}{2^{10}} = 1 - \frac{1}{1024} = \frac{1023}{1024}.$$

Property of probability

- We can use counting techniques to find the probability of events derived from other events.

Let E_1 and E_2 be events in the sample space S. Then

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2).$$

What is the probability that a positive integer selected at random from the set of positive integers not exceeding 100 is divisible by either 2 or 5?

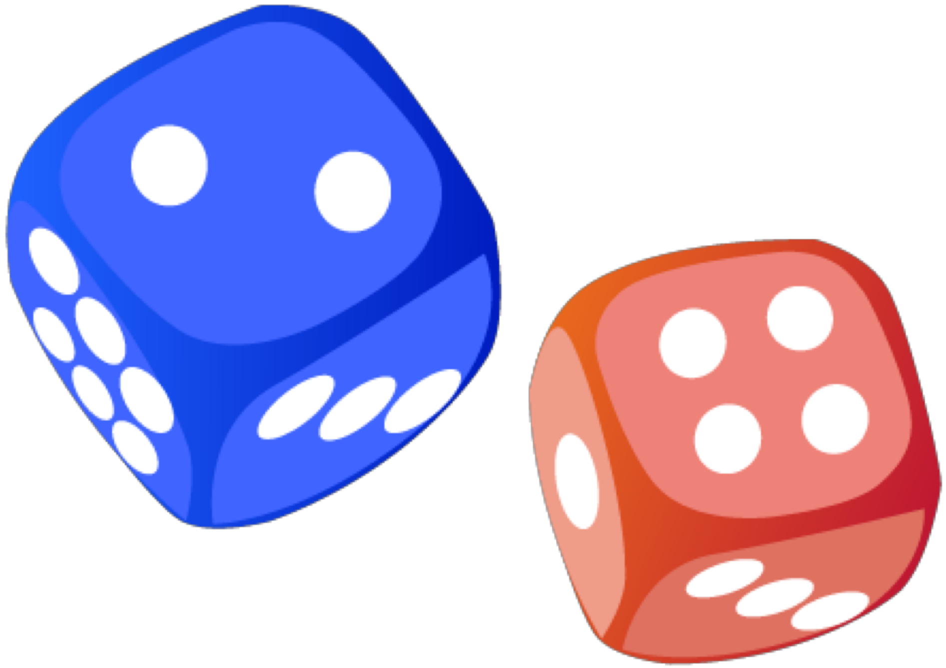
Solution: Let E_1 be the event that the integer selected at random is divisible by 2, and let E_2 be the event that it is divisible by 5.

$\overline{E}$

$$|E_1| = 50,\ |E_2| = 20, \text{ and } |E_1 \cap E_2| = 10$$

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) = \frac{50}{100} + \frac{20}{100} - \frac{10}{100} = \frac{3}{5}.$$

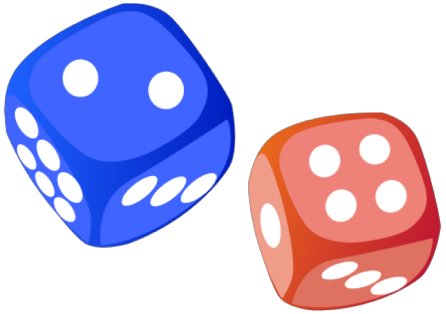
Examples



X is the Sum of Two Dice. What is the probability of X?

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						



This sequence provides an example of a discrete random variable. Suppose that you have a red die which, when thrown, takes the numbers from 1 to 6 with equal probability.

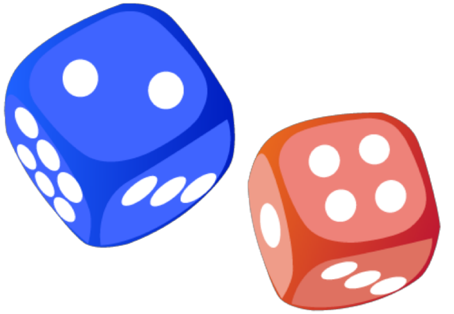
Probability Distribution Example: X is the Sum of Two Dice

red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Suppose that you also have a green die that can take the numbers from 1 to 6 with equal probability.

Probability Distribution Example: X is the Sum of Two Dice

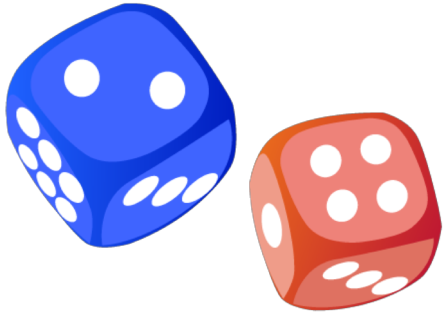
red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						



We will define a random variable X as the sum of the numbers when the dice are thrown.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6				10		



For example, if the red die is 4 and the green one is 6, X is equal to 10.

Probability Distribution Example: X is the Sum of Two Dice

red green	1	2	3	4	5	6
1						
2						
3						
4						
5		7				
6						

Similarly, if the red die is 2 and the green one is 5, X is equal to 7.

Probability Distribution Example: X is the Sum of Two Dice

red green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

The table shows all the possible outcomes.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

We will now define f, the frequencies associated with the possible values of X.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	4
6	
7	
8	
9	
10	
11	
12	

For example, there are four outcomes which make X equal to 5.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Similarly you can work out the frequencies for all the other values of X.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

Finally we will derive the probability of obtaining each value of X.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

If there is 1/6 probability of obtaining each number on the red die, and the same on the green die, each outcome in the table will occur with 1/36 probability.

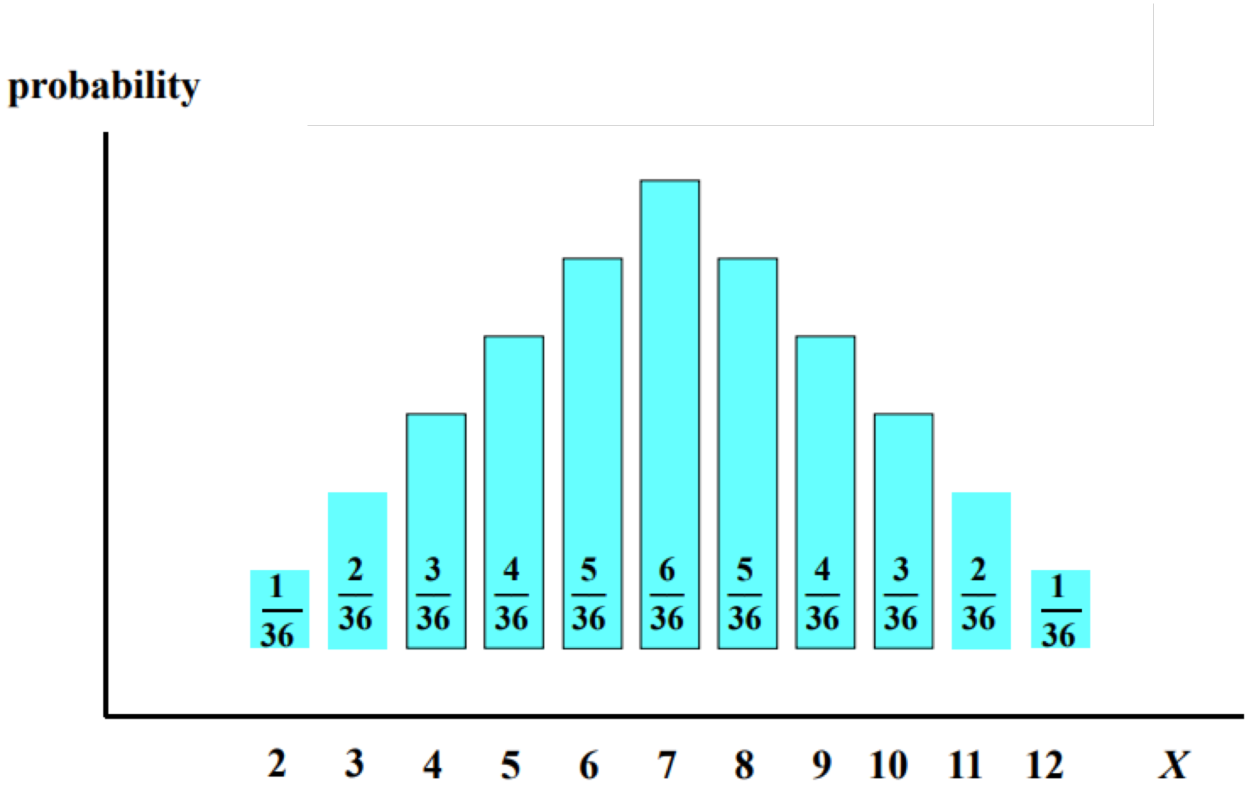
Probability Distribution Example: X is the Sum of Two Dice

green \ red	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

Hence to obtain the probabilities associated with the different values of X, we divide the frequencies by 36.

Probability Distribution Example: X is the Sum of Two Dice



The distribution is shown graphically. in this example it is symmetrical, highest for X equal to 7 and declining on either side.

Continuous case

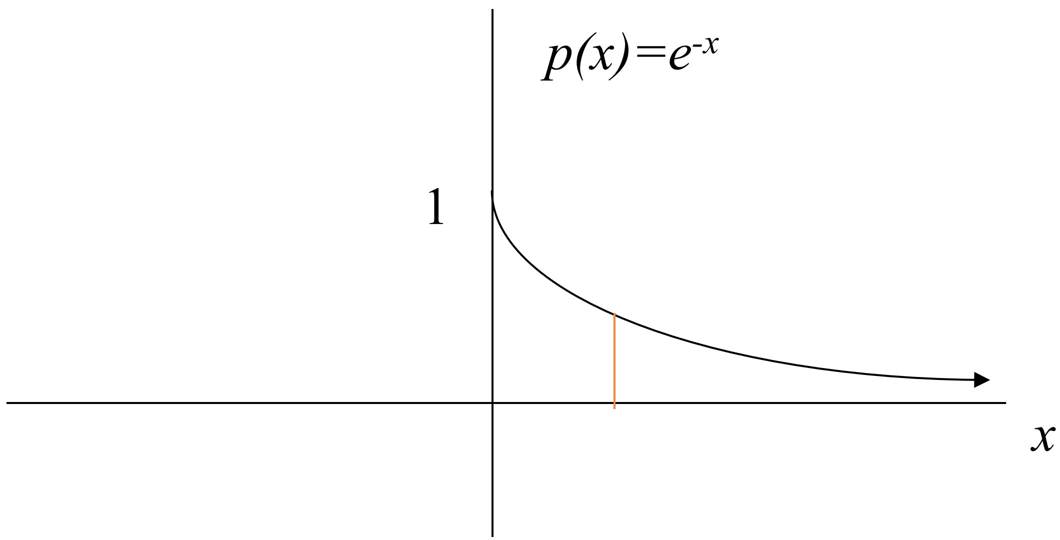
- The probability function that accompanies a continuous random variable is a continuous mathematical function that integrates to 1.
  - For example, recall the negative exponential function (in probability, this is called an "exponential distribution"):

$$f(x) = e^{-x}$$

- This function integrates to 1:

$$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$$

Probability density function for continuous case



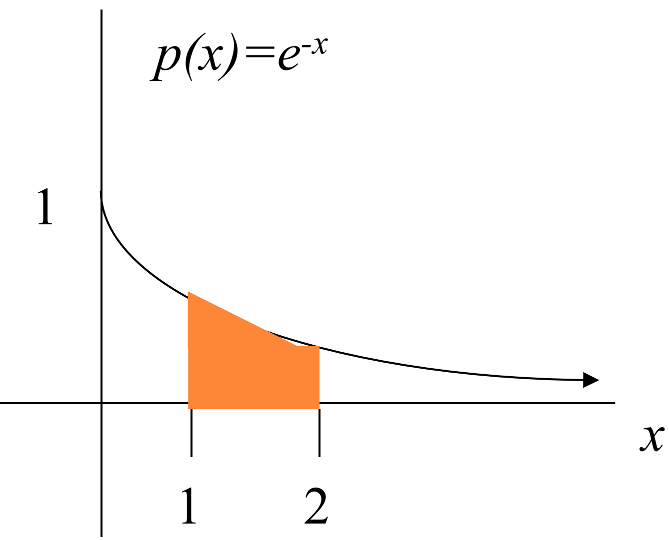
The probability that x is any exact particular value (such as 1.9976) is 0; we can only assign probabilities to possible ranges of x.

Probability density function for continuous case

For example, the probability of *x* falling within 1 to 2:

Clinical example: Survival times after lung transplant may roughly follow an exponential function.

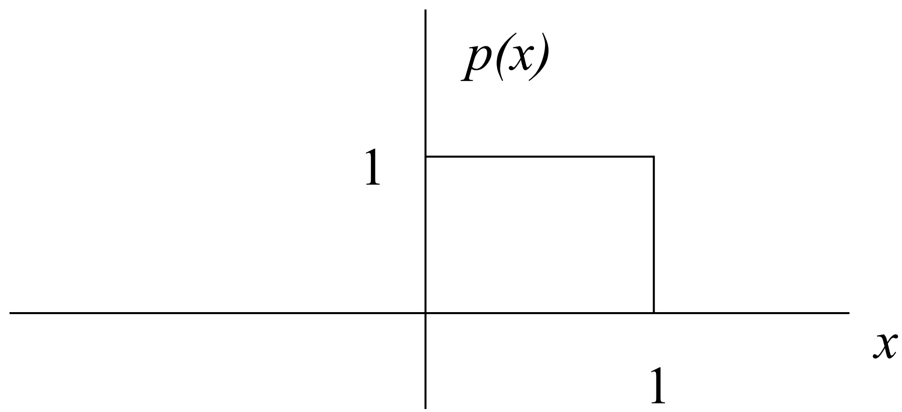
Then, the probability that a patient will die in the second year after surgery (between years 1 and 2) is 23%.



$$P(1 \le x \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

Example: Uniform distribution

The uniform distribution: all values are equally likely.
$f(x) = 1$, for $1 \geq x \geq 0$

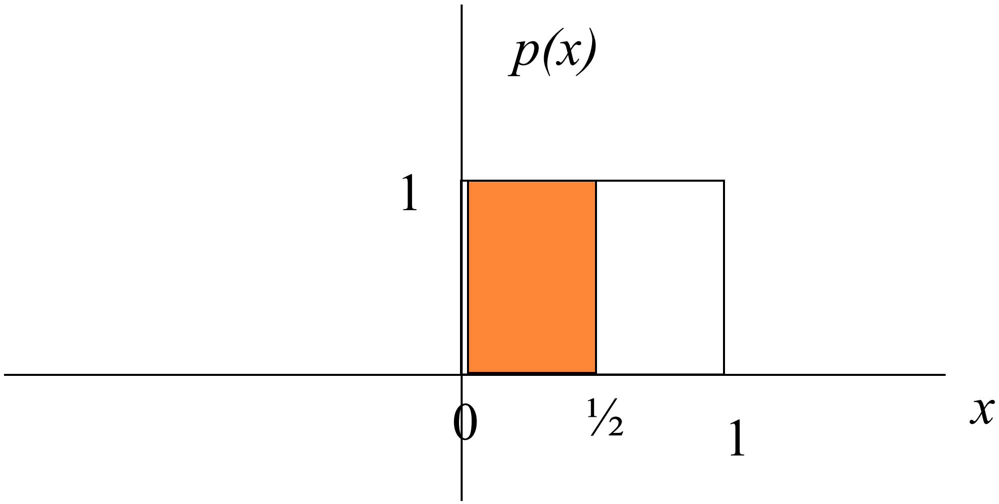


We can see it's a probability distribution because it integrates to 1 (the area under the curve is 1):

$$\int_0^1 1 = x \Big|_0^1 = 1 - 0 = 1$$

Example: Uniform distribution

What's the probability that x is between 0 and ½?



Clinical Research Example: When randomizing patients in an RCT, we often use a random number generator on the computer. These programs work by randomly generating a number between 0 and 1 (with equal probability of every number in between). Then a subject who gets X<.5 is control and a subject who gets X>.5 is treatment.

$$P(\tfrac{1}{2} \geq x \geq 0) = \tfrac{1}{2}$$

Conditional Probability

Probability functions

If S is a finite nonempty sample space of equally likely outcomes, and E is an event, that is, a subset of S, then the *probability* of E is $p(E) = \frac{|E|}{|S|}$.

- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*
- *p(x)* is a number from 0 to 1.0.
- The area under a probability function is always 1.

Pairs of Discrete Random Variables

- Let x and y be two discrete r.v.
- For each possible pair of values, we can define a joint probability ***P(x, y)***
- We can also define a joint probability mass function ***P(x,y)*** which offers a complete characterization of the pair of

$$P_x(x) = \sum_{y \in Y} P(x, y)$$

$$P_y(y) = \sum_{x \in X} P(x, y)$$

Marginal distributions

Example

Two dice are rolled and the sum of the face values is six? What is the probability that at least one of the dice came up a 3?

How can you get a 6 on two dice? 1-5, 5-1, 2-4, 4-2, 3-3

One of these five has a 3.

$\therefore$ 1/5

Statistical Independence

- Two random variables x and y are said to be independent, if and only if

$$P(x,y)=P_x(x)\,P_y(y)$$

that is, when knowing the value of x does not give us additional information for the value of y.

- Or, equivalently

$$E[f(x)g(y)] = E[f(x)]\,E[g(y)]$$

for any functions f(x) and g(y).

Conditional Probability

- When two r.v. are not independent, knowing one allows better estimate of the other (e.g. outside temperature, season)

$$\Pr[x = x_i \mid y = y_j] = \frac{\Pr[x = x_i , y = y_j]}{\Pr[y = y_j]}$$

- If independent P(x|y)=P(x)

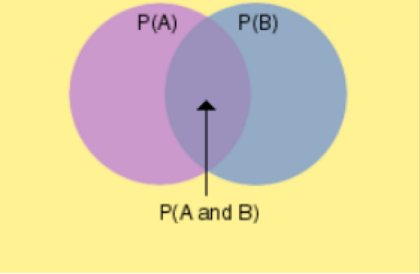
Conditional Probability Example

- A jar contains black and white marbles.
  - Two marbles are chosen without replacement.
  - The probability of selecting a black marble and then a white marble is 0.34.
  - The probability of selecting a black marble on the first draw is 0.47.

- What is the probability of selecting a white marble on the second draw, given that the first marble draw was black?

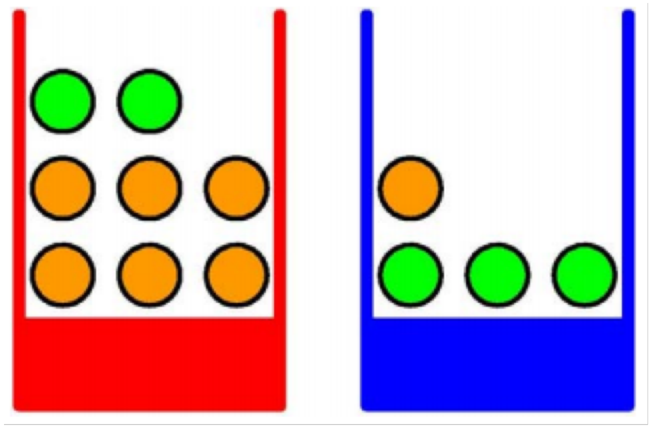
$$P(White \mid Black) = \frac{P(Black \wedge White)}{P(Black)} = \frac{0.34}{0.47} = 0.72$$

A is black in first draw, B is white in second draw



Sum and Product Rules

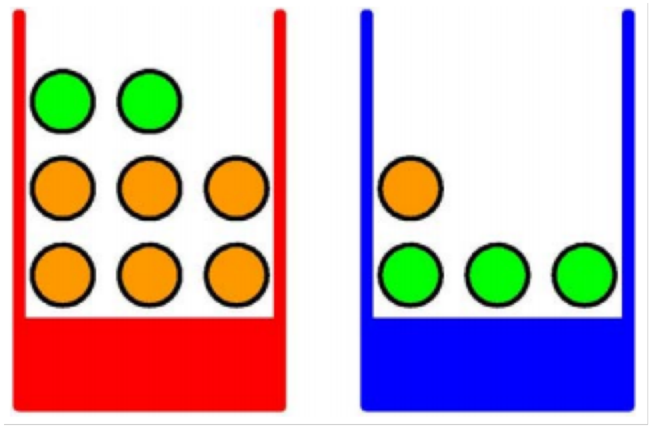
- Example:
  - We have two boxes: one red and one blue
  - Red box: 2 apples and 6 oranges
  - Blue box: 3 apples and 1 orange



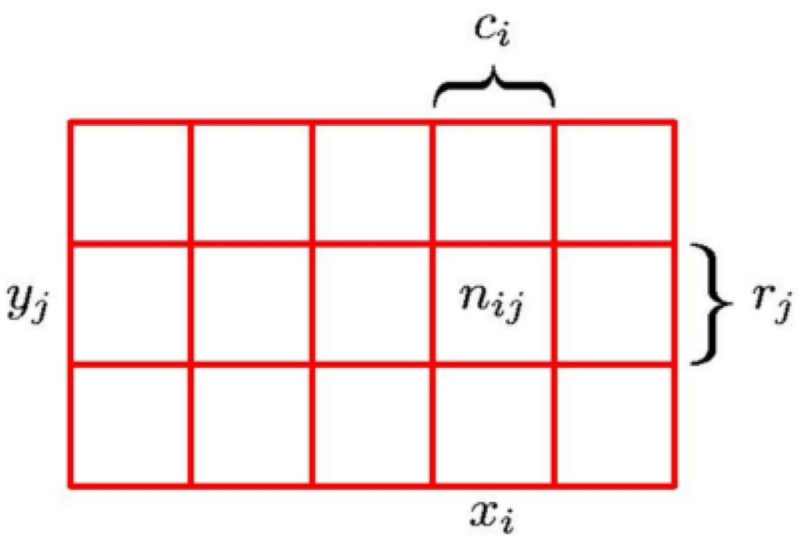
[*C.M. Bishop, "Pattern Recognition and Machine Learning", 2006*]

Sum and Product Rules

- Define:
  - B random variable for box picked (r or b)
  - F identity of fruit (a or o)
- p(B=r)=8/12 and p(B=b)=4/12
  - Events are mutually exclusive and include all possible outcomes ⇨ their probabilities must sum to 1.



Sum and Product Rules



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

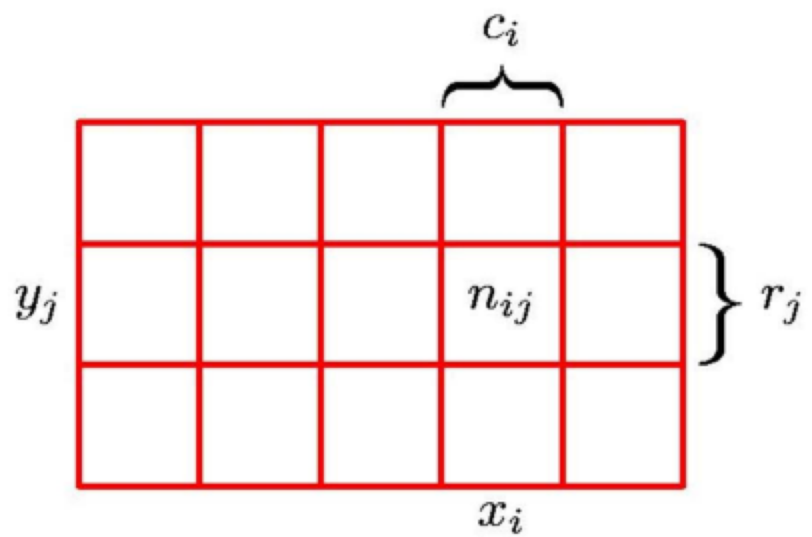
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Sum and Product Rules



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$
$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$
$$= p(Y = y_j | X = x_i) p(X = x_i)$$

Sum and Product Rules

- Sum Rule $$p(X) = \sum_{Y} p(X, Y)$$
- Product Rule $$p(X, Y) = p(Y|X)p(X)$$

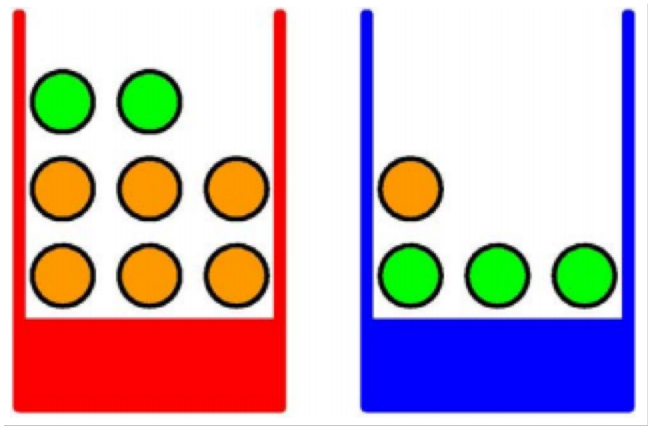
Law of Total Probability

- If an event A can occur in m different ways and if these m different ways are mutually exclusive, then the probability of A occurring is the sum of the probabilities of the sub-events

$$P(X = x_i) = \sum_j P(X = x_i \mid Y = y_j) P(Y = y_j)$$

Sum and Product Rules

- Back to the fruit baskets
  - p(B=r)=4/10 and p(B=b)=6/10
  - p(B=r) + p(B=b) = 1
- Conditional probabilities
  - p(F=a | B = r) = 1/4
  - p(F=o | B = r) = 3/4
  - p(F=a | B = b) = 3/4
  - p(F=o | B = b) = 1/4



Sum and Product Rules

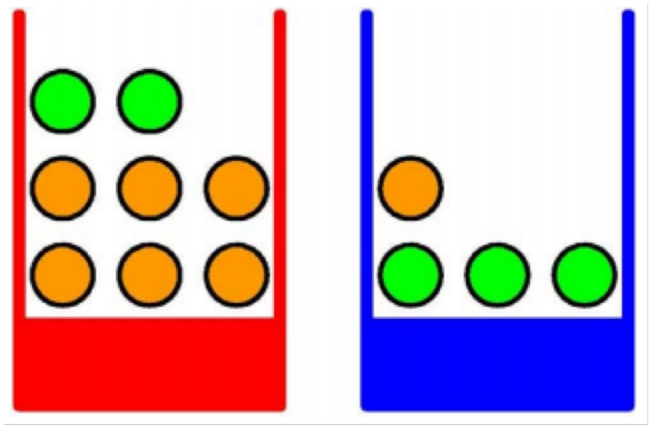
- Note:

$p(F=a \mid B = r) + p(F=o \mid B = r) = 1$

$p(F=a) = p(F=a \mid B = r)\, p(B=r) + p(F=a \mid B = b) p(B=b)$

$= 1/4 * 4/10 + 3/4 * 6/10 = 11/20$

- Sum rule: $p(F=o) = ?$



Bayes Rules, Expected Value, Variances

Law of Total Probability

$$P_x(x) = \sum_{y \in Y} P(x, y)$$

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

Bayes Rule

$$P(x \mid y) = \frac{P(x,y)}{P(y)} = \frac{P(y \mid x)P(x)}{\sum_{x \in X} P(x,y)}$$

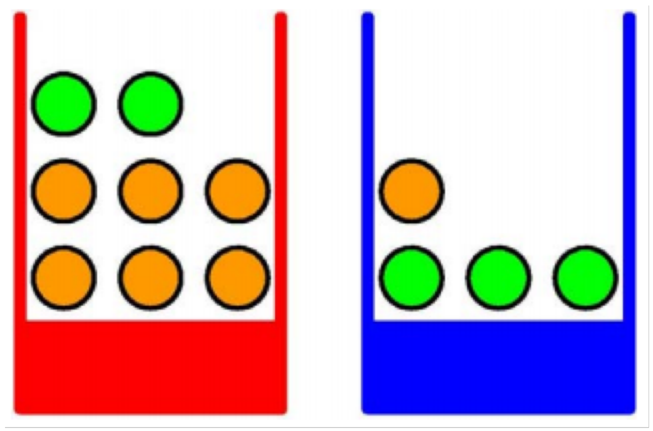
$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

- x is the unknown cause
- y is the observed evidence
- Bayes rule shows how probability of x changes after we have observed y

Bayes Rule on the Fruit Example

- Suppose we have selected an orange. Which box did it come from?

$$p(B=r \mid F=o) = \frac{p(F=o \mid B=r)\,p(B=r)}{p(F=o)} = \frac{\frac{3}{4} \times \frac{4}{10}}{\frac{9}{20}} = \frac{2}{3}$$



Continuous Random Variables

- Examples: room temperature, time to run100m, weight of child at birth…
- Cannot talk about probability of that x has a particular value
- Instead, probability that x falls in an interval ⇨ **probability density function**

$$\Pr[x \in (a,b)] = \int_a^b p(x)dx$$

$$p(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} p(x)dx = 1$$

Bayes Rule for Continuous Case

- Bayes rule $$p(x \mid y) = \frac{p(y \mid x)p(x)}{\int_{-\infty}^{\infty} p(y \mid x)p(x)dx}$$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

Expected Value and Variance

- All probability distributions are characterized by an expected value (mean) and a variance (standard deviation squared).

Expected value of a random variable

- Expected value is just the average or mean (μ) of random variable *x*.
- It's sometimes called a "weighted average" because more frequent values of X are weighted more highly in the average.
- It's also how we expect X to behave on-average over the long run ("frequentist" view again).

Expected value, formally

Discrete case:

$$E(X) = \sum_{\text{all } x} x_i p(x_i)$$

Continuous case:

$$E(X) = \int_{\text{all } x} x_i p(x_i) dx$$

- E(X) = μ. These symbols are used interchangeably

Example: expected value

- Recall the following probability distribution of ER arrivals:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

$$\sum_{i=1}^{5} x_i\, p(x) = 10(.4) + 11(.2) + 12(.2) + 13(.1) + 14(.1) = 11.3$$

Sample Mean is a special case of Expected Value…

- Sample mean, for a sample of n subjects:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)$$

The probability (frequency) of each person in the sample is 1/n.

Example: the lottery

- A certain lottery works by picking 6 numbers from 1 to 49. It costs $1.00 to play the lottery, and if you win, you win $2 million after taxes.

- *If you play the lottery once, what are your expected winnings or losses?*

Expected Value

- Expected value is an extremely useful concept for good decision-making!

Lottery

Calculate the probability of winning in 1 try:

$$\frac{1}{\binom{49}{6}} = \frac{1}{\frac{49!}{43!6!}} = \frac{1}{13{,}983{,}816} = 7.2 \text{ x } 10^{-8}$$

"49 choose 6"

Out of 49 numbers, this is the number of distinct combinations of 6.

The probability function (note, sums to 1.0):

x$	*p(x)*
-1	.999999928
+ 2 million	*7.2 x 10*$^{--8}$

Expected Value

The probability function

$x\$$	$p(x)$
-1	.999999928
+ 2 million	$7.2 \times 10^{--8}$

Expected Value

E(X) = P(win)*\$2,000,000 + P(lose)*-\$1.00

$= 2.0 \times 10^{6} * 7.2 \times 10^{-8} + .999999928\ (-1)$

$= .144 - .999999928 = -\$.86$

Negative expected value is never good!
You shouldn’t play if you expect to lose money!

Expected Value

If you play the lottery every week for 10 years, what are your expected winnings or losses?

520 x (-.86) = -$447.20

Variance/standard deviation

σ^2=Var(x) =E(x-μ)^2

"The expected (or average) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 \, p(x_i)$$

Variance, continuous

Discrete case:

$$Var(X) = \sum_{\text{all x}} (x_i - \mu)^2 p(x_i)$$

Continuous case?:

$$Var(X) = \int_{\text{all x}} (x_i - \mu)^2 p(x_i) dx$$

- Var(X)= σ^2
- SD(X) = σ
- These symbols are used interchangeably.

Similarity to empirical variance

- The variance of a sample: $s^2 =$

$$\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{n-1} = \sum_{i=1}^{N}(x_i - \bar{x})^2 \left(\frac{1}{n-1}\right)$$

Division by n-1 reflects the fact that we have lost a "degree of freedom" (piece of information) because we had to estimate the sample mean before we could estimate the sample variance.

Variance

$$\sigma^2 = \sum_{\text{all x}} (x_i - \mu)^2 p(x_i)$$

$$\sigma^2 = \sum_{\text{all x}} (x_i - \mu)^2 p(x_i) =$$

$$= (1 - 200{,}000)^2(.5) + (400{,}000 - 200{,}000)^2(.5) = 200{,}000^2$$

$$\sigma = \sqrt{200{,}000^2} = 200{,}000$$

Now you examine your personal risk tolerance…

Practice Problem

On the roulette wheel, X=1 with probability 18/38 and X= -1 with probability 20/38.

We already calculated the mean to be = -\$.053. What's the variance of X?

Answer

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

$$= (+1 - -.053)^2 (18/38) + (-1 - -.053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-1 + .053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-.947)^2 (20/38)$$

$$= .997$$

$$\sigma = \sqrt{.997} = .99$$

Standard deviation is $.99. Interpretation: On average, you're either 1 dollar above or 1 dollar below the mean, which is just under zero. Makes sense!

Binominal Distribution

Binomial Probability Distribution

- A fixed number of observations (trials), n
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary outcome
  - e.g., head or tail in each toss of a coin; disease or no disease
  - Generally called "success" and "failure"
  - Probability of success is p, probability of failure is 1 – p
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin

Binomial distribution

Take the example of 5 coin tosses. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Solution: One way to get exactly 3 heads: HHHTT

What's the probability of this exact arrangement?
P(H) x P(H) x P(H) x P(T) x P(T) = $(1/2)^3 \times (1/2)^2$

Another way to get exactly 3 heads: THHHT
Probability of this exact outcome = $(1/2)^1 \times (1/2)^3 \times (1/2)^1$
= $(1/2)^3 \times (1/2)^2$

Binomial distribution

- In fact, $(1/2)^3$ *x* $(1/2)^2$ is the probability of each unique outcome that has exactly 3 heads and 2 tails.

- So, the overall probability of 3 heads and 2 tails is:

  $(1/2)^3 \; x \; (1/2)^2 + (1/2)^3 \; x \; (1/2)^2 + (1/2)^3 \; x \; (1/2)^2 + \ldots..$ for as many unique arrangements as there are—but how many are there?

Binomial distribution

$\binom{5}{3}$ ways to arrange 3 heads in 5 trials

Outcome	Probability
THHHT	$(1/2)^3 \, x \, (1/2)^2$
HHHTT	$(1/2)^3 \, x \, (1/2)^2$
TTHHH	$(1/2)^3 \, x \, (1/2)^2$
HTTHH	$(1/2)^3 \, x \, (1/2)^2$
HHTTH	$(1/2)^3 \, x \, (1/2)^2$
HTHHT	$(1/2)^3 \, x \, (1/2)^2$
THTHH	$(1/2)^3 \, x \, (1/2)^2$
HTHTH	$(1/2)^3 \, x \, (1/2)^2$
HHTHT	$(1/2)^3 \, x \, (1/2)^2$
THHTH	$(1/2)^3 \, x \, (1/2)^2$

10 arrangements $x \, (1/2)^3 \, x \, (1/2)^2$

The probability of each unique outcome (note: they are all equal)

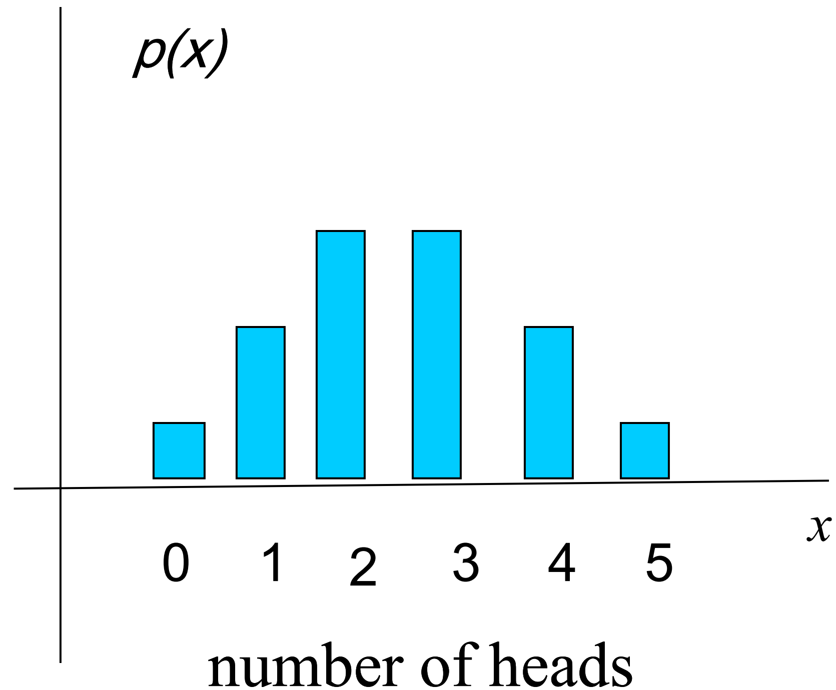
C(5,3) = 5!/3!2! = 10

Factorial review: n! = n(n-1)(n-2)…

Binomial distribution

$\therefore$ P(3 heads and 2 tails) = $\binom{5}{3}$ $x\ P(heads)^3\ x\ P(tails)^2$ =

$10\ x\ (½)^{5=}31.25\%$



Binomial distribution function:
X= the number of heads tossed in 5 coin tosses

Binomial distribution, generally

Note the general pattern emerging → if you have only two possible outcomes (call them 1/0 or yes/no or success/failure) in *n* independent trials, then the probability of exactly *X* "successes"=

$$\binom{n}{X} p^X (1-p)^{n-X}$$

n = number of trials

X = # successes out of *n* trials

p = probability of success

1-p = probability of failure

Example

- If I toss a coin 20 times, what's the probability of getting exactly 10 heads?

$$\binom{20}{10}(.5)^{10}(.5)^{10} = .176$$

Example

- If I toss a coin 20 times, what's the probability of getting of getting 2 or fewer heads?

$$\binom{20}{0}(.5)^0(.5)^{20} = \frac{20!}{20!0!}(.5)^{20} = 9.5x10^{-7} +$$

$$\binom{20}{1}(.5)^1(.5)^{19} = \frac{20!}{19!1!}(.5)^{20} = 20x9.5x10^{-7} = 1.9x10^{-5} +$$

$$\binom{20}{2}(.5)^2(.5)^{18} = \frac{20!}{18!2!}(.5)^{20} = 190x9.5x10^{-7} = 1.8x10^{-4}$$

$$= 1.8x10^{-4}$$

Expected Value and Variance for Binormal Distribution

- **All probability distributions are characterized by an expected value and a variance:**

If X follows a binomial distribution with parameters n and p: ***X ~ Bin (n, p)***

Then:

$E(X) = np$

$Var(X) = np(1-p)$

$SD(X) = \sqrt{np(1-p)}$

Note: the variance will always lie between

0*N-.25 *N

p(1-p) reaches maximum at p=.5

P(1-p)=.25

Example

- 1. You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration, then if you (randomly) sample 500 exposed people, how many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

$$X \sim \text{binomial}\ (500, .05)$$

$$E(X) = 500\ (.05) = 25$$

$$Var(X) = 500\ (.05)\ (.95) = 23.75$$

$$StdDev(X) = \text{square root}\ (23.75) = 4.87$$

$$\therefore 25 \pm 4.87$$

Example

2. What's the probability that **at most** 10 exposed subjects develop the disease?

This is asking for a CUMULATIVE PROBABILITY: the probability of 0 getting the disease or 1 or 2 or 3 or 4 or up to 10.

P(X≤10) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)+….+ P(X=10)=

$$\binom{500}{0}(.05)^0(.95)^{500}+\binom{500}{1}(.05)^1(.95)^{499}+\binom{500}{2}(.05)^2(.95)^{498}+\ldots+\binom{500}{10}(.05)^{10}(.95)^{490}<.01$$

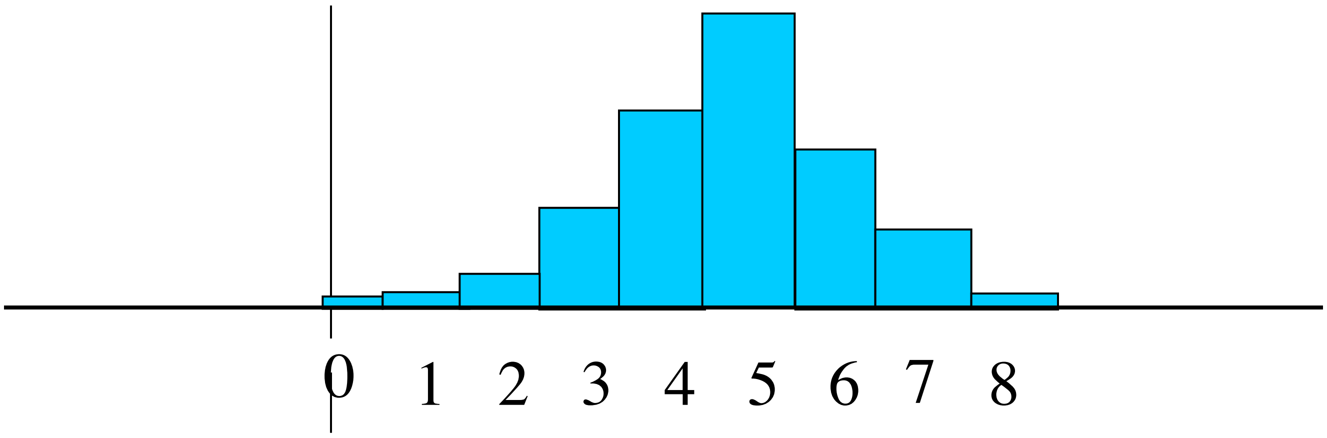
Practice Problem:

You are conducting a case-control study of smoking and lung cancer. If the probability of being a smoker among lung cancer cases is .6, what's the probability that in a group of 8 cases you have:

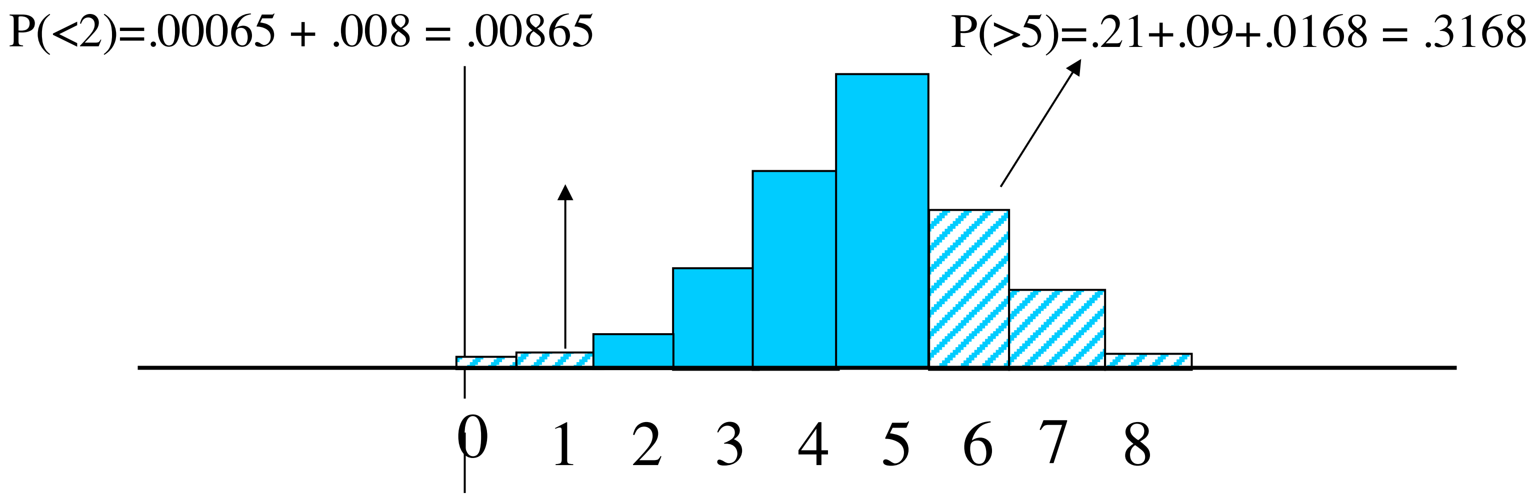
a. Less than 2 smokers?
b. More than 5?
c. What are the expected value and variance of the number of smokers?

Answer

X	P(X)
0	$1(.4)^8=.00065$
1	$8(.6)^1\,(.4)^7=.008$
2	$28(.6)^2\,(.4)^6=.04$
3	$56(.6)^3\,(.4)^5=.12$
4	$70(.6)^4\,(.4)^4=.23$
5	$56(.6)^5\,(.4)^3=.28$
6	$28(.6)^6\,(.4)^2=.21$
7	$8(.6)^7\,(.4)^1=.090$
8	$1(.6)^8=.0168$



Answer, continued



E(X) = 8 (.6) = 4.8
Var(X) = 8 (.6) (.4) =1.92
StdDev(X) = 1.38

Proportions…

- The binomial distribution forms the basis of statistics for proportions.
- A proportion is just a binomial count divided by n.
  - For example, if we sample 200 cases and find 60 smokers, X=60 but the observed proportion=.30.
- Statistics for proportions are similar to binomial counts, but differ by a factor of n.

Stats for proportions

For binomial:

$$\mu_x = np$$

$$\sigma_x^2 = np(1-p)$$

$$\sigma_x = \sqrt{np(1-p)}$$

Differs by a factor of n.

Differs by a factor of n.

For proportion:

$$\mu_{\hat{p}} = p$$

$$\sigma_{\hat{p}}^2 = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

P-hat stands for "sample proportion."

It all comes back to normal…

- Statistics for proportions are based on a normal distribution, because the binomial can be approximated as normal if $np>5$

Next class

- Topic: Advanced Counting Techniques
- Pre-class reading: Chap 8