

A Random Walk through CS70, Pt. II: Probability

CS70 Summer 2016 - Lecture 8C

David Dinh
10 August 2016
UC Berkeley

1

Conditional Probability

Definition:

$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Live demo.

From definition: $\Pr[A \cap B] = \Pr[A] \Pr[B|A]$.

Or, generally: $\Pr[A_1 \cap \dots \cap A_n] = \Pr[A_1] \Pr[A_2|A_1] \dots \Pr[A_n|A_1 \cap \dots \cap A_{n-1}]$.

4

Today

Same as yesterday (and tomorrow). Review, applications, gigs, cool examples, research questions...

Probability today!

2

Bayes' Theorem

$$\Pr[A|B] = \frac{\Pr[A] \Pr[B|A]}{\Pr[B]}$$

Or if I know for sure that exactly one of A_1, \dots, A_n hold, then:

$$\Pr[A_k|B] = \frac{\Pr[A_k] \Pr[B|A_k]}{\sum_k \Pr[A_k] \Pr[B|A_k]}.$$

Useful theorem for inference (updating beliefs). Heavily used in AI.
[CS188](#).

5

Fundamentals

Map of outcomes in a probability space Ω to values in $[0, 1]$:
 $\sum_{\omega \in \Omega} \Pr[\omega] = 1$

Events: set of outcomes. $\Pr[E] = \sum_{\omega \in E} \Pr[\omega]$.

Inclusion-Exclusion: $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$.

Union bound: $\Pr[A_1 \cup A_2 \cup \dots \cup A_n] \leq \Pr[A_1] + \Pr[A_2] + \dots + \Pr[A_n]$.

Total probability: if A_1, \dots, A_n partition the entire sample space (disjoint, covers all of it), then $\Pr[B] = \Pr[B \cap A_1] + \dots + \Pr[B \cap A_n]$.

3

Random Variables: Discrete

Random variable: function that assigns a real number $X(\omega)$ to each outcome ω in a probability space.

Random variables X, Y are independent if the events $Y = a$ and $X = b$ are independent for all a, b . If X, Y independent, then $f(X), g(Y)$ independent for all f, g .

Expectation: $E[X] = \sum_t t \Pr[X = t]$

Tail sum: for nonnegative r.v. X : $E[X] = \sum_{i=0}^{\infty} \Pr[X > i]$.

Expectation of function: $E[g(X)] = \sum_t g(t) \Pr[X = t]$

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

Standard deviation: square root of variance.

Linearity of expectation: $E[aX + bY] = aE[X] + bE[Y]$

For independent RV: $E[XY] = E[X]E[Y]$, $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$

6

Example: Random-SAT

Let's say I have some Boolean clause that looks like this ("3-CNF")

$$(a \vee b \vee \bar{c}) \wedge (\bar{b} \wedge d \wedge e) \wedge \dots$$

n clauses (three boolean variables, some may be negated). What is expected number of clauses that I satisfy with a random assignment? $7n/8$.

Doesn't matter if variables are repeated! Expectation is linear.

Also proves (by probabilistic method) that there exists some assignment satisfying at least $7/8$ of the clauses.

Turns out that we don't know any better constant-factor approximation for this. $7/8$ is the best we can do! If we can efficiently do better (i.e. $7/8 + \epsilon$ fraction of clauses satisfied, for constant ϵ) this would prove $P = NP$ which would, among many other things, render public key cryptography impossible!

"Hardness of approximation". Ongoing topic of research.

7

Application: Streaming Algorithm for Counting Uniques

Let's say that you're building a server that wants to count unique visitors. But you only have a very small amount of memory - enough to remember one number. How do you distinguish between a million unique visitors and a single IP address sending a million requests to your site?

Map each IP address to a single number between 0 and 1 uniformly.

Keep the minimum number of all the visitors to your website (only requires space for one number!).

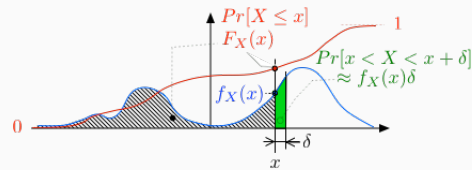
What's the number you get? Minimum of $Uniform(0,1)$. Distribution? CDF:

$$\Pr(\min_i X_i \leq x) = 1 - \Pr(\text{all } x_i \text{ at least } x) = 1 - (1-x)^n$$

So PDF is $f(x) = n(1-x)^{n-1}$. Expectation: $\int_0^1 xn(1-x)^{n-1}dx = 1/(n+1)$. Just invert the minimum number to estimate number of unique visitors!

10

Random Variables: Continuous



Distributions represented with a pdf

$$f_X(t) = \lim_{\delta \rightarrow 0} \frac{\Pr[X \in [t, t+\delta]]}{\delta}$$

...or, equivalently, a cdf:

$$F_X(t) = \Pr[X \leq t] = \int_{-\infty}^t f_X(z) dz.$$

$$\Pr[X \in [a, b]] = \int_a^b f_X(t) dt = F_X(b) - F_X(a)$$

8

Distributions

Discrete: Uniform, Bernoulli, geometric, binomial, Poisson

Continuous: Exponential, normal, uniform.

Make sure you know what they mean intuitively (although formula sheet will have the formulas for them).

For instance: What's the distribution of the sum of two independent binomial random variables? What's the distribution of the minimum of two independent geometric random variables? Prove these formally for practice!

11

Expectation/Variance for Continuous

Sum \rightarrow Integral. Most properties carry over.

$$E[X] = \int_{-\infty}^{\infty} tf_X(t) dt$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(t)f_X(t) dt$$

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

Linearity of expectation: $E[aX + bY] = aE[X] + bE[Y]$

For independent RV: $E[XY] = E[X]E[Y]$, $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$

9

Tail Bounds

Markov: For X non-negative, a positive,

$$\Pr[X \geq a] \leq \frac{E[X]}{a}.$$

Chebyshev: For all a positive,

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.$$

Chernoff: Family of exponential bounds for sum of mutually independent 0-1 random variables. Derive by noting that $\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}]$, and then applying Markov to bound

$$\Pr[e^{tX} \geq e^{ta}] \leq \frac{E[e^{tX}]}{e^{ta}}$$

for a good value of t .

12

Law of Large Numbers and CLT

If X_1, X_2, \dots are pairwise independent, and identically distributed with mean μ : $\Pr\left[\left|\frac{\sum_{i=1}^n X_i}{n} - \mu\right| \geq \epsilon\right] \rightarrow 0$ as $n \rightarrow \infty$.

With many i.i.d. samples we converge not only to the mean, but also to a normal distribution with the same variance.

CLT: Suppose X_1, X_2, \dots are i.i.d. random variables with expectation μ and variance σ^2 . Let

$$S_n := \frac{(\sum_i X_i) - n\mu}{\sigma\sqrt{n}}$$

Then S_n tends towards $\mathcal{N}(0, 1)$ as $n \rightarrow \infty$.

Or:

$$\Pr[S_n \leq a] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-x^2/2} dx$$

This is an approximation, not a bound.

13

Example: Markov Proof

Here's a theorem and a proof of the sort that we might ask you to do on the test.

Theorem: If a transient state j is accessible from state i , then state i is transient.

Proof: Suppose i is not accessible from j . Then there is a nonzero probability that, starting at i , we will go to j , at which point we will never be able to see i again. So i is transient.

On the other hand, suppose i is accessible from j . Suppose for contradiction that i is recurrent. Then if we're at j , we *have* to hit i again (because i is recurrent, so we have to go back to i if we go from i to j). But when we're at i , we know that we're definitely going to hit j sometime (because there's a nonzero chance of going to j from i , and we'll be back at i infinitely many times due to it being recurrent). So j is recurrent. Contradiction! So i has to be transient.

16

Markov Chains

Live Demo

Transition matrix P . Timesteps correspond to matrix multiplication: $\pi \rightarrow \pi P$.

Hitting time: How long does it take us to get to some state j ?

Strategy: let $\beta(i)$ be the time it takes to get to j from i , for each state i . $\beta(j) = 0$. Set up system of linear equations and solve.

14

Markov Chain Classifications

Irreducible Markov chain: all states communicate with every other state. Equivalently: graph representation is strongly connected.

Periodic Markov chain: any state is periodic.

Ergodic Markov chain: every state is ergodic. Any finite, irreducible, aperiodic Markov chain is ergodic.

17

State Classifications

State j is **accessible** from i : can get from i to j with nonzero probability. Equivalently: exists path from i to j .

i accessible from j and j accessible from i : i, j **communicate**.

If, given that we're at some state, we will see that state again sometime in the future with probability 1, state is **recurrent**. If there is a nonzero probability that we don't ever see state again, state is **transient**.

Every finite chain has a recurrent state.

State is **periodic** if, given that we're currently at that state, the probability that we are at that state s steps later is zero unless s divides some integer $\Delta > 1$.

Ergodic state: aperiodic + recurrent.

15

Stationary Distributions

Distribution is unchanged by state. Intuitively: if I have a lot (approaching infinity) of people on the same MC: the number of people at each state is constant (even if the individual people may move around).

To find limiting distribution? Solve **balance equations**: $\pi = \pi P$.

Let $f_{i,j}^t$ be the probability that we first (if $i = j$, we don't count the zeroth timestep) hit j exactly t timesteps after we start at i . Then $h_{i,j} = \sum_{t \geq 1} t f_{i,j}^t$.

Suppose we are given a finite, irreducible, aperiodic Markov chain. Then:

- There is a unique stationary distribution π .
- For all j, i , the limit $\lim_{t \rightarrow \infty} P_{j,i}^t$ exists and is independent of j .
- $\pi_i = \lim_{t \rightarrow \infty} P_{j,i}^t = 1/h_{i,i}$

18

Random Walks

Markov chain on an undirected graph. At a vertex, pick edge with uniform probability and walk down it.

For undirected graphs: aperiodic if and only if graph is not bipartite.

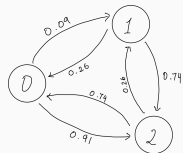
Stationary distribution: $\pi_v = d(v)/(2|E|)$.

Cover time (expected time that it takes to hit all the vertices, starting from the worst vertex possible): bounded above by $4|V||E|$.

19

Parrondo's Paradox III

So how often do we end up with case A? Here's the approach: one state for each value of $d \pmod{3}$.



Aperiodic? Irreducible? Yep! Limiting distribution = stationary distribution! Just solve for the stationary distribution with $\pi = \pi P$.

Result: $\pi = [0.382604, 0.154728, 0.462668]$. Plug in:

$$0.3826(0.09) + (0.1547 + 0.4627)(0.74) = 0.4913 < \frac{1}{2}$$

So I lose money in the long run.

22

Example/Gig: Parrondo's Paradox

Let's say I have two slot machines. Each one takes some amount of money and then spits out some amount of money.

Suppose that the expected return of each machine is negative - I get less money than I put in... the house always wins, after all. If I play machine 1 for a while, I expect to end up broke. Same with machine 2.

So if I play machine 1 and machine 2 alternately, I should expect to end up broke too, right? Hmm...

20

Parrondo's Paradox III

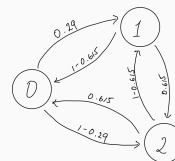
So now, what if I decide to flip a fair coin to figure out which machine to play?

I have d dollars... if d is a multiple of 3, probability of winning is:

$$\frac{1}{2}(0.49) + \frac{1}{2}(0.09) = 0.29$$

If d isn't a multiple of 3, probability of winning is:

$$\frac{1}{2}(0.49) + \frac{1}{2}(0.74) = 0.615$$



23

Parrondo's Paradox II

Let's say that the slot machines work as follows:

Machine 1: Put in some money. You gain a dollar w.p. 0.49 and lose a dollar w.p. 0.51. Pretty obvious that you lose money playing this game.

Machine 2: You put in d dollars.

- Case A: If d is a multiple of 3 then you gain a dollar w.p. 0.09 and lose a dollar w.p. 0.91.
- Case B: Otherwise, you gain a dollar w.p. 0.74 and lose a dollar w.p. 0.26.

What's the probability of winning a round? $1/3$ probability of case A happening, so it would be

$$\frac{1}{3}(0.09) + \frac{2}{3}(0.74) = \frac{157}{300} > \frac{1}{2}$$

right? Are you sure? **No!** Probability of case A happening is not $1/3$! (be careful about nonuniform probability spaces. MT2 1.1/1.2!)

21

Parrondo's Paradox IV

Stationary distribution: $\pi = [0.344583, 0.254343, 0.401075]$.

Probability of winning:

$$0.3446(0.29) + (0.2543 + 0.4011)(0.615) = 0.503011 > \frac{1}{2}$$

So we expect to... gain money??!?!?!?!?!?

Did we just break linearity of expectation? No! It doesn't make a whole lot of sense to talk about "expected winnings" for a state without taking into account the current state. Our distribution across states changes between the two games!

24

Questions?