THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

Lecture 12: Review

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Conditional Expectation

- Recall the **expectation** of X. $E[X] = \sum_x xP(X = x)$.
- The **conditional expectation** of random variable X given event A with $P(A) > 0$ is defined as: $E[X|A] = \sum_x xP(X = x|A)$.
- For a function $g(X)$, $E[g(X)|A] = \sum_x g(x)P(X = x|A)$.

Total expectation theorem

For a partition $\{A_1, \ldots, A_n\}$

$$P(X = k) = \sum_{i}^{n} P(X = k|A_i)P(A_i)$$

$$E[X] = \sum_{i=1}^{n} E[X|A_i]P(A_i)$$

Example

I have two coins: one fair and one biased ($p = .25$). I pick one at random and toss it a hundred times. On an average how many heads will I see?

- $A = \{$I pick fair coin$\}$.
- $X = \{$Number ofheads in 100 tosses$\}$.
- $E[X] =?$

Example

I have two coins: one fair and one biased ($p = .25$). I pick one at random and toss it a hundred times. On an average how many heads will I see?

- $A = \{\text{I pick fair coin}\}$.
- $X = \{\text{Number ofheads in 100 tosses}\}$.
- $E[X] =?$
- $E[X] = E[X|A]P(A) + E[X|A^c]P(A^c) = 100 \times .5 \times .5 + 100 \times .25 \times .5 = 25 + 12.5 = 37.5$

Example

I have two coins: one fair and one biased ($p = .25$). I pick one at random and toss it until I see a head. On an average how long do I have to wait until I see a head?

- $A = \{\text{I pick fair coin}\}$.
- $X = \{\text{Number of tosses to get a head}\}$.
- $E[X] =?$

Example

I have two coins: one fair and one biased ($p = .25$). I pick one at random and toss it until I see a head. On an average how long do I have to wait until I see a head?

- $A = \{\text{I pick fair coin}\}$.
- $X = \{\text{Number of tosses to get a head}\}$.
- $E[X] = ?$
- $E[X] = E[X|A]P(A) + E[X|A^c]P(A^c) = 1/.5 \times .5 + 1/.25 \times .5 = 2.5$

Example

Four buses carrying 148 students from the same school arrive at a football stadium. The buses carry, respectively, 40, 33, 25, and 50 students. One of the students is randomly selected. Let X denote the number of students that were on the bus carrying the randomly selected student. One of the 4 bus drivers is also randomly selected. Let Y denote the number of students on her bus.

Axioms of probability

Our probability law must follow three axioms:

1. **Nonnegativity**: $\mathbf{P}(A) \geq 0$, for every event A.
2. **Additivity**: If A and B are two disjoint events, then the probability of their union satisfies $P(A \cup B) = P(A) + P(B)$
   This extends to the union of infinitely many disjoint events:
   $$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots$$
3. **Normalization**: The probability of the entire sample space Ω is equal to 1, i.e. $P(\Omega) = 1$

From these axioms, and from basic properties of sets, we can derive a number of other useful results, including:

- For any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A^c) = 1 - P(A)$

Practice problem—Probability 2

A total of 28 percent of American males smoke cigarettes, 7 percent smoke cigars, and 5 percent smoke both cigars and cigarettes.

- What percentage of males smokes neither cigars nor cigarettes?

- What percentage smokes cigars but not cigarettes?

Practice problem—Probability 2

A total of 28 percent of American males smoke cigarettes, 7 percent smoke cigars, and 5 percent smoke both cigars and cigarettes.

- What percentage of males smokes neither cigars nor cigarettes?
  - $P(C_1) = .28$ and $P(C_2) = .07$, and $P(C_1 \cap C_2) = .05$
  - $P(C_1^c \cap C_2^c) = P((C_1 \cup C_2)^c) = 1 - P(C_1 \cup C_2) = 1 - (.28 + .07 - .05) = 1 - 0.3 = 0.7$
  - The last part used De-morgan's law.
- What percentage smokes cigars but not cigarettes?
  - $P(C_1 \cap C_2^c) = P(C_1) - P(C_1 \cap C_2) = .07 - .05 = .02$

Counting

Find a formula for the probability that among a set of n people, **at least two have their birthdays in the same month** of the year (assuming the months are equally likely for birthdays). Assume $n \leq 12$ (for $n > 13$ the probability is equal to 1.0). In your solution let $A =$"at least two matching birthday months."

- Easier to think about $P(A^c) = 1 - P(A)$
- $A^c =$\{No matching b-day months\}
- How many elements are there in A^c? $|A^c| =$?
  $12 \times 11 \times 10 \times \ldots (12 - n + 1) = (12)_n$
- The number of possible n-tuples of birthday months is 12^n.
- $P(A^c) = \frac{(12)_n}{12^n}$

Things to remember

- Probability axioms
- Conditional probability
  - Independence
  - Conditional independence
  - Mutual vs. Pairwise independence
  - Bayes rule
  - Rule of total probability

Counting-Occupancy numbers (stars and bars)

- Say you want to distribute n fruits among k children so that everyone gets at least 1.
- Say x_i is the number of fruits going to child i. So we are looking for number of k-tuples of positive integers such that $x_1 + x_2 + \cdots + x_k = n$.
- Writing in stars and bars, you want to place $(k-1)$ bars in $(n-1)$ spaces between the stars.
- **number of k-tuples of positive integers such that $\mathbf{x_1 + x_2 + \cdots + x_k = n}$. This is $\binom{\mathbf{n-1}}{\mathbf{k-1}}$**

Occupancy numbers (stars and bars)

Now I want to divide 10 fruits among 4 children. A child may or may not get any fruit. How many ways to do this?

- We want to use our former idea, but how?
- Why not divide 14 fruits to 4 children so that everyone has at least one, and then remove one fruit from each? The fruits are indistinguishable and so it does't matter which one you take out.
- This way everyone will have zero or more fruits, and total number of fruits is 10.
- So the answer is $\binom{13}{3}$.
- **number of k-tuples of non-negative integers such that $x_1 + x_2 + \cdots + x_k = n$. This is $\binom{n+k-1}{k-1}$**

Occupancy problem

How many ways can we distribute 7 identical pieces of candy to 4 children?

- Same as asking for the total number of solutions to $x_1 + x_2 + x_3 + x_4 = 7$, where x_i are non-negative integers.
- Stars and bars: $\binom{7+4-1}{4-1} = \binom{10}{3}$

How many ways can we distribute 7 identical pieces of candy to 4 children so that each has at least one candy?

- Same as asking for the total number of solutions to $x_1 + x_2 + x_3 + x_4 = 7$, where x_i are **positive** integers.
- Stars and bars again: I want to put 3 bars between 7 stars. There are $7 - 1$ places to choose from and so the answer is $\binom{6}{3}$

Occupancy problem

How many different combinations of 5 sweaters can you buy, if you have 8 different colors?

- Let x_i be number of sweaters bought with the i^{th} color. $i \in \{1, \ldots 8\}$.
- I want to count the number of ways I can have $\sum_{i=1}^{8} x_i = 5$, where $x_i \geq 0$.
- But this is $\binom{5+8-1}{8-1} = \binom{12}{7}$

Number of permutations when you have repetitions

How many distinct arrangements are there of the word "MISSISSIPPI"?

- Total number of letters? 11.
- M appears once, I and S appears 4 times, P twice.
- So number of distinct arrangements: $\frac{11!}{1!4!4!2!}$

How many distinct arrangements are there of the word "MISSISSIPPI" where the first and last letter are vowels?

- How many distinct arrangements are of ($\underline{I}$, , , , , , , ,$\underline{I}$)? Remaining letters (MISSISSPPI) can be arranged in $10!/(3!4!2!)$ ways.

Practice problem—Combinatorics 1a

There is a bucket of 10 red and 10 blue balls. I pick 5 balls without replacement. What is the probability that there will be 3 red and 2 blue balls?

Practice problem—Combinatorics 1a

There is a bucket of 10 red and 10 blue balls. I pick 5 balls without replacement. What is the probability that there will be 3 red and 2 blue balls?

- Total number of ways to pick 5 balls: $\binom{20}{5}$

Practice problem—Combinatorics 1a

There is a bucket of 10 red and 10 blue balls. I pick 5 balls without replacement. What is the probability that there will be 3 red and 2 blue balls?

- Total number of ways to pick 5 balls: $\binom{20}{5}$
- Total number of ways to pick 3 red $\binom{10}{3}$
- Total number of ways to pick 2 blue $\binom{10}{2}$

Practice problem—Combinatorics 1a

There is a bucket of 10 red and 10 blue balls. I pick 5 balls without replacement. What is the probability that there will be 3 red and 2 blue balls?

- Total number of ways to pick 5 balls: $\binom{20}{5}$
- Total number of ways to pick 3 red $\binom{10}{3}$
- Total number of ways to pick 2 blue $\binom{10}{2}$
- $\frac{\binom{10}{3}\binom{10}{2}}{\binom{20}{5}}$

Practice problem—Probability 3

A closet contains 10 pairs of shoes. If 8 shoes are randomly selected, what is the probability that there will be

- no complete pair?
- exactly 1 complete pair?

Practice problem—Probability 3

A closet contains 10 pairs of shoes. If 8 shoes are randomly selected, what is the probability that there will be

- no complete pair?
  - Total number of ways $\binom{20}{8}$.
  - Choose 8 out of 10 pairs. Now look at all possible ways to choose a shoe from a pair. $\binom{10}{8}2^8$
  - So the probability is $\frac{\binom{10}{8}2^8}{\binom{20}{8}}$
- exactly 1 complete pair?
  - Pick 1 pair in 10 ways.
  - Now there are 6 more to pick from remaining 9 pairs, but there can be no complete pair.
  - This is $\binom{9}{6}2^6$.
  - So probability is $\frac{10 \times \binom{9}{6}2^6}{\binom{20}{8}}$

Things to remember

- Permutations
- Combinations (binomial coefficients)
- Occupancy numbers (dividing indistinguishable objects among distinguishable people/bins/children)
- Counting with repetition
- Counting and probability

Cumulative distribution functions

- For a **discrete** random variable, to get the probability of X being in a range B, we **sum** the PMF over that range:

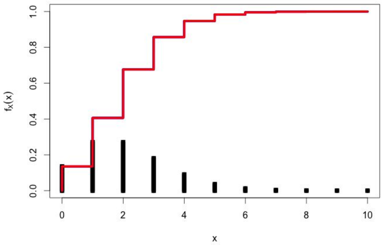
$$P(X \in B) = \sum_{x \in B} p_X(x)$$

e.g. if $X \sim Binomial(10, 0.2)$

$$P(2 < X \leq 5) = p_X(3) + p_X(4) + p_X(5) = \sum_{k=3}^{5} \binom{10}{k} 0.2^k (1 - 0.2)^{10-k}$$

Cumulative distribution functions

- In both cases, we call the probability $P(X \leq x)$ the **cumulative distribution function** (CDF) $F_X(x)$ of X



Common discrete random variables

We have looked at four main types of discrete random variable:

- **Bernoulli**: We have a biased coin with probability of heads p. A Bernoulli random variable is 1 if we get heads, 0 if we get tails.
  - If $X \sim Bernoulli(p)$, $p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{otherwise.} \end{cases}$
  - Examples: Has disease, hits target.
- **Binomial**: We have a sequence of n biased coin flips, each with probability of heads p – i.e. a sequence of n independent $Bernoulli(p)$ trials. A Binomial random variable returns the number of heads.
  - If $X \sim Binomial(n, p)$, $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - p^k because we have heads (prob. p) k times, $(1-p)^{n-k}$ because we have tails (prob. $1-p$) $n-k$ times.
  - Why $\binom{n}{k}$? Because this is the number of sequences of length n that have exactly k heads.
  - Examples: How many people will vote for a candidate, how many of my seeds will sprout.

Common discrete random variables

- **Geometric**: We have a biased coin with probability of heads p. A geometric random variable returns the number of times we have to throw the coin before we get heads. e.g. If our sequence is $(T, T, H, T, \dots)$, then $X = 3$.
  - If $X \sim \mathit{Geometric}(p)$, $p_X(k) = (1-p)^{k-1}p$
  - Prob. of getting a sequence of $k-1$ tails and then a head.
  - $E[X] = 1/p$ (should know this) and $\mathrm{var}(X) = (1-p)/p^2$ (don't need to know this)
  - $P(X > k) = 1 - P(X \leq k) = (1-p)^k$
  - Memoryless property: $P(X > k + j | X > j) = P(X > k)$
- **Poisson**: Independent events occur, on average, λ times over a given period/distance/area. A Poisson random variable returns the number of times they actually happen.
  - If $X \sim \mathit{Poisson}(\lambda)$, $p_X(k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$.
  - $E[X] = \lambda$ and $\mathrm{var}(X) = \lambda$.
- The Poisson distribution with $\lambda = np$ is a good approximation to the $\mathit{Binomial}(n, p)$ distribution, when n is large and p is small.

Things to remember

- What is a PMF?
- What is a CDF?
- How to get expectation and variance
- PMF, Expectation and variance of Bernoulli, Binomial, Uniform, Geometric, Poisson
- Conditional PMF
  - conditional expectation
  - total expectation theorem

Practice problem—Random variables 4a

The suicide rate in a certain state is 1 suicide per $100,000$ inhabitants per month. You can treat the number of suicides in each month as independent.

1. Find the probability that, in a city of $400,000$ inhabitants within this state, there will be 8 or more suicides in a given month.
   - $X \sim Poisson(4)$. We want $P(X \geq 8)$.
   - $q = P(X \geq 8) = 1 - \sum_{k=0}^{7} \frac{e^{-4}4^k}{k!}$

Practice problem—Random variables 4b

The suicide rate in a certain state is 1 suicide per $100,000$ inhabitants per month. You can treat the number of suicides in each month as independent.

1. Find the probability that, in a city of $400,000$ inhabitants within this state, there will be 8 or more suicides in a given month.
2. What is the probability that there will be at least 2 months during the year that will have 8 or more suicides?

Practice problem—Random variables 4b

The suicide rate in a certain state is 1 suicide per $100,000$ inhabitants per month. You can treat the number of suicides in each month as independent.

1. Find the probability that, in a city of $400,000$ inhabitants within this state, there will be 8 or more suicides in a given month.
2. What is the probability that there will be at least 2 months during the year that will have 8 or more suicides?
   - Define $Y_i, i \in \{1, \ldots, 8\}$ as

$$Y_i = \begin{cases} 1 & \text{If } X_i \geq 8 \\ 0 & \text{o.w.} \end{cases}$$

$$P(Y_i = 1) = P(X_i \geq 8) = q$$

   - Let $Z = \sum_i Y_i$. We want $P(Z \geq 2)$.
   - But $Z \sim Bin(12, q)$. So $P(Z \geq 2) = 1 - P(Z = 0) - P(Z = 1) = 1 - (1-q)^{12} - 12q(1-q)^{11}$