Stat310

Conditional probability

Hadley Wickham

Housekeeping

How many of you have a cellphone?

How many of you now have the textbook?

Office hours: after class, tomorrow 4-5pm (location: TBA)

Mailing list: http://groups.google.com/group/stat310

Class party earnings update

Events vs. outcomes

Earnings Report Totals

Glossary

January 1, 2010 to January 17, 2010

	Items Shipped	Revenue	Referral Fees
Total Amazon.com Items Shipped	23	$1,420.43	$81.92
Total Third Party Items Shipped	7	$548.89	$32.93
Total Items Shipped	**30**	**$1,969.32**	**$114.85**
Total Items Returned	**0**	**$0.00**	**$0.00**
Total Refunds	**0**	**$0.00**	**$0.00**
TOTAL REFERRAL FEES	**30**	**$1,969.32**	**$114.85**

Books						
Backwards & Forwards: A Technical Manual for Reading Plays	Amazon.com	$14.13	6.02%	1	$14.13	**$0.85**
Causality: Models, Reasoning and Inference	Amazon.com	$39.20	5.99%	1	$39.20	**$2.35**
Data Manipulation with R (Use R)	Amazon.com	$44.46	6.01%	1	$44.46	**$2.67**
Interactive and Dynamic Graphics for Data Analysis: With R and GGobi (Use R)	Amazon.com	$48.89	5.99%	1	$48.89	**$2.93**
Law and Economics (5th Edition)	Amazon.com	$131.63	6.00%	1	$131.63	**$7.90**
Mathematical Statistics with Applications	Third Party	$79.97	6.00%	1	$79.97	**$4.80**
Mathematical Statistics with Applications	Third Party	$83.00	6.00%	1	$83.00	**$4.98**
Mathematical Statistics with Applications	Third Party	$86.28	6.00%	1	$86.28	**$5.18**
Mathematical Statistics with Applications	Third Party	$89.88	6.00%	3	$269.64	**$16.17**
Mathematical Statistics with Applications	Amazon.com	$94.03	6.00%	2	$188.06	**$11.28**
Modern Labor Economics: Theory and Public Policy (10th Edition)	Amazon.com	$133.25	6.00%	1	$133.25	**$8.00**
Notes on Directing: 130 Lessons in Leadership from the Director's Chair	Amazon.com	$10.20	5.98%	1	$10.20	**$0.61**
R Programming for Bioinformatics (Chapman & Hall/CRC Computer Science & Data Analysis)	Amazon.com	$47.21	5.99%	1	$47.21	**$2.83**
Robert Rauschenberg (October Files)	Amazon.com	$14.04	5.98%	1	$14.04	**$0.84**
Thermodynamics: An Engineering Approach with Student Resource DVD	Amazon.com	$147.26	6.00%	1	$147.26	**$8.84**
You May Ask Yourself: An Introduction to Thinking Like a Sociologist	Third Party	$30.00	6.00%	1	$30.00	**$1.80**
ggplot2: Elegant Graphics for Data Analysis (Use R)	Amazon.com	$48.31	6.00%	6	$289.86	**$17.40**
ggplot2: Elegant Graphics for Data Analysis (Use R)	Amazon.com	$48.80	6.00%	1	$48.80	**$2.93**
ggplot2: Elegant Graphics for Data Analysis (Use R)	Amazon.com	$48.85	6.00%	2	$97.70	**$5.86**
Electronics						
Apple MA348LL/A 15-inch MacBook Pro Battery (Retail Packaging)	Amazon.com	$121.76	4.00%	1	$121.76	**$4.87**
Kensington 64343 MicroSaver DS Notebook Computer Lock with Keys (PC/Mac)	Amazon.com	$43.98	4.00%	1	$43.98	**$1.76**

Events vs. outcomes

S = {a, b, c}.
A = {a, b}. B = {b, c}, C = {c}
Set of all events = {{a, b, c}, {a, b}, {a, c}, {b, c}, {a}, {b}, {c}, ∅}

S = {o_1, o_2, o_3, ...}
A = {o_1, o_2} B = {o_2, o_4, ...} C = {o_{154}}
Set of all events = ?

Event is a subset. Outcome is an element.

Motivation

Imagine you select a person at random and give them an HIV test. If the HIV test is positive what's the probability that they have HIV?

a) 95-100%

b) 90-94%

c) 80-89%

d) 65-79%

e) < 64%

It’s 50%!

1. Probability paradox
2. Conditional probability
3. Law of total probability
4. Independence

Your turn

A family has two children. One of the children is a boy. What is the probability that the other is a girl?

Your turn

A family has two children. They do not have two girls. What is the probability they have a boy and a girl?

What is a paradox?

A paradox is a statement or group of statements that leads to a contradiction or a situation which **defies intuition**.
(http://en.wikipedia.org/wiki/Paradox)

Our statistical intuition is often wrong, so we need a careful mathematical process to check

Process

1. Write out complete sample space
2. Remove excluded events
3. Define event of interest
4. Calculate probability

Conditional probability

Definition. Sample space interpretation.

Is the conditional probability a probability function?

Multiplication “rule”

Example

What is the probability that a randomly picked person out of this class is pregnant in a years time?

(~1.7 million pregnancies to girls aged 20-24, ~30 million girls)

http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5806a1.htm?s_cid=ss5806a1_e#tab29

Steps

1. Define basic events.

2. Write problem as probability of conditional event.

3. Think about the easiest way to calculate that probability (complements? intersection? conditioning?)

Law of total probability

Example

Randomly pick a guy from this class. What is the probability he gets a girl pregnant in the next year?

How might we want to partition the sample space?

Data

P(no sex) = 0.2

P(gay sex) = 0.05

P(no contraception) = 0.18

P(uses condom) = 0.71

P(withdrawal) = 0.11

P(pregnant | no contraception) = 0.85

P(pregnant | use condom) = 0.15

P(pregnant | withdrawal) = 0.27

http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5806a1.htm?s_cid=ss5806a1_e#tab29

Independence

If A and B are **independent**, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$.

In words: The fact that B occurred does not give us any information about A (and vice versa)

This implies $P(A \cap B) = P(A)\,P(B)$. (Why?)

Mutually independent if ...

Your turn

Are **mutually exclusive** events **independent**? Why/why not?

Toss three fair coins. Consider the events A = {all heads or all tails}, B = {at least two heads}, C = {at most two tails}. Which pairs of events are independent?