CH 3.1 Random Variables

Definition: Random Variable

Given an experiment with sample space S, a random variable is a function from the sample space S to the real numbers $\mathbb{R}$.

A random variable assigns a numerical value $X(s)$ to each possible outcome s of the experiment.

CH 3.2 Distributions & Mass prob. functions

Definition: (3.2.1) Discrete Random Variables

A random variable X is said to be discrete if there is a finite list of values $a_1, a_2, \dots, a_n$ or an infinite list of values $a_1, a_2, \dots$ such that $P(X = a_j \text{ for some } j) = 1$. If X is a discrete r.v., then finite or countably infinite set of values x s.t. $P(X=x) > 0$ is called the support of X.

In contrast, a continuous r.v. can take on any real value in an interval.

Definition: (3.2.2) Probability Mass Function

The PMF of a discrete r.v. X is the function p_X given by $p_X(x) = P(X=x)$.

Remark: This is positive if x is in the support of X, and 0 otherwise.

In writing $P(X=x)$, we use $X=x$ to denote an event, consisting of all outcomes s to which X assigns the number x.

Formally, $\{X=x\}$ is defined as $\{s \in S : X(s) = x\}$.

Theorem: (3.2.7) Valid PMFs

Let X be a discrete r.v. with support $x_1, x_2, \dots$ (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). The PMF p_X must satisfy:

(i) Nonnegative: $p_X(x) > 0$ if $x = x_j$ for some j, and $p_X(x) = 0$ otherwise;
(ii) Sums to 1: $\sum_{j=1}^{\infty} p_X(x_j) = 1$

Proof:

First criterion is true since probability is non-negative. Second is true since X must take on some value, and the events $\{X = x_j\}$ are disjoint, so

$$\sum_{j=1}^{\infty} P(X = x_j) = P\left(\bigcup_{j=1}^{\infty} \{X = x_j\}\right) = P(X = x_1 \text{ or } X = x_2 \text{ or } \dots) = 1$$

In general, giving a discrete r.v. X and a set B of real numbers, if we know the PMF of X we can find $P(X \in B)$, the probability that X is in B, by summing up the heights of the vertical bars at points in B in the plot of the PMF of X.

CH 3.3 Bernoulli & Binomial

Definition: (3.3.1) Bernoulli distribution

An r.v. X is said to have the Bernoulli distribution with parameter p if $P(X=1)=p$ and $P(X=0)=1-p$, where $0<p<1$ We write this as $X \sim \text{Bern}(p)$. The $\sim$ is read "is distributed as".

Definition: (3.3.2) Indicator Random Variable

The Indicator Random Variable of an event A is the rv which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by I_A or $I(A)$. Note that $I_A \sim \text{Bern}(p)$ with $p=P(A)$.

Theorem: (3.3.5) Binomial PMF

If $X \sim \text{Bin}(n,p)$, then the PMF of X is

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

for $k = 0, 1, \cdots, n$ (and $P(X=k)=0$ otherwise).

Remark: (3.3.6)

If two discrete r.v.s have the same PMF, then they must have the same support.

Proof: (of 3.3.5)

An experiment consisting of n independent Bernoulli trials produce a sequence of successes and failures. The probability of any specific sequence of k successes and $n-k$ failures is $p^k(1-p)^{n-k}$. There are $\binom{n}{k}$ such sequences, since we just need to select where the successes are. Therefore, letting X be the number of successes,

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

for $k=0,1,\cdots,n$, and $P(X=k)=0$ and otherwise. This is a valid PMF because it is nonnegative and sums up to 1 by the binomial theorem.

Theorem: (3.3.7)

Let $X \sim \text{Bin}(n,p)$, and $q=1-p$. Then $n-X \sim \text{Bin}(n,q)$

Proof:

Interpret X as the # of successes in n independent Bernoulli trials. Then $n-X$ is the # of failures in those trials. Then we have $n-X \sim \text{Bin}(n,q)$.

Alternatively, let $Y = n-X$. Then

$$P(Y=k) = P(X=n-k) = \binom{n}{n-k} p^{n-k} q^{k} = \binom{n}{k} q^{k} p^{n-k}$$

for $k = 0, 1, \cdots, n$

Corollary: (3.3.8)

Let $X \sim \text{Bin}(n,p)$, with $p = 1/2$ and n even. Then the distribution of X is symmetric about $n/2$, in the sense that $P(X = n/2 + j) = P(X = n/2 - j)$ for all nonnegative integers j.

Proof:

By theorem 3.3.7, $n-X$ is also $\text{Bin}(n, 1/2)$, so

$$P(X=k) = P(n-X=k) = P(X=n-k)$$

for all nonnegative integers k. Letting $k = n/2 + j$, the desired result follows.

CH 3.4 Hypergeometric

Story: (3.4.1) Hypergeometric Distribution

Consider an urn with w white balls and b black balls. We draw n balls out of the urn at random without replacement, such that all $\binom{w+b}{n}$ samples are equally likely. Let X be the # of white balls in the sample. Then X is said to have the Hypergeometric distribution with parameters $w, b,$ and n; we denote this by $X \sim \text{HGeom}(w,b,n)$.

Theorem: (3.4.2) Hypergeometric PMF

If $X \sim \text{HGeom}(w,b,n)$, then the PMF of X is

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

for integers k satisfying $0 \le k \le w$ and $0 \le n-k \le b$, and $P(X=k)=0$ otherwise.

Proof:

To get $P(X=k)$, we first count the # of possible ways to draw exactly k white balls and $n-k$ black balls from the urn (without distinguishing between different orderings for getting the same set of balls). There are $\binom{w}{k}\binom{b}{n-k}$ ways to draw k white and $n-k$ black balls by the multiplication rule, and there are $\binom{w+b}{n}$ total ways to draw n balls.

Proof: (continued...)

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

Theorem:

The HGeom(w,b,n) and HGeom$(n,w+b-n,w)$ distributions are identical. That is, if $X \sim \text{HGeom}(w,b,n)$ and $Y \sim \text{HGeom}(n,w+b-n,w)$, then X and Y have the same distribution.

Proof:

They have the same PMF. We can check this algebraically with X and Y.

CH 3.5 Discrete Uniform

Story: (3.5.1) Discrete Uniform Distribution

Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e. all values in C are equally likely). Then $X \sim \text{DUnif}(C)$.

The PMF of $X \sim \text{DUnif}(C)$ is

$$P(X=x)=\frac{1}{|C|}$$

for $x \in C$, since a PMF must sum to 1.

CH 3.6 CDF

Definition: (3.6.1)

The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x)=P(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F for a CDF.