Math 3081 (Probability and Statistics)

Lecture #6 of 27 $\sim$ July 14th, 2021

Bayes' Formula and Discrete Random Variables (Part 1)

- Bayes' Formula and Applications
- The Prosecutor's Fallacy
- Discrete Random Variables
- Expected Value

This material represents §1.4.3 + §2.1.1 from the course notes and problems 15-20 on WeBWorK 2.

Recall, I

Recall the definitions of conditional probability and independence:

Definition

If A and B are events and $P(B) > 0$, we define the conditional probability $P(A|B)$, the probability that A occurs given that B occurred, as $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$.

Definition

We say that two events A and B are independent if $P(A|B) = P(A)$, or equivalently if $P(B|A) = P(B)$.

In practice, we usually use the equivalent formulation that says A and B are independent precisely when $P(A \cap B) = P(A) \cdot P(B)$.

More Bayes' Formula, I

At the end of the last lecture, we also introduced Bayes' formula:

Theorem (Bayes' Formula)

If A and B are any events, then

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}.$$

More generally, if events $B_1, B_2, \dots, B_k$ are mutually exclusive and have union the entire sample space, then

$$P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A|B_1) \cdot P(B_1) + \cdots + P(A|B_k) \cdot P(B_k)}.$$

This result is named after Rev. Thomas Bayes who used conditional probability to give bounds on an unknown parameter (a topic we will discuss a bit later in the course) in the 1760s.

More Bayes' Formula, II

Proof:

- By definition, $P(B|A) = P(A \cap B)/P(A)$, and we also have $P(A \cap B) = P(A|B) \cdot P(B)$.
- Also, as we have noted several times already, $P(A) = P(A \cap B) + P(A \cap B^c)$ since these events are mutually exclusive and have union A.
- Then $P(A \cap B) = P(A|B) \cdot P(B)$ and $P(A \cap B^c) = P(A|B^c) \cdot P(B^c)$ from the definition of conditional probability. Plugging all of these values in yields the formula immediately.
- The second formula follows in the same way by writing $P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$.

More Bayes' Formula, III

Example: Alex has two urns, one labeled H which has 5 black and 4 red balls, and one labeled T which has 4 black and 11 red balls. Alex flips an unfair coin that has a $2/3$ probability of landing heads, and then draws one ball at random from the correspondingly-labeled urn (H for heads, T for tails). If Alex draws a red ball, what is the probability that the coin flip was heads?

More Bayes' Formula, III

Example: Alex has two urns, one labeled H which has 5 black and 4 red balls, and one labeled T which has 4 black and 11 red balls. Alex flips an unfair coin that has a 2/3 probability of landing heads, and then draws one ball at random from the correspondingly-labeled urn (H for heads, T for tails). If Alex draws a red ball, what is the probability that the coin flip was heads?

- Let R, H, T be the events of "red ball", "heads", and "tails".
- We want $P(H|R)$, which we may get via Bayes' formula:
$$P(H|R) = \frac{P(R|H) \cdot P(H)}{P(R|H) \cdot P(H) + P(R|T) \cdot P(T)}$$
- We have $P(H) = 2/3$, $P(T) = 1/3$, and also $P(R|H) = 4/9$ and $P(R|T) = 11/15$. Therefore,
$$P(H|R) = \frac{P(R|H) \cdot P(H)}{P(R|H) \cdot P(H) + P(R|T) \cdot P(T)} = \frac{40}{73} \approx 54.8\%.$$

More Bayes' Formula, IV

Bayes' formula can also be used to give another solution to the Monty Hall Problem:

- Recall that P_1, P_2, P_3 identify the prize location and H_2, H_3 identify which door the host opened: we want $P(P_1|H_2)$.

More Bayes' Formula, IV

Bayes' formula can also be used to give another solution to the Monty Hall Problem:

- Recall that P_1, P_2, P_3 identify the prize location and H_2, H_3 identify which door the host opened: we want $P(P_1|H_2)$.
- By symmetry, $P(P_1) = P(P_2) = P(P_3) = 1/3$.
- Also, from the setup, $P(H_2|P_1) = P(H_3|P_1) = 1/2$, $P(H_3|P_2) = P(H_2|P_3) = 1$, and $P(H_2|P_2) = P(H_3|P_3) = 0$.
- Therefore, since P_1, P_2, P_3 are disjoint and have union the entire sample space, by Bayes' formula we see $P(P_1|H_2)$
$$= \frac{P(H_2|P_1) \cdot P(P_1)}{P(H_2|P_1) \cdot P(P_1) + P(H_2|P_2) \cdot P(P_2) + P(H_2|P_3) \cdot P(P_3)}$$
$$= \frac{(1/2) \cdot (1/3)}{(1/2) \cdot (1/3) + 0 \cdot (1/3) + 1 \cdot (1/3)} = \frac{1/6}{1/2} = \frac{1}{3},$$ as before.

The Prosecutor's Fallacy, I

<u>Example</u> (Prosecutor's Fallacy): A DNA sample from a minor crime is compared to a state forensic database containing 100,000 records.

- A single suspect is identified on this basis alone, with no other evidence suggesting guilt or innocence.
- From analysis of human genetic variation, it is determined that the probability that a randomly-selected innocent person would match the DNA sample is 1 in 10000.
- At the trial, the prosecutor states that the probability that the suspect is innocent is only 1 in 10000, and observes that this figure means that it is overwhelmingly likely (a probability of 99.99%) that the suspect is guilty.

Critique the prosecutor's statement.

The Prosecutor's Fallacy, II

We will write everything using more careful notation.

- Suppose M is the event that there is a DNA match, and I is the event that the suspect is innocent.

The Prosecutor's Fallacy, II

We will write everything using more careful notation.

- Suppose M is the event that there is a DNA match, and I is the event that the suspect is innocent.
- The conditional probability $P(I|M)$ is the probability that the suspect is innocent given that there is a DNA match, which is what the prosecutor is claiming is equal to $1/10000$.
- However, the 1-in-10000 figure is actually the probability that there is a match given that the suspect is innocent: this is the conditional probability $P(M|I)$, which as we have seen is quite different from $P(I|M)$.
- Thus, the prosecutor has made a serious error: namely, mixing up the probability $P(M|I)$ with the probability $P(I|M)$.

The Prosecutor's Fallacy, III

We can use Bayes' formula to find the actual value of $P(I|M)$:

- We have $P(I|M) = \dfrac{P(M|I) \cdot P(I)}{P(M|I) \cdot P(I) + P(M|I^c) \cdot P(I^c)}$.
- A priori, there is no reason to believe that the given suspect is any more likely to be guilty than any other person in the database, so we will take $P(I^c) = 1/100000 = 0.00001$ so that $P(I) = 0.99999$.
- We also take $P(M|I) = 1/10000 = 0.0001$ as indicated, and $P(M|I^c) = 1$ since (we presume) the DNA analysis will always identify a guilty suspect.
- Then $P(I|M) = \dfrac{0.0001 \cdot 0.99999}{0.0001 \cdot 0.99999 + 1 \cdot 0.00001} \approx 90.9\%$.
- The conditional probability in this case states that there is a 90.9% chance that the suspect is innocent, given the existence of a positive match and no other evidence: quite a far cry from the prosecutor's claim of 99.999% guilt!

The Prosecutor's Fallacy, IV

The confusion of the probability of innocence given a positive match with the probability of a positive match given innocence is called the prosecutor's fallacy.

- As we just showed with this dramatic example, it is a very serious error.
- This error (and others of a similar nature) have actually led to erroneous convictions in several famous cases.
- In our example, one would expect roughly $100000/10000 = 10$ DNA matches to come from the database, and given the lack of evidence to say otherwise, it is no more likely that the given suspect is guilty than any of these 10 people.

If you take one thing away from this discussion, it should be the importance of computing the correct probability! (This is one reason we have used algebraic language as much as possible.)

Overview of §2: Random Variables

We now begin our next chapter of the course: §2: Random Variables, which is the study of functions defined on sample spaces.

- We will begin by discussing discrete random variables, which are random variables defined on finite sample spaces (or countably infinite ones, like the positive integers).
- We will develop basic measures of the behavior of random variables, such as its expected value, variance, and standard deviation, and also study the situation of having more than random variable defined on the same sample space.
- We then discuss continuous random variables (defined on the real line), along with the many analogies between discrete and continuous random variables.
- Finally, we discuss a number of important random variables and some applications to modeling real-world phenomena.

Random Variables: Motivation

When we observe the result of an experiment, we are often interested in some specific property of the outcome rather than the entire outcome itself.

- For example, if we flip a fair coin 5 times, we may want to know only the total number of heads obtained, rather than the exact sequence of all 5 flips.
- As another example, if we roll a pair of dice, we may want to know the sum of the outcomes rather than the results of each individual roll.
- As a third example, if an archer fires an arrow at a target, we may want to know the number of points she scores, rather than the exact position of the arrow.

Each of these quantities is a function on the outcomes comprising the underlying sample space.

Random Variables

Formally, properties of outcomes can be thought of as functions defined on the outcomes of a sample space:

Definition

A random variable is a (real-valued) function defined on the outcomes in a sample space.

Random Variables

Formally, properties of outcomes can be thought of as functions defined on the outcomes of a sample space:

Definition

A random variable is a (real-valued) function defined on the outcomes in a sample space.

Since we have spent most of our time analyzing finite sample spaces, we will first discuss random variables on such sample spaces:

Definition

A discrete random variable is a random variable is one whose underlying sample space is finite or countably infinite.

Discrete Random Variables, I

Examples: Consider the the experiment of flipping a coin 5 times, with corresponding sample space S.

- One random variable X on S is the total number of heads obtained. The value of X on the outcome $HHHHT$ is 4, while the value of X on the outcome $TTTTT$ is 0.

Discrete Random Variables, I

<u>Examples</u>: Consider the the experiment of flipping a coin 5 times, with corresponding sample space S.

- One random variable X on S is the total number of heads obtained. The value of X on the outcome $HHHHT$ is 4, while the value of X on the outcome $TTTTT$ is 0.
- Another random variable Y is the length of the longest consecutive run of heads. The value of Y on the outcome $HTHHT$ is 2, while the value on $THHHH$ is 4, and the value on $HTHHH$ is 3.

Discrete Random Variables, I

Examples: Consider the the experiment of flipping a coin 5 times, with corresponding sample space S.

- One random variable X on S is the total number of heads obtained. The value of X on the outcome $HHHHT$ is 4, while the value of X on the outcome $TTTTT$ is 0.
- Another random variable Y is the length of the longest consecutive run of heads. The value of Y on the outcome $HTHHT$ is 2, while the value on $THHHH$ is 4, and the value on $HTHHH$ is 3.
- Because random variables are merely functions on outcomes, we can define a random variable however we like simply by specifying all its values.
- We can define a third random variable Z to be 1 on the outcome $TTTTT$ and 0 on all other outcomes. (It identifies whether we flipped all tails.)

Discrete Random Variables, II

Examples: Consider the experiment of rolling a pair of dice, with corresponding sample space S.

- One random variable X is the largest die rolled. The value of X on the outcome $(1,4)$ is 4, while the value on $(6,6)$ is 6 and the value on $(1,5)$ is 5.
- Another random variable S is the sum of the outcomes. The value of S on the outcome $(1,4)$ is 5, while the value on $(6,6)$ is 12 and the value on $(1,5)$ is 6.

Discrete Random Variables, II

Examples: Consider the experiment of rolling a pair of dice, with corresponding sample space S.

- One random variable X is the largest die rolled. The value of X on the outcome $(1, 4)$ is 4, while the value on $(6, 6)$ is 6 and the value on $(1, 5)$ is 5.
- Another random variable S is the sum of the outcomes. The value of S on the outcome $(1, 4)$ is 5, while the value on $(6, 6)$ is 12 and the value on $(1, 5)$ is 6.
- A third random variable P is the product of the outcomes.
- A fourth random variable D is the difference of the outcomes.
- A fifth random variable R_1 is the number on the first die.
- A sixth random variable R_2 is the number on the second die.
- A seventh random variable is [insert your favorite idea here].

Discrete Random Variables, III

If X is a random variable, the set of outcomes on which X takes a particular value (or range of values) is a subset of the sample space, which is to say, it is an event.

Discrete Random Variables, III

If X is a random variable, the set of outcomes on which X takes a particular value (or range of values) is a subset of the sample space, which is to say, it is an event.

Thus, if we have a probability distribution on the sample space, we may therefore ask about quantities like the following:

- $P(X = n)$, the probability that X takes the value n.
- $P(X \geq 5)$, the probability that the value of X is at least 5.
- $P(2 < X < 4)$, the probability that the value of X is strictly between 2 and 4.

Discrete Random Variables, IV

A common way to tabulate all of this information is to make a list or table of all the possible values of X along with their corresponding probabilities, which we can package conveniently as a function:

Definition

If X is a random variable on the sample space S, then the function p_X such that $p_X(E) = P(X \in E)$ for any event E is called the probability density function (pdf) of X.

For discrete random variables, we will usually be interested in the values $p_X(a)$ on a real number a, which gives the probability that the random variable X takes the value a.

Discrete Random Variables, V

For discrete random variables with a small number of outcomes, we usually describe the probability density function using a table of values (i.e., by simply listing all the possible results and their probabilities).

- In certain situations, we can find a convenient formula for the values of the probability density function on arbitrary events, but in many other cases, the best we can do is simply to tabulate all the different values.
- We will retain the notation for the underlying probability density function for applications later, when we will manipulate general distributions.

Examples, I

Example: If a fair coin is flipped 4 times, let X be the random variable counting the total number of heads obtained.

1. Find the probability distribution for X.
2. Find $P(X = 2)$.
3. Find $P(X \geq 3)$.
4. Find $P(1 < X < 4)$.

Examples, II

Example: If a fair coin is flipped 4 times, let X be the random variable counting the total number of heads obtained.

1. Find the probability distribution for X.

Examples, II

Example: If a fair coin is flipped 4 times, let X be the random variable counting the total number of heads obtained.

1. Find the probability distribution for X.

- First, we list all possible outcomes and the corresponding value of X:

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

Examples, III

Example: If a fair coin is flipped 4 times, let X be the random variable counting the total number of heads obtained.

1. Find the probability distribution for X.

- Using the table on the previous slide, we can then easily find the probability distribution of X: we just need to compute the probability that X takes each particular value.
- The results are in the (much more compact) table below:

n	0	1	2	3	4
$P(X = n)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

Examples, IV

Example: If a fair coin is flipped 4 times, let X be the random variable counting the total number of heads obtained.

2. Find $P(X = 2)$.
3. Find $P(X \geq 3)$.
4. Find $P(1 < X < 4)$.

n	0	1	2	3	4
$P(X = n)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

Examples, IV

Example: If a fair coin is flipped 4 times, let X be the random variable counting the total number of heads obtained.

2. Find $P(X = 2)$.
3. Find $P(X \geq 3)$.
4. Find $P(1 < X < 4)$.

n	0	1	2	3	4
$P(X = n)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

- We can now just read off these values from the table: $P(X = 2) = 6/16$, $P(X \geq 3) = (4/16) + (1/16) = 5/16$, and $P(1 < X < 4) = (6/16) + (4/16) = 10/16$.

Examples, V

Example: If three loaded dice each with a probability $2/7$ of landing 6 and $1/7$ probability for each other outcome are rolled, let X be the random variable counting the total number of sixes obtained.

1. Find the probability distribution for X.
2. Find $P(X = 1)$.
3. Find $P(X \geq 1)$.
4. Find $P(X \neq 2)$.

Examples, V

Example: If three loaded dice each with a probability $2/7$ of landing 6 and $1/7$ probability for each other outcome are rolled, let X be the random variable counting the total number of sixes obtained.

1. Find the probability distribution for X.
2. Find $P(X = 1)$.
3. Find $P(X \geq 1)$.
4. Find $P(X \neq 2)$.

- As in the last example, we simply need to tabulate the probability of each possible value of X, and then use the resulting probability distribution table to answer the other questions.

Examples, VI

Example: If three loaded dice each with a probability $2/7$ of landing 6 and $1/7$ probability for each other outcome are rolled, let X be the random variable counting the total number of sixes obtained.

1. Find the probability distribution for X.

Examples, VI

<u>Example</u>: If three loaded dice each with a probability $2/7$ of landing 6 and $1/7$ probability for each other outcome are rolled, let X be the random variable counting the total number of sixes obtained.

1. Find the probability distribution for X.

- We compute the probability that X takes each possible value.
- $X = 0$ means no dice roll sixes: $(5/7)^3 = 125/343$.
- $X = 1$ means one die rolls a six: $3 \cdot (2/7) \cdot (5/7)^2 = 150/343$.
- $X = 2$ means two dice roll a six: $3 \cdot (2/7)^2 \cdot (5/7) = 60/343$.
- $X = 3$ means all three dice roll sixes: $(2/7)^3 = 8/343$.
- This yields the following probability distribution table:

n	0	1	2	3
$P(X = n)$	$125/343$	$150/343$	$60/343$	$8/343$

Examples, VII

Example: If three loaded dice each with a probability $2/7$ of landing 6 and $1/7$ probability for each other outcome are rolled, let X be the random variable counting the total number of sixes obtained.

2. Find $P(X = 1)$.
3. Find $P(X \geq 1)$.
4. Find $P(X \neq 2)$.

Examples, VII

Example: If three loaded dice each with a probability $2/7$ of landing 6 and $1/7$ probability for each other outcome are rolled, let X be the random variable counting the total number of sixes obtained.

2. Find $P(X = 1)$.
3. Find $P(X \geq 1)$.
4. Find $P(X \neq 2)$.

n	0	1	2	3
$P(X = n)$	$125/343$	$150/343$	$60/343$	$8/343$

- We can now easily compute each of the given probabilities by reading the appropriate entries from the table.
- We see $P(X = 1) = 150/343$,
  $P(X \geq 1) = (150/343) + (60/343) + (8/343) = 218/343$, and
  $P(X \neq 2) = (125/343) + (150/343) + (8/343) = 283/343$.

Examples, VIII

Example: If two standard 6-sided dice are rolled, let S be the random variable giving the sum of the outcomes.

1. Find the probability distribution for S.
2. Find $P(S = 7)$.
3. Find $P(3 < S < 8)$.
4. Find $P(S \geq 10)$.

Examples, IX

Example: If two standard 6-sided dice are rolled, let S be the random variable giving the sum of the outcomes.

1. Find the probability distribution for S.

Examples, IX

<u>Example</u>: If two standard 6-sided dice are rolled, let S be the random variable giving the sum of the outcomes.

1. Find the probability distribution for S.

- We can count the possible outcomes based on the associated value of S that they yield.
- For the respective values $S = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ there are $1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1$ outcomes yielding that value of S. This yields the table below:

n	2	3	4	5	6	7	8	9	10	11	12
$P(S = n)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Examples, X

<u>Example</u>: If two standard 6-sided dice are rolled, let S be the random variable giving the sum of the outcomes.

2. Find $P(S = 7)$.
3. Find $P(3 < S < 8)$.
4. Find $P(S \geq 10)$.

Examples, X

Example: If two standard 6-sided dice are rolled, let S be the random variable giving the sum of the outcomes.

2. Find $P(S = 7)$.
3. Find $P(3 < S < 8)$.
4. Find $P(S \geq 10)$.

n	2	3	4	5	6	7	8	9	10	11	12
$P(S = n)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- Reading the appropriate entries from the table yields $P(S = 7) = 6/36 = 1/6$, $P(3 < S < 8) = (3/36) + (4/36) + (5/36) + (6/36) = 1/2$, and $P(S \geq 10) = (3/36) + (2/36) + (1/36) = 1/6$.

Bernoulli Trials, I

A particularly simple random variable is the one that identifies whether a specific event has occurred:

Definition

If E is any event, we define the Bernoulli random variable for E to be $X_E = \begin{cases} 1 & \text{if } E \text{ occurs} \\ 0 & \text{if } E \text{ does not occur} \end{cases}$.

- The name for this random variable comes from the idea of a Bernoulli trial, which is an experiment having only two possible outcomes, success (with probability p) and failure (with probability $1 - p$).
- We think of E as being the event of success, while E^c is the event of failure.

Bernoulli Trials, II

Many experiments consist of a sequence of independent Bernoulli trials, in which the outcome of each trial is independent from the outcomes of all of the others. Here are a few examples:

- Flipping a fair or unfair coin 10 times and testing whether heads is obtained for each flip.
- Recording the results (hit or no hit) from a sequence of 20 of a baseball player's at-bats.
- Recording the results (positive or negative) from a random screening of 250 people for a disease.
- Rolling a fair or unfair die 30 times and recording whether or not a 6 was rolled.

Binomial Distributions, I

We can describe explicitly the probability distribution of the random variable X giving the total number of successes in a sequence of Bernoulli trials:

Proposition (The Binomial Distribution)

Let X be the random variable representing the total number of successes obtained by performing n independent Bernoulli trials each of which has a success probability p. Then the probability distribution of X is the binomial distribution, in which $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for integers k with $0 \le k \le n$, and $P(X = k) = 0$ for other k.

The binomial distribution is so named because of the presence of the binomial coefficients $\binom{n}{k}$.

Binomial Distributions, II

Proof:

- From our results on counting combinations with binomial coefficients, we can see that there are $\binom{n}{k}$ ways to choose k trials yielding success out of a total of n.
- Furthermore, since all of the trials are independent, the probability of obtaining any given pattern of k successes and $n-k$ failures is equal to $p^k(1-p)^{n-k}$.
- Thus, since the probability of obtaining any given one of the $\binom{n}{k}$ outcomes with exactly k successes is $p^k(1-p)^{n-k}$, the probability of obtaining exactly k successes is $\binom{n}{k}p^k(1-p)^{n-k}$, as claimed.

Binomial Distributions, III

Example: A baseball player's batting average is 0.378, meaning that she has a probability of 0.378 of getting a hit on any given at-bat, independently of any other at-bat. Find the probabilities of the following events:

1. In her first 100 at-bats, she gets exactly 37 hits.
2. In her first 100 at-bats, she gets exactly 40 hits.
3. In her first 100 at-bats, she gets exactly 50 hits.

Binomial Distributions, III

Example: A baseball player's batting average is 0.378, meaning that she has a probability of 0.378 of getting a hit on any given at-bat, independently of any other at-bat. Find the probabilities of the following events:

1. In her first 100 at-bats, she gets exactly 37 hits.
2. In her first 100 at-bats, she gets exactly 40 hits.
3. In her first 100 at-bats, she gets exactly 50 hits.

- We can view each at-bat as an independent Bernoulli trial (with a hit being considered a success), so the total number of hits will be binomially distributed with $n = 100$ and $p = 0.378$.

Binomial Distributions, IV

Example: A baseball player's batting average is 0.378, meaning that she has a probability of 0.378 of getting a hit on any given at-bat, independently of any other at-bat. Find the probabilities of the following events:

1. In her first 100 at-bats, she gets exactly 37 hits.
2. In her first 100 at-bats, she gets exactly 40 hits.
3. In her first 100 at-bats, she gets exactly 50 hits.

Binomial Distributions, IV

Example: A baseball player's batting average is 0.378, meaning that she has a probability of 0.378 of getting a hit on any given at-bat, independently of any other at-bat. Find the probabilities of the following events:

1. In her first 100 at-bats, she gets exactly 37 hits.
2. In her first 100 at-bats, she gets exactly 40 hits.
3. In her first 100 at-bats, she gets exactly 50 hits.

- Thus, since the distribution is binomial, the probability of k hits will be $\binom{100}{k} \cdot 0.378^k \cdot 0.622^{100-k}$.
- The probability of 37 hits is $\binom{100}{37} \cdot 0.378^{37} \cdot 0.622^{63} \approx 8.13\%$.
- The probability of 40 hits is $\binom{100}{40} \cdot 0.378^{40} \cdot 0.622^{60} \approx 7.33\%$.
- The probability of 50 hits is $\binom{100}{50} \cdot 0.378^{50} \cdot 0.622^{50} \approx 0.37\%$.

Binomial Distributions, V

Example: An unfair coin with probability $3/5$ of coming up heads on each flip is flipped 60 times. Find the probabilities of

1. Obtaining 30 heads.
2. Obtaining 36 heads.
3. Obtaining 40 heads.
4. Obtaining between 30 and 36 heads (inclusive).

Binomial Distributions, V

Example: An unfair coin with probability $3/5$ of coming up heads on each flip is flipped 60 times. Find the probabilities of

1. Obtaining 30 heads.
2. Obtaining 36 heads.
3. Obtaining 40 heads.
4. Obtaining between 30 and 36 heads (inclusive).

- The total number of heads will be binomially distributed with $n = 60$ and $p = 3/5$, so the probability of obtaining exactly k heads will be $\binom{60}{k}(3/5)^k(2/5)^{60-k}$.
- The probability of 30 heads is $\binom{60}{30}(3/5)^{30}(2/5)^{30} \approx 3.01\%$.
- The probability of 36 heads is $\binom{60}{36}(3/5)^{36}(2/5)^{24} \approx 10.46\%$.
- The probability of 40 heads is $\binom{60}{40}(3/5)^{40}(2/5)^{20} \approx 6.16\%$.
- The probability of 30 to 36 heads (inclusive) is $\binom{60}{30}(3/5)^{30}(2/5)^{30} + \binom{60}{31}(3/5)^{31}(2/5)^{29} + \cdots + \binom{60}{36}(3/5)^{36}(2/5)^{24} \approx 50.44\%$.

Functions of Random Variables, I

If we have a random variable X defined on the sample space, then since X is a function on outcomes, we can define various new random variables in terms of X.

- If g is any real-valued function, we can define a new random variable $g(X)$ by evaluating g on all of the results of X. Some possibilities include $g(X) = 2X$, which doubles every value of X, or $g(X) = X^2$, which squares every value of X.
- More generally, if we have a collection of random variables $X_1, X_2, \ldots, X_n$ defined on the same sample space, we can construct new functions in terms of them, such as the sum $X_1 + X_2 + \cdots + X_n$ that returns the sum of the values of $X_1, \ldots, X_n$ on any given outcome.

Functions of Random Variables, II

Example: Suppose X is the random variable with probability distribution given below. Find the probability distributions for $2X+1$ and for X^2.

n	1	3	4	8
$P(X=n)$	0.2	0.5	0.1	0.2

Functions of Random Variables, II

Example: Suppose X is the random variable with probability distribution given below. Find the probability distributions for $2X+1$ and for X^2.

n	1	3	4	8
$P(X=n)$	0.2	0.5	0.1	0.2

- If $X = 1, 3, 4, 8$ then $2X+1 = 3, 7, 9, 17$ respectively, so we obtain the following distribution for $2X+1$:

n	3	7	9	17
$P(2X+1=n)$	0.2	0.5	0.1	0.2

Functions of Random Variables, II

Example: Suppose X is the random variable with probability distribution given below. Find the probability distributions for $2X + 1$ and for X^2.

n	1	3	4	8
$P(X = n)$	0.2	0.5	0.1	0.2

- If $X = 1, 3, 4, 8$ then $2X + 1 = 3, 7, 9, 17$ respectively, so we obtain the following distribution for $2X + 1$:

n	3	7	9	17
$P(2X + 1 = n)$	0.2	0.5	0.1	0.2

- In the same way, here is the distribution for X^2:

n	1	9	16	64
$P(X^2 = n)$	0.2	0.5	0.1	0.2

Expected Value, I

If we repeat an experiment many times and record the different values of a random variable X each time, a useful statistic summarizing the outcomes is the average value of the outcomes.

- We would like a way to describe the "average value" of a random variable X.
- Suppose that the sample space has outcomes $s_1, s_2, \ldots, s_n$ on which the random variable X takes on the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$, where $p_1 + \cdots + p_n = 1$.
- Under our interpretation of these probabilities as giving the relative frequencies of events when we repeat an experiment, if we perform the experiment N times (N large), we should obtain the outcome s_i (on which $X = x_i$) approximately $p_i N$ times for each $1 \le i \le n$. The average value is then $\dfrac{(p_1 N)x_1 + (p_2 N)x_2 + \cdots + (p_n N)x_n}{N} = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$.

Expected Value, II

We can use the calculation on the previous slide to give a definition of the "average value" for an arbitrary discrete random variable:

Definition

If X is a discrete random variable, the expected value of X, written $E(X)$, is the sum $E(X) = \sum_{s_i \in S} P(s_i)X(s_i)$ over all outcomes s_i in the sample space S.

In words, the expected value $E(X)$ is the average of the values that X takes on the outcomes in the sample space, weighted by the probability of each outcome.

The expected value is also sometimes called the mean or the average value of X, and is often also written as μ_X ("mu-X") or as $\overline{X}$ ("X-bar").

Expected Value, III

<u>Example</u>: If a fair coin is flipped once, the expected value of the random variable X giving the number of total heads obtained is equal to $E(X) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$, because there are two possible outcomes, 0 heads and 1 head, each of probability $1/2$.

In this example, notice that since the coin comes up heads half the time, it is quite reasonable to say that the average number of heads per flip is $1/2$, which is exactly what this expected value calculation gives.

Expected Value, IV

Examples:

- If an unfair coin with a probability $2/3$ of landing heads is flipped once, the expected value of the random variable X giving the number of total heads obtained is equal to $E(X) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$, because there are two possible outcomes, 0 heads and 1 head, of respective probabilities $1/3$ and $2/3$.

Expected Value, IV

Examples:

- If an unfair coin with a probability $2/3$ of landing heads is flipped once, the expected value of the random variable X giving the number of total heads obtained is equal to $E(X) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot 1 = \frac{2}{3}$, because there are two possible outcomes, 0 heads and 1 head, of respective probabilities $1/3$ and $2/3$.
- If a standard 6-sided die is rolled once, the expected value of the random variable X giving the result is equal to $E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{7}{2}$, because each of the 6 possible outcomes 1,2,3,4,5,6 has probability $1/6$ of occurring.

Expected Value, V

Example: Find $E(X)$ and $E(Y)$ if X and Y are the discrete random variables whose probability distributions appear below.

n	1	3	4	8
$P(X = n)$	0.2	0.5	0.1	0.2
$P(Y = n)$	0.1	0	0.4	0.5

Expected Value, V

<u>Example</u>: Find $E(X)$ and $E(Y)$ if X and Y are the discrete random variables whose probability distributions appear below.

n	1	3	4	8
$P(X = n)$	0.2	0.5	0.1	0.2
$P(Y = n)$	0.1	0	0.4	0.5

- We simply apply the formula in each case.
- For X we get $E(X) = 0.2 \cdot 1 + 0.5 \cdot 3 + 0.1 \cdot 4 + 0.2 \cdot 8 = 3.7$.
- For Y we get $E(Y) = 0.1 \cdot 1 + 0 \cdot 3 + 0.4 \cdot 4 + 0.5 \cdot 8 = 5.7$.

Expected Value, VI

The expected value of a discrete random variable can be infinite or even not defined at all.

Examples:

- If X is the discrete random variable whose value is 2^n occurring with probability 2^{-n} for $n \geq 1$, then its expected value is
$E(X) = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + 16 \cdot \frac{1}{16} + \cdots = 1 + 1 + 1 + 1 + \cdots = \infty.$
- If Y is the discrete random variable whose value is $(-2)^n$ occurring with the probability 2^{-n} for $n \geq 1$, then its expected value is the sum
$2 \cdot (-\frac{1}{2}) + 4 \cdot \frac{1}{4} + 8 \cdot (-\frac{1}{8}) + 16 \cdot \frac{1}{16} + \cdots = -1 + 1 - 1 + 1 - \cdots.$
This sum does not converge, so the expected value of this random variable is not defined.

Expected Value, VII

A common application of expected value is to calculate the expected winnings from a game of chance.

- In such scenarios, we view the outcomes from the game of chance as our sample space, and the random variable we are studying represents the total amount won or lost.
- The expected value of this random variable measures the average amount one should expect to win (or lose!) per game, upon playing the game many times.

Expected Value, VIII

Example: In one version of a "Pick 3" lottery, a single entry ticket costs \$1. In this lottery, 3 single digits are drawn at random, and a ticket must match all 3 digits in the correct order to win the \$500 prize. What is the expected value of one ticket for this lottery?

Expected Value, VIII

Example: In one version of a "Pick 3" lottery, a single entry ticket costs \$1. In this lottery, 3 single digits are drawn at random, and a ticket must match all 3 digits in the correct order to win the \$500 prize. What is the expected value of one ticket for this lottery?

- From the description, we can see that there is a $1/1000$ probability of winning the prize and a $999/1000$ probability of winning nothing.
- Since winning the prize nets a total of \$499 (the prize minus the \$1 entry fee), and winning nothing nets a total of $-\$1$, the expected value of the random variable giving the net winnings is equal to $\frac{1}{1000}(\$499) + \frac{999}{1000}(-\$1) = -\$0.50$.
- The expected value of $-\$0.50$, in this case, indicates that if one plays this lottery many times, on average one should expect to lose 50 cents on every ticket.

Summary

We established Bayes's formula and described some of its uses, such as analyzing the prosecutor's fallacy.

We discussed discrete random variables and gave some examples, including the binomial distribution.

We introduced the expected value of a discrete random variable.

Next lecture: Discrete random variables (part 2).