GOV 2001/ 1002/ Stat E-200 Section 1
Probability Review

Solé Prillaman

Harvard University

January 28, 2015

LOGISTICS

Course Website: j.mp/G2001
lecture notes, videos, announcements

Canvas: problem sets, discussion board

Learning Catalytics: in class activities

Logistics

Got a question?

Post it to the discussion board

Come to office hours: Solé - Fridays 2-4pm; Stephen - Mondays 11am-1pm

Email us: Make sure to email both Stephen and I

LOGISTICS - COURSE REQUIREMENTS

Problem Sets

- Work in groups
- 5 attempts

Assessment Problems

- Work **alone**
- 1 attempt

Replication Paper

- Coauthored
- Deadlines throughout semester - more to come

Logistics - To Do

Go to R Section- Tomorrow at 4pm in CGIS S-153.

Reading Assignment- Unifying Political Methodology chs 1-3.

Problem Set 1- Due by 6pm next Wednesday on Canvas.

First Week Survey- Fill out on Canvas.

Learning Catalytics- Register before class on Monday.

Discussion board- Post and get to know each other! (Change notifications)

Logistics - Discussion Board

Everyone should change their notification preferences in Canvas! This makes sure that when updates are made on the discussion board, you know about them.

- Go to 'Settings' in the top right, then under 'Ways to Contact' you make sure your correct email address is there.
- Go to 'Settings' in the top right, and then 'Notifications', and choose 'Notify me right away' for 'Discussion' and 'Discussion Post'.

Outline

Basic Probability

Random Variables & Probability Distributions

Simulation

PROBABILITY

Probability is a formal model of uncertainty.

PROBABILITY

Probability is a formal model of uncertainty.

Suppose I roll a die.

What is the probability we roll a 2?

PROBABILITY

Probability is a formal model of uncertainty.

Suppose I roll a die.

What is the probability we roll a 2?
$P(X = 2) = \frac{1}{6}$

PROBABILITY

Probability is a formal model of uncertainty.

Suppose I roll a die.

What is the probability I roll something less than 4?

PROBABILITY

Probability is a formal model of uncertainty.

Suppose I roll a die.

What is the probability I roll something less than 4?
$P(X < 4) = P(X = 1) + P(X = 2) + P(X + 3) = \frac{3}{6}$

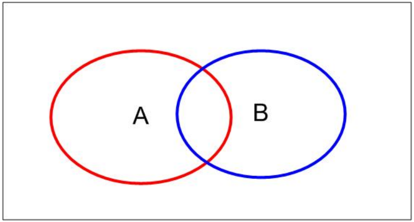
Notation and Definitions

Sample Space (S): The set of all possible outcomes from some process.

Event: Any subset of the sample space.

Union: The union of two events A and B, $A \cup B$, is the set containing all elements of A **or** B.

Intersection: The intersection of two events A and B, $A \cap B$, is the set containing all elements in both A **and** B.



Axioms of Probability

Let S be the sample space and A be an event in S.

1. For any event A, $P(A) \geq 0$

Axioms of Probability

Let S be the sample space and A be an event in S.

1. For any event A, $P(A) \geq 0$

   e.g. The probability of getting a heads cannot be -1.

Axioms of Probability

Let S be the sample space and A be an event in S.

1. For any event A, $P(A) \geq 0$

   e.g. The probability of getting a heads cannot be -1.

2. $P(S) = 1$ where S is the sample space

Axioms of Probability

Let S be the sample space and A be an event in S.

1. For any event A, $P(A) \geq 0$
   e.g. The probability of getting a heads cannot be -1.

2. $P(S) = 1$ where S is the sample space
   e.g. $P(heads) + P(tails) = 1$

Axioms of Probability

Let S be the sample space and A be an event in S.

1. For any event A, $P(A) \geq 0$
   e.g. The probability of getting a heads cannot be -1.

2. $P(S) = 1$ where S is the sample space
   e.g. $P(heads) + P(tails) = 1$

3. For any sequence of disjoint events (A and B),
   $P(A \cup B) = P(A) + P(B)$

Axioms of Probability

Let S be the sample space and A be an event in S.

1. For any event A, $P(A) \geq 0$

   e.g. The probability of getting a heads cannot be -1.

2. $P(S) = 1$ where S is the sample space

   e.g. $P(heads) + P(tails) = 1$

3. For any sequence of disjoint events (A and B),
   $P(A \cup B) = P(A) + P(B)$

   e.g. P(rolling a 3 or rolling a 2) $= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$

Probability Example

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Let's define C to be the event of having a car phone and S to be the event of receiving a speeding violation.

Marginal Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Marginal Probability: The probability of an event without including any additional information.

Marginal Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Marginal Probability: The probability of an event without including any additional information. Look at the margins!

Marginal Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Marginal Probability: The probability of an event without including any additional information. Look at the margins!

What is the probability of having a car phone?

Marginal Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Marginal Probability: The probability of an event without including any additional information. Look at the margins!

What is the probability of having a car phone? $P(C) = \frac{305}{755}$

Joint Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Joint Probability: The probability of some event A and another event B both occurring ($P(A \cap B)$ or $P(A,B)$).

Joint Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Joint Probability: The probability of some event A and another event B both occurring (P(A∩B) or P(A,B)). Look inside the table.

Joint Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Joint Probability: The probability of some event A and another event B both occurring (P(A∩B) or P(A,B)). Look inside the table.

What is the probability of getting a speeding violation and having a car phone?

Joint Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Joint Probability: The probability of some event A and another event B both occurring (P(A∩B) or P(A,B)). Look inside the table.

What is the probability of getting a speeding violation and having a car phone? $P(S \cap C) = \frac{25}{755}$

Conditional Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Conditional Probability: The probability of some event A, given that another event B has occurred (P(A|B)).

Conditional Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Conditional Probability: The probability of some event A, given that another event B has occurred ($P(A|B)$). Condition on the row/column and then calculate the probability.

Conditional Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

Conditional Probability: The probability of some event A, given that another event B has occurred (P(A|B)).Condition on the row/column and then calculate the probability.

What is the probability of getting a speeding violation given that you have a car phone?

Conditional Probability

	Speeding Violation	No Speeding Violation	Total
Car Phone	25	280	305
No Car Phone	45	405	450
Total	70	685	755

Source: `http://cnx.org/content/m16835/1.12/`

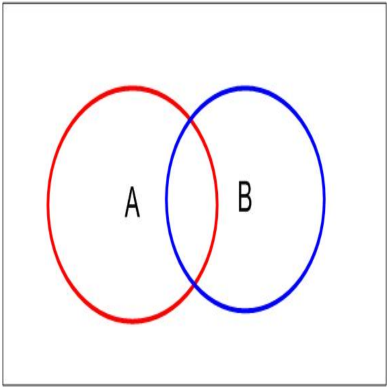
Conditional Probability: The probability of some event A, given that another event B has occurred (P(A|B)).Condition on the row/column and then calculate the probability.

What is the probability of getting a speeding violation given that you have a car phone? $P(S|C) = \frac{25}{305}$

Conditional Probability

What if we don't have such a nice table?

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$



CONDITIONAL PROBABILITY

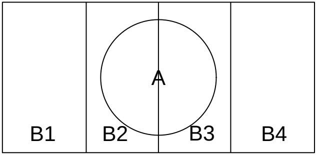
Note that if:

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A|B)P(B) &= P(A \cap B)
\end{aligned}
$$

Law of Total Probability

Law of Total Probability: Let S be the sample space and let the disjoint k events $B_1, ..., B_k$ partition S such that $P(B_1 \cup ... \cup B_k) = P(S) = 1$. If A is some other event in S, then

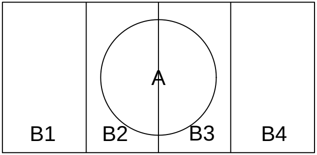
$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$



Law of Total Probability

Law of Total Probability: Let S be the sample space and let the disjoint k events $B_1, ..., B_k$ partition S such that $P(B_1 \cup ... \cup B_k) = P(S) = 1$. If A is some other event in S, then

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$

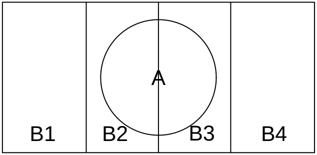


What is the P(A) in this figure?

Law of Total Probability

Law of Total Probability: Let S be the sample space and let the disjoint k events $B_1, \ldots, B_k$ partition S such that $P(B_1 \cup \ldots \cup B_k) = P(S) = 1$. If A is some other event in S, then

$$P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B_i)P(A|B_i)$$



What is the P(A) in this figure?

$P(A) =$
$P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) + P(B_4)P(A|B_4)$

Bayes' Rule

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$ Definition of Conditional Prob.

Bayes' Rule

$$
\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} && \text{Definition of Conditional Prob.} \\
&= \frac{P(B)P(A|B)}{P(A)} && \text{Rearranged Conditional Prob.}
\end{aligned}
$$

Bayes' Rule

$$\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} && \text{Definition of Conditional Prob.} \\
&= \frac{P(B)P(A|B)}{P(A)} && \text{Rearranged Conditional Prob.} \\
&= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^C)P(A|B^C)} && \text{Law of Total Prob.}
\end{aligned}$$

Side Note: Bayes' Rule

Bayes' rule will be extremely important going forward.

Why?

- Often we want to know $P(\theta|\text{data})$.
- But what we *do* know is $P(\text{data}|\theta)$.
- We'll be able to *infer* θ by using a variant of Bayes rule. Stay tuned.

Bayes' Rule:

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

Outline

Basic Probability

Random Variables & Probability Distributions

Simulation

RANDOM VARIABLES

A **random variable** is a function from S, the sample space, to R the real line, in other words a numerical value calculated from the outcome of a random experiment.

Random Variables

A **random variable** is a function from S, the sample space, to R the real line, in other words a numerical value calculated from the outcome of a random experiment.

e.g. When rolling a two dice, we may be interested in whether or not the sum of the two dice is 7. Or we might be interested in the actual sum of the two dice.

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

QUIZ BREAK!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_{\mathbf{i}}, \sigma^{\mathbf{2}})$$
$$\mu_{\mathbf{i}} = \mathbf{x_i}\beta$$

Which of the components of this model is a random variable?

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

Which of the components of this model is a random variable?
Y_i

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

Which of the components of this model is a random variable?
Y_i

Why isn't β a random variable?

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

Which of the components of this model is a random variable?
Y_i

Why isn't β a random variable? β is the *true* parameter value since this is the model, our estimators of β are random variables.

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

Which of the components of this model is a random variable?
Y_i

Why isn't β a random variable? β is the *true* parameter value since this is the model, our estimators of β are random variables.

What does this tell us about which is the systematic and stochastic components?

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_{\mathbf{i}}, \sigma^{\mathbf{2}})$$
$$\mu_{\mathbf{i}} = \mathbf{x_i}\beta$$

What does this tell us about our forms of uncertainty: fundamental v. estimation?

Quiz Break!

$$\mathbf{Y_i} \sim \mathbf{f_N}(\mu_\mathbf{i}, \sigma^\mathbf{2})$$
$$\mu_\mathbf{i} = \mathbf{x_i}\beta$$

What does this tell us about our forms of uncertainty: fundamental v. estimation?

Estimation uncertainty comes from having to use an estimator to estimate β and σ^2.

Fundamental uncertainty comes from the fact that Y is a random variable.

Ex. Waiting for the Redline – How long will it take for the next T to get here?

$$X = \begin{cases} 1 & \text{if the redline arrives within 1 minute} \\ 2 & \text{if } 1-2 \text{ minutes} \\ 3 & \text{if } 2-3 \text{ minutes} \\ 4 & \text{if } 3-4 \text{ minutes} \\ \vdots & \vdots \end{cases}$$

Discrete Example

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)?

DISCRETE EXAMPLE

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)? $Pr(X = 1) = \pi = .2$.

DISCRETE EXAMPLE

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)? $Pr(X = 1) = \pi = .2.$

What is Pr(X=2)?

Discrete Example

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)? $Pr(X = 1) = \pi = .2.$

What is Pr(X=2)? $Pr(X = 2) = (1 - \pi)\pi = .8 \cdot .2 = .16.$

Discrete Example

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)? $Pr(X = 1) = \pi = .2.$

What is Pr(X=2)? $Pr(X = 2) = (1 - \pi)\pi = .8 \cdot .2 = .16.$

What is Pr(X=3)?

Discrete Example

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)? $Pr(X = 1) = \pi = .2.$

What is Pr(X=2)? $Pr(X = 2) = (1 - \pi)\pi = .8 \cdot .2 = .16.$

What is Pr(X=3)? $Pr(X = 3) = (1 - \pi)^2\pi = .8^2 \cdot .2 = .128$

Discrete Example

Now, suppose the probability that the T comes in any given minute is a constant $\pi = .2$, and whether the T comes is independent of what has happened in previous periods.

What is Pr(X=1)? $Pr(X = 1) = \pi = .2.$

What is Pr(X=2)? $Pr(X = 2) = (1 - \pi)\pi = .8 \cdot .2 = .16.$

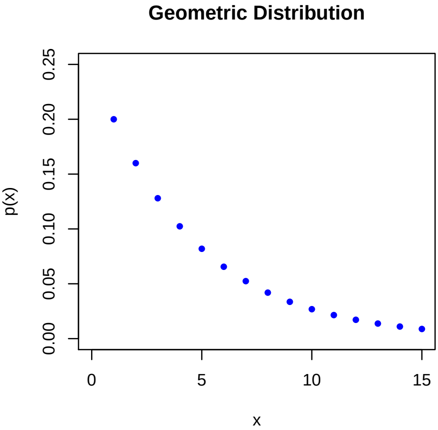
What is Pr(X=3)? $Pr(X = 3) = (1 - \pi)^2\pi = .8^2 \cdot .2 = .128$

And generally...

$$Pr(X = x) = (1 - \pi)^{x-1}\pi = .8^{x-1} \cdot .2$$

Which is the Probability Mass Function.

Discrete Example



PMF of X, a Geometric random variable with parameter $\pi = .2$.

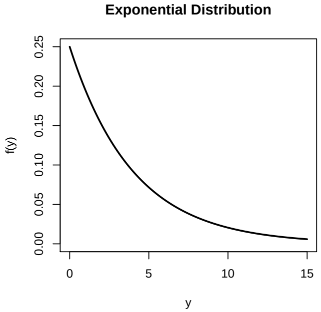
Continuous Example

Ex. Waiting for the Redline: an alternative model where Y is the *exact* time the T arrives.

Continuous Example

Ex. Waiting for the Redline: an alternative model where Y is the *exact* time the T arrives.

$$f(y) = \lambda e^{-\lambda y} = .2e^{-.2y}$$



Probability density function (PDF) of Y, an Exponential random variable with parameter $\lambda = .2$.

Continuous Example

Recall that to identify probabilities for continuous random variables we have to use:

$$P(Y \in A) = \int_A f(y)dy.$$

$$Pr(2 \leq y \leq 10) \quad =$$

Continuous Example

Recall that to identify probabilities for continuous random variables we have to use:

$$P(Y \in A) = \int_A f(y)dy.$$

$$Pr(2 \leq y \leq 10) \quad = \quad \int_2^{10} .2e^{-.5y}dy$$

Continuous Example

Recall that to identify probabilities for continuous random variables we have to use:

$$P(Y \in A) = \int_A f(y)dy.$$

$$\begin{aligned} Pr(2 \leq y \leq 10) &= \int_2^{10} .2e^{-.5y}dy \\ &= -e^{-.2y}|_2^{10} \end{aligned}$$

Continuous Example

Recall that to identify probabilities for continuous random variables we have to use:

$$P(Y \in A) = \int_A f(y)dy.$$

$$\begin{aligned} Pr(2 \leq y \leq 10) &= \int_2^{10} .2e^{-.5y}dy \\ &= -e^{-.2y}|_2^{10} \\ &= -e^{-.2\cdot 10} + e^{-.2\cdot 2} \end{aligned}$$

Continuous Example

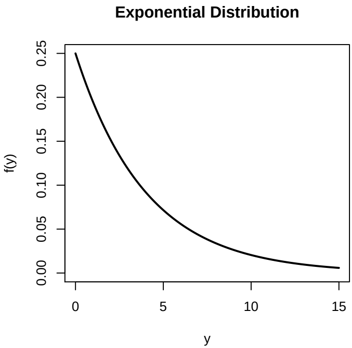
Recall that to identify probabilities for continuous random variables we have to use:

$$P(Y \in A) = \int_A f(y)dy.$$

$$\begin{aligned} Pr(2 \le y \le 10) &= \int_2^{10} .2e^{-.5y}dy \\ &= -e^{-.2y}|_2^{10} \\ &= -e^{-.2 \cdot 10} + e^{-.2 \cdot 2} \\ &\approx .535 \end{aligned}$$

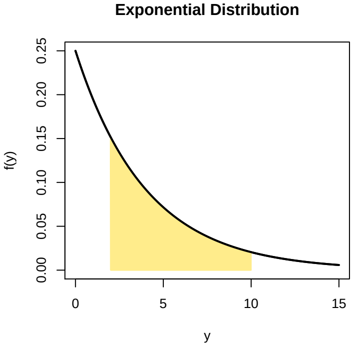
Continuous Example

$$f(y) = \lambda e^{-\lambda y} = .25e^{-.25y}$$



Continuous Example

$$f(y) = \lambda e^{-\lambda y} = .25e^{-.25y}$$



PMFs and PDFs

Characteristics of all PDFs and PMFs:

1. The support is all values of y where $P(Y = y) > 0$.
2. Probability mass/density must sum/integrate to 1.

PMFs and PDFs

Characteristics of all PDFs and PMFs:

1. The support is all values of y where $P(Y = y) > 0$.
2. Probability mass/density must sum/integrate to 1.
3. $P(Y = y) \geq 0$ for all y

PMFs and PDFs

Characteristics of all PDFs and PMFs:

1. The support is all values of y where $P(Y = y) > 0$.
2. Probability mass/density must sum/integrate to 1.
3. $P(Y = y) \geq 0$ for all y

Ex.:

1. $\int_0^\infty .25e^{-.25y}dy = -e^{-.25y}|_0^\infty = 0 + 1 = 1$

PMFs and PDFs

Characteristics of all PDFs and PMFs:

1. The support is all values of y where $P(Y = y) > 0$.
2. Probability mass/density must sum/integrate to 1.
3. $P(Y = y) \geq 0$ for all y

Ex.:

1. $\int_0^\infty .25e^{-.25y}dy = -e^{-.25y}|_0^\infty = 0 + 1 = 1$
2. $.25e^{-.25y} \geq 0$ for all $y \in (0, \infty)$

EXPECTATION OF A RANDOM VARIABLE

Discrete Case:

$$E(X) = \sum_i x_i P(X = x_i)$$

where $P(X = x)$ is the probability mass function (PMF).

Expectation of a random variable

Discrete Case:

$$E(X) = \sum_i x_i P(X = x_i)$$

where $P(X = x)$ is the probability mass function (PMF).

Continuous Case:

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy,$$

where $f(y)$ is the probability density function (PDF).

Our running examples...

Discrete Time:

$$E(X) = \sum_i x_i P(X = x_i)$$

Our running examples...

Discrete Time:

$$
\begin{aligned}
E(X) &= \sum_i x_i P(X = x_i) \\
&= \sum_{x_i=1}^{\infty} x_i (1 - .2)^{x_i - 1} \cdot .2
\end{aligned}
$$

OUR RUNNING EXAMPLES...

Discrete Time:

$$
\begin{aligned}
E(X) &= \sum_i x_i P(X = x_i) \\
&= \sum_{x_i=1}^{\infty} x_i (1 - .2)^{x_i - 1} \cdot .2 \\
&= 5
\end{aligned}
$$

OUR RUNNING EXAMPLES...

Discrete Time:

$$\begin{aligned} E(X) &= \sum_i x_i P(X = x_i) \\ &= \sum_{x_i=1}^{\infty} x_i (1 - .2)^{x_i - 1} \cdot .2 \\ &= 5 \end{aligned}$$

Continuous Time:

$$E(Y) = \int_0^{\infty} y f(y) dy$$

Our running examples...

Discrete Time:

$$\begin{aligned} E(X) &= \sum_i x_i P(X = x_i) \\ &= \sum_{x_i=1}^{\infty} x_i (1 - .2)^{x_i - 1} \cdot .2 \\ &= 5 \end{aligned}$$

Continuous Time:

$$\begin{aligned} E(Y) &= \int_0^{\infty} y f(y) dy \\ &= \int_0^{\infty} y \cdot .25 e^{-.25y} dy \end{aligned}$$

Our running examples...

Discrete Time:

$$\begin{aligned} E(X) &= \sum_i x_i P(X = x_i) \\ &= \sum_{x_i=1}^{\infty} x_i (1 - .2)^{x_i - 1} \cdot .2 \\ &= 5 \end{aligned}$$

Continuous Time:

$$\begin{aligned} E(Y) &= \int_0^{\infty} y f(y) dy \\ &= \int_0^{\infty} y \cdot .25 e^{-.25y} dy \\ &= 5 \end{aligned}$$

Expectation of a Function of a Random Variable

Now let's complicate things: suppose we want to find $E[g(X)]$, where $g(X)$ is any function of X.

Expectation of a Function of a Random Variable

Now let's complicate things: suppose we want to find $E[g(X)]$, where $g(X)$ is any function of X.

Discrete Case:

$$E[g(X)] = \sum_i g(x_i)P(X = x_i)$$

Expectation of a Function of a Random Variable

Now let's complicate things: suppose we want to find $E[g(X)]$, where $g(X)$ is any function of X.

Discrete Case:

$$E[g(X)] = \sum_i g(x_i)P(X = x_i)$$

Continuous Case:

$$E[g(y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$$

OUR RUNNING EXAMPLES...

Suppose we want to find $E[g(X)]$, where $g(X) = \sqrt{1+x}$.

$$E[g(X)] \quad = \quad \sum_i g(x_i) P(X = x_i)$$

Our running examples...

Suppose we want to find $E[g(X)]$, where $g(X) = \sqrt{1+x}$.

$$
\begin{aligned}
E[g(X)] &= \sum_i g(x_i) P(X = x_i) \\
&= \sum_{x=1}^{\infty} \sqrt{1+x}(1-.2)^{x-1} \cdot .2
\end{aligned}
$$

Our running examples...

Suppose we want to find $E[g(X)]$, where $g(X) = \sqrt{1+x}$.

$$E[g(X)] = \sum_i g(x_i)P(X = x_i) = \sum_{x=1}^{\infty} \sqrt{1+x}(1-.2)^{x-1} \cdot .2$$

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)f(y)dy$$

Our running examples...

Suppose we want to find $E[g(X)]$, where $g(X) = \sqrt{1+x}$.

$$
\begin{aligned}
E[g(X)] &= \sum_i g(x_i) P(X = x_i) \\
&= \sum_{x=1}^{\infty} \sqrt{1+x}(1-.2)^{x-1} \cdot .2
\end{aligned}
$$

$$
\begin{aligned}
E[g(Y)] &= \int_{-\infty}^{\infty} g(y) f(y) dy \\
&= \int_{0}^{\infty} \sqrt{1+y}.25e^{-.25y} dy
\end{aligned}
$$

Outline

THE MONTE CARLO METHOD

Basic idea: rather than calculate quantities analytically using deterministic formulae, approximate quantities using random sampling.


Simulating an expectation

$$E(X) = \sum_{x_i=1}^{\infty} x_i(1-.2)^{x_i-1} \cdot .2$$

```
set.seed(02138)
draws <- rgeom(n = 100000, prob = .2)
mean(draws)
[1] 4.98607
```

SIMULATING AN EXPECTATION

$$E(X) = \sum_{x_i=1}^{\infty} x_i(1-.2)^{x_i-1} \cdot .2$$

```
set.seed(02138)
draws <- rgeom(n = 100000, prob = .2)
mean(draws)
[1] 4.98607
```

$$E(Y) = \int_0^{\infty} y \cdot .25e^{-.25y} dy$$

```
draws <- rexp(n = 100000, rate = .25)
mean(draws)
[1] 4.99337
```

Neither of these are perfectly accurate but they become arbitrarily accurate as $n \to \infty$. Try it!

Monte Carlo Integration

What we just did was called **Monte Carlo Integration**, which means exactly what it sounds like (doing integrals via Monte Carlo simulation).

If we need to take an integral of the following form:

$$I = \int g(x)f(x)dx$$

Monte Carlo Integration allows us to approximate it by simulating M values from $f(x)$ and calculating:

$$\hat{I}_M = \frac{1}{M}\sum_{i=1}^{M} g(x^{(i)})$$

By the Strong Law of Large Numbers, our estimate $\hat{I}_M$ is a simulation consistent estimator of I as $M \to \infty$(our estimate gets better as we increase the number of simulations).

Simulating An Expectation of a Function

$$E[g(X)] = \sum_{x=1}^{\infty} \sqrt{1+x}(1-.2)^{x-1}.2$$

Approach:

1. Draw $M = 100000$ samples from the geometric distribution.
2. Calculate $g(x^{(i)})$ for each.
3. Find the mean of these.

```
draws <- rgeom(n = 100000, prob = .2)
g.draws <- sqrt(1 + draws)
mean(g.draws)
[1] 2.31169
```

Simulating An Expectation of a Function

$$E[g(X)] = \int_0^\infty \sqrt{1+y}.2e^{-.2y}dy$$

Approach:

1. Draw $M = 100000$ samples from the exponential distribution.
2. Calculate $g(x^{(i)})$ for each.
3. Find the mean of these.

```
set.seed(2001)
draws <- rexp(n = 100000, rate = .2)
g.draws <- sqrt(1 + draws)
mean(g.draws)
[1] 2.274344
```

Other Cool Things in R

What is f(20000)?

```
## dnorm provides the PDF of the normal distribution
dnorm(20000, mean=40000, sd=10000)
```

What is the probability that a salary is less that $20,000?

```
## pnorm provides the CDF of the normal distribution
pnorm(20000, mean=40000, sd=10000)
```

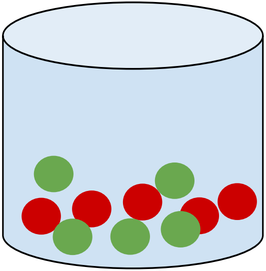
What value of p satisfies $P(X \leq p) = .95$?

```
## qnorm provides the inverse CDF of the normal distribution
qnorm(0.95, mean=40000, sd=10000)
```

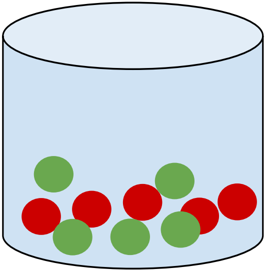
OTHER COOL THINGS IN R

For a list of the distributions that have similar commands in R, go to `http://www.stat.umn.edu/geyer/old/5101/rlook.html`.

SIMULATION FOR PROBABILITY PROBLEMS



I have an urn composed of 5 red balls and 5 green balls. If I sample 4 balls without replacement from the urn, what is the probability of drawing 4 balls all of the same color?



Approach via simulation:

1. Construct our population (aka our urn).
2. Figure out how to take one sample from it.
3. Figure out a programming rule for determining whether our condition was met i.e. we drew "4 red balls or 4 green balls".
4. Throw a for loop around it and sample repeatedly.
5. Determine what proportion of times our condition was successful.

1. Construct our population

Here are some ways to do this:

```
urn <- c("G","G","G","G","G","R","R","R","R","R")
urn
[1] "G" "G" "G" "G" "G" "R" "R" "R" "R" "R"

# Or
urn <- c(rep("red",5),rep("green",5))
urn
[1] "red"   "red"   "red"   "red"   "red"
  "green" "green" "green" "green"  "green"

# Or
urn <- c(rep(1,5),rep(0,5))
urn
[1] 1 1 1 1 1 0 0 0 0 0
```

I'll use the last one because numbers will be easier to use later on.

2. Take one sample

We need to use the `sample()` function with the `replace = FALSE` argument:

```
set.seed(1217)
draw <- sample(x = urn, size = 4, replace = FALSE)
draw
[1] 1 1 1 1
```

3. Determine if a success or failure

3. Determine if a success or failure

Because we used numeric signifiers for red and green, there is an easy test for whether or not we have drawn balls of all one color. If the numbers sum up to either 0 or 4, then we have a 'success'.

```
sum(draw) == 4
[1] TRUE
sum(draw) == 0
[1] FALSE
```

We can combine these test using '|' which means 'or'

```
success <- (sum(draw) == 4) | (sum(draw) == 0)
success
[1] TRUE
```

4. Repeat with a for-loop

Here's our guts so far:

```
draw <- sample(x = urn, size = 4, replace = FALSE)
success <- (sum(draw) == 4) | (sum(draw) == 0)
```

We can repeat this over and over 100000 times:

```
set.seed(1217)
sims <- 100000
success <- NULL
for(i in 1:sims){
  draw <- sample(x = urn, size = 4, replace = FALSE)
  success[i] <- sum(draw) == 4 | sum(draw) == 0
}
head(success)
[1]  TRUE  TRUE FALSE FALSE FALSE FALSE
```

(4. Repeat with replicate)

For-loops in R are slow. To get fancy:

```
eval <- function(draw){
  success <- sum(draw) == 4 | sum(draw) == 0
  return(success)
}
```

We can use replicate with our function to make the loop faster:

```
set.seed(1217)
sims <- 100000
success <- replicate(sims, eval(sample(urn, 4, replace=F)))
head(success)
[1]  TRUE  TRUE FALSE FALSE FALSE FALSE
```

5. Determine proportion of success

Two equivalent approaches:

```
sum(success)/sims
[1] 0.04846
mean(success)
[1] 0.04846
```

So the probability of selecting 4 balls of the same color is approximately 0.05.

Why is this useful?

- Math is hard or impossible and takes too long.
- Consider trying to integrate

$$I(f) = \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-\frac{x^2}{2}} dx$$

Which is the standard normal density and cannot be evaluated in closed form.

- How could we solve this?
    1. Sample 1000 points, $X_1, \ldots, X_{1000}$, uniformly distributed over the interval $(0, 1)$.
    2. Evaluate the function at each of these points and take the mean.

$$\frac{1}{1000} \left(\frac{1}{\sqrt{2\pi}} \right) \sum_{i=1}^{1000} e^{-\frac{X_i^2}{2}}$$

QUESTIONS

Questions?