18.440: Lecture 16

Lectures 1-15 Review

Scott Sheffield

MIT
Counting tricks and basic principles of probability

Discrete random variables
Counting tricks and basic principles of probability

Discrete random variables
Selected counting tricks

- Break “choosing one of the items to be counted” into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.
Selected counting tricks

- Break “choosing one of the items to be counted” into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.
- Overcount by a fixed factor.
Selected counting tricks

- Break “choosing one of the items to be counted” into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.
- Overcount by a fixed factor.
- If you have $n$ elements you wish to divide into $r$ distinct piles of sizes $n_1, n_2 \ldots n_r$, how many ways to do that?

\[
\left( n, n_1, n_2, \ldots, n_r \right) := \frac{n!}{n_1! n_2! \ldots n_r!}.
\]

How many sequences $a_1, \ldots, a_k$ of non-negative integers satisfy $a_1 + a_2 + \ldots + a_k = n$?

\[
\text{Answer: } \binom{n+k-1}{n-1}.
\]

Represent partition by $k-1$ bars and $n$ stars, e.g., as $\ast\ast|\ast\ast||\ast\ast\ast\ast|\ast$. 

18.440 Lecture 16
- Break “choosing one of the items to be counted” into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.

- Overcount by a fixed factor.

- If you have $n$ elements you wish to divide into $r$ distinct piles of sizes $n_1, n_2 \ldots n_r$, how many ways to do that?

- Answer $(\binom{n}{n_1,n_2,\ldots,n_r}) := \frac{n!}{n_1!n_2!\ldots n_r!}$. 

How many sequences $a_1, \ldots, a_k$ of non-negative integers satisfy $a_1 + a_2 + \ldots + a_k = n$?

- Answer: $(\binom{n+k-1}{k-1})$. Represent partition by $k-1$ bars and $n$ stars, e.g., as $\star\star|\star\star||\star\star\star\star|\star$. 

Selected counting tricks
Selected counting tricks

- Break “choosing one of the items to be counted” into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.
- Overcount by a fixed factor.
- If you have $n$ elements you wish to divide into $r$ distinct piles of sizes $n_1, n_2 \ldots n_r$, how many ways to do that?
  - Answer: $\binom{n}{n_1,n_2,\ldots,n_r} := \frac{n!}{n_1!n_2!\ldots n_r!}$.
- How many sequences $a_1, \ldots, a_k$ of non-negative integers satisfy $a_1 + a_2 + \ldots + a_k = n$?
Selected counting tricks

- Break “choosing one of the items to be counted” into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.

- Overcount by a fixed factor.

- If you have $n$ elements you wish to divide into $r$ distinct piles of sizes $n_1, n_2 \ldots n_r$, how many ways to do that?

  - **Answer**: \[ \binom{n}{n_1,n_2,\ldots,n_r} := \frac{n!}{n_1!n_2!\ldots n_r!}. \]

- How many sequences $a_1, \ldots, a_k$ of non-negative integers satisfy $a_1 + a_2 + \ldots + a_k = n$?

  - **Answer**: \[ \binom{n+k-1}{n}. \] Represent partition by $k - 1$ bars and $n$ stars, e.g., as $**|***||**.*$. 

18.440 Lecture 16
Axioms of probability

- Have a set $S$ called *sample space*. 

$P(A) \in [0, 1]$ for all (measurable) $A \subset S$.

$P(S) = 1$.

Finite additivity: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.

Countable additivity: $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ if $E_i \cap E_j = \emptyset$ for each pair $i$ and $j$. 

18.440 Lecture 16
Axioms of probability

- Have a set $S$ called *sample space*.
- $P(A) \in [0, 1]$ for all (measurable) $A \subset S$. 
Axioms of probability

- Have a set $S$ called *sample space*.
- $P(A) \in [0, 1]$ for all (measurable) $A \subseteq S$.
- $P(S) = 1$. 
Axioms of probability

- Have a set $S$ called *sample space*.
- $P(A) \in [0, 1]$ for all (measurable) $A \subset S$.
- $P(S) = 1$.
- Finite additivity: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.
Axioms of probability

- Have a set $S$ called sample space.
- $P(A) \in [0, 1]$ for all (measurable) $A \subset S$.
- $P(S) = 1$.
- Finite additivity: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.
- Countable additivity: $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ if $E_i \cap E_j = \emptyset$ for each pair $i$ and $j$. 

Consequences of axioms

- $P(A^c) = 1 - P(A)$
Consequences of axioms

- $P(A^c) = 1 - P(A)$
- $A \subset B$ implies $P(A) \leq P(B)$
Consequences of axioms

- $P(A^c) = 1 - P(A)$
- $A \subset B$ implies $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(AB)$
Consequences of axioms

- $P(A^c) = 1 - P(A)$
- $A \subset B$ implies $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(AB)$
- $P(AB) \leq P(A)$
Inclusion-exclusion identity

- Observe $P(A \cup B) = P(A) + P(B) - P(AB)$.
Inclusion-exclusion identity

- Observe $P(A \cup B) = P(A) + P(B) - P(AB)$.
- Also, $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)$. 
Observe $P(A \cup B) = P(A) + P(B) - P(AB)$.

Also, $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)$.

More generally,

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \ldots + (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r})$$

$$= + \ldots + (-1)^{n+1} P(E_1 E_2 \ldots E_n).$$
Inclusion-exclusion identity

> Observe $P(A \cup B) = P(A) + P(B) - P(AB)$.

> Also, $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)$.

> More generally,

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \ldots$$

$$+ (-1)^{r+1} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r})$$

$$= + \ldots + (-1)^{n+1} P(E_1 E_2 \ldots E_n).$$

> The notation $\sum_{i_1 < i_2 < \ldots < i_r}$ means a sum over all of the $\binom{n}{r}$ subsets of size $r$ of the set $\{1, 2, \ldots, n\}$. 
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.
- What is $P(E_{i_1}E_{i_2} \ldots E_{i_r})$?

$1 - P(E_{i_1}E_{i_2} \ldots E_{i_r}) \approx \frac{1}{e} \approx 0.36788$
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.
- What is $P(E_1E_2\ldots E_r)$?
- Answer: $\frac{(n-r)!}{n!}$.

There are $\binom{n}{r}$ terms like that in the inclusion-exclusion sum.

What is $\binom{n}{r}\frac{(n-r)!}{n!}$?

Answer: $\frac{1}{r!}$.

$P(\bigcup_{i=1}^n E_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots + (-1)^n \frac{1}{n!} \approx 1/e \approx 0.36788$. 
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.
- What is $P(E_1 E_2 \ldots E_i)$?
- Answer: $\frac{(n-r)!}{n!}$.
- There are $\binom{n}{r}$ terms like that in the inclusion exclusion sum. What is $\binom{n}{r} \frac{(n-r)!}{n!}$?
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.
- What is $P(E_{i_1}E_{i_2} \ldots E_{i_r})$?
- Answer: $\frac{(n-r)!}{n!}$.
- There are $\binom{n}{r}$ terms like that in the inclusion exclusion sum.
- What is $\binom{n}{r} \frac{(n-r)!}{n!}$?
- Answer: $\frac{1}{r!}$. 
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.

- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.

- What is $P(E_{i_1}E_{i_2}\ldots E_{i_r})$?

- Answer: $\frac{(n-r)!}{n!}$.

- There are $\binom{n}{r}$ terms like that in the inclusion exclusion sum. What is $\binom{n}{r}\frac{(n-r)!}{n!}$?

- Answer: $\frac{1}{r!}$.

- $P(\bigcup_{i=1}^{n} E_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots \pm \frac{1}{n!}$
Famous hat problem

- $n$ people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let $E_i$ be the event that $i$th person gets own hat.
- What is $P(E_1E_2 \ldots E_r)$?
- Answer: $\frac{(n-r)!}{n!}$.
- There are $\binom{n}{r}$ terms like that in the inclusion exclusion sum.
- What is $\binom{n}{r}\frac{(n-r)!}{n!}$?
- Answer: $\frac{1}{r!}$.
- $P(\bigcup_{i=1}^{n} E_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \ldots \pm \frac{1}{n!}$
- $1 - P(\bigcup_{i=1}^{n} E_i) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \ldots \pm \frac{1}{n!} \approx 1/e \approx .36788$
Definition: $P(E|F) = \frac{P(EF)}{P(F)}$.
Definition: $P(E|F) = \frac{P(EF)}{P(F)}$.

Call $P(E|F)$ the “conditional probability of $E$ given $F$” or “probability of $E$ conditioned on $F$”.
Definition: \( P(E|F) = \frac{P(EF)}{P(F)} \).

Call \( P(E|F) \) the “conditional probability of \( E \) given \( F \)” or “probability of \( E \) conditioned on \( F \”).

Nice fact: \( P(E_1E_2E_3\ldots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\ldots P(E_n|E_1\ldots E_{n-1}) \)

Useful when we think about multi-step experiments.

For example, let \( E_i \) be event \( i \)th person gets own hat in the \( n \)-hat shuffle problem.
Definition: $P(E|F) = P(EF)/P(F)$.

Call $P(E|F)$ the “conditional probability of $E$ given $F$” or “probability of $E$ conditioned on $F$”.

Nice fact: $P(E_1E_2E_3\ldots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\ldots P(E_n|E_1\ldots E_{n-1})$

Useful when we think about multi-step experiments.
Definition: $P(E|F) = P(EF)/P(F)$.

Call $P(E|F)$ the “conditional probability of $E$ given $F$” or “probability of $E$ conditioned on $F$”.

Nice fact: $P(E_1E_2E_3\ldots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\ldots P(E_n|E_1\ldots E_{n-1})$

Useful when we think about multi-step experiments.

For example, let $E_i$ be event $i$th person gets own hat in the $n$-hat shuffle problem.
Dividing probability into two cases

\[ P(E) = P(EF) + P(EF^c) \]
\[ = P(E|F)P(F) + P(E|F^c)P(F^c) \]
Dividing probability into two cases

\[ P(E) = P(EF) + P(EF^c) \]
\[ = P(E|F)P(F) + P(E|F^c)P(F^c) \]

- In words: want to know the probability of \( E \). There are two scenarios \( F \) and \( F^c \). If I know the probabilities of the two scenarios and the probability of \( E \) conditioned on each scenario, I can work out the probability of \( E \).
Bayes’ theorem


\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \]
Bayes’ theorem

- Bayes’ theorem/law/rule states the following:
  \[ P(A|B) = \frac{P(B|A)P(A)}{P(B)} \].

- Follows from definition of conditional probability:
  \[ P(AB) = P(B)P(A|B) = P(A)P(B|A). \]
Bayes’ theorem

Bayes’ theorem/law/rule states the following:
\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \]

Follows from definition of conditional probability:
\[ P(AB) = P(B)P(A|B) = P(A)P(B|A). \]

Tells how to update estimate of probability of \( A \) when new evidence restricts your sample space to \( B \).
Bayes’ theorem

Bayes’ theorem/law/rule states the following:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \]

Follows from definition of conditional probability:

\[ P(AB) = P(B)P(A|B) = P(A)P(B|A). \]

 Tells how to update estimate of probability of \( A \) when new evidence restricts your sample space to \( B \).

So \( P(A|B) \) is \( \frac{P(B|A)}{P(B)} \) times \( P(A) \).
Bayes’ theorem/law/rule states the following:

\[ P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \]

Follows from definition of conditional probability:

\[ P(AB) = P(B)P(A|B) = P(A)P(B|A). \]

Tells how to update estimate of probability of \( A \) when new evidence restricts your sample space to \( B \).

So \( P(A|B) \) is \( \frac{P(B|A)}{P(B)} \) times \( P(A) \).

Ratio \( \frac{P(B|A)}{P(B)} \) determines “how compelling new evidence is”.
We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\bigcup E_i) = \sum P(E_i|F)$, if $i$ ranges over a countable set and the $E_i$ are disjoint.
$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\cup E_i) = \sum P(E_i|F)$, if $i$ ranges over a countable set and the $E_i$ are disjoint.
- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.
$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\bigcup E_i) = \sum P(E_i|F)$, if $i$ ranges over a countable set and the $E_i$ are disjoint.

- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.

- To get former from latter, we set probabilities of elements outside of $F$ to zero and multiply probabilities of events inside of $F$ by $1/P(F)$. 

18.440 Lecture 16
$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\bigcup E_i) = \sum P(E_i|F)$, if $i$ ranges over a countable set and the $E_i$ are disjoint.
- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.
- To get former from latter, we set probabilities of elements outside of $F$ to zero and multiply probabilities of events inside of $F$ by $1/P(F)$.
- $P(\cdot)$ is the *prior* probability measure and $P(\cdot|F)$ is the *posterior* measure (revised after discovering that $F$ occurs).
Say $E$ and $F$ are independent if $P(EF) = P(E)P(F)$. Equivalent statements are:

- $P(E|F) = P(E)$
- $P(F|E) = P(F)$
Say $E$ and $F$ are **independent** if $P(EF) = P(E)P(F)$.

Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$. 
Say $E_1 \ldots E_n$ are independent if for each 
\[ \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\} \] we have 
\[ P(E_{i_1}E_{i_2} \ldots E_{i_k}) = P(E_{i_1})P(E_{i_2}) \ldots P(E_{i_k}). \]
Say $E_1 \ldots E_n$ are independent if for each 
$\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\}$ we have 
$P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1})P(E_{i_2}) \ldots P(E_{i_k})$.

In other words, the product rule works.
Independence of multiple events

- Say $E_1 \ldots E_n$ are independent if for each
  \[
  \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots, n\}
  \]
  we have
  \[
  P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1}) P(E_{i_2}) \ldots P(E_{i_k}).
  \]
- In other words, the product rule works.
- Independence implies
  \[
  P(E_1 E_2 E_3 \mid E_4 E_5 E_6) = \frac{P(E_1) P(E_2) P(E_3) P(E_4) P(E_5) P(E_6)}{P(E_4) P(E_5) P(E_6)} = P(E_1 E_2 E_3),
  \]
  and other similar statements.
Independence of multiple events

Say \( E_1 \ldots E_n \) are independent if for each \( \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\} \) we have
\[
P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1})P(E_{i_2})\ldots P(E_{i_k}).
\]

In other words, the product rule works.

Independence implies
\[
P(E_1 E_2 E_3 | E_4 E_5 E_6) = \frac{P(E_1)P(E_2)P(E_3)P(E_4)P(E_5)P(E_6)}{P(E_4)P(E_5)P(E_6)} = P(E_1 E_2 E_3), \text{ and other similar statements.}
\]

Does pairwise independence imply independence?
Independence of multiple events

Say \( E_1 \ldots E_n \) are independent if for each \( \{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\} \) we have
\[
P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1}) P(E_{i_2}) \ldots P(E_{i_k}).
\]

In other words, the product rule works.

Independence implies
\[
P(E_1 E_2 E_3 | E_4 E_5 E_6) = \frac{P(E_1) P(E_2) P(E_3) P(E_4) P(E_5) P(E_6)}{P(E_4) P(E_5) P(E_6)} = P(E_1 E_2 E_3),
\]
and other similar statements.

Does pairwise independence imply independence?

No. Consider these three events: first coin heads, second coin heads, odd number heads. Pairwise independent, not independent.
Outline

Counting tricks and basic principles of probability

Discrete random variables
Counting tricks and basic principles of probability

Discrete random variables
A random variable $X$ is a function from the state space to the real numbers.
A random variable $X$ is a function from the state space to the real numbers.

Can interpret $X$ as a quantity whose value depends on the outcome of an experiment.
Random variables

- A random variable $X$ is a function from the state space to the real numbers.
- Can interpret $X$ as a quantity whose value depends on the outcome of an experiment.
- Say $X$ is a **discrete** random variable if (with probability one) if it takes one of a countable set of values.
Random variables

- A random variable $X$ is a function from the state space to the real numbers.
- Can interpret $X$ as a quantity whose value depends on the outcome of an experiment.
- Say $X$ is a **discrete** random variable if (with probability one) it takes one of a countable set of values.
- For each $a$ in this countable set, write $p(a) := P\{X = a\}$. Call $p$ the **probability mass function**.
Random variables

- A random variable $X$ is a function from the state space to the real numbers.

- Can interpret $X$ as a quantity whose value depends on the outcome of an experiment.

- Say $X$ is a **discrete** random variable if (with probability one) it takes one of a countable set of values.

- For each $a$ in this countable set, write $p(a) := P\{X = a\}$. Call $p$ the **probability mass function**.

- Write $F(a) = P\{X \leq a\} = \sum_{x\leq a} p(x)$. Call $F$ the **cumulative distribution function**.
Indicators

Given any event $E$, can define an **indicator** random variable, i.e., let $X$ be random variable equal to 1 on the event $E$ and 0 otherwise. Write this as $X = 1_E$. 

The value of $1_E$ (either 1 or 0) indicates whether the event $E$ has occurred.

If $E_1, E_2, \ldots, E_k$ are events then $X = \sum_{i=1}^{k} 1_{E_i}$ is the number of these events that occur.

Example: in $n$-hat shuffle problem, let $E_i$ be the event $i$th person gets own hat.

Then $\sum_{i=1}^{n} 1_{E_i}$ is total number of people who get own hats.
Given any event $E$, can define an **indicator** random variable, i.e., let $X$ be random variable equal to 1 on the event $E$ and 0 otherwise. Write this as $X = 1_E$.

The value of $1_E$ (either 1 or 0) *indicates* whether the event has occurred.
Given any event $E$, can define an **indicator** random variable, i.e., let $X$ be random variable equal to 1 on the event $E$ and 0 otherwise. Write this as $X = 1_E$.

The value of $1_E$ (either 1 or 0) *indicates* whether the event has occurred.

If $E_1, E_2, \ldots, E_k$ are events then $X = \sum_{i=1}^{k} 1_{E_i}$ is the number of these events that occur.
Indicators

- Given any event $E$, can define an **indicator** random variable, i.e., let $X$ be random variable equal to 1 on the event $E$ and 0 otherwise. Write this as $X = 1_E$.
- The value of $1_E$ (either 1 or 0) *indicates* whether the event has occurred.
- If $E_1, E_2, \ldots, E_k$ are events then $X = \sum_{i=1}^{k} 1_{E_i}$ is the number of these events that occur.
- Example: in $n$-hat shuffle problem, let $E_i$ be the event $i$th person gets own hat.
Indicators

- Given any event $E$, can define an **indicator** random variable, i.e., let $X$ be random variable equal to 1 on the event $E$ and 0 otherwise. Write this as $X = 1_E$.
- The value of $1_E$ (either 1 or 0) **indicates** whether the event has occurred.
- If $E_1, E_2, \ldots, E_k$ are events then $X = \sum_{i=1}^{k} 1_{E_i}$ is the number of these events that occur.
- Example: in $n$-hat shuffle problem, let $E_i$ be the event $i$th person gets own hat.
- Then $\sum_{i=1}^{n} 1_{E_i}$ is total number of people who get own hats.
Say $X$ is a **discrete** random variable if (with probability one) it takes one of a countable set of values.
Say $X$ is a **discrete** random variable if (with probability one) it takes one of a countable set of values.

For each $a$ in this countable set, write $p(a) := P\{X = a\}$. Call $p$ the **probability mass function**.
Say $X$ is a **discrete** random variable if (with probability one) it takes one of a countable set of values.

For each $a$ in this countable set, write $p(a) := P\{X = a\}$. Call $p$ the **probability mass function**.

The **expectation** of $X$, written $E[X]$, is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x).$$
Expectation of a discrete random variable

- Say \( X \) is a **discrete** random variable if (with probability one) it takes one of a countable set of values.
- For each \( a \) in this countable set, write \( p(a) := P\{X = a\} \). Call \( p \) the **probability mass function**.
- The **expectation** of \( X \), written \( E[X] \), is defined by
  \[
  E[X] = \sum_{x: p(x) > 0} xp(x).
  \]
- Represents weighted average of possible values \( X \) can take, each value being weighted by its probability.
If the state space $S$ is countable, we can give a SUM OVER STATE SPACE definition of expectation:

$$E[X] = \sum_{s \in S} P\{s\} X(s).$$
If the state space $S$ is countable, we can give \textbf{SUM OVER STATE SPACE} definition of expectation:

$$E[X] = \sum_{s \in S} P\{s\} X(s).$$

Agrees with the \textbf{SUM OVER POSSIBLE $X$ VALUES} definition:

$$E[X] = \sum_{x: p(x) > 0} xp(x).$$
Expectation of a function of a random variable

- If $X$ is a random variable and $g$ is a function from the real numbers to the real numbers then $g(X)$ is also a random variable.
If $X$ is a random variable and $g$ is a function from the real numbers to the real numbers then $g(X)$ is also a random variable.

How can we compute $E[g(X)]$?
If $X$ is a random variable and $g$ is a function from the real numbers to the real numbers then $g(X)$ is also a random variable.

How can we compute $E[g(X)]$?

Answer:

$$E[g(X)] = \sum_{x: p(x) > 0} g(x)p(x).$$
Additivity of expectation

If $X$ and $Y$ are distinct random variables, then

$$E[X + Y] = E[X] + E[Y].$$
Additivity of expectation

- If $X$ and $Y$ are distinct random variables, then

- In fact, for real constants $a$ and $b$, we have
Additivity of expectation

- If $X$ and $Y$ are distinct random variables, then $E[X + Y] = E[X] + E[Y]$.
- In fact, for real constants $a$ and $b$, we have $E[aX + bY] = aE[X] + bE[Y]$.
- This is called the **linearity of expectation**.
Additivity of expectation

- If $X$ and $Y$ are distinct random variables, then 
  \[ E[X + Y] = E[X] + E[Y]. \]
- In fact, for real constants $a$ and $b$, we have 
  \[ E[aX + bY] = aE[X] + bE[Y]. \]
- This is called the **linearity of expectation**.
- Can extend to more variables 
  \[ E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]. \]
Let $X$ be a random variable with mean $\mu$. 
Defining variance in discrete case

- Let $X$ be a random variable with mean $\mu$.
- The variance of $X$, denoted $\text{Var}(X)$, is defined by $\text{Var}(X) = E[(X - \mu)^2]$. 

Variance is one way to measure the amount a random variable "varies" from its mean over successive trials.

Very important alternate formula: $\text{Var}(X) = E[X^2] - (E[X])^2$. 

18.440 Lecture 16
Let $X$ be a random variable with mean $\mu$.

The variance of $X$, denoted $\text{Var}(X)$, is defined by $\text{Var}(X) = E[(X - \mu)^2]$.

Taking $g(x) = (x - \mu)^2$, and recalling that $E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$, we find that

$$\text{Var}[X] = \sum_{x:p(x)>0} (x - \mu)^2 p(x).$$
Let $X$ be a random variable with mean $\mu$.

The variance of $X$, denoted $\text{Var}(X)$, is defined by

$$\text{Var}(X) = E[(X - \mu)^2].$$

Taking $g(x) = (x - \mu)^2$, and recalling that

$$E[g(X)] = \sum_{x: p(x) > 0} g(x)p(x),$$

we find that

$$\text{Var}[X] = \sum_{x: p(x) > 0} (x - \mu)^2 p(x).$$

Variance is one way to measure the amount a random variable “varies” from its mean over successive trials.
Let $X$ be a random variable with mean $\mu$.

The variance of $X$, denoted $\text{Var}(X)$, is defined by

$$\text{Var}(X) = E[(X - \mu)^2].$$

Taking $g(x) = (x - \mu)^2$, and recalling that $E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$, we find that

$$\text{Var}[X] = \sum_{x:p(x)>0} (x - \mu)^2 p(x).$$

Variance is one way to measure the amount a random variable “varies” from its mean over successive trials.

Very important alternate formula: $\text{Var}[X] = E[X^2] - (E[X])^2$. 
Identity

- If $Y = X + b$, where $b$ is constant, then $\text{Var}[Y] = \text{Var}[X]$. 
Identity

- If $Y = X + b$, where $b$ is constant, then $\text{Var}[Y] = \text{Var}[X]$.
- Also, $\text{Var}[aX] = a^2 \text{Var}[X]$. 
Identity

- If $Y = X + b$, where $b$ is constant, then $\text{Var}[Y] = \text{Var}[X]$.
- Also, $\text{Var}[aX] = a^2\text{Var}[X]$.

18.440 Lecture 16
Write $SD[X] = \sqrt{Var[X]}$. 
- Write $SD[X] = \sqrt{\text{Var}[X]}$.
- Satisfies identity $SD[aX] = aSD[X]$. 
Standard deviation

- Write $SD[X] = \sqrt{Var[X]}$.
- Satisfies identity $SD[aX] = aSD[X]$.
- Uses the same units as $X$ itself.
Standard deviation

- Write $SD[X] = \sqrt{Var[X]}$.
- Satisfies identity $SD[aX] = aSD[X]$.
- Uses the same units as $X$ itself.
- If we switch from feet to inches in our “height of randomly chosen person” example, then $X$, $E[X]$, and $SD[X]$ each get multiplied by 12, but $Var[X]$ gets multiplied by 144.
Toss fair coin $n$ times. (Tosses are independent.) What is the probability of $k$ heads?

Answer: \[ \binom{n}{k} / 2^n. \]

What if coin has $p$ probability to be heads?

Answer: \[ \binom{n}{k} p^k (1-p)^{n-k}. \]

Writing $q = 1-p$, we can write this as \[ \binom{n}{k} p^k q^{n-k}. \]

Can use binomial theorem to show probabilities sum to one:

\[ 1 = 1^n = (p + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}. \]

Number of heads is binomial random variable with parameters \((n, p)\).
Bernoulli random variables

- Toss fair coin $n$ times. (Tosses are independent.) What is the probability of $k$ heads?
- Answer: $\binom{n}{k}/2^n$. 

- What if coin has $p$ probability to be heads?
- Answer: $\binom{n}{k} p^k (1-p)^{n-k}$.

- Writing $q = 1-p$, we can write this as $\binom{n}{k} p^k q^{n-k}$.

- Can use binomial theorem to show probabilities sum to one:

$$1 = (p+q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.$$ 

- Number of heads is binomial random variable with parameters $(n,p)$. 

Toss fair coin \( n \) times. (Tosses are independent.) What is the probability of \( k \) heads?

Answer: \( \binom{n}{k}/2^n \).

What if coin has \( p \) probability to be heads?

Writing \( q = 1 - p \), we can write this as

\[
\binom{n}{k} p^k q^{n-k}.
\]

Can use binomial theorem to show probabilities sum to one:

\[
1 = \left(p + q\right)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.
\]

Number of heads is binomial random variable with parameters \((n, p)\).
Bernoulli random variables

- Toss fair coin $n$ times. (Tosses are independent.) What is the probability of $k$ heads?
  - Answer: $\binom{n}{k}/2^n$.
- What if coin has $p$ probability to be heads?
  - Answer: $\binom{n}{k}p^k(1-p)^{n-k}$.
Toss fair coin $n$ times. (Tosses are independent.) What is the probability of $k$ heads?

Answer: $\binom{n}{k}/2^n$.

What if coin has $p$ probability to be heads?

Answer: $\binom{n}{k}p^k(1-p)^{n-k}$.

Writing $q = 1 - p$, we can write this as $\binom{n}{k}p^k q^{n-k}$.

Number of heads is binomial random variable with parameters $(n, p)$.
Toss fair coin $n$ times. (Tosses are independent.) What is the probability of $k$ heads?

Answer: $\left(\begin{array}{c} n \\ k \end{array}\right) / 2^n$.

What if coin has $p$ probability to be heads?

Answer: $\left(\begin{array}{c} n \\ k \end{array}\right) p^k (1 - p)^{n-k}$.

Writing $q = 1 - p$, we can write this as $\left(\begin{array}{c} n \\ k \end{array}\right) p^k q^{n-k}$.

Can use binomial theorem to show probabilities sum to one:
Toss fair coin \( n \) times. (Tosses are independent.) What is the probability of \( k \) heads?

Answer: \( \binom{n}{k} / 2^n \).

What if coin has \( p \) probability to be heads?

Answer: \( \binom{n}{k} p^k (1 - p)^{n-k} \).

Writing \( q = 1 - p \), we can write this as \( \binom{n}{k} p^k q^{n-k} \).

Can use binomial theorem to show probabilities sum to one:

\[
1 = 1^n = (p + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}.
\]
Bernoulli random variables

- Toss fair coin $n$ times. (Tosses are independent.) What is the probability of $k$ heads?
  - Answer: $\binom{n}{k}/2^n$.
- What if coin has $p$ probability to be heads?
  - Answer: $\binom{n}{k}p^k(1-p)^{n-k}$.
  - Writing $q = 1 - p$, we can write this as $\binom{n}{k}p^k q^{n-k}$.
  - Can use binomial theorem to show probabilities sum to one:
    - $1 = 1^n = (p + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$.
  - Number of heads is **binomial random variable with parameters** $(n, p)$. 
Decomposition approach to computing expectation

Let $X$ be a binomial random variable with parameters $(n, p)$. Here is one way to compute $E[X]$.

- Write $X = \sum_{j=1}^{n} X_j$, where $X_j$ is 1 if the $j$th coin is heads, 0 otherwise.
- In other words, $X_j$ is the number of heads (zero or one) on the $j$th toss.
- Note that $E[X_j] = p \cdot 1 + (1-p) \cdot 0 = p$ for each $j$.
- Conclude by additivity of expectation that $E[X] = n \sum_{j=1}^{n} E[X_j] = np$. 
Let $X$ be a binomial random variable with parameters $(n, p)$. Here is one way to compute $E[X]$.

Think of $X$ as representing number of heads in $n$ tosses of a coin that is heads with probability $p$. 

Note that $E[X_j] = p \cdot 1 + (1 - p) \cdot 0 = p$ for each $j$. 

Conclude by additivity of expectation that 

$$E[X] = n \sum_{j=1}^{n} E[X_j] = n \sum_{j=1}^{n} p = np.$$
Let $X$ be a binomial random variable with parameters $(n, p)$. Here is one way to compute $E[X]$.

Think of $X$ as representing number of heads in $n$ tosses of a coin that is heads with probability $p$.

Write $X = \sum_{j=1}^{n} X_j$, where $X_j$ is 1 if the $j$th coin is heads, 0 otherwise.
Decomposition approach to computing expectation

- Let $X$ be a binomial random variable with parameters $(n, p)$. Here is one way to compute $E[X]$.
- Think of $X$ as representing number of heads in $n$ tosses of coin that is heads with probability $p$.
- Write $X = \sum_{j=1}^{n} X_j$, where $X_j$ is 1 if the $j$th coin is heads, 0 otherwise.
- In other words, $X_j$ is the number of heads (zero or one) on the $j$th toss.
Let $X$ be a binomial random variable with parameters $(n, p)$. Here is one way to compute $E[X]$.

- Think of $X$ as representing number of heads in $n$ tosses of coin that is heads with probability $p$.
- Write $X = \sum_{j=1}^{n} X_j$, where $X_j$ is 1 if the $j$th coin is heads, 0 otherwise.
- In other words, $X_j$ is the number of heads (zero or one) on the $j$th toss.
- Note that $E[X_j] = p \cdot 1 + (1 - p) \cdot 0 = p$ for each $j$. 
Decomposition approach to computing expectation

Let $X$ be a binomial random variable with parameters $(n, p)$. Here is one way to compute $E[X]$.

Think of $X$ as representing number of heads in $n$ tosses of coin that is heads with probability $p$.

Write $X = \sum_{j=1}^{n} X_j$, where $X_j$ is 1 if the $j$th coin is heads, 0 otherwise.

In other words, $X_j$ is the number of heads (zero or one) on the $j$th toss.

Note that $E[X_j] = p \cdot 1 + (1 - p) \cdot 0 = p$ for each $j$.

Conclude by additivity of expectation that

$$E[X] = \sum_{j=1}^{n} E[X_j] = \sum_{j=1}^{n} p = np.$$
Compute variance with decomposition trick

\[ X = \sum_{j=1}^{n} X_j, \text{ so} \]
\[ E[X^2] = E[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_iX_j] \]
Compute variance with decomposition trick

- $X = \sum_{j=1}^{n} X_j$, so
  
  $E[X^2] = E[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j]$ 

- $E[X_i X_j]$ is $p$ if $i = j$, $p^2$ otherwise.
Compute variance with decomposition trick

- \[ X = \sum_{j=1}^{n} X_j, \text{ so} \]
  \[ E[X^2] = E[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] \]

- \( E[X_i X_j] \) is \( p \) if \( i = j \), \( p^2 \) otherwise.

- \( \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] \) has \( n \) terms equal to \( p \) and \( (n - 1)n \) terms equal to \( p^2 \).
Compute variance with decomposition trick

- \( X = \sum_{j=1}^{n} X_j \), so
  \[
  E[X^2] = E[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j]
  \]
- \( E[X_i X_j] \) is \( p \) if \( i = j \), \( p^2 \) otherwise.
- \( \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] \) has \( n \) terms equal to \( p \) and \((n - 1)n\) terms equal to \( p^2 \).
- So \( E[X^2] = np + (n - 1)np^2 = np + (np)^2 - np^2 \).
Compute variance with decomposition trick

- \( X = \sum_{j=1}^{n} X_j \), so
  \[
  E[X^2] = E[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j]
  \]

- \( E[X_i X_j] \) is \( p \) if \( i = j \), \( p^2 \) otherwise.

- \( \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] \) has \( n \) terms equal to \( p \) and \( (n - 1)n \) terms equal to \( p^2 \).

- So \( E[X^2] = np + (n - 1)np^2 = np + (np)^2 - np^2 \).

- Thus
  \[
  \]
Compute variance with decomposition trick

- \( X = \sum_{j=1}^{n} X_j \), so
  \[
  E[X^2] = E[\sum_{i=1}^{n} X_i \sum_{j=1}^{n} X_j] = \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j]
  \]
- \( E[X_i X_j] \) is \( p \) if \( i = j \), \( p^2 \) otherwise.
- \( \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j] \) has \( n \) terms equal to \( p \) and \((n-1)n\) terms equal to \( p^2\).
- So \( E[X^2] = np + (n-1)np^2 = np + (np)^2 - np^2 \).
- Thus \( \text{Var}[X] = E[X^2] - E[X]^2 = np - np^2 = np(1 - p) = npq \).
- Can show generally that if \( X_1, \ldots, X_n \) independent then \( \text{Var}[\sum_{j=1}^{n} X_j] = \sum_{j=1}^{n} \text{Var}[X_j] \).
Let \( \lambda \) be some moderate-sized number. Say \( \lambda = 2 \) or \( \lambda = 3 \). Let \( n \) be a huge number, say \( n = 10^6 \).
Bernoulli random variable with \( n \) large and \( np = \lambda \)

- Let \( \lambda \) be some moderate-sized number. Say \( \lambda = 2 \) or \( \lambda = 3 \). Let \( n \) be a huge number, say \( n = 10^6 \).
- Suppose I have a coin that comes on heads with probability \( \lambda/n \) and I toss it \( n \) times.

\[
\text{Answer: } np = \lambda.
\]

- Let \( k \) be some moderate-sized number (say \( k = 4 \)). What is the probability that I see exactly \( k \) heads?

\[
\text{Binomial formula: } \binom{n}{k} p^k (1-p)^{n-k}.
\]

- This is approximately \( \lambda \frac{k}{k!} (1-p)^{n-k} \approx \lambda \frac{k}{k!} e^{-\lambda} \).

- A Poisson random variable \( X \) with parameter \( \lambda \) satisfies \( \text{P}\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \) for integer \( k \geq 0 \).
Let $\lambda$ be some moderate-sized number. Say $\lambda = 2$ or $\lambda = 3$. Let $n$ be a huge number, say $n = 10^6$.

Suppose I have a coin that comes on heads with probability $\lambda/n$ and I toss it $n$ times.

How many heads do I expect to see?

Answer: $np = \lambda$.

Let $k$ be some moderate sized number (say $k = 4$). What is the probability that I see exactly $k$ heads?

Binomial formula:

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)(n-2)...(n-k+1)}{k!} p^k (1-p)^{n-k}.$$

This is approximately $\frac{\lambda^k}{k!} (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$.

A Poisson random variable $X$ with parameter $\lambda$ satisfies $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$ for integer $k \geq 0$. 
Let $\lambda$ be some moderate-sized number. Say $\lambda = 2$ or $\lambda = 3$. Let $n$ be a huge number, say $n = 10^6$.

Suppose I have a coin that comes on heads with probability $\lambda/n$ and I toss it $n$ times.

How many heads do I expect to see?

Answer: $np = \lambda$. 

A Poisson random variable $X$ with parameter $\lambda$ satisfies $P\{X = k\} = \frac{\lambda^k}{k!}e^{-\lambda}$ for integer $k \geq 0$. 


Bernoulli random variable with $n$ large and $np = \lambda$

- Let $\lambda$ be some moderate-sized number. Say $\lambda = 2$ or $\lambda = 3$. Let $n$ be a huge number, say $n = 10^6$.
- Suppose I have a coin that comes on heads with probability $\lambda/n$ and I toss it $n$ times.
- How many heads do I expect to see?
- Answer: $np = \lambda$.
- Let $k$ be some moderate sized number (say $k = 4$). What is the probability that I see exactly $k$ heads?
Bernoulli random variable with \( n \) large and \( np = \lambda \)

- Let \( \lambda \) be some moderate-sized number. Say \( \lambda = 2 \) or \( \lambda = 3 \). Let \( n \) be a huge number, say \( n = 10^6 \).
- Suppose I have a coin that comes on heads with probability \( \lambda/n \) and I toss it \( n \) times.
- How many heads do I expect to see?
- Answer: \( np = \lambda \).
- Let \( k \) be some moderate sized number (say \( k = 4 \)). What is the probability that I see exactly \( k \) heads?
- Binomial formula:
  \[
  \binom{n}{k} p^k (1 - p)^{n-k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} p^k (1 - p)^{n-k}.
  \]
Let $\lambda$ be some moderate-sized number. Say $\lambda = 2$ or $\lambda = 3$. Let $n$ be a huge number, say $n = 10^6$.

Suppose I have a coin that comes on heads with probability $\lambda/n$ and I toss it $n$ times.

How many heads do I expect to see?

Answer: $np = \lambda$.

Let $k$ be some moderate sized number (say $k = 4$). What is the probability that I see exactly $k$ heads?

Binomial formula:

$$\binom{n}{k} p^k (1 - p)^{n-k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} p^k (1 - p)^{n-k}. $$

This is approximately $\frac{\lambda^k}{k!} (1 - p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$. 

Let $\lambda$ be some moderate-sized number. Say $\lambda = 2$ or $\lambda = 3$. Let $n$ be a huge number, say $n = 10^6$.

Suppose I have a coin that comes on heads with probability $\lambda/n$ and I toss it $n$ times.

How many heads do I expect to see?

Answer: $np = \lambda$.

Let $k$ be some moderate sized number (say $k = 4$). What is the probability that I see exactly $k$ heads?

Binomial formula:

$$\binom{n}{k} p^k (1 - p)^{n-k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} p^k (1 - p)^{n-k}.$$  

This is approximately $\frac{\lambda^k}{k!} (1 - p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$.

A Poisson random variable $X$ with parameter $\lambda$ satisfies $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$ for integer $k \geq 0$. 

A Poisson random variable $X$ with parameter $\lambda$ satisfies

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$

for integer $k \geq 0$. 

Clever computation tricks yield

$$E[X] = \lambda$$

and

$$\text{Var}[X] = \lambda.$$
A Poisson random variable $X$ with parameter $\lambda$ satisfies
$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$$ for integer $k \geq 0$.

Clever computation tricks yield $E[X] = \lambda$ and $\text{Var}[X] = \lambda$. 

We think of a Poisson random variable as being (roughly) a Bernoulli $(n, p)$ random variable with $n$ very large and $p = \lambda/n$.

This also suggests $E[X] = np = \lambda$ and $\text{Var}[X] = npq \approx \lambda$. 
A Poisson random variable $X$ with parameter $\lambda$ satisfies $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$ for integer $k \geq 0$.

Clever computation tricks yield $E[X] = \lambda$ and $\text{Var}[X] = \lambda$.

We think of a Poisson random variable as being (roughly) a Bernoulli $(n, p)$ random variable with $n$ very large and $p = \lambda/n$. 


A Poisson random variable $X$ with parameter $\lambda$ satisfies

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \text{ for integer } k \geq 0.$$ 

Clever computation tricks yield $E[X] = \lambda$ and $\text{Var}[X] = \lambda$.

We think of a Poisson random variable as being (roughly) a Bernoulli $(n, p)$ random variable with $n$ very large and $p = \lambda/n$.

This also suggests $E[X] = np = \lambda$ and $\text{Var}[X] = npq \approx \lambda$. 

18.440 Lecture 16
A Poisson point process is a random function $N(t)$ called a Poisson process of rate $\lambda$. For each $t > s \geq 0$, the value $N(t) - N(s)$ describes the number of events occurring in the time interval $(s, t)$ and is Poisson with rate $(t - s)\lambda$. The numbers of events occurring in disjoint intervals are independent random variables. Probability to see zero events in first $t$ time units is $e^{-\lambda t}$. Let $T_k$ be time elapsed, since the previous event, until the $k$th event occurs. Then the $T_k$ are independent random variables, each of which is exponential with parameter $\lambda$. 
A Poisson point process is a random function $N(t)$ called a Poisson process of rate $\lambda$.

For each $t > s \geq 0$, the value $N(t) - N(s)$ describes the number of events occurring in the time interval $(s, t)$ and is Poisson with rate $(t - s)\lambda$. 

The numbers of events occurring in disjoint intervals are independent random variables.

Probability to see zero events in first $t$ time units is $e^{-\lambda t}$.

Let $T_k$ be time elapsed, since the previous event, until the $k$th event occurs. Then the $T_k$ are independent random variables, each of which is exponential with parameter $\lambda$. 


A Poisson point process is a random function $N(t)$ called a Poisson process of rate $\lambda$.

For each $t > s \geq 0$, the value $N(t) - N(s)$ describes the number of events occurring in the time interval $(s, t)$ and is Poisson with rate $(t - s)\lambda$.

The numbers of events occurring in disjoint intervals are independent random variables.
A Poisson point process is a random function $N(t)$ called a Poisson process of rate $\lambda$.

For each $t > s \geq 0$, the value $N(t) - N(s)$ describes the number of events occurring in the time interval $(s, t)$ and is Poisson with rate $(t - s)\lambda$.

The numbers of events occurring in disjoint intervals are independent random variables.

Probability to see zero events in first $t$ time units is $e^{-\lambda t}$.

Let $T_k$ be time elapsed, since the previous event, until the $k$th event occurs. Then the $T_k$ are independent random variables, each of which is exponential with parameter $\lambda$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$. Let $X$ be such that the first heads is on the $X$th toss. Answer:

$$P\{X = k\} = (1 - p)^{k-1}p = q^k, \quad \text{where} \quad q = 1 - p$$

Say $X$ is a geometric random variable with parameter $p$. Some cool calculation tricks show that $E[X] = \frac{1}{p}$. And $\text{Var}[X] = \frac{q}{p^2}$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability \( p \).

Let \( X \) be such that the first heads is on the \( X \)th toss.

\[
P\{X = k\} = (1 - p)^{k-1} p = q^{k-1} p, \quad \text{where} \quad q = 1 - p
\]

\( X \) is a geometric random variable with parameter \( p \).

Some cool calculation tricks show that \( E[X] = \frac{1}{p} \).

And \( \text{Var}[X] = \frac{q}{p^2} \).
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the first heads is on the $X$th toss.

Answer: $P\{X = k\} = (1 - p)^{k-1}p = q^{k-1}p$, where $q = 1 - p$ is tails probability.
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the first heads is on the $X$th toss.

Answer: $P\{X = k\} = (1 - p)^{k-1}p = q^{k-1}p$, where $q = 1 - p$ is tails probability.

Say $X$ is a geometric random variable with parameter $p$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the first heads is on the $X$th toss.

Answer: $P\{X = k\} = (1 - p)^{k-1}p = q^{k-1}p$, where $q = 1 - p$ is tails probability.

Say $X$ is a geometric random variable with parameter $p$.

Some cool calculation tricks show that $E[X] = 1/p$. 
Geometric random variables

- Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.
- Let $X$ be such that the first heads is on the $X$th toss.
- Answer: $P\{X = k\} = (1 - p)^{k-1}p = q^{k-1}p$, where $q = 1 - p$ is tails probability.
- Say $X$ is a geometric random variable with parameter $p$.
- Some cool calculation tricks show that $E[X] = 1/p$.
- And $\text{Var}[X] = q/p^2$. 

18.440 Lecture 16
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$. Let $X$ be such that the $r$th heads is on the $X$th toss. Then

$$P\{X = k\} = \binom{k - 1}{r - 1} p^{r-1} (1-p)^{k-r}.$$ 

Call $X$ negative binomial random variable with parameters $(r, p)$. So $E[X] = r/p$. And $Var[X] = rq/p^2$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the $r$th heads is on the $X$th toss.
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the $r$th heads is on the $X$th toss.

Then $P\{X = k\} = \binom{k-1}{r-1} p^{r-1} (1 - p)^{k-r} p$. 

Call $X$ negative binomial random variable with parameters $(r, p)$.

So $E[X] = \frac{r}{p}$.

And $Var[X] = \frac{rq}{p^2}$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the $r$th heads is on the $X$th toss.

Then $P\{X = k\} = \binom{k-1}{r-1} p^{r-1} (1 - p)^{k-r} p$.

Call $X$ **negative binomial random variable with parameters** $(r, p)$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the $r$th heads is on the $X$th toss.

Then $P\{X = k\} = \binom{k-1}{r-1} p^{r-1} (1 - p)^{k-r} p$.

Call $X$ **negative binomial random variable with parameters** $(r, p)$.

So $E[X] = r/p$. 
Consider an infinite sequence of independent tosses of a coin that comes up heads with probability $p$.

Let $X$ be such that the $r$th heads is on the $X$th toss.

Then $P\{X = k\} = \binom{k-1}{r-1} p^{r-1} (1 - p)^{k-r} p$.

Call $X$ **negative binomial random variable with parameters** $(r, p)$.

So $E[X] = r/p$.

And $\text{Var}[X] = rq/p^2$. 