COMPSCI 240: Reasoning Under Uncertainty

Nic Herndon and Andrew Lan

University of Massachusetts at Amherst

Spring 2019

Lecture 9: Common Discrete Random Variables

Announcements

- Homework submissions – be kind to the graders if you want them to be kind to you (if we can't read it, we can't give you points for your work)

4.7 Homework Submission

All homework assignments will be submitted as PDF files on Gradescope. You can scan your homework and convert it to PDF with the Evernote Scannable app available for iOS and Android devices, or by using a scanner on campus. When submitting on Gradescope, be sure to select the pages where the answers to the questions are, and save the selection. Otherwise, we will have to search through your homework for the answer and may end up deducting points or not viewing your solution. Also, be sure to submit to Gradescope early (not at 11:54PM if due at 11:55PM), because there may be issues with uploading. For more details on using these apps with Gradescope, please read the Gradescope guide. The gradescope course site is at `https://www.gradescope.com/courses/37854/` and the entry code is MVD8KX.

1

Problem 1 $(3+3+3=9)$
Suppose you draw two cards from a deck of 52 cards without replacement.

1) What's the probability that the first draw is a heart and the second draw is not a heart?
2) What's the probability that exactly one of the cards are hearts?
3) If you draw two cards *with* replacement, what's the probability that none of the cards are hearts?

Problem 2 (4)
A factory produces 85 T-shirts and 10 sweatshirts each hour. If 3 shirts (either T-shirts or sweatshirts) are picked at random then what is the probability that all of them are T-shirts?

Problem 3 (3+4=7)
There are 4 bags each containing 100 marbles. Bag 1 has 40 red and 60 blue marbles. Bag 2 has 30 red and 70 blue marbles. Bag 3 has 75 red and 25 blue marbles. Bag 4 has 50 red and 50 blue marbles. Now a bag is chosen at random and a marble is also picked at random.

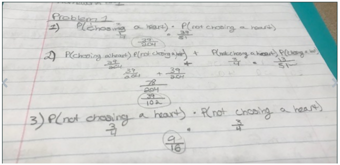
1) What is the probability that the marble is red?
2) What happens when the first two bags are chosen with probability 0.3 each and other two bags are chosen with probability 0.2 each?

Probem 4 (10)
The disc containing the only copy of your homework just got corrupted, and the disk got mixed up with two other corrupted discs that were lying around. It is equally likely that any of the three discs holds the corrupted remains of your homework. Your computer expert friend offers to have a look, and you know from past experience that his probability of finding your homework from any disc is 0.35 (assuming the homework is there). Given that he searches on disc 1 but cannot find your homework, what is the probability that your homework is on disc i, for $i = 1, 2, 3$?

Problem 5 (4+4+4=12)
We roll two fair 6-sided dice, A and B. Each one of the 36 possible outcomes is assumed to be equally likely.

1) Find the probability that dice A is larger than dice B.
2) Given that the roll resulted in a sum of 5 or less, find the conditional probability that the two dice were equal.
3) Given that the two dice land on different numbers, find the conditional probability that the two dice differed by 2.

Problem 6 (8)
For any events A, B, C, and $D = A \cap B \cap C$ prove the following equality:



Discrete Uniform Random Variables

- A *discrete uniform random variable* X with range $[a, b]$ takes on any integer value between (and including) a and b with the same probability
- For example, the random variable that maps a fair six-sided dice roll to the number that comes up is a uniform random variable with $a = 1$, $b = 6$ and $P(X = k) = 1/6$ for $k = 1, ..., 6$.
- The PMF of a discrete uniform random variable X is

$$P(X = k) = \frac{1}{b - a + 1} \text{ for } k = a, \ldots, b$$

- Used to model probabilistic situations where each of the values $a, ..., b$ are equally likely.

Bernoulli Random Variables

- Suppose we have an experiment with two outcomes H and T. H happens with probability p and T with probability $1 - p$, $0 < p < 1$.
- We define a random variable X such that $X(H) = 1$ and $X(T) = 0$.
- This is called a Bernoulli random variable X that takes the two values 0 or 1.
- Its PMF looks like

$$P(X = k) = \begin{cases} 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

- You can also define $X(H) = 0$ and $X(T) = 1$, with $P(X = 1) = p' = 1 - p$

Bernoulli Random Variables: Examples

- Whether a coin lands heads or tails.
- Whether a server is online or offline.
- Whether an email is spam or not.
- Whether a pixel in a black and white image is black or white.
- Whether a patient has a disease or not.

Binomial Random Variable

- A binomial random variable is the combination of independent and identically distributed Bernoulli random variables
- Suppose we flip n coins independently, where each coin has probability p of being heads
- The set of outcomes is:

$$\Omega = \{(TTT \ldots TT), (TTT \ldots TH), \ldots, (HHH \ldots HH)\}$$

- Define a random variable X where for each $o \in \Omega$,

$$X(o) = \text{ "the number of heads in outcome } o\text{"}$$

- We're already shown that $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Binomial Random Variables: Examples

- The number of heads in N coin tosses.
- The number of servers that fail in a cluster of N servers.
- The number of games a football team wins in a season of N games (assuming i.i.d.).
- The number of True/False questions you get correct if you guess each of N questions.

Possible Exam Question

$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ is the formula for two possible outcomes of coin toss: Head and Tail.

If now we have three different outcomes say Head, Tail and Edge with a probability of p, q, and r, respectively ($p + q + r = 1$).

Suppose you toss a coin n times. What is the probability of k heads, ℓ tails, m edges and in n tosses ($n = k + \ell + m$)?

(Note that this is NOT a binomial random variable as we have three possible outcomes.)

$$P(k \text{ heads}, l \text{ tails}, \text{ and } m \text{ edges}) = \frac{n!}{k! \cdot l! \cdot m!} \cdot p^k \cdot q^l \cdot r^m$$

Geometric Random Variables

- Suppose we flip a biased coin repeatedly until it lands heads. Let X be the number of tosses needed for a head to come up for the first time.
- The PMF of a *geometric random* variable X is

$$P(X = k) = (1 - p)^{k-1} \cdot p \quad \text{for } k = 1, 2, 3, \ldots$$

- Used to model the number of repeated independent trials up to (and including) the *first "successful" trial*.
- Example: the number of patients we test before the first one we find who has a given disease.

Geometric Random Variables

- Prove the normalization of a geometric random variable.
  Hint: Infinite Geometric Series (IGS):
  $1 + r + r^2 + r^3 + \cdots = \frac{1}{1-r}$, where $-1 < r < 1$.

$$\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty}(1-p)^{k-1}p = p\sum_{k=0}^{\infty}(1-p)^k \overset{\text{IGS}}{=} p\cdot\frac{1}{1-(1-p)} = 1$$

Geometric Random Variables: Examples

- Products made by a machine have a 3% defective rate.
- What is the probability that the first defect occurs in the fifth item inspected?

$$P(X = k) = (1 - p)^{k-1} \cdot p = (1 - 0.03)^{5-1} \cdot 0.03 = 0.0265\ldots$$

Poisson Random Variables

- A *Poisson random variable* X is a random variable that has the following PMF

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

- **WHAT?!** Where did this come from?

Poisson Random Variables

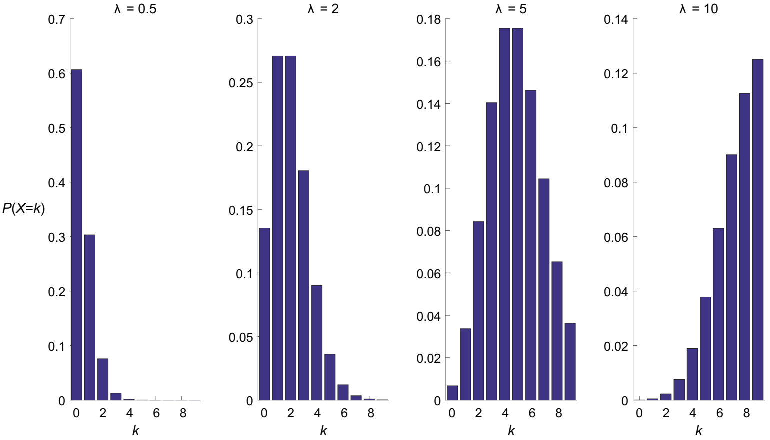
- A *Poisson random variable* X is a random variable that has the following PMF

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

- The Poisson distribution is one of the most widely used probability distributions.
- Built based on Taylor series: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.
- Think about Poisson RV as a framework that provides approximation of a real-life random variable as a function of λ.

Poisson Random Variables

- Think about Poisson RV as a framework that provides approximation of different PMFs as a function of λ.



Poisson Random Variables

- It is generally used in scenarios where we are counting the occurrences of certain events within an interval of time or space.
  - The number of typos in a book with n words.
  - The number of cars that crash in a city on a given day.
  - The number of phone calls arriving at a call center per minute etc.
- λ represents the expected number of events (we will learn more about this).
  - The average number of typos in a book.
  - The average number of car crash per day.
  - The average number of phone calls per minute.

Poisson Random Variables: Example

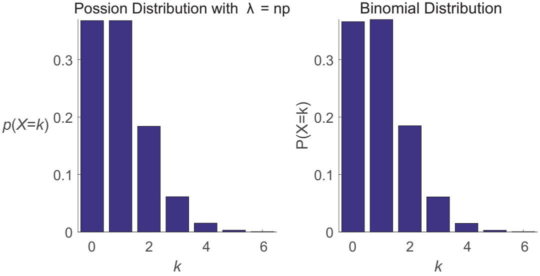
- Suppose that the number of phone calls arriving at a call center per minute can be modeled by a discrete Poisson PMF.
- In average, the call center receives 10 calls.
- What is the probability that the center will receive 5 calls?

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$P_X(k) = e^{-10} \frac{10^5}{5!} = 0.0378...$$

Poisson Random Variables

- A Poisson PMF with $\lambda = np$ is a *good approximation* for a binomial PMF with very small p and very large n if $k \ll n$
  - A bionomial RV X is the number of heads (k) in the n-toss sequence, where the coin comes up a head with probability p.
- Example: $n = 100$ and $p = 0.01$ for the binomial r.v. where as $\lambda = np$ for the Poisson r.v.



- Poisson PMF provides much simpler models and calculations: $\binom{n}{k} p^k (1-p)^{n-k}$ **vs.** $e^{-\lambda} \frac{\lambda^k}{k!}$

Summary: Discrete Random Variables

- **Uniform:** For $k = a, \ldots, b$:

$$P(X = k) = \frac{1}{b - a + 1}$$

- **Bernoulli:** For $k = 0$ or 1:

$$P(X = k) = \begin{cases} 1 - p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$$

- **Binomial:** For $k = 0, \ldots, N$

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

- **Geometric:** For $k = 1, 2, 3, \ldots$, $P(X = k) = (1 - p)^{k-1} \cdot p$
- **Poisson:** $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ for $k = 0, 1, 2, \ldots$

Example

- Let X be the number of times you toss a dice until you see a six. Then X is

  A)Binomial B)Geometric C)Uniform D)Poisson E)Bernoulli

  - The answer is Geometric.

- Suppose you toss a dice ten times and let X be the number of times you saw a six. Then X is

  A)Binomial B)Geometric C)Uniform D)Poisson E)Bernoulli

  - The answer is Binomial.

Challenging Problem

Textbook Problem 11 (reworded): A smoker carries one matchbox in his right pocket and one in his left pocket. Each time he wants to light a cigarette, he selects a matchbox from either pocket with equal probability, independent of earlier selections. The two matchboxes have initially n matches each.

Suppose that the smoker reached for a match and discovered that the corresponding matchbox is empty. What is the PMF of the number of remaining matches in the matchbox at the opposite pocket?

Challenging Problem

Solution (reworded):

1. If k matches are remaining in one pocket (say left) and 0 remaining in the other pocket (say right), it is equivalent as if we have selected $n-k$ matches from the left and n matches from the right pocket, respectively.
2. Similarly to how we computed the Binomial PMF, the probability of a single sequence that we have selected $n-k$ from the left and n from the right pocket is:
$$\frac{1}{2}^{n-k} \times \frac{1}{2}^{n} = \frac{1}{2}^{2n-k}.$$
3. Then, out of a total of $(n-k)+n = 2n-k$ matches that have been selected, we want to count the number of combinations where exactly n of them are from the right pocket (so that it is empty), which is:
$$\binom{2n-k}{n}.$$
4. Thus,
$$P(X=k) = \binom{2n-k}{n}\frac{1}{2}^{2n-k},$$
where $k = 0, 1, 2, \cdots, n$.

Next Lecture: Can you make money from Roulette?