4: Discrete Distributions

PSTAT 120A: Summer 2022

Ethan P. Marzban

June 29, 2022

University of California, Santa Barbara

Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)

Discrete Distributions

Bernoulli Trial

Definition: Bernoulli Trial

A **Bernoulli Trial** is an experiment in which:

- There is a well-defined notion of “success” and “failure” (i.e. non-success)
- The probability of success remains a constant value p over all repetitions of this trial.

- Tossing a coin is an example of a Bernoulli Trial; “sucess” could be “lands heads”, and whether or not the coin is fair we assume there to be a fixed probability p of the coin landing heads.

Leadup

- Consider the following three random variables:
  1. Toss a fair coin 100 times, and let X denote the number of heads.
  2. Roll a fair six-sided die 27 times, and let Y denote the number of times the die lands on an even number.
  3. From a population of size 1000, in which 4 people have a particular disease, take a sample of 100 people with replacement and let Z denote the number of individuals with diseases I observe.
- For each of these experiments and associated random variables, we could follow the same steps as we did when dealing with our two-coin example: in other words, we could construct Ω, find the mapping X (or Y or Z), construct the p.m.f., and find $\mathbb{E}[X]$ (or $\mathbb{E}[Y]$ or $\mathbb{E}[Z]$).
- But, notice that each of the scenarios listed above are all just special cases of the following:

  In n independent Bernoulli trials, where each trial results in a "success" with probability p, let W denote the number of successes.

- So, if we can deal with this general case, we can simply plug in different values of n and p.

Distributions

- This is how I like to think about distributions: as a "package" which deals with some general question in generality, from which we can glean information on individual situations.
- The true technical definition of a distribution is much more technical! (But, for the purposes of this class, this notion of a distribution as a "package" will suffice.)

- So, for example: if W denotes the number of successes in n independent Bernoulli trials, and where the probability on any given trial is p, we say W follows the **Binomial** distribution with **parameters** n and p, and notate this $W \sim \text{Bin}(n, p)$.

Binomial Distribution

- Let's try an example.
- Suppose $W \sim \text{Bin}(n, p)$; in other words, W denotes the number of successes in n independent Bernoulli trials with probability of success p.
- We can derive the p.m.f. of W using some counting arguments:
  - When computing $p_W(k)$, we are computing the probability of exactly k successes in n trials.
  - Suppose that these k trials occurred consecutively, as my first k trials. The probability of this is simply $p^k(1-p)^{n-k}$.
  - But, the event $\{W = k\}$ doesn't mean "k successes all at the beginning," but rather "k successes across all n trials." Thus, we need to multiply by all of the ways in which we can distribute the k successes among the n trials: $\binom{n}{k}$.
  - That is:
  $$p_W(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k = 0, 1, 2, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$
  - This is the p.m.f. of the Binomial distribution with parameters n and p.

Binomial Distribution

- With the p.m.f. of W, we can now compute $\mathbb{E}[W]$:

$$
\begin{aligned}
\mathbb{E}[W] := \sum_k k p_W(k) &= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(k-1!)(n-k)!} \cdot p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} n \cdot \frac{(n-1)!}{(k-1)!([n-1]-[k-1])!} p^k (1-p)^{n-k} \\
&= n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k} \\
&= n \sum_{m=0}^{n-1} \binom{n-1}{m} p^{m+1} (1-p)^{n-m-1} \\
&= np \sum_{m=0}^{n-1} \binom{n-1}{m} p^m (1-p)^{(n-1)-m} = np\cancel{(p+1-p)^{n-1}} = np
\end{aligned}
$$

Binomial Distribution

- With a bit of work, one can show that $\text{Var}(W) = np(1-p)$

- So, to summarize: if W counts the number of successes in n independent Bernoulli trials, then $W \sim \text{Bin}(n, p)$ and:
  - $S_W = \{0, 1, \cdots, n\}$
  - $p_W(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{if } k = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$
  - $\mathbb{E}[W] = np$
  - $\text{Var}(W) = np(1-p)$

Example

Suppose I simultaneously roll 10 fair six-sided dice, and let X denote the number of even numbers showing.

(a) What is the probability that X is 2?

(b) What is $\mathbb{E}[X]$?

(c) What is $\text{Var}(X)$?

- We have a well-defined notion of success: "die lands on an even number."
- Since the coin is fair, we can use the classical definition of probability to say $p := \mathbb{P}(\text{success}) = \mathbb{P}(\text{even number}) = \mathbb{P}(\{2, 4, 6\}) = 1/2$
- Additionally, we have $n = 10$ Bernoulli Trials (one corresponding to each die roll), meaning $X \sim \text{Bin}(10, 1/2)$
- From here, we can easily answer each of the subquestions using our information on the Binomial distribution!

  (a) $\mathbb{P}(X = 2) = \binom{10}{2} \left(\frac{1}{2} \right)^2 \left(\frac{1}{2} \right)^{10-2} = \boxed{\frac{45}{1024}}$

  (b) $\mathbb{E}[X] = (10)\left(\frac{1}{2} \right) = \boxed{5}$; $\text{Var}(X) = (10)\left(\frac{1}{2} \right)\left(1 - \frac{1}{2} \right) = \boxed{5/2}$

Another Distribution:

- Consider again a sequence of Bernoulli trials.
- Now, however, let X denote the number of trials needed to observe our first success? (Let's include the final successful trial when counting). So, for example, if we observe

  $$\text{(Failure) (Failure) (Failure) (Success)}$$

  then $X = 4$.
- What is the state space of X? $S_X = \{1, 2, 3, \cdots\}$
- To find the p.m.f., we can construct a modified slot diagram. Specifically, when $X = k$ we must have $(k-1)$ failures followed by one success:

  $$\underbrace{\underline{\text{Failure}}\&\underline{\text{Failure}}\&\cdots\&\underline{\text{Failure}}}_{k-1 \text{ trials}}\&\underline{\text{Success}}$$

- Therefore $\mathbb{P}(X = k) = (1-p)^{k-1} \cdot p$, meaning

  $$p_X(k) = \begin{cases} (1-p)^{k-1} \cdot p & \text{if } k = 1, 2, 3, \cdots \\ 0 & \text{otherwise} \end{cases}$$

- This is called the **Geometric Distribution**, with parameter p.

Geometric Distribution: Expectation and Variance

- We can now find $\mathbb{E}[X]$, if $X \sim \text{Geom}(p)$

$$
\begin{aligned}
\mathbb{E}[X] = \sum_k p_X(k) &= \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p \\
&= \frac{p}{1-p} \sum_{k=1}^{\infty} k \cdot (1-p)^k \\
&= \frac{p}{1-p} \sum_{k=0}^{\infty} k \cdot (1-p)^k \\
&= \frac{p}{1-p} \times \frac{1-p}{[1-(1-p)]^2} = \frac{p}{1-p} \times \frac{1-p}{p^2} = \boxed{\frac{1}{p}}
\end{aligned}
$$

- You will also show that $\text{Var}(X) = \frac{1-p}{p^2}$ (there is a very neat trick to this computation!)

Geometric Distribution

- So, to summarize: if X counts the number of independent Bernoulli trials (including the final successful trial) needed to observe the first success, we have $X \sim \text{Geom}(p)$ and:
  - $S_X = \{1, 2, 3, \cdots\}$
  - $p_X(k) = \begin{cases} (1-p)^{k-1} \cdot p & \text{if } k = 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$
  - $\mathbb{E}[X] = \dfrac{1}{p}$
  - $\text{Var}(X) = \dfrac{1-p}{p^2}$

- As an example: suppose we want to know the average number of rolls of a fair six-sided die needed to observe the number "1" for the first time. Letting X denote the number of rolls until we observe our first "1" we have $X \sim \text{Geom}(1/6)$, meaning

$$\mathbb{E}[X] = \frac{1}{(1/6)} = \boxed{6}$$

Extending the Geometric Distribution

- We have seen that the Geometric distribution arises when counting the number of trials until our first success.
- What if we wanted to count the number of trials until our second success? or our third success?
- Let X denote the number of independent Bernoulli trials needed to observe the r^{th} success, where $r \in \mathbb{N}$.
- The state space of X is $S_X = \{r,\ r+1,\ r+2,\ \cdots\}$
- For the event $\{X = k\}$ to have occurred, we require $(r-1)$ successes among the first $k-1$ trials, followed by a success on the k^{th} trial:

$$\underbrace{\underline{\qquad}\ \ \underline{\qquad}\ \ \cdots\ \ \underline{\qquad}}_{(r-1)\text{ successes in }(k-1)\text{ trials}}\ \ \underline{\text{success}}$$

- The probability of observing $(r-1)$ successes in $(k-1)$ trials can be computed using the Binomial distribution! The probability of this is

$$\binom{k-1}{r-1} \cdot p^{r-1} \cdot (1-p)^{k-r}$$

- Therefore, $\mathbb{P}(X = k)$ is given by

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} \cdot p^{r-1} \cdot (1-p)^{k-r} \cdot p = \binom{k-1}{r-1} \cdot p^{r} \cdot (1-p)^{k-r}$$

The Negative Binomial Distribution

- Because of the presence of the Binomial distribution in our computation above, this new distribution is called the **Negative Binomial** distribution with parameters r and p.
- So, to summarize: if X counts the number of independent Bernoulli trials needed to observe r^{th} success then $X \sim \text{NegBin}(r, p)$ and:
  - $S_X = \{r,\ r+1,\ r+2, \cdots\}$
  - $p_X(k) = \begin{cases} \binom{k-1}{r-1} \cdot p^r \cdot (1-p)^{k-r} & \text{if } k = r,\ r+1,\ r+2,\ \cdots \\ 0 & \text{otherwise} \end{cases}$
  - $\mathbb{E}[X] = \dfrac{r}{p}$
  - $\text{Var}(X) = \dfrac{r}{p^2}$

The Negative Binomial Distribution

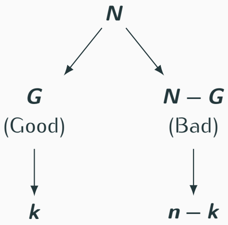
When tossing a fair coin, what is the probability that the fourth heads occurs on the 12^{th} toss?

- Let X denote the number of tosses needed to observe the fourth heads; then $X \sim \text{NegBin}(4, 1/2)$
- We seek $\mathbb{P}(W = 12)$; by the formula for the p.m.f. of the Negative Binomial distribution we have

$$\mathbb{P}(W = 12) = \binom{12-1}{4-1} \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{12-4} = \binom{11}{3} \left(\frac{1}{2}\right)^{12}$$

- By the way, the $\text{NegBin}(1, p)$ distribution has another name. What is that name? The Geometric(p) distribution.

- Now, suppose we have a lot of N items; G of which are good and the remaining $B := N - G$ of which are bad. If I take a sample of size n without replacement, I can let X denote the number of good elements in my sample.
- We have actually already found the p.m.f. of X, back when we did tree diagrams!
- In other words, to compute $\mathbb{P}(X = k)$ we have



$$\Longrightarrow \frac{\binom{G}{k}\binom{N-G}{n-k}}{\binom{N}{n}}$$

- X is said to follow the **Hypergeometric Distribution**, with parameters $N,\ G$, and n: $X \sim \text{HyperGeom}(N, G, n)$.
  - Note that the hypgergeometric distribution has three parameters! It may be difficult to remember what those three are; here's how I remember them. The first parameter is the population size, the second is the number of good elements, and the final parameter is the sample size.

Hypergeometric Distribution

- With a bit of work, one can see that if $X \sim \text{HyperGeom}(N, G, n)$ we have:
  - $S_X = \{\max\{0,\ n + G - N\},\ \cdots,\ \min\{n, G\}\}$
  - $p_X(k) = \begin{cases} \frac{\binom{G}{k}\binom{N-G}{n-k}}{\binom{N}{n}} & \text{if } k \in S_X \\ 0 & \text{otherwise} \end{cases}$
  - $\mathbb{E}[X] = n \cdot \frac{G}{N}$
  - $\text{Var}(X) = n \cdot \left(\frac{G}{N}\right) \cdot \left(1 - \frac{G}{N}\right) \cdot \left(\frac{N-n}{N-1}\right)$

Discrete Uniform Distribution

- Another distribution arises in the following context: suppose I have a box with n tickets, labelled x_1 through x_n. If I draw one ticket at random and let X denote the number showing on the ticket, then X follows the so-called **Discrete Uniform Distribution**, on the set $\{x_1, \cdots, x_n\}$. We notate this

$$X \sim \text{DiscUnif}\{x_1, \cdots, x_n\}$$

  - A key point is that $x_1, \cdots, x_n$ needn't be consecutive numbers! For example, it makes perfect sense to write $X \sim \text{DiscUnif}\{1, 4, 5, 7.8, 10\}$.

- One can show:
  - $S_X = \{x_1, \cdots, x_n\}$
  - $\mathbb{P}(X = k) = \begin{cases} \frac{1}{n} & \text{if } k \in S_X \\ 0 & \text{otherwise} \end{cases}$
  - No simple general form for $\mathbb{E}[X]$ and $\text{Var}(X)$; consider on a case-by-case basis.

- We can get a bit more specific if we consider the DiscUnif$\{a, a+1, a+2, \cdots, b-1, b\}$ distribution for fixed numbers a, b with $a < b$: firstly, for notational convenience, let $n := b - a + 1$ denote the numbers in the state space of X. Then:
  - $S_X = \{a, a+1, a+2, \cdots, b-1, b\}$
  - $\mathbb{P}(X = k) = \begin{cases} \frac{1}{n} & \text{if } k \in S_X \\ 0 & \text{otherwise} \end{cases}$
  - $\mathbb{E}[X] = \dfrac{a+b}{2}$
  - $\text{Var}(X) = \dfrac{n^2 - 1}{12}$

Practice Makes Perfect

- I know that's a lot of distributions!
- I can't stress it enough- practice makes perfect.
- Over the next few discussion worksheets I'll try and incorporate more problems that test your knowledge on discrete distributions.
- I highly encourage you to consult the textbook for problems as well!

I Want to Play A Game...

bit.ly/distmatch

Poisson Point Processes

Poisson Point Process

Definition: Poisson Point Process

The **Poisson Point Process with rate $\lambda > 0$** (or simply **Poisson Process**) counts the number of events occurring in a fixed time or space, subject to the following assumptions:

(1) The number of events occurring in non-overlapping intervals are independent,

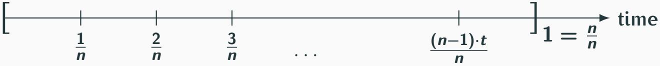
(2) Events occur at a constant rate of λ per unit time,

(3) Events cannot occur simultaneously.

- Some Examples:
    - The number of cars arriving at a traffic light
    - The number of telephone calls arriving at a switchboard
    - The number of blueberries in a 1 in^3 piece of muffin

Poisson Point Process

- Let X denote the number of arrivals in an interval of length 1. What is the distribution of X?
- First, let's discretize our notion of time. In other words, let's divide our time interval into n subintervals of equal length:



- By assumption (3), we can make n large enough (i.e. we can make our interval small enough) so that the probability of observing two or more arrivals in any of these subintervals is 0.
- Furthermore, by assumption (2) there is a constant rate λ of arrivals, meaning the probability of observing an arrival in any subinterval of length $1/n$ is simply λ/n.
- Therefore, X effectively counts the number of successes in n subintervals, where a "success" is observing an arrival... In other words, $X \sim \text{Bin}(n = n,\ \lambda = \frac{p}{n})$.

- Now, of course, time is not in actuality discrete; it is continuous. So, the true distribution of X results in taking the limit as $n \to \infty$ of our approximation to X above. That is:

$$
\begin{aligned}
\mathbb{P}(X = k) & \lim_{n\to\infty} \mathbb{P}\left(X = k \text{ under our discretized approximation}\right) \\
&= \lim_{n\to\infty} \left[\binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \right] \\
&= \lim_{n\to\infty} \left[\frac{n!}{k!(n-k)!} \cdot \frac{1}{n^k} \cdot (\lambda)^k \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \right] \\
&= \lim_{n\to\infty} \left[\frac{n \times (n-1) \times \cdots \times (n-k+1)}{n \times n \times \cdots \times n} \cdot \frac{(\lambda)^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k} \right] \\
&= \frac{(\lambda)^k}{k!} \cdot \lim_{n\to\infty} \left[\frac{n \times (n-1) \times \cdots \times (n-k+1)}{n \times n \times \cdots \times n} \right] \cdot \\
&\qquad \lim_{n\to\infty} \left[\left(1 - \frac{\lambda}{n}\right)^n \right] \cdot \lim_{n\to\infty} \left[\left(1 - \frac{\lambda}{n}\right)^{-k} \right]
\end{aligned}
$$

Poisson Point Process

- Let us examine each of the terms on the RHS separately.
  - Let's start with the rightmost term. As $n \to \infty$, $(\lambda/n) \to 0$ and so $[1 - (\lambda/n)] \to 1$, and thus $[1 - (\lambda/n)]^{-k} \to 1$.
  - Let's now examine the first term. We first rewrite the quantity inside the limit as:

$$(1) \times \left(\frac{n-1}{n} \right) \times \cdots \times \left(\frac{n-k+1}{n} \right) = (1) \cdot \left(1 - \frac{1}{n} \right) \times \cdots \times \left(1 - \frac{n-k+1}{n} \right)$$

    The key to note is that, in the rightmost formulation above, the numerators are always smaller than the denominators. This means that, when we let $n \to \infty$, the fractional terms all go to 0 and we are left with

$$\lim_{n\to\infty} \left[(1) \cdot \left(1 - \frac{1}{n} \right) \times \cdots \times \left(1 - \frac{n-k+1}{n} \right) \right] = 1 \times 1 \times \cdots \times 1 = 1$$

  - Finally, we examine the inner limit. It will be useful to recall the following definition from calculus:

$$e^a = \lim_{n\to\infty} \left(1 + \frac{a}{n} \right)^n$$

    Therefore, we immediately see that

$$\lim_{n\to\infty} \left[\left(1 - \frac{\lambda}{n} \right)^n \right] = e^{-\lambda}$$

- Putting everything together, we find that:

$$\mathbb{P}(k \text{ occurrences in the interval } [0,1]) = \frac{(\lambda)^k}{k!} \cdot e^{-\lambda}$$

The Poisson Distribution

- We call this distribution the **Poisson Distribution**, with parameter λ.
- So, if X counts the number of arrivals in a unit time interval in a Poisson Point Process with rate λ, then $X \sim \text{Pois}(\lambda)$ and:
  - $S_X = \{0, 1, 2, \cdots\}$
  - $p_X(k) = \begin{cases} e^{-\lambda} \cdot \dfrac{\lambda^k}{k!} & \text{if } k = 0, 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$
  - $\mathbb{E}[X] = \lambda$
  - $\text{Var}(X) = \lambda$
- Another very useful property: if arrivals occur according to a Poisson Process with rate λ, then the number of arrivals in an interval of length t follows the $\text{Pois}(\lambda \cdot t)$ distribution.
  - Intuitively, this makes sense: if cars arrive at an average rate of 2 per minute, then the average number of cars arriving in a 30-second interval should be 1.

Example

Suppose calls arrive at a call center according to a Poisson Process with an average rate of 2 calls per minute.

(a) What is the probability of observing exactly 2 calls between 1pm and 1:01pm?

(b) What is the expected number of calls arriving between 2pm and 2:10pm?

Part(a)

- Let X denote the number of calls arriving between 1pm and 1:01pm. Then $X \sim \text{Pois}(2)$ and

$$\mathbb{P}(X = 2) = \boxed{e^{-2} \cdot \frac{2^2}{2!}}$$

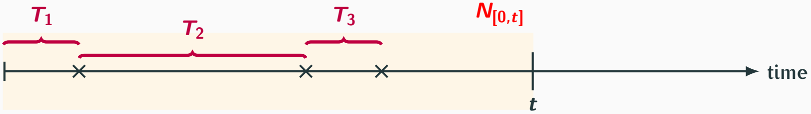
Part(b)

- Let Y denote the number of calls arriving between 2:00pm and 2:10pm. Since there are 10 minutes between 2:00pm and 2:10pm we have $Y \sim \text{Pois}(2 \cdot 10) = 20$ and so

$$\mathbb{E}[Y] = \boxed{20}$$

Leadup

- With Poisson Point Processes, drawing a timeline can often be very useful:



  - $N_{[0,t]}$; number of arrivals in $[0, t]$.
    - Discrete; $N_{[0,t]} \sim \text{Pois}(\lambda t)$

  - T_i; time between $(i-1)^{\text{th}}$and i^{th}arrivals. Sometimes called **interarrival times**.
    - State space: $S_{T_i} = [0, \infty)$
    - So, T_i is continuous!