Lecture 5: Random Variables

- Random Variable
- Distributions
- The Bernoulli and Binomial Distributions
- The Hypergeometric Distributions
- The Multinomial Distributions

Introduction

– In last lectures we reviewed some counting methods and talked about conditional probability, independence, and Bayes' rule.

– We now know how to calculate the probability of events if we know what the sample space looks like.

– Today we will move on to define real-valued functions on sample spaces and learn how to calculate their probabilities. Specifically, we will learn about real-valued functions of sample spaces, called **random variables**, and their **distributions**.

– Lastly, we will learn about some special and very useful **discrete** distributions.

Random Variable: Some Examples

1. **Elementary random variables:** a normal distributed random number
2. **Empirical random variables:** taking a sample of stocks and tabulating their prices
3. **Complex random variables:** solution of a stochastic PDE with a white noise driving term
4. **A doctor's diagnosis** can be viewed as a random sample from his posterior probability distribution on the state of your body, given the combination of
    1. his personal experience
    2. his knowledge from books, papers, and other doctors
    3. your case history
    4. your test results
5. **A novel** can be viewed as a random sample from the author's posterior probability distribution on stories, conditioned on all the things the author has observed or learned about the nature of the real world

David Mumford, *Mathematics: Frontiers and Perspectives, 2000*

The Sample Space?

1. **Elementary random variables:** $\Omega = \mathbb{R}$
2. **Empirical random variables:** Ω is essentially unknowable
3. **Complex random variables:** Ω is some big product of the probability spaces from which all the random elements in the construction have been drawn
4. **The novelist or doctor:** Ω is the full probability model that he/she has constructed of how the world works

Random Variables

– A (real-valued) **random variable** is a function that maps the sample space to the real line. That is, if X is a random variable, then X maps every element of Ω to a real number. If "X" is a random variable, we will write "x" as its observed value.

– Let C be a subset of the real line such that $\{X \in C\}$ is an event, and let $P(X \in C)$ denote the probability that the value of X will belong to the subset C, which equals to the probability that the outcome ω of the experiment will be such that $X(\omega) \in C$. In symbols,

$$P(X \in C) = P(\{\omega \in \Omega | X(\omega) \in C\}).$$

Examples

– *Example 1:* Back to our toy example. Suppose you toss a fair coin twice and let X be the number of heads observed. Then $\Omega = \{HH, HT, TH, TT\}$, and

$$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0$$

$$\text{Then,}\ \ \mathbb{P}(X = 0) = \mathbb{P}(\{TT\}) = \frac{1}{4},\ \mathbb{P}(X = 2) = \mathbb{P}(\{HH\}) = \frac{1}{4}$$

$$\mathbb{P}(X = 1) = \mathbb{P}(\{HT, TH\}) = \frac{2}{4} = \frac{1}{2}$$

– Clearly, we can define several random variables on the same sample space. Can you think of another random variable defined on Ω?

Examples

Example 2: Suppose you roll a fair die five times and let X be the number of 4's observed. Then writing out the sample space is tedious but clearly, $X \in \{0, 1, 2, 3, 4, 5\}$. So that if for example we observe $\omega = (1, 1, 3, 4, 2) \in \Omega$, $X(\omega) = 1$, but if we observe $\omega = (4, 4, 5, 4, 2) \in \Omega$, $X(\omega) = 3$.

What then is the probability that $X = 2$? Since this is a fair die, $\mathbb{P}$(observing 4 on any roll) $= \frac{1}{6}$ and $\mathbb{P}$(observing any other number on any roll) $= \frac{5}{6}$. For X to be 2, exactly two of the rolls must be 4 and exactly three must be any number besides 4. Lastly, there are $\binom{5}{2}$ ways of getting two rolls with "4" from five rolls.

$$\Rightarrow \mathbb{P}(X = 2) = \binom{5}{2} \times \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = \binom{5}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^3 = 0.1608$$

– Turns out there's a name for the collection of the probabilities for this random variable. We'll get to that soon. First, a few definitions.

Distributions

– The **distribution of a random variable** X is the collection of all probabilities of the form $\mathbb{P}(X \in C)$ for all sets C of real numbers such that $\{X \in C\}$ is an event.

– A random variable X has a **discrete distribution** if X can only take on **a finite number of different values** or a sequence of **countably infinite values**. For example, $X = 0, 1, 2$ and 3 implies X is a discrete random variable.

– A random variable X has a **continuous distribution** if X can take on **infinite** values. For example, $X \in [0, 1]$ implies X is a continuous random variable.

Probability Mass Function

– **Probability Mass Function (PMF):** If X has a discrete distribution, the PMF is defined as a function f such that for every real number x,

$$f(x) = \mathbb{P}(X = x); \quad \text{where} \quad \sum_{i=1}^{\infty} f(x_i) = 1$$

– I will continue to use $f(x)$ and $\mathbb{P}(X = x)$ interchangeably. Note that,

$$\mathbb{P}(X \leq c) = \sum_{i:x_i \leq c} \mathbb{P}(X = x_i)$$

Bernoulli Distribution

Suppose X is a random variable that can only take on the values 1 or 0 with probabilities $P(X = 1) = p$ and $P(X = 0) = 1 - p$. Then X is said to have a Bernoulli distribution with probability of "success" p, denoted $X \sim \text{Bernoulli}(p)$

- We're flipping a coin once and our random variable is $X_1 = 1$ if the outcome is heads and $X_1 = 0$ if the outcome is tails. Then, $X_1 \sim \text{Bernoulli}(1/2)$
- We're rolling a die and our random variable takes on the value $X_2 = 1$ if the outcome is strictly greater than 4 and $X_2 = 0$ otherwise. Then, $X_2 \sim \text{Bernoulli}(1/3)$

Binomial Distribution

– Some distributions of random variables are very useful in statistics. One of them is the **binomial distribution**.

– The binomial formula gives the probability of exactly y successes in n tries, where each try has the same probability of success p and each try is independent.

$$\mathbb{P}(\text{exactly } y \text{ successes}) = \mathbb{P}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y};\ y = 0, 1, \ldots, n.$$

– This is exactly the situation one has when trying to find the probability of y heads in n tosses of a coin that has probability p of coming up heads. The random variable X is said to have the binomial distribution $\mathbf{Y} \sim \text{Binomial}(\mathbf{n}, \mathbf{p})$.

– When $n = 1$, the distribution is the **Bernoulli distribution**. Thus, the binomial distribution results from independent Bernoulli trials.

Repeated Bernoulli Experiment

- We're flipping a fair coin 4 times and we want to count the total number of tails. The coin flips (X_1, X_2, X_3, X_4) are Bernoulli$(1/2)$ random variables and they are independent by assumption, so the total number of tails is $Y = X_1 + X_2 + X_3 + X_4 \sim \text{Binomial}(4, 1/2)$
- Our best friend Bobbie is taking a multiple choice test with 10 questions and 3 choices per question. For each question, there're only one correct answer. He hasn't studies for the test and he decides to choose the answers "at random", so he has a $1/3$ chance of getting each question right. Let $X_i = 1$ if his answer to the i-th question is right, so $X_i \sim \text{Bernoulli}(1/3)$. The total number of right answers in his test is

$$Y = X_1 + X_2 + \cdots + X_{10} \sim \text{Binomial}(10, 1/3)$$

Is it a Binomial?

Suppose that we flip a fair coin. The random variable X_1 equals 1 if it comes up heads and $X_1 = 0$ if it comes up tails (so $X_1 \sim \text{Bernoulli}(1/2)$). If $X_1 = 1$, we will use a loaded coin with a probability of coming up heads equal to 2/3 for our next flip (X_2). If $X_1 = 0$, we will use a fair coin for X_2. The random variable X_2 is also Bernoulli, since it can only take on the values 0 or 1. The probability of success is

$$
\begin{aligned}
P(X_2 = 1) &= P(X_1 = 0)P(X_2 = 1|X_1 = 0) + P(X_1 = 1)P(X_2 = 1|X_1 = 1) \\
&= \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{2}{3} = \frac{7}{12} \approx 0.583
\end{aligned}
$$

So $X_1 \sim \text{Bernoulli}(1/2)$ and $X_2 \sim \text{Bernoulli}(7/12)$. Is $Y = X_1 + X_2$ a Binomial?

The answer is no because (1) the probabilities of success for X_1 and X_2 are different, (2) X_1 and X_2 are not independent. The coin we flip in X_2 depends on the outcome of X_1, so X_1 and X_2 are clearly dependent.

The Intuition behind the Binomial Formula

– Why does the Binomial formula work? We already discussed this idea in example 2, but we need to make the idea general.

- How many arrangements are there that give x heads in n tries? From the previous class, we know the answer is $\binom{n}{x}$.
- Each arrangement is different from the other arrangements. Thus the probability of exactly x successes is the sum over all possible arrangements, and there are $\binom{n}{x}$ of those.
- Each arrangement has been the same probability: $p^x(1-p)^{n-x}$. To see this, consider the sequences HTHT and THTH from four independent coin tosses. The first has probability $p \times (1-p) \times p \times (1-p) = p^2(1-p)^2$. The second arrangement has probability $(1-p) \times p \times (1-p) \times p = p^2(1-p)^2$.
- Combining all pieces gives $\mathbb{P}(\text{exactly } x \text{ successes}) = \binom{n}{x} p^x (1-p)^{n-x}$

Examples of Binomial Probabilities

Example 3: What is the probability of exactly three fives in six rolls of a fair die?

$$\begin{aligned}\mathbb{P}(\text{exactly 3 successes}) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \binom{6}{3}\left(\frac{1}{6}\right)^3 \left(1-\frac{1}{6}\right)^{6-3} \\ &= 20 \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)^3 = 0.0536\end{aligned}$$

– *Example 2 again:* Suppose you roll a fair die five times and let X be the number of fours observed. Then X has the binomial distribution $\text{Bin}\left(5, \frac{1}{6}\right)$. Using the formula,

$$\mathbb{P}(X=2) = \binom{5}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^3 = 0.1608 \text{ Same as before.}$$

Examples of Binomial Probabilities (Cont'd)

- Suppose we're flipping a fair coin 4 times and we want to count the total number of tails, which we denote Y. What is the probability that we get 2 tails? We have that $Y \sim \text{Binomial}(4, 1/2)$, so

$$P(Y = 2) = \binom{4}{2}(1/2)^2(1/2)^2 = 0.375$$

- Our best friend Bobbie is taking a multiple choice test with 10 questions and 3 choices per question. For each question, there're only one correct answer. He hasn't studies for the test and he decides to choose the answers "at random", so he has a $1/3$ chance of getting each question right. What is the probability that he gets at least half of them right? Let Y be the total number of right answers in his test. Then $Y \sim \text{Binomial}(10, 1/3)$. We're interested in finding $P(Y \geq 5)$, which is equal to

$$\begin{aligned} P(Y \geq 5) &= P(Y = 5) + P(Y = 6) + \dots P(Y = 10) \\ &= \binom{10}{5}(1/3)^5(2/3)^5 + \binom{10}{6}(1/3)^6(2/3)^4 + \dots (1/3)^{10} \approx 0.213 \end{aligned}$$

Spin Wheel Game

– Suppose you are on a spin wheel game show and the spin wheel has numbers 1 to 10.

– You will win \$2 every time the spin wheel stops at either numbers 2, 6 or 7.

– If even numbers are twice as likely to show up as odd numbers, but all even numbers are equally likely between them and all odd numbers are also equally likely between them: (1) What is the probability that you will win \$6 in 6 independent spins? (2) What about the probability that you will win at least \$4?

Answers: (1) ≈ 0.21948 (2) ≈ 0.64882

Hypergeometric Distribution

– Suppose we have a population of N elements where M elements have a certain characteristic and $N - M$ don't. Suppose that we select n elements of the population **without replacement**. If X is the number of elements in the sample that have the characteristic, then

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

and X is said to have a Hypergeometric distribution, denoted $X \sim \text{Hypergeometric}(N, M, n)$. We've seen this type of random variable before in homeworks!

– This pmf is more intuitive than it looks at first. First, there are $\binom{N}{n}$ ways to pick n elements from a population of size N. This is the total number of elements in our sample space. All we need to do next is figure out how many of those will give us x successes.

– Well, if there are M total successes, there are $\binom{M}{k}$ ways to choose k successes from M possible successes. Lastly, to have k successes, there must be $n - k$ failures as well with $\binom{N-M}{n-k}$ ways to choose those.

Examples

– *Example 7:* Suppose a shipment of 100 lightbulbs contains 10 defective bulbs. You draw five at random, and test them. What is the probability that all five work?

$$\mathbb{P}[X = x] = \frac{\binom{90}{5}\binom{100-90}{5-5}}{\binom{100}{5}} = 0.5838$$

– *Example 8:* You are dealt five card from a standard deck. What is the probability of getting two or more clubs?

$$\begin{aligned}\mathbb{P}[X \geq 2] &= 1 - \mathbb{P}[X = 0] - \mathbb{P}[X = 1] \\ &= 1 - \frac{\binom{13}{0}\binom{39}{5}}{\binom{52}{5}} - \frac{\binom{13}{1}\binom{39}{4}}{\binom{52}{5}} \\ &= 1 - 0.2215 - 0.4115 = 0.3671.\end{aligned}$$

Binomial or Hypergeometric ?

Suppose that you have 20 really good friends, 10 of which like Broccoli. You want to host a dinner party, but your apartment is too small and can only fit 5 friends. You're a nice person, so you decide that the right thing to do is selecting 5 of them at random. What is the probability that all of your randomly selected guest like Broccoli? What is the probability that at least one of them doesn't?

– The key difference between Binomial and Hypergeometric is **with or without replacement**. If you're in a Hypergeometric scenario, your draws are not independent (and recall that independence is an assumption of the Binomial!)

– If the size of the population N is big relative to the sample size n, the Binomial and the Hypergeometric give similar answers. Can you see why?

The Multinomial Distributions

Remember that the Binomial distribution describes a random variable that represents success (or failure) based on n independent trials of an experiments with two outcomes, where the probability of success is the same for each trial. Our toy examples have been tossing a coin n times and rolling a die n times.

An extension to the binomial distribution which allows for three or more outcomes (k outcomes) for each trial, where the probability of each outcome is the same for each trial is called the **Multinomial Distribution**.

The Multinomial Distribution Cont'd

A random variable (vector) $\mathbf{X} = (X_1, \ldots, X_k)$ has the multinomial distribution with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_k)$ if its pmf is given by

$$f(\boldsymbol{x}) = \mathbb{P}(\boldsymbol{X} = \boldsymbol{x}) = \binom{n}{x_1, \ldots, x_k} p_i^{x_1}, \cdots, p_n^{x_k} \quad \text{for} \quad x_1 + \ldots + x_k = n$$

where $\binom{n}{x_1, \ldots, x_k} = \dfrac{n!}{x_1!x_2!\ldots x_k!}$ is called the multinomial coefficient.

Example

Example 4: Suppose that for a single roll of a loaded die we can observe one with probability 0.25, two or three with probability 0.1 each, four or five with probability 0.2 each and six with probability 0.15.

What is the probability that in 15 independent rolls, we will observe one four times, two thrice, three once, four once, five four times and six twice.

Example (Cont'd)

Let $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5, X_6)$ be the observed combination for the die. Then $\boldsymbol{p} = (0.25, 0.1, 0.1, 0.2, 0.2, 0.15)$ and

$$\begin{aligned}
\mathbb{P}\left[\boldsymbol{X} = (4,3,1,1,4,2)\right] &= \binom{n}{x_1, \ldots, x_k} p_i^{x_1}, \cdots, p_n^{x_k} \\
&= \binom{15}{4,3,1,1,4,2} 0.25^4 0.1^3 0.1^1 0.2^1 0.2^4 0.15^2 \\
&= \frac{15!}{4!3!1!1!4!2!} 0.25^4 0.1^3 0.1^1 0.2^1 0.2^4 0.15^2 \\
&= 0.0005
\end{aligned}$$

Properties of the Multinomial Distribution

A few important points:

1. The vector of probabilities $\boldsymbol{p}$ should sum to 1.
2. Any $X_i \in (X_1, \ldots, X_k)$ has a binomial distribution with parameters n and p_i. ***This is why treating the die roll as a binomial when we only care about one of the sides works!***
3. Collapsing the k different outcomes to two outcomes gets you back to a binomial distribution as well. If $\mathbf{X} = (X_1, \ldots, X_k)$ has a multinomial distribution with parameters n and $\boldsymbol{p} = (p_1, \ldots, p_k)$ and $l < k$ where $i_1, \ldots, i_l$ are distinct elements of the set $\{1, \ldots, k\}$, then $Y = X_{i_1} + \cdots + X_{i_l}$ has a binomial distribution with parameters n and $p_{i_1} + \cdots + p_{i_l}$
4. With replacement

Recap

We discussed the following:

- What random variables are
- What distributions and pmfs are
- Some important discrete distributions

Suggested reading:

- D.S. Sec. 3.1, 5.2, 5.3, 5.9
- OpenIntro3: Sec. 2.4, 3.4.1