18.440: Lecture 16

Lectures 1-15 Review

Scott Sheffield

MIT

Outline

Counting tricks and basic principles of probability

Discrete random variables

Outline

Counting tricks and basic principles of probability

Discrete random variables

Selected counting tricks

- Break "choosing one of the items to be counted" into a sequence of stages so that one always has the same number of choices to make at each stage. Then the total count becomes a product of number of choices available at each stage.
- Overcount by a fixed factor.
- If you have n elements you wish to divide into r distinct piles of sizes $n_1, n_2 \ldots n_r$, how many ways to do that?
- Answer $\binom{n}{n_1, n_2, \ldots, n_r} := \frac{n!}{n_1! n_2! \ldots n_r!}$.
- How many sequences $a_1, \ldots, a_k$ of non-negative integers satisfy $a_1 + a_2 + \ldots + a_k = n$?
- Answer: $\binom{n+k-1}{n}$. Represent partition by $k-1$ bars and n stars, e.g., as $** \mid **||****|*$.

Axioms of probability

- Have a set S called *sample space*.
- $P(A) \in [0,1]$ for all (measurable) $A \subset S$.
- $P(S) = 1$.
- Finite additivity: $P(A \cup B) = P(A) + P(B)$ if $A \cap B = \emptyset$.
- Countable additivity: $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ if $E_i \cap E_j = \emptyset$ for each pair i and j.

Consequences of axioms

- $P(A^c) = 1 - P(A)$
- $A \subset B$ implies $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(AB)$
- $P(AB) \leq P(A)$

Inclusion-exclusion identity

- Observe $P(A \cup B) = P(A) + P(B) - P(AB)$.
- Also, $P(E \cup F \cup G) =$
  $P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)$.
- More generally,

$$P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \ldots$$

$$+ (-1)^{(r+1)} \sum_{i_1 < i_2 < \ldots < i_r} P(E_{i_1} E_{i_2} \ldots E_{i_r})$$

$$= + \ldots + (-1)^{n+1} P(E_1 E_2 \ldots E_n).$$

- The notation $\sum_{i_1 < i_2 < \ldots < i_r}$ means a sum over all of the $\binom{n}{r}$ subsets of size r of the set $\{1, 2, \ldots, n\}$.

Famous hat problem

- n people toss hats into a bin, randomly shuffle, return one hat to each person. Find probability nobody gets own hat.
- Inclusion-exclusion. Let E_i be the event that ith person gets own hat.
- What is $P(E_{i_1} E_{i_2} \dots E_{i_r})$?
- Answer: $\frac{(n-r)!}{n!}$.
- There are $\binom{n}{r}$ terms like that in the inclusion exclusion sum. What is $\binom{n}{r}\frac{(n-r)!}{n!}$?
- Answer: $\frac{1}{r!}$.
- $P(\cup_{i=1}^n E_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots \pm \frac{1}{n!}$
- $1 - P(\cup_{i=1}^n E_i) = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots \pm \frac{1}{n!} \approx 1/e \approx .36788$

Conditional probability

- Definition: $P(E|F) = P(EF)/P(F)$.
- Call $P(E|F)$ the "conditional probability of E given F" or "probability of E conditioned on F".
- Nice fact: $P(E_1E_2E_3 \dots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2) \dots P(E_n|E_1 \dots E_{n-1})$
- Useful when we think about multi-step experiments.
- For example, let E_i be event ith person gets own hat in the n-hat shuffle problem.

Dividing probability into two cases

- $$P(E) = P(EF) + P(EF^c)$$
$$= P(E|F)P(F) + P(E|F^c)P(F^c)$$

- In words: want to know the probability of E. There are two scenarios F and F^c. If I know the probabilities of the two scenarios and the probability of E conditioned on each scenario, I can work out the probability of E.

Bayes' theorem

- Bayes' theorem/law/rule states the following: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$.
- Follows from definition of conditional probability: $P(AB) = P(B)P(A|B) = P(A)P(B|A)$.
- Tells how to update estimate of probability of A when new evidence restricts your sample space to B.
- So $P(A|B)$ is $\frac{P(B|A)}{P(B)}$ times $P(A)$.
- Ratio $\frac{P(B|A)}{P(B)}$ determines "how compelling new evidence is".

$P(\cdot|F)$ is a probability measure

- We can check the probability axioms: $0 \leq P(E|F) \leq 1$, $P(S|F) = 1$, and $P(\cup E_i) = \sum P(E_i|F)$, if i ranges over a countable set and the E_i are disjoint.
- The probability measure $P(\cdot|F)$ is related to $P(\cdot)$.
- To get former from latter, we set probabilities of elements outside of F to zero and multiply probabilities of events inside of F by $1/P(F)$.
- $P(\cdot)$ is the *prior* probability measure and $P(\cdot|F)$ is the *posterior* measure (revised after discovering that F occurs).

Independence

- Say E and F are **independent** if $P(EF) = P(E)P(F)$.
- Equivalent statement: $P(E|F) = P(E)$. Also equivalent: $P(F|E) = P(F)$.

Independence of multiple events

- Say $E_1 \ldots E_n$ are independent if for each $\{i_1, i_2, \ldots, i_k\} \subset \{1, 2, \ldots n\}$ we have $P(E_{i_1} E_{i_2} \ldots E_{i_k}) = P(E_{i_1})P(E_{i_2}) \ldots P(E_{i_k})$.
- In other words, the product rule works.
- Independence implies $P(E_1 E_2 E_3 | E_4 E_5 E_6) = \frac{P(E_1)P(E_2)P(E_3)P(E_4)P(E_5)P(E_6)}{P(E_4)P(E_5)P(E_6)} = P(E_1 E_2 E_3)$, and other similar statements.
- Does pairwise independence imply independence?
- No. Consider these three events: first coin heads, second coin heads, odd number heads. Pairwise independent, not independent.

Outline

Counting tricks and basic principles of probability

Discrete random variables

Outline

Counting tricks and basic principles of probability

Discrete random variables

Random variables

- A random variable X is a function from the state space to the real numbers.
- Can interpret X as a quantity whose value depends on the outcome of an experiment.
- Say X is a **discrete** random variable if (with probability one) if it takes one of a countable set of values.
- For each a in this countable set, write $p(a) := P\{X = a\}$. Call p the **probability mass function**.
- Write $F(a) = P\{X \leq a\} = \sum_{x \leq a} p(x)$. Call F the **cumulative distribution function**.

Indicators

- Given any event E, can define an **indicator** random variable, i.e., let X be random variable equal to 1 on the event E and 0 otherwise. Write this as $X = 1_E$.
- The value of 1_E (either 1 or 0) *indicates* whether the event has occurred.
- If $E_1, E_2, \ldots, E_k$ are events then $X = \sum_{i=1}^k 1_{E_i}$ is the number of these events that occur.
- Example: in n-hat shuffle problem, let E_i be the event ith person gets own hat.
- Then $\sum_{i=1}^n 1_{E_i}$ is total number of people who get own hats.

Expectation of a discrete random variable

- Say X is a **discrete** random variable if (with probability one) it takes one of a countable set of values.
- For each a in this countable set, write $p(a) := P\{X = a\}$. Call p the **probability mass function**.
- The **expectation** of X, written $E[X]$, is defined by

$$E[X] = \sum_{x:p(x)>0} xp(x).$$

- Represents weighted average of possible values X can take, each value being weighted by its probability.

Expectation when state space is countable

- If the state space S is countable, we can give **SUM OVER STATE SPACE** definition of expectation:

$$E[X] = \sum_{s \in S} P\{s\} X(s).$$

- Agrees with the **SUM OVER POSSIBLE X VALUES** definition:

$$E[X] = \sum_{x: p(x) > 0} x p(x).$$

Expectation of a function of a random variable

- If X is a random variable and g is a function from the real numbers to the real numbers then $g(X)$ is also a random variable.
- How can we compute $E[g(X)]$?
- Answer:
$$E[g(X)] = \sum_{x:p(x)>0} g(x)p(x).$$

Additivity of expectation

- If X and Y are distinct random variables, then $E[X+Y] = E[X] + E[Y]$.
- In fact, for real constants a and b, we have $E[aX + bY] = aE[X] + bE[Y]$.
- This is called the **linearity of expectation**.
- Can extend to more variables $E[X_1 + X_2 + \ldots + X_n] = E[X_1] + E[X_2] + \ldots + E[X_n]$.

Defining variance in discrete case

- Let X be a random variable with mean μ.
- The variance of X, denoted $\mathrm{Var}(X)$, is defined by $\mathrm{Var}(X) = E[(X-\mu)^2]$.
- Taking $g(x) = (x-\mu)^2$, and recalling that $E[g(X)] = \sum_{x:p(x)>0} g(x)p(x)$, we find that

$$\mathrm{Var}[X] = \sum_{x:p(x)>0} (x-\mu)^2 p(x).$$

- Variance is one way to measure the amount a random variable "varies" from its mean over successive trials.
- Very important alternate formula: $\mathrm{Var}[X] = E[X^2] - (E[X])^2$.

Identity

- If $Y = X + b$, where b is constant, then $\mathrm{Var}[Y] = \mathrm{Var}[X]$.
- Also, $\mathrm{Var}[aX] = a^2\mathrm{Var}[X]$.
- Proof: $\mathrm{Var}[aX] = E[a^2X^2] - E[aX]^2 = a^2E[X^2] - a^2E[X]^2 = a^2\mathrm{Var}[X]$.

Standard deviation

- Write $\mathrm{SD}[X] = \sqrt{\mathrm{Var}[X]}$.
- Satisfies identity $\mathrm{SD}[aX] = a\mathrm{SD}[X]$.
- Uses the same units as X itself.
- If we switch from feet to inches in our "height of randomly chosen person" example, then X, $E[X]$, and $\mathrm{SD}[X]$ each get multiplied by 12, but $\mathrm{Var}[X]$ gets multiplied by 144.

Bernoulli random variables

- Toss fair coin n times. (Tosses are independent.) What is the probability of k heads?
- Answer: $\binom{n}{k}/2^n$.
- What if coin has p probability to be heads?
- Answer: $\binom{n}{k}p^k(1-p)^{n-k}$.
- Writing $q = 1 - p$, we can write this as $\binom{n}{k}p^k q^{n-k}$
- Can use binomial theorem to show probabilities sum to one:
- $1 = 1^n = (p+q)^n = \sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}$.
- Number of heads is **binomial random variable with parameters** (n, p).

Decomposition approach to computing expectation

- Let X be a binomial random variable with parameters (n, p). Here is one way to compute $E[X]$.
- Think of X as representing number of heads in n tosses of coin that is heads with probability p.
- Write $X = \sum_{j=1}^{n} X_j$, where X_j is 1 if the jth coin is heads, 0 otherwise.
- In other words, X_j is the number of heads (zero or one) on the jth toss.
- Note that $E[X_j] = p \cdot 1 + (1-p) \cdot 0 = p$ for each j.
- Conclude by additivity of expectation that

$$E[X] = \sum_{j=1}^{n} E[X_j] = \sum_{j=1}^{n} p = np.$$

Compute variance with decomposition trick

- $X = \sum_{j=1}^n X_j$, so
  $E[X^2] = E[\sum_{i=1}^n X_i \sum_{j=1}^n X_j] = \sum_{i=1}^n \sum_{j=1}^n E[X_i X_j]$
- $E[X_i X_j]$ is p if $i = j$, p^2 otherwise.
- $\sum_{i=1}^n \sum_{j=1}^n E[X_i X_j]$ has n terms equal to p and $(n-1)n$ terms equal to p^2.
- So $E[X^2] = np + (n-1)np^2 = np + (np)^2 - np^2$.
- Thus
  $\mathrm{Var}[X] = E[X^2] - E[X]^2 = np - np^2 = np(1-p) = npq$.
- Can show generally that if $X_1, \ldots, X_n$ independent then
  $\mathrm{Var}[\sum_{j=1}^n X_j] = \sum_{j=1}^n \mathrm{Var}[X_j]$

Bernoulli random variable with n large and $np = \lambda$

- Let λ be some moderate-sized number. Say $\lambda = 2$ or $\lambda = 3$. Let n be a huge number, say $n = 10^6$.
- Suppose I have a coin that comes on heads with probability λ/n and I toss it n times.
- How many heads do I expect to see?
- Answer: $np = \lambda$.
- Let k be some moderate sized number (say $k = 4$). What is the probability that I see exactly k heads?
- Binomial formula:
  $\binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} p^k (1-p)^{n-k}$.
- This is approximately $\frac{\lambda^k}{k!}(1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}$.
- A **Poisson random variable** X with parameter λ satisfies $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$ for integer $k \geq 0$.

Expectation and variance

- A **Poisson random variable** X with parameter λ satisfies $P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}$ for integer $k \geq 0$.
- Clever computation tricks yield $E[X] = \lambda$ and $\mathrm{Var}[X] = \lambda$.
- We think of a Poisson random variable as being (roughly) a Bernoulli (n, p) random variable with n very large and $p = \lambda/n$.
- This also suggests $E[X] = np = \lambda$ and $\mathrm{Var}[X] = npq \approx \lambda$.

Poisson point process

- A Poisson point process is a random function $N(t)$ called a Poisson process of rate λ.
- For each $t > s \geq 0$, the value $N(t) - N(s)$ describes the number of events occurring in the time interval (s, t) and is Poisson with rate $(t - s)\lambda$.
- The numbers of events occurring in disjoint intervals are independent random variables.
- Probability to see zero events in first t time units is $e^{-\lambda t}$.
- Let T_k be time elapsed, since the previous event, until the kth event occurs. Then the T_k are independent random variables, each of which is exponential with parameter λ.

Geometric random variables

- Consider an infinite sequence of independent tosses of a coin that comes up heads with probability p.
- Let X be such that the first heads is on the Xth toss.
- Answer: $P\{X = k\} = (1-p)^{k-1}p = q^{k-1}p$, where $q = 1-p$ is tails probability.
- Say X is a **geometric random variable with parameter** p.
- Some cool calculation tricks show that $E[X] = 1/p$.
- And $\mathrm{Var}[X] = q/p^2$.

Negative binomial random variables

- Consider an infinite sequence of independent tosses of a coin that comes up heads with probability p.
- Let X be such that the rth heads is on the Xth toss.
- Then $P\{X = k\} = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}p$.
- Call X **negative binomial random variable with parameters** (r, p).
- So $E[X] = r/p$.
- And $\mathrm{Var}[X] = rq/p^2$.

18.440 Probability and Random Variables
Spring 2014