Statistics 110: Introduction to Probability — Section 4, Fall 2018

Expected Value and Indicators

TFs: Justin Zhu, Michael Zhang

Important Distributions

Bernoulli Distribution

Bernoulli The Bernoulli distribution is the simplest case of the Binomial distribution, where we only have one trial, or $n = 1$. Let us say that X is distributed Bern(p). We know the following:

Story. X "succeeds" (is 1) with probability p, and X "fails" (is 0) with probability $1 - p$.

Example. A fair coin flip is distributed Bern($\frac{1}{2}$).

PMF. The probability mass function of a Bernoulli is:

$$P(X = x) = p^x(1 - p)^{1-x}$$

or simply

$$P(X = x) = \begin{cases} p, & x = 1 \\ 1 - p, & x = 0 \end{cases}$$

Binomial Distribution

Binomial Let us say that X is distributed Bin(n, p). We know the following:

Story X is the number of "successes" that we will achieve in n independent trials, where each trial can be either a success or a failure, each with the same probability p of success.

Example If Jeremy Lin makes 10 free throws and each one independently has a $\frac{3}{4}$ chance of getting in, then the number of free throws he makes is distributed Bin(10, $\frac{3}{4}$), or, letting X be the number of free throws that he makes, X is a Binomial Random Variable distributed Bin(10, $\frac{3}{4}$).

PMF The probability mass function of a Binomial is:

$$P(X = x) = \binom{n}{x} p^x(1 - p)^{n-x}$$

Hypergeometric

Let us say that X is distributed HGeom(w, b, n). We know the following:

Story In a population of b undesired objects and w desired objects, X is the number of "successes" we will have in a draw of n objects, without replacement.

Example 1) Let's say that we have only b Weedles (failure) and w Pikachus (success) in Viridian Forest. We encounter n of the Pokemon in the forest, and X is the number of Pikachus in our encounters. 2) The number of aces that you draw in 5 cards (without replacement). 3) You have w white balls and b black balls, and you draw b balls. X is the number of white balls you will draw in your sample.

PMF The probability mass function of a Hypergeometric is:

$$P(X = k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

Geometric

Let us say that X is distributed Geom(p). We know the following:

Story X is the number of "failures" that we will achieve before we achieve our first success. Our successes have probability p.

Example If each pokeball we throw has a $\frac{1}{10}$ probability to catch Mew, the number of failed pokeballs will be distributed Geom$(\frac{1}{10})$.

PMF With $q = 1 - p$, the probability mass function of a Geometric is:

$$P(X = k) = q^k p$$

Negative Binomial

Let us say that X is distributed NBin(r, p). We know the following:

Story X is the number of "failures" that we will achieve before we achieve our rth success. Our successes have probability p.

Example Thundershock has 60% accuracy and can faint a wild Raticate in 3 hits. The number of misses before Pikachu faints Raticate with Thundershock is distributed NBin$(3, .6)$.

PMF With $q = 1 - p$, the probability mass function of a Negative Binomial is:

$$P(X = n) = \binom{n + r - 1}{r - 1} p^r q^n$$

Poisson Distribution (Discrete)

Let us say that X is distributed Pois(λ). We know the following:

Story There are rare events (low probability events) that occur many different ways (high possibilities of occurences) at an average rate of λ occurrences per unit space or time. The number of events that occur in that unit of space or time is X.

Example A certain busy intersection has an average of 2 accidents per month. Since an accident is a low probability event that can happen many different ways, the number of accidents in a month at that intersection is distributed Pois(2). The number of accidents that happen in two months at that intersection is distributed Pois(4)

PMF The PMF of a Poisson is:

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

Chicken and Egg Problem A very important example to know for the Poisson distribution is the Chicken and Egg problem, which says that if the number of eggs laid is Pois(λ) and each egg independently hatches with probability p, then the number of chicks will be distributed Pois(λp).

Limit of the Binomial If we take the binomial distribution, and let n be really large, and p be very small, then the distribution can be apprxomiated as $\text{Pois}(np)$.

Discrete Distributions Overview

Distribution	PMF and Support	E(X)	Variance	Equivalent To
Bernoulli $\text{Bern}(p)$	$P(X=1)=p$ $P(X=0)=q$	p	pq	$\text{Bin}(1,p)$
Binomial $\text{Bin}(n,p)$	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ $k \in \{0,1,2,\dots n\}$	np	npq	Sum of n $\text{Bern}(p)$
Geometric $\text{Geom}(p)$	$P(X=k)=q^k p$ $k \in \{0,1,2,\dots\}$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\text{NBin}(1,p)$
First Success $\mathcal{FS}(p)$	$P(X=k)=q^{k-1}p$ $k \in \{1,2,3,\dots\}$	$\frac{1}{p}$	$\frac{q}{p^2}$	Geom(p) + 1
Negative Binomial $\text{NBin}(r,p)$	$P(X=n)=\binom{n+r-1}{r-1}p^r q^n$ $n \in \{0,1,2,\dots\}$	$r\frac{q}{p}$	$r\frac{q}{p^2}$	Sum of r $\text{Geom}(p)$
Hypergeometric $\text{HGeom}(w,b,n)$	$P(X=k)=\binom{w}{k}\binom{b}{n-k}/\binom{w+b}{n}$ $k \in \{0,1,2,\dots,\min(n,w)\}$	$n\frac{w}{b+w}$		
Poisson $\text{Pois}(\lambda)$	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$ $k \in \{0,1,2,\dots\}$	λ	λ	

Expected Value

The **Expected Value** (or *expectation, mean*) of a random variable can be thought of as the "weighted average" of the possible outcomes of the random variable. Mathematically, if $x_1, x_2, x_3, \dots$ are all of the possible values that X can take, the expected value of X can be calculated as follows:

$$E(X) = \sum_i x_i P(X = x_i)$$

Linearity of Expectation

The most important property of expected value is **Linearity of Expectation**. For **any** two random variables X and Y, a and b scaling coefficients and c is our constant, the following property of holds:

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

The above is true regardless of whether X and Y are independent.

Conditional Expected Value

Conditional distributions are still distributions. Treating them as a whole and applying the definition of expectation gives:

$$E(X|A) = \sum_i x_i P(X = x_i|A)$$

Indicator Random Variables

Indicator Random Variables are random variables whose value is 1 when a particular event happens, or 0 when it does not. Let I_A be an indicator random variable for the event A. Then, we have:

$$I_A = \begin{cases} 1 & A \text{ occurs} \\ 0 & A \text{ does not occur} \end{cases}$$

Suppose $P(A) = p$. Then, $I \sim \text{Bern}(p)$ because I has a p chance of being 1, and a $1 - p$ chance of being 0.

Properties of Indicators

- $(I_A)^2 = I_A$, and $(I_A)^k = I_A$ for any power k.
- $I_{A^c} = 1 - I_A$
- $I_{A \cap B} = I_A I_B$ is the indicator for the event $A \cap B$ (that is, $I_A I_B = 1$ if and only if A and B occur, and 0 otherwise)
- $I_{A \cup B} = I_A + I_B - I_A I_B$

Fundamental Bridge

The fundamental bridge is the idea that $E(I_A) = P(A)$. When we want to calculate the expected value of a complicated event, sometimes we can break it down into many indicator random variables, and then apply linearity of expectation on that. For example, if $X = I_1 + I_2 + \ldots + I_n$, then:

$$\begin{aligned} E(X) &= E(I_1) + E(I_2) + \ldots + E(I_n) \\ &= P(I_1) + P(I_2) + \ldots + P(I_n) \end{aligned}$$

Variance

Variance tells us how spread out the distribution of a random variable is. It is defined as

$$\text{Var}(X) = E\left[(X - E(X))^2\right] = E(X^2) - (E(X))^2$$

Properties of Variance

- $\text{Var}(cX) = c^2 \text{Var}(X)$
- $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y)$ if X and Y are independent

Practice Problems

Example 1. Dice collector.
What is the expected number of times a die must be rolled until the numbers 1 through 6 have all shown up at least once?

Example 2. Mutual Friends.
Alice and Bob have just met, and wonder whether they have a mutual friend. Each has 50 friends, out of 1000 other people who live in their town. They think that it's unlikely that they have a friend in common, saying each of us is only friends with 5% of the people here, so it would be very unlikely that our two 5%'s overlap.

Assume that Alice's 50 friends are a random sample of the 1000 people (equally likely to be any 50 of the 1000), and similarly for Bob. Also assume that knowing who Alice's friends are gives no information about who Bob's friends are. Let X be the number of mutual friends they have.

(a) Compute $E(X)$

(b) Find the PMF of X.

(c) Is the distribution of X one of the important distributions we have looked at? If so, which?

Example 3. Min and Max.
Let $X \sim \text{Bin}\left(n, \frac{1}{2}\right)$ and $Y \sim \text{Bin}\left(n+1, \frac{1}{2}\right)$ independently.

(a) Let V = $\min(X, Y)$ be the smaller of X and Y, and let $W = \max(X, Y)$. So if X crystalizes to x and Y crystalizes to y, then V crystalizes to $\min(x, y)$ and W crystalizes to $\max(x, y)$. Find $E(V) + E(W)$.

(b) Show that $E|X - Y| = E(W) - E(V)$, with notation as in (a).

(c) Compute $\text{Var}(n - X)$.

Example 4. Expected Number and Variance of Matches.
Suppose 100 people, each with a hat. We mix the hats and hand them out randomly to each person.

(a) What is the expected number of people who get their own hat?

(b) What is the variance of the number of people who get their own hat?

Example 5. Prizes.
There are n prizes, with values \$1, \$2 ... \$$n$. You get to choose k random prizes, without replacement. What is the expected total value of the prizes you get?

Example 6. Coin Runs.
A coin with probability p of Heads is flipped n times. The sequence of outcomes canbe divided into runs (blocks of H's or blocks of T's), e.g., $HHHTTHTTTH$ becomes $\boxed{HHH}\boxed{TT}\boxed{H}\boxed{TTT}\boxed{H}$, which has 5 runs. Find the expected number of runs.

Hint: Start by finding the expected number of tosses (other than the first) where the outcome is different from the previous one.

Example 7. True/False.

(a) If X and Y have the same CDF, they have the same expectation.

(b) If X and Y have the same expectation, they have the same CDF.

(c) If X and Y have the same CDF, they must be dependent.

(d) If X and Y have the same CDF, they must be independent.

(e) If X and Y have the same distribution, $P(X < Y)$ is at most $1/2$.

(f) If X and Y are independent and have the same distribution, $P(X < Y)$ is at most $1/2$.