Probability

Probability

- Random variables

  weather {rain, sun}

  AAPL {up, same, down}

  HW1G {0 ... 100}

- Atomic events

  (rain, same, 93)

  (rain, same, 94) ...

- Sample space

  $\underline{\Omega}$ = {all atomic events}

Probability

- Events

weather = rain

$HW1 \geq 90$

- Combining events

$\Omega -$ weather = rain (NOT)

= weather = sun

weather = rain $\cap$ $HW1 \geq 90$ (AND)

$HW1 \geq 90 \cup H1I \leq 10$ (OR)

Probability

- Measure: $\mu : 2^\Omega \mapsto \mathbb{R}_+$
  - disjoint union: $\mu(E_1 \cup E_2 \cup \ldots E_k) = \mu(E_1) + \ldots$
  - e.g.: count $\mu(E) = |E|$
  - interpretation: how big is an event
- Distribution: μ s.t. $\mu(\Omega) = 1$
  - interpretation: $\mu(E)$ = probability of E
  - e.g.: uniform: $\mu(E) = \frac{|E|}{|\Omega|}$

Example

AAPL price

Weather

	up	same	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

$P(AAPL = up) = 0.3$

Bigger example

AAPL price

Weather	up	same	down
sun	0.03	0.05	0.02
rain	0.07	0.12	0.05

Weather	up	same	down
sun	0.14	0.23	0.09
rain	0.06	0.10	0.04

P(up) = 0.3

P(down, sun

Notation

- X=x: event that r.v. X is realized as value x
- P(X=x) means probability of event X=x
  - if clear from context, may omit "X=" P(x
  - instead of P(Weather=rain), just P(rain)
  - complex events too: e.g., P(X=x, Y≠y)
- P(X) means a function: x → P(X=x)

Functions of RVs

- Extend definition: any deterministic function of RVs is also an RV
- E.g., 3[sunny] + 5[same]

  (1 if sunny, 0 if not)

AAPL price

Weather

	up	same	down
sun	3	8	3
[illegible]	[illegible]	[illegible]	[illegible]

Sample v. population

Suppose we watch for 100 days and count up our observations

AAPL price

Weather

	up	same	d
sun	0.09	0.15	
rain	0.21	0.35	

AAPL price

Weather

	up	same	d
sun	7	12	
rain	22	41	

Law of large numbers

- If we take a sample of size N from distribution P, count up frequencies of atomic events, and normalize (divide by N) to get a distribution $\tilde{P}$
- Then $\tilde{P} \rightarrow P$ as $N \rightarrow \infty$

Working w/ distributions

- Marginals → dist'n as it would have been if we ignored an R.V.
- Joint → before forgetting

Marginals

AAPL price

Weather	up	same	down	
sun	0.09	0.15	0.06	0.3
rain	0.21	0.35	0.14	0.7
	0.3	0.5	0.2	

marginal of AAPL

Marginals

AAPL price

Weather	up	same	down
sun	0.03	0.05	0.02
rain	0.07	0.12	0.05

Weather	up	same	down
sun	0.14	0.23	0.09
rain	0.06	0.10	0.04

P(Weather, A

AAPL

W	up	same
sun	.17	.28
rain	.13	.22

W	
sun	.56
rain	.44

Law of total probability

- Two RVs, X and Y
- Y has values $y_1, y_2, \ldots, y_k$
- $P(X) =$ $P(X \wedge y_1) + P(X \wedge y_2) + \ldots + P(X \wedge$

Working w/ distribution

Coin

Weather	H
sun	0.15
rain	0.35

- Conditional:
  - Observation

    event that happened
  - Consistency
  - Renormalization
- Notation:

P(Weather | Coin = H)

Weather = sun

0.15
0.35

.3
.7

.3

Conditionals in the literature

When you have eliminated the impossible, whatever remains, however improbable, must be the truth.

—Sir Arthur Conan Doyle, as Sherlock Holmes

Conditionals

AAPL price

Weather	up	same	down
sun	0.03	0.05	0.02
rain	0.07	0.12	0.05

Weather	up	same	down
sun	0.14	0.23	0.09
rain	0.06	0.10	0.04

In general

- Zero out all but some slice of high-D table
    - or an irregular set of entries
- Throw away zeros
    - unless irregular structure makes it inconvenient
- Renormalize
    - normalizer for P(. | event) is P(event)

Conditionals

- Thought experiment: what happens if we condition on an event of zero probability?

NaN

Notation

- $P(X \mid Y)$ is a function: $x, y \rightarrow P(X{=}x \mid Y{=}y)$
- As is standard, expressions are evaluated separately for each realization:
  - $P(X \mid Y)\ P(Y)$ means the function $x, y \rightarrow$ $P(X{=}x \mid Y{=}y)\ P(Y{=}y)$

Exercise

change ↓

Hall opens
#2
→ no prize

Independence

- X and Y are **independent** if, for all possible values of y, $P(X) = P(X \mid Y=y)$
    - equivalently, for all possible values of x, $P(Y) = P(Y \mid X=x)$
    - equivalently, $P(X,Y) = P(X)\,P(Y)$
- Knowing X or Y gives us no information about the other

= product of marginals

AAPL price

Weather	up	same	down	
sun	0.09	0.15	0.06	0.3
rain	0.21	0.35	0.14	0.7
	0.3	0.5	0.2	

Expectations

How much should we
xpect to earn from
ur AAPL stock?

AAPL price

Weather	up	same	d
sun	0.09	0.15	
rain	0.21	0.35	

Weather	up	same	d
sun	+1	0	
rain	+1	0	

Linearity of expectation

- Expectation is a linear function of numbers in bottom table
- E.g., change -1s to 0s or to -2s

AAPL price

Weather

	up	same	d
sun	0.09	0.15	
rain	0.21	0.35	

Weather

	up	same	d
sun	+1	0	
rain	+1	0	

Conditional expectation

What if we know it's
unny?

AAPL price

Weather

	up	same	d
sun	0.09	0.15	
rain	0.21	0.35	

Weather

	up	same	d
sun	+1	0	
rain	+1	0	

Independence and expectation

- If X and Y are independent, then:
- Proof:

Variance

- Two stocks: one as above, other always earns 0.1 each day
- Same expectation, but one is much more variable
- Measure of variability: variance

Variance

- If zero-mean: variance = $E(X^2)$
  - Ex: constant 0 v. coin-flip ±1
- In general: $E((X - E(X))^2)$

Exercise: simplify the expression for variance

- $E((X - E(X))^2)$

Covariance

- Suppose we want an approximate numeric measure of (in)dependence
- Consider the r.v. XY
  - if X, Y are typically both +ve or both -ve
  - if X, Y are independent

Covariance

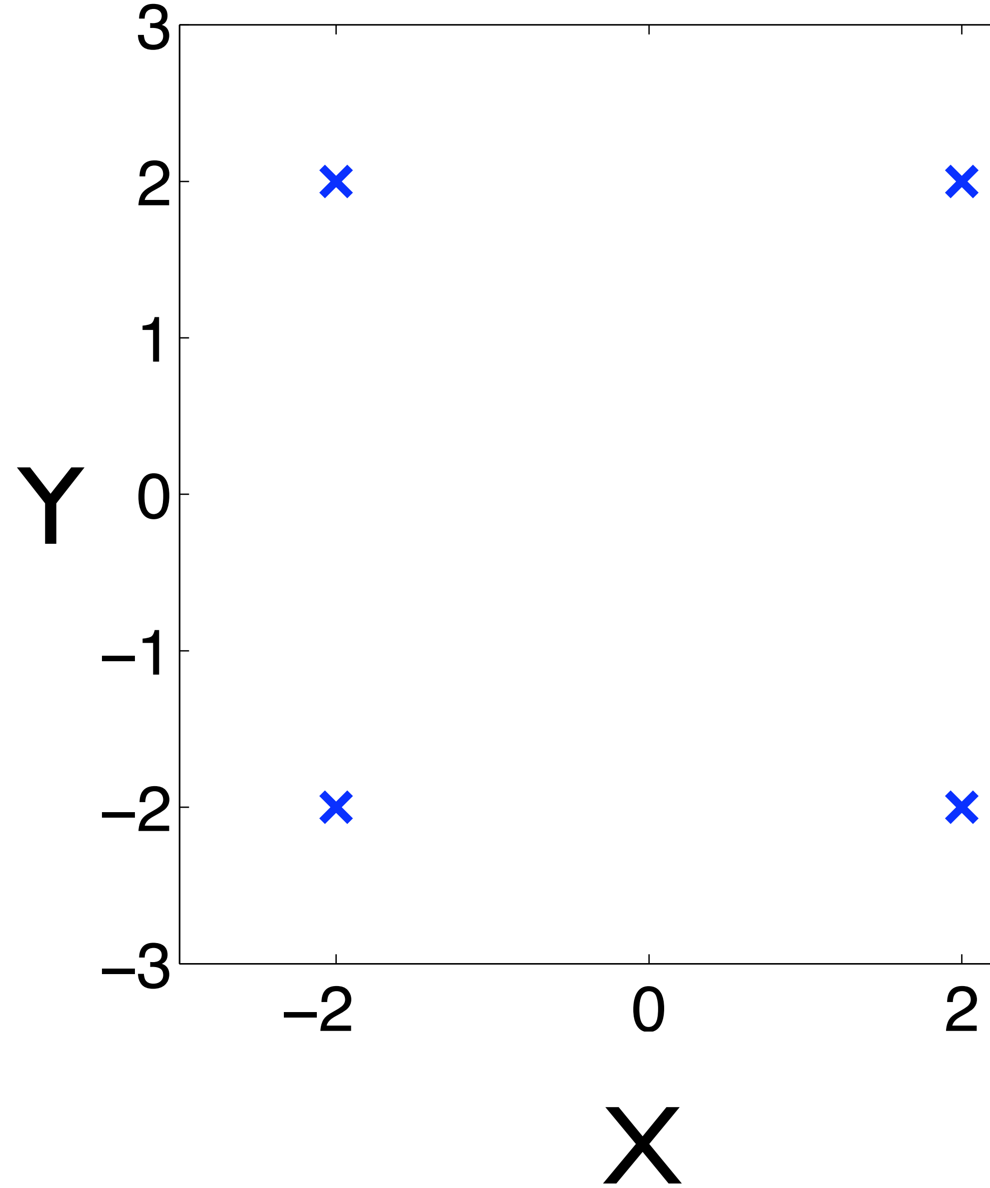
- cov(X,Y) =
- Is this a good measure of dependence?
  - Suppose we scale X by 10:

Correlation

- Like covariance, but control for variance of individual r.v.s
- cor(X, Y) =

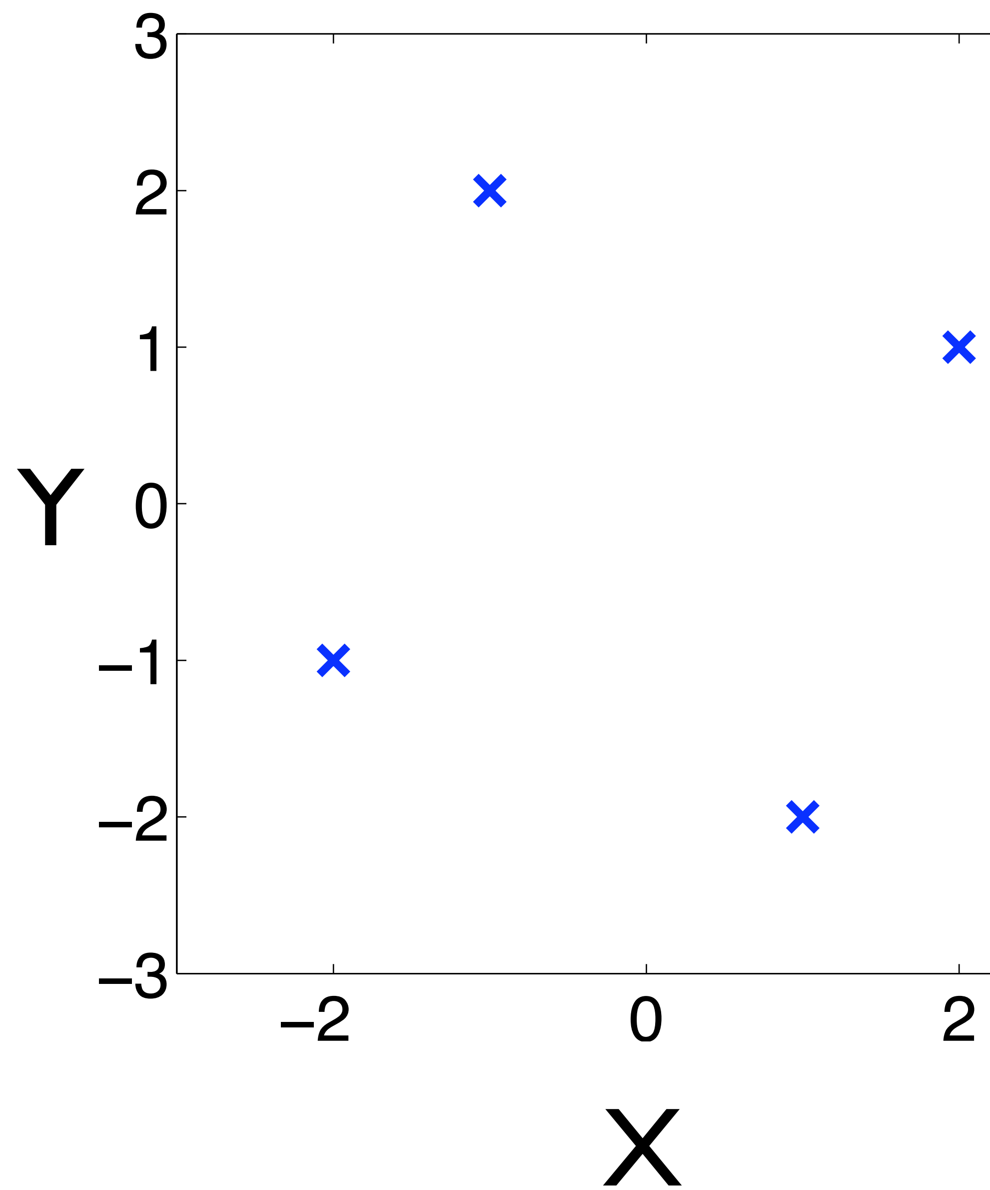
Correlation v. independence

- Equal probability on each point
- Are X and Y independent?
- Are X and Y uncorrelated?



Correlation v. independence

- Equal probability on each point
- Are X and Y independent?
- Are X and Y uncorrelated?



Law of large numbers

- Sample mean = expectation calculated from a sample =
- More general form of law:
- If we take a sample of size N from distribution P with mean μ and compute sample mean $\tilde{\mu}$
- Then $\tilde{\mu} \rightarrow \mu$ as $N \rightarrow \infty$

CLT

- **Central limit theorem**: for a sample of size N, population mean μ, population variance σ^2, the sample average has
  - mean
  - variance

CLT proof

- Assume mu = 0 for simplicity

Bayes Rule

Rev. Thomas Ba
1702–1761

- For any X, Y, C
  - $P(X \mid Y, C)\, P(Y \mid C) = P(Y \mid X, C)\, P(X \mid C)$
- Simple version (without context)
  - $P(X \mid Y)\, P(Y) = P(Y \mid X)\, P(X)$
- Can be taken as definition of conditioning

Bayes rule: usual form

- Take symmetric form
  - $P(X \mid Y)\, P(Y) = P(Y \mid X)\, P(X)$
- Divide by $P(Y)$

Exercise

- You are tested for a rare disease, emacsitis—prevalence 3 in 100,000
- Your receive a test that is 99% ***ensitive*** and 99% ***specific***
  - sensitivity = P(yes | emacsitis)
  - specificity = P(no | ~emacsitis)
- The test comes out ***positive***
- Do you have emacsitis?