Midterm 2 Review

CS70 Summer 2016 - Lecture 6D

Grace Dinh
28 July 2016
UC Berkeley

Midterm 2: Format

8 questions, 190 points, 110 minutes (same as MT1).

Two pages (one double-sided sheet) of **handwritten** notes.

Coverage: we will assume knowledge of all the material from the beginning of the class to yesterday, but we will only explicitly test for material seen after MT1.

We will give you a formula sheet (see MT2 logistics post on Piazza to see it). On it: all the distributions we'll expect you to know (with expectation + variance), and Chernoff bounds.

Probability Basics

Events and Sample Spaces

Probability space: set of outcomes, denoted with Ω. Each outcome in the probability space ω occurs with some probability.

$$\Pr[\omega] \in [0,1]$$

$$\sum_{\omega \in \Omega} \Pr[\omega] = 1$$

Uniform probability space: each outcome has the same probability.

An event E is a set of outcomes; the probability that an event happens is

$$\Pr[E] = \sum_{\omega \in E} \Pr[\omega] \ .$$

Events can be combined using standard set operations.

Disjointness and Additivity

If A, B disjoint (no intersection): $\Pr[A \cup B] = \Pr[A] + \Pr[B]$. Pairwise disjoint events (any two are disjoint) can also be summed.

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

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

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

What are the probabilities?

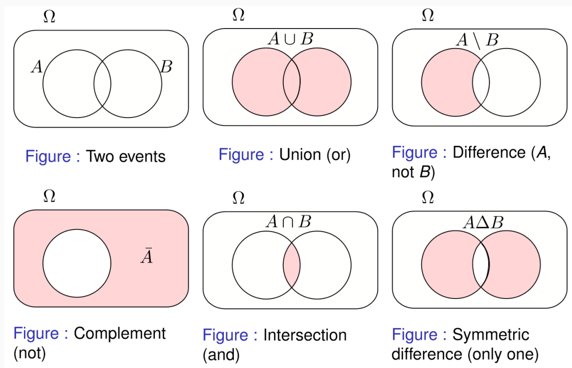


Figure : Two events

Figure : Union (or)

Figure : Difference (A, not B)

Figure : Complement (not)

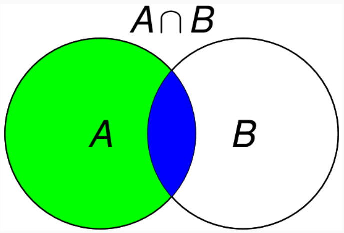
Figure : Intersection (and)

Figure : Symmetric difference (only one)

Conditional Probability and the Product Rule

Intuitively: "If I know B is true, what's the probability that A is true?
Definition:

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



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

$$\Pr[A_1 \cap \ldots \cap A_n] = \Pr[A_1]\Pr[A_2|A_1] \ldots \Pr[A_n|A_1 \cap \ldots \cap A_{n-1}] .$$

Correlation and Independence

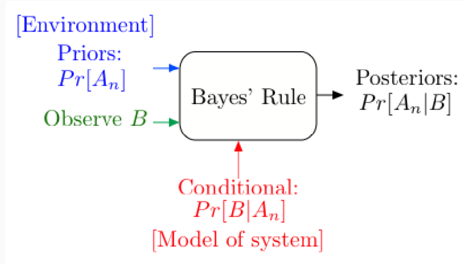
"Knowing that A is true tells you nothing about B." Independence: $\Pr[A \cap B] = \Pr[A]\Pr[B]$. Equivalently: $\Pr[A|B] = \Pr[A]$.

Or maybe knowing that one is true tells you that the other is likely to be true, too.

Positive Correlation: $\Pr[A \cap B] > \Pr[A]\Pr[B]$.

Negative Correlation: $\Pr[A \cap B] < \Pr[A]\Pr[B]$.

Bayes' Theorem



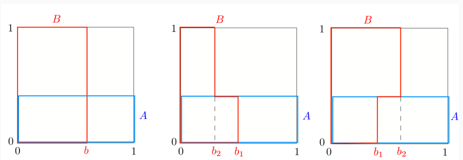
You know you will get a good grade in CS70 with some probability (prior). You take midterm 2 and get a good grade (observation). With this new information, figure out the probability that you get a good grade in CS70 (posterior).

Bayes' Theorem II

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

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

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



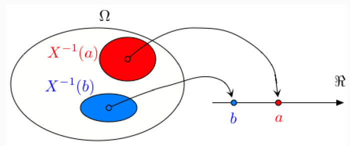
Random Variables

Fundamentals

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

Example: I catch 5 Pokemon. How many different kinds of Pokemon do I catch? Outcome: the exact type of each Pokemon I catch. Random variable: maps outcome to a number, e.g. {Psyduck, Typhlosion, Psyduck, Dratini, Typhlosion} $\rightarrow$ 3.

Discrete distributions: when there are a finite number of values an R.V. can take: pairs of values and probabilities. Probability of R.V. taking on a value: probability that an event that maps onto that value occurs.



Continuous RVs

Probability space is infinite and maps onto a continuous set of reals.

Distributions represented with a pdf

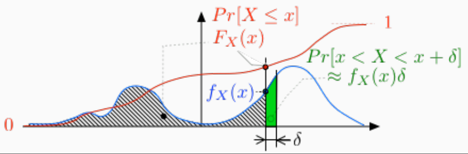
$$f_X(t) = \lim_{\delta \to 0} \frac{\Pr[X \in [t, t+\delta]]}{\delta}$$

...or, equivalently, a cdf:

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

.

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



Independence

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

Average over a huge (approaching ∞) number of trials.

Discrete:

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

Continuous:

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

Functions of variables:

Discrete:

$$E[g(X)] = \sum_t g(t) \Pr[X = t]$$

Continuous:

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

Expectation: Properties

Linearity of expectation: $E[a \sum_i X_i] = a_i \sum_i E[X_i]$ for **any** random variables X_i!

For independent X, Y: $E[XY] = E[X]E[Y]$.

Variance

“How spread-out is my distribution?”

$$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$

For any X: $Var[aX] = a^2 Var[X]$

For independent X, Y: $Var[X + Y] = Var[X] + Var[Y]$

Standard deviation is defined as square root of variance.

Indicators

If A is an event, let indicator r.v. be defined as 1 if our outcome is in the event and 0 otherwise.

Expectation? Same as probability that event happened!

Common Discrete Distributions

Uniform: Choose random integer in some finite interval.

Bernoulli: 1 w.p. p and 0 w.p. $1 - p$.

Binomial: I catch n Pokemon. Each Pokemon is a Lucario with probability p. How many Lucarios do I catch? Or: sum of binomials!

Geometric: Each Pokemon is a Shiftry with probability p. How many Pokemon do I need to catch until I first run into a Shiftry?. Memorylessness.

Poisson: I catch, on average, one Pokemon every minute. How many Pokemon do I catch in an hour?

We'll give the exact distribution functions, expectation and variance for these distributions to you on the exam... but you should intuitively understand them.

Common Continuous Distributions

Uniform: Pick random real number in some interval.

Exponential: I catch, on average, one Pokemon every minute. When do I catch my first Pokemon? Continuous analog of geometric.

Normal: Continuous analog of binomial. Models sums of lots of i.i.d. random variables (CLT).

We'll give the exact pdf, expectation and variance for these distributions to you on the exam... but you should intuitively understand them.

Tail Bounds and LLN

Confidence Intervals

Confidence intervals: if X falls in $[a, b]$ with probability $1 - \alpha$, then we say that $[a, b]$ is an $1 - \alpha$ confidence interval for X.

Markov

For X non-negative, a positive:

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

Not a very tight bound most of the time!

Or: for monotone non-decreasing function f that takes non-negative values, and non-negative X:

$$\Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}$$

for all a s.t. f(a) > 0.

Chebyshev

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

for all $a > 0$.

How did we get this? Just use Markov and use $f(x) = x^2$ as our function.

Chernoff

Family of exponential bounds for sum of mutually independent 0-1 random variables.

General approach to derive these: note that

$$\Pr[Xgea] = \Pr[e^{tX} \geq e^{ta}] \ .$$

Bound

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

using Markov. Choose a good t.

All the bounds you need are on the equation sheet on the exam.

LLN and CLT

If $X_1, X_2, \ldots$ are pairwise independent, and identically distributed with mean μ: $\Pr[\left|\frac{\sum_i X_i}{n} - \mu\right| \geq \epsilon] \to 0$ as $n \to \infty$.

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 $X_1, X_2, \ldots$ are i.i.d. random variables with expectation μ and variance σ^2. Let

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

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

Or:

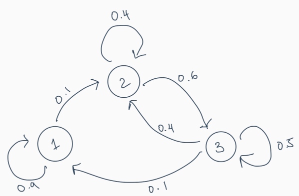
$$\Pr[S_n \leq a] \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx$$

This is an approximation, not a bound.

Markov Chains

Definitions

Set of states, transition probabilities, and initial distribution.



Also representable as a transition matrix.

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.4 & 0.6 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Distributions are row vectors. Timesteps correspond to matrix multiplication: $\pi \to \pi P$.

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.

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 with probability 1, state is **recurrent**. If there is a nonzero probability that we don't see state again, state is **transient**. Every finite chain has a recurrent state.

State is **periodic** if, once we're at a state, we can only return to that state at evenly spaced timesteps.

Ergodic state: aperiodic + recurrent.

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.

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 $r_{i,j}^t$ 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 $h_{i,j} = \sum_{t \geq 1} t r_{i,j}^t$.

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

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

Random Walks on Undirected Graphs

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 $4|V||E|$.

Good luck on the midterm!