LECTURE 6

Discrete Random Variables and Probability Distributions

Go to "BACKGROUND COURSE NOTES" at the end of my web page and download the file *distributions*.

Today we say goodbye to the elementary theory of probability and start Chapter 3. We will open the door to the application of algebra to probability theory by introducing the concept "random variable". What you will need to get from it (at a minimum) is the ability to do the following *"Good Citizen Problems"*. I will give you a *probability mass function* $p(x)$. You will be asked to compute

(i) The *expected value (or mean)* $E(X)$.

(ii) The *variance* $V(X)$.

(iii) The *cumulative distribution function* $F(x)$.

You will learn what these words mean shortly.

Mathematical Definition. Let S be the sample space of some experiment (mathematically a set S with a probability measure P). A random variable X is a real-value function on S.

Intuitive Idea. A random variable is a function whose values have probabilities attached.

Remark . *To go from the mathematical definition to the "intuitive idea" is tricky and not really that important at this stage.*

Basic Example. Flip a fair coin three times so

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Let X be function on X given by

$$X = \text{number of heads}$$

so X is the function given by

$$\begin{array}{cccccccc} \{HHH, & HHT, & HTH, & HTT, & THH, & THT, & TTH, & TTT\}. \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 3 & 2 & 2 & 1 & 2 & 1 & 1 & 0 \end{array}$$

What is

$$P(X = 0), \quad P(X = 3), \quad P(X = 1), \quad P(X = 2)$$

Answers. Note $\#(S) = 8$

$$P(X=0) = P(TTT) = \frac{1}{8}$$
$$P(X=1) = P(HTT) + P(THT) + P(TTH) = \frac{3}{8}$$
$$P(X=2) = P(HHT) + P(HTH) + P(THH) = \frac{3}{8}$$
$$P(X=3) = P(HHH) = \frac{3}{8}$$

We will tabulate this

Value $\longrightarrow x$	0	1	2	3
Probability of that value $\longrightarrow P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Get use to such tabular presentations.

Rolling a Die. Roll a fair die, let

$$X = \text{the number that comes up.}$$

So X takes values 1, 2, 3, 4, 5, 6 each with probability $\frac{1}{6}$.

x	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

This is a special case of the *discrete uniform distribution* where X takes values $1, 2, 3, \cdots, n$ each with probability $\frac{1}{n}$ (so "roll a fair die with n faces").

Bernoulli Random Variables. Usually random variables are introduced to make things numerical. We illustrate this by an important example – next page. First meet some random variables.

Definition (The simplest random variable(s))**.** *The actual simplest random variable is a random variable in the technical sense but isn't really random. It takes one value (let's suppose it is 0) with probability one*

x	*0*
$P(X = x)$	*1*

Nobody ever mentions this because it is too simple – it is deterministic. The simplest random variable that actually is random takes ***two*** *values, let's suppose they are 1 and 0 with probabilities* p *and* q*. Since* X *has to be either 1 or 0 we must have*

$$p + q = 1.$$

So we get

x	*1*	*0*
$P(X = x)$	p	q

This is called the ***Bernoulli random variable with parameter*** p*. So a Bernoulli random variable is a random variable that takes only* ***two*** *values 0 and 1.*

Where do Bernoulli random variables come from?

We go back to elementary probability.

Definition. *A* **Bernoulli experiment** *is an experiment which has two outcomes which we call (by convention) "success"* S *and failure* F*.*

Example (Flipping a coin)**.** *We will call a head a success and a tail a failure.*

Often we call a "success" something that is in fact far from an actual success – eg., a machine breaking down. By convention we let $P(S) = p$ *and* $P(F) = q$*, so again* $p + q = 1$*. Thus, the sample space* $\mathcal{S}$ *of a Bernoulli experiment is given by*

$$\mathcal{S} = \{S, F\}.$$

To join up pages 7 and 9 we define a random variable X *on* $\mathcal{S}$ *by* $X(S) = 1$ *and* $X(F) = 0$ *so* $P(X = 1) = P(S) = p$ *and* $P(X = 0) = P(F) = q$*.*

Discrete Random Variables

Definition. *A subset* $\mathcal{S}$ *of the real line* $\mathbb{R}$ *is said to be discrete if for every whole number* n *there are only finitely many elements of* $\mathcal{S}$ *in the interval* $[-n, n]$*. So a* ***finite*** *subset of* $\mathbf{R}$ *is* ***discrete*** *but so is the set of integers* $\mathbf{Z}$*.*

Definition. *A random variable is said to be discrete if its set of possible values is a discrete set.*

A possible value means a value x_0 so that $P(X = x_o) \neq 0$.

Definition. *The probability mass function (abbreviated pmf) of a discrete random variable X is the function P_X defined by*

$$P_X(x) = p(x = X)$$

We will often write $p(x)$ instead of $P_X(x)$.

Note:

(i) $p(x) \geq 0$

(ii) $\sum_{all\ possible_x} p(x) = 1$

(iii) $p(x) = 0$ *for all x outside a countable set.*

Graphical Representations of pmf's

There are two kinds of graphical representations of pmf's, the "line graph" and the "probability histogram". We will illustrate them with the Bernoulli distribution with parameter p.

x	1	0
$P(X = x)$	p	q

table

q {| |}p

0 1

line graph

q {| |}p

$-\frac{1}{2}$ 0 $\frac{1}{2}$ 1 $\frac{3}{2}$

histogram

We also illustrate these for the basic example (pg.5)

x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

table

In the next picture (for the probability histogram) the bases of all four rectangles have length 1 and they are centered at 0,1,2 and 3 respectively. The point is to make the area of each box equal to the corresponding probability, so the first box has area 1/8, the second has area 3/8, the third has area 3/8 and the fourth has area 1/8.

The Cumulative Distribution Function

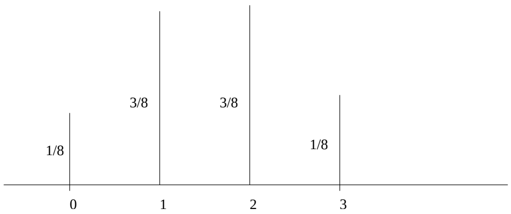

Figure 1: The probability mass function

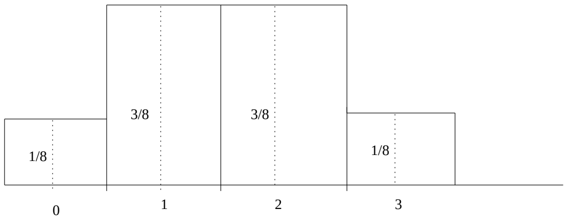

Figure 2: The probability histogram

The cumulative distribution function F_X (abbreviated cdf) of a discrete random variable X is defined by

$$F_X(x) = P(X \leq x).$$

We will often write $F(x)$ instead of $F_X(x)$.

Bank account analogy. Suppose you deposit \$1000 at the beginning of every month. The "line graph" of your deposits is on the previous page. We will use t (time as our variable). Let

$$F(t) = \text{the amount you have accumulated at time } t$$

What does the graph of F look like?

It is critical to observe that whereas the deposit function is zero for all real numbers except $0, 1, 2, \cdots, 11$ the cumulation function is never zero between 1 and ∞.

You would be very upset if you walked into the bank on July 5th and they told you your balance was zero – you never took any money out. Once your balance was nonzero it was never thereafter.

Back to Probability. The cumulative distribution $F(x)$ is "the total probability you have accumulated when you get to x". Once it is nonzero it is never zero again ($p(x) \geq 0$ means "you never take any probability out").

To write out $F(x)$ in formulas you will need several (many) formulas. There should never be *equalities* in your formulas, only *inequalities*.

The cdf for the Basic Example.

We have

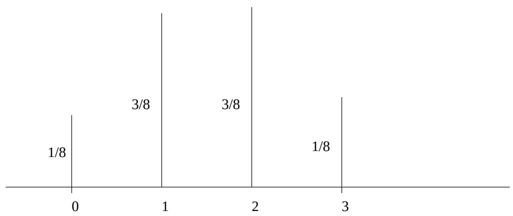

Figure 3: pmf for Bin(3,1/2).

so we start accumulation proability at $x = 0$.

Ordinary Graph of F

We have

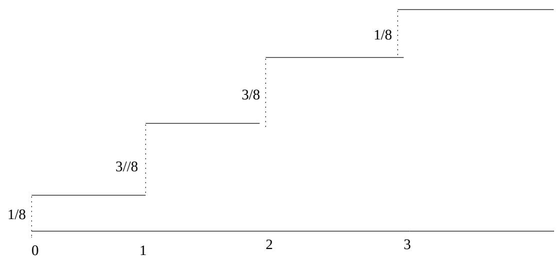

Figure 4: cdf for Bin(3,1/2).

Formulas for F

$$F(x) = \left\{ \begin{array}{ll} 0 & x < 0 \\ \frac{1}{8} & 0 \leq x < 1 \\ \frac{4}{8} & 1 \leq x < 2 \\ \frac{7}{8} & 1 \leq x < 3 \\ 1 & 3 \leq x \end{array} \right\}$$

You can see you have to be careful about the inequalities on the right-hand side.

Expected Value

Definition. *Let X be a discrete random variable with set of possible values D and pmf $p(x)$. The expected value or mean value of X denoted $E(X)$ or μ is defined by*

$$E(X) = \sum_{x \in D} xP(X = x) = \sum_{x \in D} xp(x).$$

Remark . *$E(X)$ is the whole point for monetary games of chance, e.g., lotteries, blackjack, slot machines.*

If $X =$ your payoff, the operators of these games make sure $E(X) < 0$. Thorp's cord-counting strategy in blackjack changed $E(X) < 0$ (because ties went to the dealer) to $E(X) > 0$ to the dismay of the casinos. See "How to Beat the Dea;er" by Edward Thorp (a math professor at UC Irvine).

Example . *The expected value of the Bernoulli distribution*

$$E(X) = \sum_{x} xP(X = x) = (0)(q) + (1)(p) = p.$$

The expected value for the basic example (so the expected number of heads)

$$E(X) = (0)(\frac{1}{8}) + (1)(\frac{3}{8}) + (2)(\frac{3}{8}) + (3)(\frac{1}{8}) = \frac{3}{2}.$$

The expected value is NOT the most probable value.

For the basic example the possible values of X where 0, 1, 2, 3 and so $\frac{3}{2}$ was not even a possible value

$$P(X = \frac{3}{2}) = 0.$$

The most probable values were 1 and 2 (tied) each with probability $\frac{3}{8}$.

Rolling a Die

$$\begin{aligned} E(X) &= (1)(\frac{1}{6}) + (2)(\frac{1}{6}) + (3)(\frac{1}{6}) + (4)(\frac{1}{6}) + (5)(\frac{1}{6}) + (6)(\frac{1}{6}) \\ &= \frac{1}{6}[1 + 2 + 3 + 4 + 5 + 6] \\ &= \frac{1}{6}\frac{(7)(6)}{2} = \frac{7}{2}. \end{aligned}$$

Variance

The expected value does not tell you everything you want to know about a random variable (how could it, it is just one number). Suppose you and a friend play the following game of chance. Flip a coin. If a head comes up you get \$1, if a tail comes up you pay your friend \$1. So if $X =$ your payoff

$$X(H) = +1 \quad , \quad X(T) = -1$$
$$E(X) = (+1)(\frac{1}{2}) + (-1)(\frac{1}{2}) = 0.$$

so this is a fair game. Now suppose you play the game changing \$1 to \$1000. It is still a fair game

$$E(X) = (1000)(\frac{1}{2}) + (-1000)(\frac{1}{2}) = 0$$

but I personally would be very reluctant to play this game. The notion of variance is designed to capture the difference between the two games.

Definition. *Let X be a discrete random variable with set of possible values D and expected value μ. Then the variance of X, denoted $V(X)$ or σ^2 is defined by*

$$\begin{aligned} V(X) &= \sum_{x \in D} (x-\mu)^2 P(X=x) \\ &= \sum_{x \in D} (x-\mu)^2 xp(x). \qquad (*) \end{aligned}$$

The standard deviation σ of X is defined to be the ***square-root*** *of the variance*

$$\sigma = \sqrt{V(X)} = \sqrt{\sigma^2}.$$

Check that for the two games above (with your friend) $\sigma = 1$ for the \$1 game and $\sigma = 1000$ for the \$1000 game.

The Shortcut Formula for $V(X)$

The number of arithmetic operations (subtractions) necessary to compute σ^2 can be *greatly* reduced by using

Proposition.

(i) $V(X) = E(X^2) - E(X)^2$

or

(ii) $V(X) = \sum_{x \in D} x^2 p(x) - \mu^2$

In the formula (*) you need $\#(D)$ subtractions (for each $x \in D$ you have to subtract μ then square ...). For the shortcut formula you need only one. ***Always use the shortcut formula.***

Remark . *Logically, versions (i) of the shortcut formula is not correct because we haven't yet defined the random variable X^2. We will do this soon – "change of random variable".*

Example (The fair die). *X = outcome of rolling a die. We have seen (pg. 24)*

$$\begin{aligned} E(X) &= \mu = \frac{7}{2} \\ E(X^2) &= (1)^2(\frac{1}{6}) + (2)^2(\frac{1}{6}) + (3)^2(\frac{1}{6}) + (4)^2(\frac{1}{6}) + (5)^2(\frac{1}{6}) + (6)^2(\frac{1}{6}) \\ &= \frac{1}{6}[1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2] \\ &= \frac{1}{6}[91]. \end{aligned}$$

so

$$E(X^2) = \frac{91}{6}.$$

Hence,

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4}.$$

Remark .

(i) How did I know

$$1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 = 91.$$

This because

$$\sum_{k=1} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Now plug in $n = 6$.

(ii) In the formula for $E(X^2)$ ***don't square the probabilities*** *(that is $\frac{1}{6}$)*

$$E(X^2) = (1^2)(\frac{1}{6}) + (2^2)(\frac{1}{6}) + ...$$