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 introduction the concept of “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-valued 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.

The 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 are

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

Answers

Note $\sharp(S) = 8$

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

We will tabulate this

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

Get used 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 $1, 2, 3, \ldots, n$ each with probability $\frac{1}{n}$ (so roll a fair die with n faces").

Bernoulli Random Variable

Usually random variables are introduced to make things numerical. We illustrate this by an important example - page 8. 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
$P(X = 0)$	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 on 0 we must have

$$p + q = 1.$$

So we get

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

This 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.

Z Often we call a "success" something that is in fact for from an actual success- e.g., a machine breaking down.

In order to obtain a Bernoulli random variable if we first assign probabilities to S and F by

$$P(S) = p \quad \text{and} \quad P(F) = q$$

so again $p + q = 1$.
Thus the sample space of a a Bernoulli experiment will be denoted $\mathcal{S}$ (note that that the previous caligraphic $\mathcal{S}$ is different from Roman S) and is given by

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

We then obtain a Bernoulli random variable X on $\mathcal{S}$ by defining

$$X(S) = 1 \text{ and } X(F) = 0$$

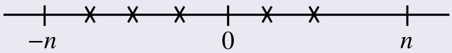
so

$$P(X = 1) = P(S) = p \text{ and } P(X = 0) = P(F) = q.$$

Discrete Random Variables

Definition

A subset S of the red line $\mathbb{R}$ is said to be discrete if for every whole number n there are only finitely many elements of S in the interval $[-n, n]$.



So a finite subset of $\mathbb{R}$ is discrete but so is the set of integers $\mathbb{Z}$.

Remark

The definition in the text on page 98 is wrong. The set of rational numbers $\mathbb{Q}$ is countably infinite but is not discrete. This is not important for this course but I find it almost unbelievablel that the editors of this text would allow such an error to run through nine editions of the text.

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 $P(X = x_0) \neq 0$.

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) $\sum_{\substack{\text{all possible} \\ X}} p(x) = 1$

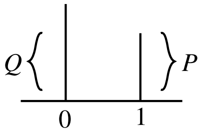
(iii) $p(x) = 0$ *for all X outside a discrete set.*

Graphical Representations of Proof's

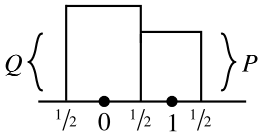
There are two kinds of graphical representations of proof's, the "line graph" and the "probability histogram". We will illustrate them with the Bernoulli distribution with parameter P.

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

table



line graph

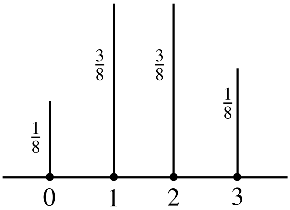


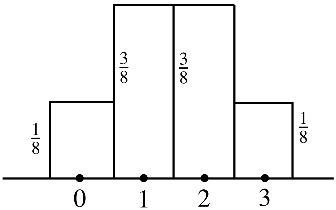
histogram

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





The Cumulative Distribution Function

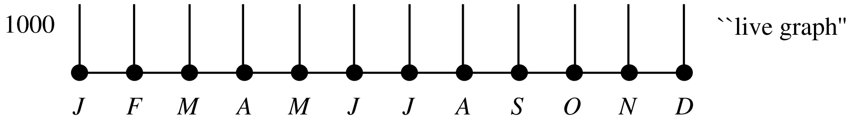
The cumulative distribution function F_X (abbreviated cdf) of a discrete random variable X is defined by

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

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

Bank account analogy

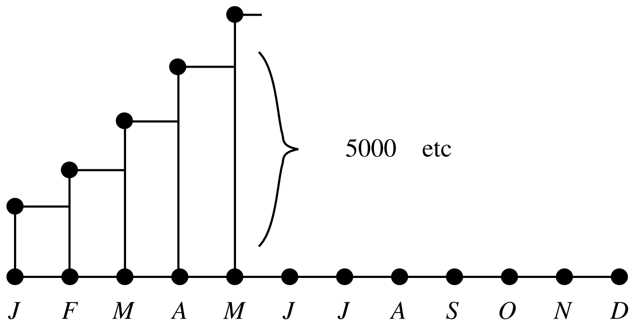
Suppose you deposit 1000 at the beginning of every month.



The "line graph" of your deposits is on the previous page. We will use t (time as our variable). Let

$F(t)$ = 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 on page 15 is zero for all real numbers except 12 numbers (Here I am replacing the symbols for the months e.g. J by the numbers 1 through 12) the cumulative distribution function is never zero between 1 and ∞. You would be very upset if you walked into the bank on July 5^{th} and they told you your balance was zero - you never took any money out. Once your balance was nonzero it was never zero thereafter.

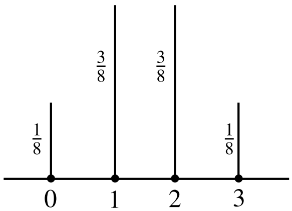
Back to Probability

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 $p(x)$ in formulas you will need several (many) formulas. There should never be EQUALITIES in you formulas only INEQUALITIES.

The cdf for the Basic Example

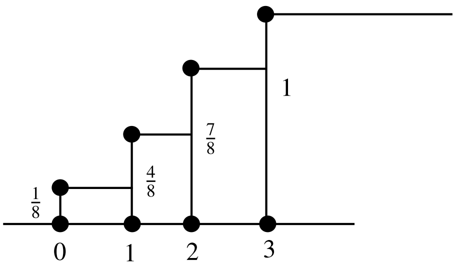
We have



line graph of P

So we start accumulation probability at $X = 0$

Ordinary Graph of F



Formulas for F

$$\left\{\begin{array}{ll} 0 & x \le 0 \\ \frac{1}{8} & 0 \le x < 1 \\ \frac{4}{8} & 1 \le x < 2 \\ \frac{7}{8} & 1 \le x < 3 \\ 1 & 3 \le x \end{array}\right\} \longleftarrow \text{be careful}$$

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

The relation between p and F

We will need the relation between the probability mass function $p(x)$ and the cumulative distribution function $F(x)$. Recall that if $F(x)$ is a function of an integer variable x then the **backward difference** function (discrete derivative) ΔF of F is defined by

$$\Delta F(x) = F(x) - F(x-1).$$

The relation we want is

Theorem 1

$$p(x) = \Delta F(x)$$

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 denote $E(X)$ or μ (Greek letter mu) is defined by

$$E(X) = \sum_{x \in D} \times P(X = x) = \sum_{x \in D} \times p(x)$$

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 card-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 Dealer" by Edward Thorp (a math professor at UCIrvine).

Examples

The expicted value of the Bernoulli distribution.

$$E(X) = \sum_x \times P(X = x) = (0)(q) + (1)(p)$$
$$= P$$

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

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

$\mathcal{Z}$ *The expected value is NOT the most probable value.*

Examples (Cont.)

For the basic example the possible values of X where $0, 1, 2, 3$ so $3/2$ was not even a possible value

$$P\left(X = 3/2\right) = 0$$

The most probable values were 1 and 2 (tied) each with probability $3/8$.

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

Rolling of a Die

Variance

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

$$X(11) = +1, \; X(T) = -1$$

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

so this is a fair game.

Now suppose you play the game changing §1 to §1000. It is still a fair game

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

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 (sigma squared) is defined by

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

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

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

Definition (Cont.)

Check that for the two games above (with your friend)
$\sigma = 1$ *for the §1 game*
$\sigma = 1000$ *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) $V(X) = \sum_{x \in D} X^2 P(X) - \mu^2$

Proposition (Cont.)

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

Remark

Logically, version (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)

$$E(X) = \mu = \frac{7}{2}$$

$$E(X^2) = (1)^2\left(\frac{1}{6}\right) + (2)^2\left(\frac{1}{6}\right) + (3)^2\left(\frac{1}{6}\right) + (4)^2\left(\frac{1}{6}\right) + (5)^2\left(\frac{1}{6}\right) + (6)^2\left(\frac{1}{6}\right)$$

$$= \frac{1}{6}\left[1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2\right]$$

$$= \frac{1}{6}[91] \longleftarrow \textit{later}$$

So

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

Here

don't forget to square μ

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

Remarks

(1) *How did I know*

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

This because

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

Now plug in $n = 6$.

(2) *In the formula for $E(X^2)$ don't square the probabilities*