Review: probability

- RVs, events, sample space Ω
- Measures, distributions
  - disjoint union property (law of total probability—book calls this "sum rule")
- Sample v. population
- Law of large numbers
- Marginals, conditionals

Monty Hall

Terminology

- Experiment =
- Prior =
- Posterior =

Example: model selection

- You're gambling to decide who has to clean the lab
- You are accused of using weighted dice!
- Two models:
    - fair dice: all 36 rolls equally likely
    - weighted: rolls summing to 7 more likely

prior:

observation:

posterior:

Philosophy

- ***Frequentist*** v. ***Bayesian***
- Frequentist view: a probability is a property of the world (the coin has P(H) = 0.62)
- Bayesian view: a probability is a representation of our internal beliefs about the world (we think P(H) = 0.62)

Difference

- Bayesian is willing to assign P(E) to any E, even one which has happened already (although it will be 1 or 0 if E or ¬E has been observed)
- Frequentist will assign probabilities ***only*** to outcomes of future experiments
- Consider the question: what is the probability that coin #273 is fair?

Which is right?

- Both!
- Bayesians can ask more questions
- But for a question that makes sense to both, answer will agree
- Can often rephrase a Bayesian question in frequentist terms
  - answer may differ
  - either may see other's answer as a reasonable approximation

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

Independence: probability = 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	

Readings

- So far: p1–4, sec 1–1.2, sec 2–2.3
- We'll put them next to relevant lectures on schedule page
- They provide extra detail beyond what's in lecture—you are responsible for knowing it
- No specific due date

Expectations

- How much should we expect to earn from our AAPL stock?

AAPL price

Weather	up	same	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

Weather	up	same	down
sun	+1	0	-1
rain	+1	0	-1

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	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

Weather	up	same	down
sun	+1	0	-1
rain	+1	0	-1

Conditional expectation

- What if we know it's sunny?

AAPL price

Weather	up	same	down
sun	0.09	0.15	0.06
rain	0.21	0.35	0.14

Weather	up	same	down
sun	+1	0	-1
rain	+1	0	-1

Independence and expectation

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

Sample means

- Sample mean =
- Expectation of sample mean:

Estimators

- Common task: given a sample, infer something about the population
- An ***estimator*** is a function of a sample that we use to tell us something about the population
- E.g., sample mean is a good estimator of population mean
- E.g., linear regression

Law of large numbers (more general form)

- If we take a sample of size N from a distribution P with mean μ and compute sample mean $\bar{x}$
- Then $\bar{x} \to \mu$ as $N \to \infty$

Bias

- Given an estimator T of a population quantity θ
- The ***bias*** of T is
- Sample mean is   estimator of population mean
- $(1 + \sum x_i) / (N+1)$ is

Variance

- Two estimators of population mean: sample mean, mean of every 2nd sample
- Both unbiased, 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

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

Exercise

- What is the variance of 3X?

Sample variance

- Sample variance =
- Expectation:
- Sample size correction:

Bias-variance decomposition

- Estimator T of population quantity θ
- ***Mean squared error*** = $E((T - \theta)^2)$ =

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

Covariance

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

  - if X, Y are independent

Covariance

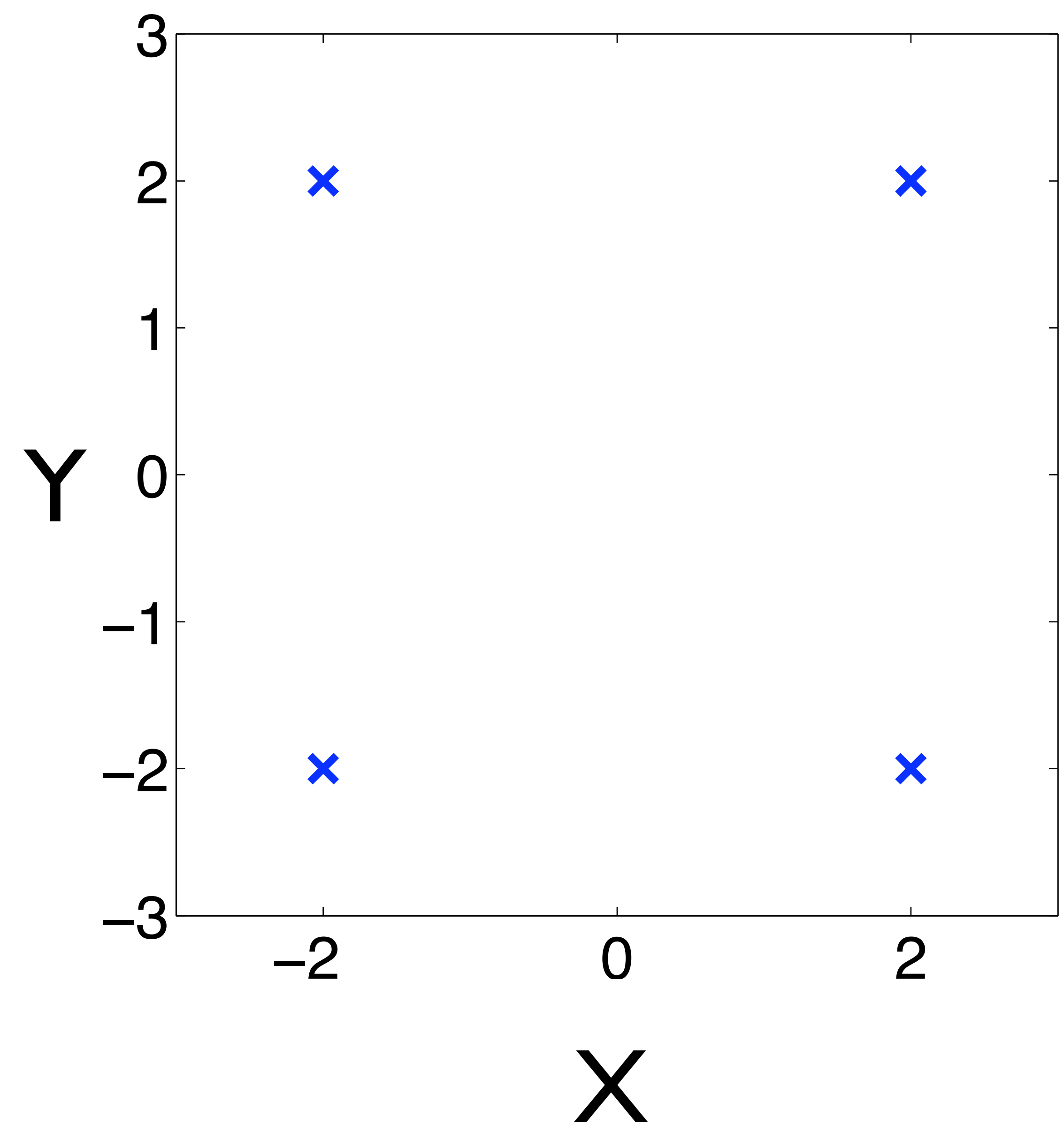
- cov(X,Y) =
- Is this a good measure of dependence?
  - Suppose we scale X by 10:

Correlation

- Like covariance, but control for variance of individual r.v.s
- cor(X,Y) =
- cor(10X,Y) =

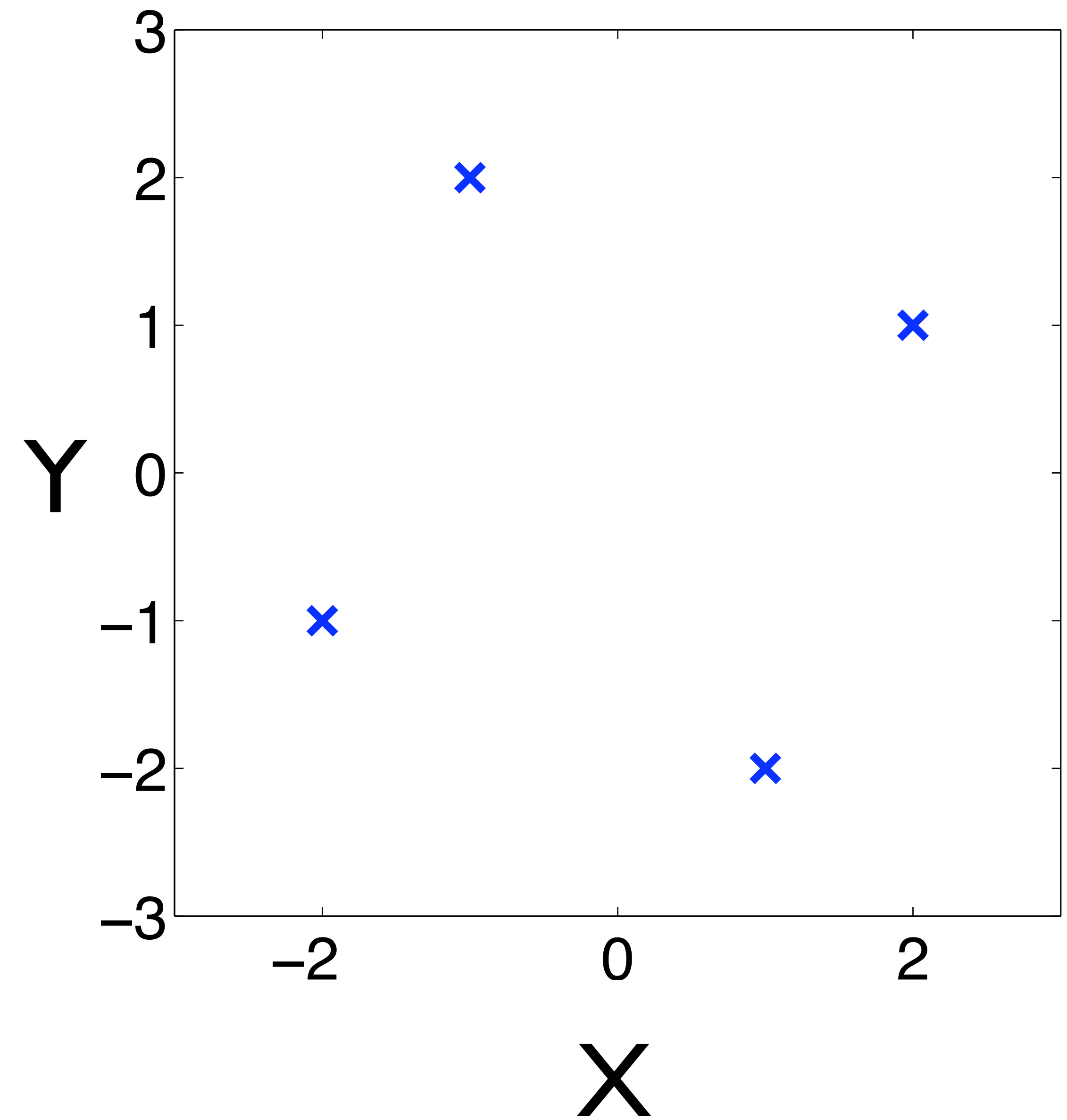
Correlation v. independence

- 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?



Bayes Rule

Rev. Thomas Bayes
1702–1761

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

Bayes rule: usual form

- Take symmetric form
  - P(X | Y) P(Y) = P(Y | X) P(X)
- Divide by P(Y)

Revisit: weighted dice

- Fair dice: all 36 rolls equally likely
- Weighted: rolls summing to 7 more likely
- Data: 1-6 2-5

Exercise

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