ISTA 116: Statistical Foundations for the Information Age

Discrete Random Variables

31 October 2011

Outline

Outline

Random Variables

- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).

Random Variables

- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., sex, weight, cancer status, etc.) using a **random variable**

Random Variables

- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., sex, weight, cancer status, etc.) using a **random variable**

(Definition: Random Variable)

A **random variable** connects each element of the sample space to a value or quantity of interest.

Random Variables

- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., sex, weight, cancer status, etc.) using a **random variable**

(Definition: Random Variable)

A **random variable** connects each element of the sample space to a value or quantity of interest.

- X: People $\mapsto$ Their Sex

Random Variables

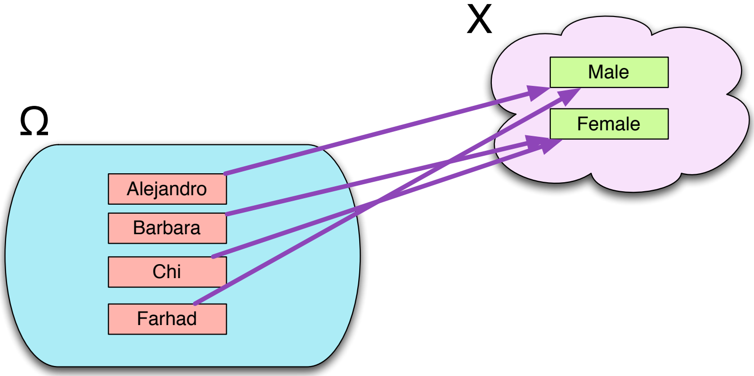
- A single random sample may have more than one characteristic that we can observe (i.e., it may be bi-/multivariate data).
- We can represent each characteristic (e.g., sex, weight, cancer status, etc.) using a **random variable**

(Definition: Random Variable)

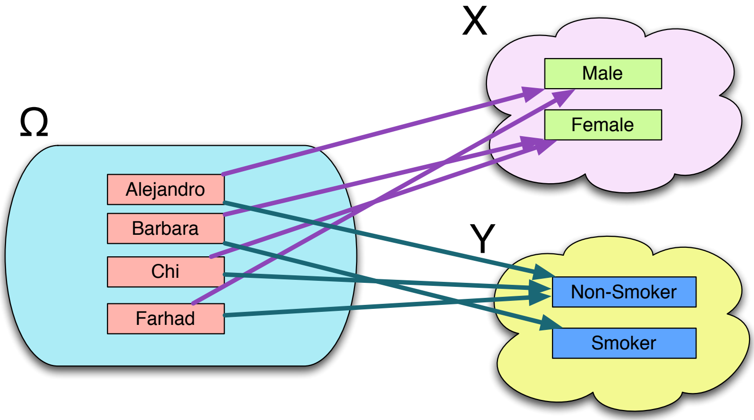
A **random variable** connects each element of the sample space to a value or quantity of interest.

- X: People $\mapsto$ Their Sex
- Y: Sequences of coin flips $\mapsto$ Number of Heads

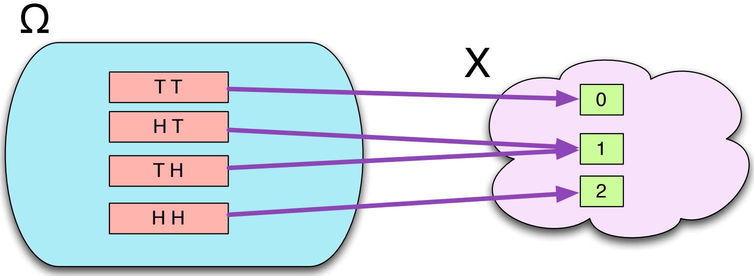
Examples of Random Variables



Examples of Random Variables



Examples of Random Variables



Outline

Types of Random Variables

- We can classify random variables based on the types of values they take on.

Types of Random Variables

- We can classify random variables based on the types of values they take on.
- **Discrete** random variables take on discrete values (e.g., categories or integers)

Types of Random Variables

- We can classify random variables based on the types of values they take on.
- **Discrete** random variables take on discrete values (e.g., categories or integers)
- **Continuous** random variables take on continuous values (e.g., real numbers).

Outline

Discrete Distributions

- When a random variable is discrete, its **distribution** is characterized by the probabilities assigned to each distinct value.

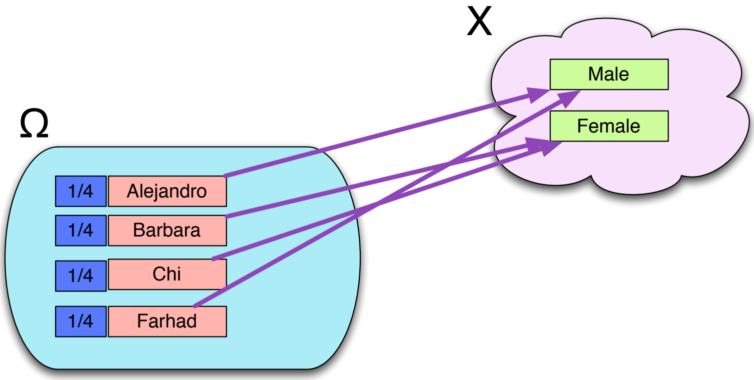
Discrete Distributions

- When a random variable is discrete, its **distribution** is characterized by the probabilities assigned to each distinct value.
- These probabilities are determined by the probabilities on the sample space itself

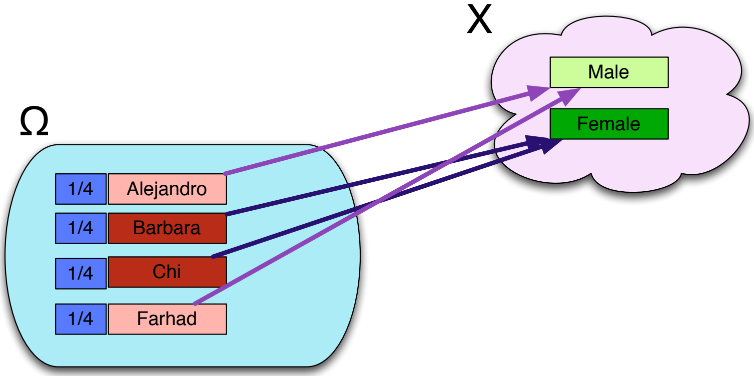
Discrete Distributions

- When a random variable is discrete, its **distribution** is characterized by the probabilities assigned to each distinct value.
- These probabilities are determined by the probabilities on the sample space itself
- If the sample space is a finite population and we make a simple random draw, then the probability of a value is the proportion of individual outcomes assigned to it.

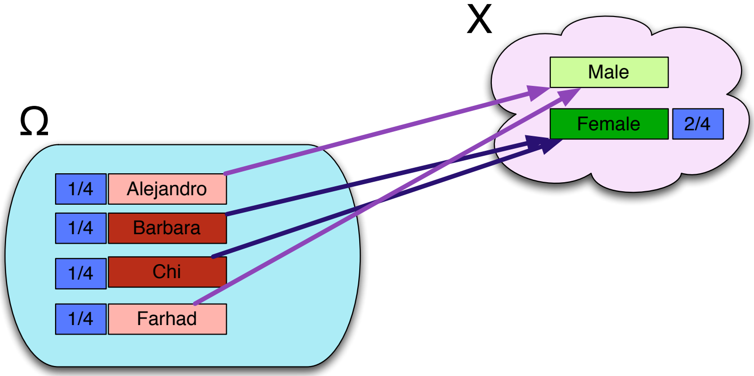
The Distribution of a Random Variable



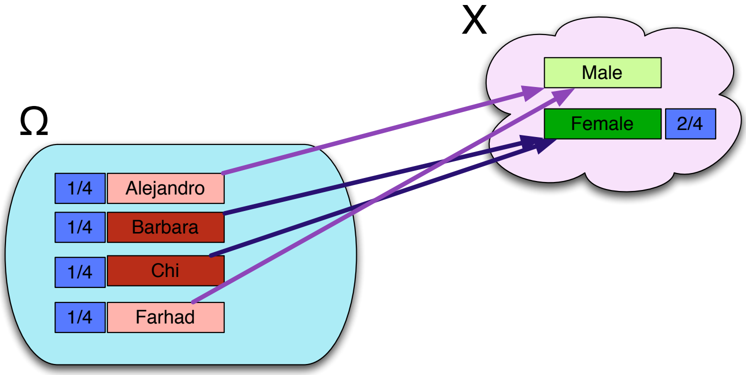
The Distribution of a Random Variable



The Distribution of a Random Variable

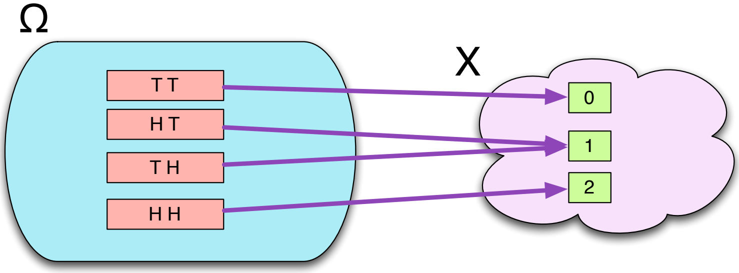


The Distribution of a Random Variable



x	Male	Female
$P(X = x)$	$1/2$	$1/2$

The Distribution of a Random Variable



x	0	1	2
$P(X = x)$			

Properties of Discrete Distributions

- Note that each value of a discrete random variable corresponds to an event in the original sample space.

Properties of Discrete Distributions

- Note that each value of a discrete random variable corresponds to an event in the original sample space.
- The probability associated with the value is the probability associated with the event.

Properties of Discrete Distributions

- Note that each value of a discrete random variable corresponds to an event in the original sample space.
- The probability associated with the value is the probability associated with the event.
- Moreover, every outcome in the sample space is associated with exactly one value of the random variable.

Properties of Discrete Distributions

- Note that each value of a discrete random variable corresponds to an event in the original sample space.
- The probability associated with the value is the probability associated with the event.
- Moreover, every outcome in the sample space is associated with exactly one value of the random variable.
- Therefore, the values of a discrete random variable give us a set of **disjoint** events whose **union** is the entire sample space.

Properties of Discrete Distributions

- What must be true of the set of probabilities then?

(Properties of Discrete Distributions)

Properties of Discrete Distributions

- What must be true of the set of probabilities then?

(Properties of Discrete Distributions)

1. For every x in the range of X, $P(X = x) \geq 0$.

Properties of Discrete Distributions

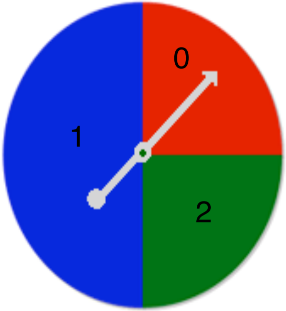
- What must be true of the set of probabilities then?

(Properties of Discrete Distributions)

1. For every x in the range of X, $P(X = x) \geq 0$.
2. $$\sum_{x \in \text{Range}(X)} P(X = x) = 1$$

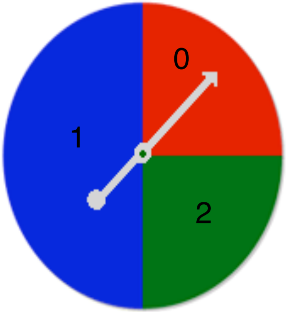
Outline

Visualizing Discrete Distributions



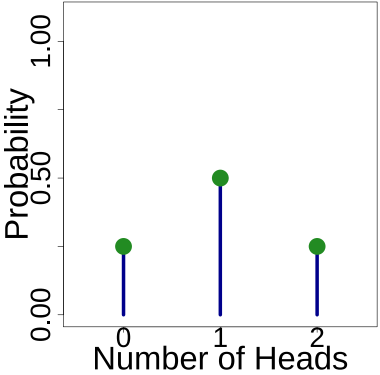
- Very simple distributions can be visualized with a pie chart.
- Can imagine a spinner mechanism that lands on a slice according to its probability.

Visualizing Discrete Distributions



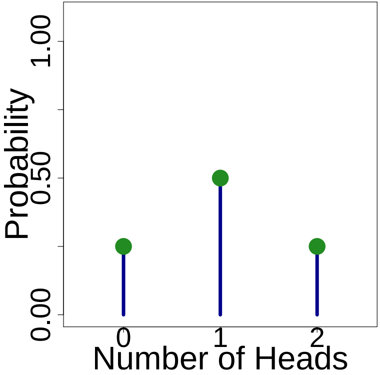
- Very simple distributions can be visualized with a pie chart.
- Can imagine a spinner mechanism that lands on a slice according to its probability.
- But, like pie charts, this is limited in its ability to convey information.

The Spike Plot



- An alternative is the **spike plot**

The Spike Plot



- An alternative is the **spike plot**
- Like a bar plot, but with probabilities, instead of frequencies or proportions, on the y-axis.

Outline

Probability Mass Function

- The **distribution** of a random variable is characterized by the set of values and their probabilities.

Probability Mass Function

- The **distribution** of a random variable is characterized by the set of values and their probabilities.
- For finite sets of values, can think of a table.

Probability Mass Function

- The **distribution** of a random variable is characterized by the set of values and their probabilities.
- For finite sets of values, can think of a table.
- One way to use such a table is to start with a value and *look up* its probability.

Probability Mass Function

- The **distribution** of a random variable is characterized by the set of values and their probabilities.
- For finite sets of values, can think of a table.
- One way to use such a table is to start with a value and *look up* its probability.
- This process is characterized by the **probability mass function**, which takes a value and returns its probability.

Probability Mass Function

(Definition: The Probability Mass Function)

A discrete random variable, X, can be characterized by its **probability mass function**, f_X, which takes values and returns probabilities:

$$f_X(x) = P(X = x)$$

Probability Mass Function

(Definition: The Probability Mass Function)

A discrete random variable, X, can be characterized by its **probability mass function**, f_X, which takes values and returns probabilities:

$$f_X(x) = P(X = x)$$

- In the most general case, we just have to consult a table.

Probability Mass Function

(Definition: The Probability Mass Function)

A discrete random variable, X, can be characterized by its **probability mass function**, f_X, which takes values and returns probabilities:

$$f_X(x) = P(X = x)$$

- In the most general case, we just have to consult a table.
- However, we will see examples later of PMFs that have algebraic expressions.

The Cumulative Distribution Function

- Often times we are interested in the probability of falling in some *range* of values.

The Cumulative Distribution Function

- Often times we are interested in the probability of falling in some *range* of values.
- For this purpose, we can use the **cumulative distribution function** (or CDF), which gives the "accumulated probability" up to a particular value.

The Cumulative Distribution Function

- Often times we are interested in the probability of falling in some *range* of values.
- For this purpose, we can use the **cumulative distribution function** (or CDF), which gives the "accumulated probability" up to a particular value.

(Definition: The Cumulative Distribution Function)

A random variable, X, can be characterized by its **cumulative distribution function**, F_X, which takes values and returns *cumulative* probabilities:

$$F_X(x) = P(X \leq x)$$

The Cumulative Distribution Function

- How can we calculate $F_X(x)$ from the distribution table?

The Cumulative Distribution Function

- How can we calculate $F_X(x)$ from the distribution table?
- How would we calculate $P(X > x)$?

The Cumulative Distribution Function

- How can we calculate $F_X(x)$ from the distribution table?
- How would we calculate $P(X > x)$?
- How about $P(X \geq x)$?

The Cumulative Distribution Function

- How can we calculate $F_X(x)$ from the distribution table?
- How would we calculate $P(X > x)$?
- How about $P(X \geq x)$?
- How would we calculate $P(a < X \leq b)$?

The Cumulative Distribution Function

- How can we calculate $F_X(x)$ from the distribution table?
- How would we calculate $P(X > x)$?
- How about $P(X \geq x)$?
- How would we calculate $P(a < X \leq b)$?
- How about $P(a \leq X \leq b)$?

The Cumulative Distribution Function

- How can we calculate $F_X(x)$ from the distribution table?
- How would we calculate $P(X > x)$?
- How about $P(X \geq x)$?
- How would we calculate $P(a < X \leq b)$?
- How about $P(a \leq X \leq b)$?
- $P(a \leq X < b)$?

Outline

Summarizing Random Variables

- As with data, it is useful to characterize the center and spread of a probability distribution.

Summarizing Random Variables

- As with data, it is useful to characterize the center and spread of a probability distribution.
- Most of the measures we've seen can be computed; but the most common are the **mean** and **variance**.

Outline

Expected Value

- The mean of a random variable is also called its **expected value**.

Expected Value

- The mean of a random variable is also called its **expected value**.
- As with a sample mean, it represents an average over the possible values; but it is **weighted** by the probabilities.

Example: Mean Number of Heads

- To compute the mean number of heads in two coin tosses:

$$\mu_X = E(X) \quad =$$

Example: Mean Number of Heads

- To compute the mean number of heads in two coin tosses:

$$\begin{aligned} \mu_X = E(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) \\ &\quad + 2 \times P(X = 2) \end{aligned}$$

Example: Mean Number of Heads

- To compute the mean number of heads in two coin tosses:

$$\begin{aligned} \mu_X = E(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) \\ &\quad + 2 \times P(X = 2) \\ &= 0 \times 1/4 + 1 \times 1/2 + 2 \times 1/4 \end{aligned}$$

Example: Mean Number of Heads

- To compute the mean number of heads in two coin tosses:

$$\begin{aligned}\mu_X = E(X) &= 0 \times P(X = 0) + 1 \times P(X = 1) \\ &\quad + 2 \times P(X = 2) \\ &= 0 \times {}^1/_4 + 1 \times {}^1/_2 + 2 \times {}^1/_4 \\ &= {}^1/_2 + {}^2/_4 = 1\end{aligned}$$

Expected Value of a Discrete Random Variable

- In general, we have:

(Definition: Expected Value of a Discrete Random Variable)

$$\mu_X = E(X) = \sum_{x \in \text{Range}(X)} xP(X = x)$$

Expected Value of a Discrete Random Variable

- In general, we have:

(Definition: Expected Value of a Discrete Random Variable)

$$\mu_X = E(X) = \sum_{x \in \mathrm{Range}(X)} xP(X = x)$$

- If X is a numeric variable with values from 0 to some number n, we have

(Expected Value of Finite, Integer-Valued Random Variable)

$$\mu_X = E(X) = \sum_{x=0}^{n} xP(X = x)$$

Outline

Variance of a Discrete Random Variable

- The variance is the expected squared deviation:

(Definition: Variance of a Discrete Random Variable)

$$\sigma_X^2 = E((X - \mu_X)^2) = \sum_{x \in \text{Range}(X)} (x - \mu_X)^2 P(X = x)$$

Variance of a Discrete Random Variable

- The variance is the expected squared deviation:

(Definition: Variance of a Discrete Random Variable)

$$\sigma_X^2 = E((X - \mu_X)^2) = \sum_{x \in \text{Range}(X)} (x - \mu_X)^2 P(X = x)$$

- If X is a numeric variable with values from 0 to some number n, we have

(Variance of Finite, Integer-Valued Random Variable)

$$\sigma_X^2 = E((X - \mu_X)^2) = \sum_{x=0}^{n} (x - \mu_x)^2 P(X = x)$$

Example: Variance of Number of Heads

- To compute the variance in the number of heads in two coin tosses:

$$\sigma_X^2 =$$

Example: Variance of Number of Heads

- To compute the variance in the number of heads in two coin tosses:

$$\begin{aligned}\sigma_X^2 &= (0-\mu_X)^2 \times P(X=0) + (1-\mu_X)^2 \times P(X=1) \\ &\quad + (2-\mu_X)^2 \times P(X=2)\end{aligned}$$

Example: Variance of Number of Heads

- To compute the variance in the number of heads in two coin tosses:

$$\begin{aligned}
\sigma_X^2 &= (0-\mu_X)^2 \times P(X=0) + (1-\mu_X)^2 \times P(X=1) \\
&\quad + (2-\mu_X)^2 \times P(X=2) \\
&= (0-1)^2 \times {}^1/_4 + (1-1)^2 \times {}^1/_2 \\
&\quad + (2-1)^2 \times {}^1/_4
\end{aligned}$$

Example: Variance of Number of Heads

- To compute the variance in the number of heads in two coin tosses:

$$\begin{aligned}
\sigma_X^2 &= (0-\mu_X)^2 \times P(X=0) + (1-\mu_X)^2 \times P(X=1) \\
&\quad +(2-\mu_X)^2 \times P(X=2) \\
&= (0-1)^2 \times {}^1\!/_4 + (1-1)^2 \times {}^1\!/_2 \\
&\quad +(2-1)^2 \times {}^1\!/_4 \\
&= {}^1\!/_4 + {}^1\!/_4 = {}^1\!/_2
\end{aligned}$$

Standard Deviation

- The standard deviation is then ____________________

Outline

Named Distributions

- Any discrete set of values, along with associated probabilities, forms a discrete probability distribution
  - as long as ________________
  - and ________________

Named Distributions

- Any discrete set of values, along with associated probabilities, forms a discrete probability distribution
  - as long as ________________
  - and ________________
- However, some sets show up so often, they are given names.

Named Distributions

- Any discrete set of values, along with associated probabilities, forms a discrete probability distribution
  - as long as ______________
  - and ______________
- However, some sets show up so often, they are given names.
- We will look at three of these:
  - The Bernoulli Distribution
  - The Discrete Uniform Distribution
  - The Binomial Distribution

Outline

The Bernoulli Distribution

- The simplest possible probability distribution is the **Bernoulli Distribution**, named after Jakob Bernoulli, who is also credited with discovering the constant e.

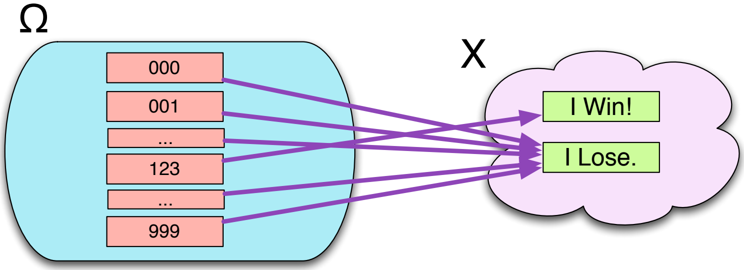
The Bernoulli Distribution

- The simplest possible probability distribution is the **Bernoulli Distribution**, named after Jakob Bernoulli, who is also credited with discovering the constant e.
- Any random variable with only two possible outcomes has a Bernoulli distribution with "success" probability p (where we choose one of the outcomes to call a "success").

The Bernoulli Distribution

- The simplest possible probability distribution is the **Bernoulli Distribution**, named after Jakob Bernoulli, who is also credited with discovering the constant e.
- Any random variable with only two possible outcomes has a Bernoulli distribution with "success" probability p (where we choose one of the outcomes to call a "success").
- Examples:
  - Flipping a single coin
  - Playing the lottery
  - Getting or not getting cancer
  - Winning or losing a baseball game
  - Graduating or not graduating from college
  - etc. etc.

Example: Playing the Lottery



x	Lose	Win
$P(X = x)$		

The Bernoulli Distribution

- If a "success" has probability p, then a failure has probability ________________

The Bernoulli Distribution

- If a "success" has probability p, then a failure has probability ________________
- Typically, we label the success outcome as "1", and the failure outcome as "0".
- We can then compute a mean and variance.

The Bernoulli Distribution

- If a "success" has probability p, then a failure has probability ________________
- Typically, we label the success outcome as "1", and the failure outcome as "0".
- We can then compute a mean and variance.
- In terms of the **parameter** p, the mean of a Bernoulli distribution is ________________

The Bernoulli Distribution

- If a “success” has probability p, then a failure has probability ________________
- Typically, we label the success outcome as “1”, and the failure outcome as “0”.
- We can then compute a mean and variance.
- In terms of the **parameter** p, the mean of a Bernoulli distribution is ________________
- The variance is ________________

The Bernoulli Distribution

- If a “success” has probability p, then a failure has probability ______________
- Typically, we label the success outcome as “1”, and the failure outcome as “0”.
- We can then compute a mean and variance.
- In terms of the **parameter** p, the mean of a Bernoulli distribution is ______________
- The variance is ______________

$$\sigma^2 = (0-p)^2(1-p) + (1-p)^2 p$$

The Bernoulli Distribution

- If a "success" has probability p, then a failure has probability ______________
- Typically, we label the success outcome as "1", and the failure outcome as "0".
- We can then compute a mean and variance.
- In terms of the **parameter** p, the mean of a Bernoulli distribution is ______________
- The variance is ______________

$$\begin{aligned} \sigma^2 &= (0-p)^2(1-p) + (1-p)^2 p \\ &= p^2(1-p) + (1-2p+p^2)p \end{aligned}$$

The Bernoulli Distribution

- If a "success" has probability p, then a failure has probability ______________
- Typically, we label the success outcome as "1", and the failure outcome as "0".
- We can then compute a mean and variance.
- In terms of the **parameter** p, the mean of a Bernoulli distribution is ______________
- The variance is ______________

$$\begin{aligned} \sigma^2 &= (0-p)^2(1-p) + (1-p)^2 p \\ &= p^2(1-p) + (1-2p+p^2)p \\ &= p^2 - p^3 + p - 2p^2 + p^3 \end{aligned}$$

The Bernoulli Distribution

- If a "success" has probability p, then a failure has probability ____________
- Typically, we label the success outcome as "1", and the failure outcome as "0".
- We can then compute a mean and variance.
- In terms of the **parameter** p, the mean of a Bernoulli distribution is ____________
- The variance is ____________

$$\begin{aligned}
\sigma^2 &= (0-p)^2(1-p) + (1-p)^2 p \\
&= p^2(1-p) + (1-2p+p^2)p \\
&= p^2 - p^3 + p - 2p^2 + p^3 \\
&= p - p^2 = p(1-p)
\end{aligned}$$

Outline

The Discrete Uniform Distribution

- Another common distribution is the **discrete uniform distribution**.

The Discrete Uniform Distribution

- Another common distribution is the **discrete uniform distribution**.
- Here, we have n possible outcomes (labeled 1 through n) which are all equally likely.

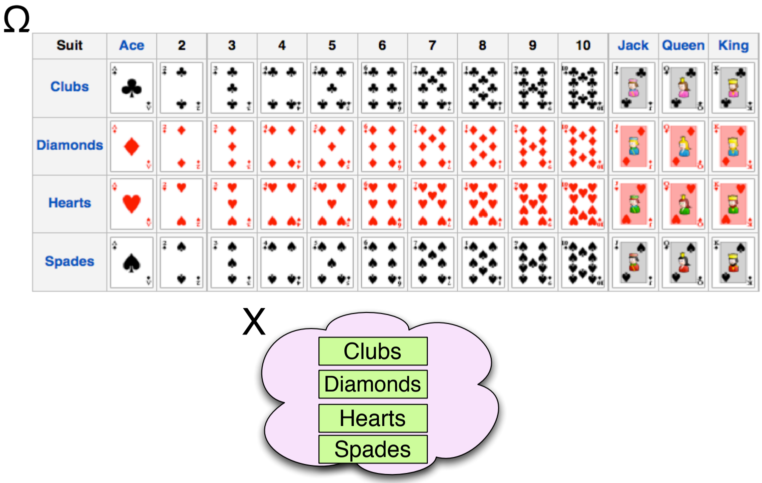
The Discrete Uniform Distribution

- Another common distribution is the **discrete uniform distribution**.
- Here, we have n possible outcomes (labeled 1 through n) which are all equally likely.
- In terms of the **parameter** n, $f(k) = ?$ for $k = 1, \ldots, n$
- What are some examples?

The Discrete Uniform Distribution

- Another common distribution is the **discrete uniform distribution**.
- Here, we have n possible outcomes (labeled 1 through n) which are all equally likely.
- In terms of the **parameter** n, $f(k) = ?$ for $k = 1, \ldots, n$
- What are some examples?
- What is $F(k)$ equal to?
- How would we find the mean and variance?

Example: Card Suit



Outline

The Binomial Distribution

- Many times we have some basic process with two outcomes (i.e., a Bernoulli process), which is repeated some number of times.

The Binomial Distribution

- Many times we have some basic process with two outcomes (i.e., a Bernoulli process), which is repeated some number of times.
- We may be interested in the *number* of "successes".

The Binomial Distribution

- Many times we have some basic process with two outcomes (i.e., a Bernoulli process), which is repeated some number of times.
- We may be interested in the *number* of "successes".
- If both
  - (a) the success probability stays constant
  - (b) success events are mutually independent
- then the number of successes has a **Binomial Distribution** with **parameters** n (number of trials) and p (individual success probability)

The Binomial Distribution

- We've seen an example already. What was it?

The Binomial Distribution

- We've seen an example already. What was it?
- The number of heads seen in 2 tosses is Binomial with $n = ?$ and $p = ?$

The Binomial Distribution

- We've seen an example already. What was it?
- The number of heads seen in 2 tosses is Binomial with $n = ?$ and $p = ?$
- Q: How do we find the probability of seeing k heads in 2 tosses?

The Binomial Distribution

- We've seen an example already. What was it?
- The number of heads seen in 2 tosses is Binomial with $n = ?$ and $p = ?$
- Q: How do we find the probability of seeing k heads in 2 tosses?
- A: Find all the individual sequences with k heads, and add their probabilities (since individual sequences are disjoint events).

The Binomial Distribution

- We've seen an example already. What was it?
- The number of heads seen in 2 tosses is Binomial with $n = ?$ and $p = ?$
- Q: How do we find the probability of seeing k heads in 2 tosses?
- A: Find all the individual sequences with k heads, and add their probabilities (since individual sequences are disjoint events).
- Q: What is the probability of any particular sequence?

The Binomial Distribution

- We've seen an example already. What was it?
- The number of heads seen in 2 tosses is Binomial with $n = ?$ and $p = ?$
- Q: How do we find the probability of seeing k heads in 2 tosses?
- A: Find all the individual sequences with k heads, and add their probabilities (since individual sequences are disjoint events).
- Q: What is the probability of any particular sequence?
- A: In two independent tosses, it's $P(\text{First Outcome}) \times P(\text{Second Outcome})$

The Binomial Distribution

- Q: In general, for a sequence of n **independent** trials, each with success probability p, how would we find the probability of a specific sequence of successes and failures?

The Binomial Distribution

- Q: In general, for a sequence of n **independent** trials, each with success probability p, how would we find the probability of a specific sequence of successes and failures?

$$\begin{aligned} P(\text{Sequence}) &= P(\text{First Outcome} \times P(\text{Second Outcome}) \\ &\quad \times \cdots \times P(\text{Last Outcome}) \\ &= \end{aligned}$$

The Binomial Distribution

- Q: In general, for a sequence of n **independent** trials, each with success probability p, how would we find the probability of a specific sequence of successes and failures?

$$\begin{aligned} P(\text{Sequence}) &= P(\text{First Outcome} \times P(\text{Second Outcome}) \\ &\quad \times \cdots \times P(\text{Last Outcome}) \\ &= p^{\#\text{ successes}} \times (1-p)^{\#\text{ failures}} \end{aligned}$$

The Binomial Distribution

- Q: In general, for a sequence of n **independent** trials, each with success probability p, how would we find the probability of a specific sequence of successes and failures?

$$\begin{aligned} P(\text{Sequence}) &= P(\text{First Outcome} \times P(\text{Second Outcome}) \\ &\quad \times \cdots \times P(\text{Last Outcome}) \\ &= p^{\#\text{ successes}} \times (1-p)^{\#\text{ failures}} \end{aligned}$$

- If there were k successes, how many failures were there?

The Binomial Distribution

- Therefore, every sequence of n trials with k successes has probability :

The Binomial Distribution

- Therefore, every sequence of n trials with k successes has probability :

$$P(\text{Each sequence with } k \text{ successes}) \quad =$$

The Binomial Distribution

- Therefore, every sequence of n trials with k successes has probability :

$$P(\text{Each sequence with } k \text{ successes}) = p^k \times (1-p)^{n-k}$$

The Binomial Distribution

- Therefore, every sequence of n trials with k successes has probability :

$$P(\text{Each sequence with } k \text{ successes}) = p^k \times (1-p)^{n-k}$$

- So, is this the probability of k successes in n (identical, independent) tries?

The Binomial Distribution

- Therefore, every sequence of n trials with k successes has probability :

$$P(\text{Each sequence with } k \text{ successes}) = p^k \times (1-p)^{n-k}$$

- So, is this the probability of k successes in n (identical, independent) tries?
- **No!** This is the probability of *each sequence*! What else do we need to know?