Chapter 2 Random Variables

§2.1 Discrete Random Variables

It is frequently the case when an experiment is performed that we are mainly interested in some function of the outcomes as opposed to the actual outcome itself.

e.g. 1. In tossing dice, we are often interested in the sum of the two dice and are not really concerned about the separate values of each die. That is, we may be interested in knowing that the sum is 7 and not concerned over whether the actual outcome was (1, 6) or (2, 5) or (3, 4) or (4, 3) or (5, 2) or (6, 1).

· e.g. 2. A coin is thrown 3 times, we may be interested in the total number of heads that occur and not care at all about the actual head-tail sequence that results. For example, 2 heads. Not $\{hht, hth, thh\}$

- These quantities of interest, or more formally, these real-valued functions defined on the sample space, are known as random variables.

- Because the value of a random variable is determined by the outcome of the experiment, we may assign probabilities to the possible values of the random variable.

Expl. 1. Suppose a coin is tossed 3 times. If we let X be the total number of heads in this experiment, then X is a random variable.

possible values for X are 0, 1, 2, 3.

Assume the coin is fair, then $P(X=0)=P(\{ttt\})=\frac{1}{8}$.

$P(X=1)=P(\{htt, tht, tth\})=3/8$, $P(X=2)=P(\{hht, hth, thh\})=3/8$,

$P(X=3)=P(\{hhh\})=1/8$

Put together:

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

$p(x_i)=P(X=x_i)$, $x_i=0, 1, 2, 3$. is called: probability mass function. p.m.f.

or frequency function.

$$\begin{cases} p(x_i)\geq 0 \\ \sum_{i=1}^{n} p(x_i)=1 \end{cases}$$

Expl. 2. Independent trials, consisting of the flipping of a coin having probability p of coming up heads are continually performed until either a head occurs or a total of n flips is made. If we let X denote the number of times the coin is flipped, then X is a r.v. find $p(x)=?$ p.m.f.?

Ans: X = number of times the coin is flipped. possible values for X are $1, 2, 3, \cdots, n$.

$P(X=1) = P(h_1) = p.$

$P(X=2) = P(t_1 h_2) = (1-p)\cdot p$

$P(X=3) = P(t_1 t_2 h_3) = (1-p)(1-p)\cdot p = (1-p)^2\cdot p.$

$\cdots$

$$
\begin{aligned}
P(X=n) &= p(t_1 t_2 \cdots t_{n-1} h_n \cup t_1 t_2 \cdots t_{n-1} t_n) \\
&= P(t_1 t_2 \cdots t_{n-1} h_n) + P(t_1 t_2 \cdots t_{n-1} t_n) \\
&= (1-p)^{n-1}\cdot p + (1-p)^n \\
&= (1-p)^{n-1}[p + (1-p)] = (1-p)^{n-1}.
\end{aligned}
$$

p.m.f.

X	1	2	3	$\cdots$	k	$\cdots$	$n-1$	n
$p(x)$	p	$(1-p)p$	$(1-p)^2\cdot p$	$\cdots$	$(1-p)^{k-1}\cdot p$	$\cdots$	$(1-p)^{n-2}\cdot p$	$(1-p)^{n-1}$

$\begin{cases} p(x_i) \geq 0 \\ \sum_i p(x_i) = 1. \end{cases}$

$$
\begin{aligned}
& p + (1-p)\cdot p + (1-p)^2 p + \cdots + (1-p)^{n-2}\cdot p + (1-p)^{n-1} \\
&= p[1 + (1-p) + (1-p)^2 + \cdots + (1-p)^{n-2}] + (1-p)^{n-1}
\end{aligned}
$$

using $1 + q + q^2 + \cdots + q^m = \dfrac{1-q^{m+1}}{1-q}$ for $q<1$, with $m = n-2$. $q = 1-p$.

$$
\begin{aligned}
&= p\cdot\frac{1-(1-p)^{n-1}}{1-(1-p)} + (1-p)^{n-1} \\
&= 1-(1-p)^{n-1} + (1-p)^{n-1} \\
&= 1.
\end{aligned}
$$

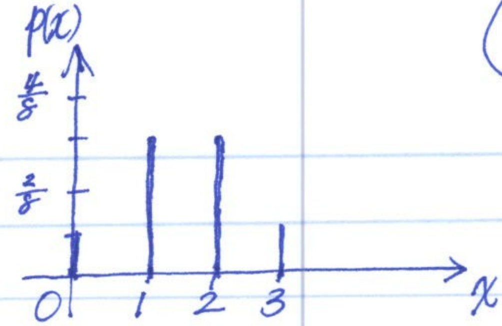
CDF (cdf) Cumulative distribution Function.

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

Expl3. Consider Expl. 1,

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

Find $F(x)$?



Ans. $F(x)=P(X\le x)$,

if $x<0$. $F(x)=P(X\le x)=0$

$x=0$. $F(0)=P(X\le 0)=P(X=0)=\frac{1}{8}$

$x=1$, $F(1)=P(X\le 1)=P(X=0 \text{ or } 1)=P(X=0)+P(X=1)=\frac{1}{8}+\frac{3}{8}=\frac{4}{8}$

$0<x<1$, $F(x)=P(X\le x)=P(X=0)=\frac{1}{8}$.

$x=2$, $F(x)=F(2)=P(X\le 2)=P(X=0 \text{ or } 1 \text{ or } 2)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}=\frac{7}{8}$.

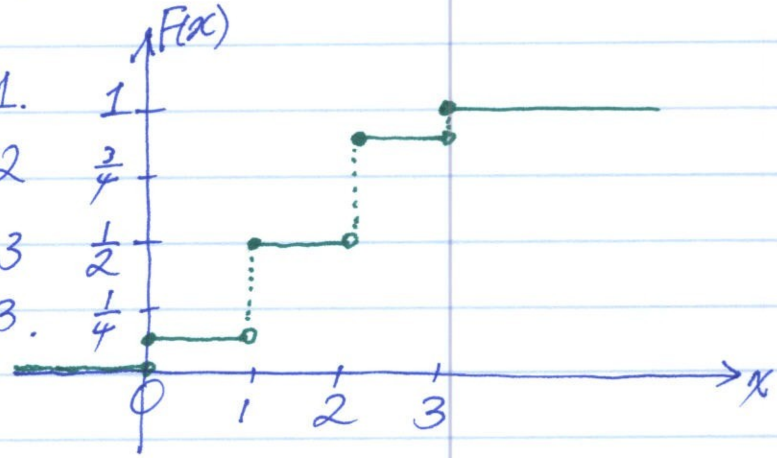
$1<x<2$, $F(x)=P(X\le x)=P(X=0 \text{ or } 1)=\frac{1}{8}+\frac{3}{8}=\frac{4}{8}$.

$x=3$, $F(x)=F(3)=P(X=0 \text{ or } 1 \text{ or } 2 \text{ or } 3)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$.

$x>3$. $F(x)=P(X\le x)=P(X=0 \text{ or } 1 \text{ or } 2 \text{ or } 3)=1$.

$2<x<3$, $F(x)=P(X=0 \text{ or } 1 \text{ or } 2)=\frac{1}{8}+\frac{3}{8}+\frac{3}{8}=\frac{7}{8}$.

$$F(x)=\begin{cases}0, & x<0\\ \frac{1}{8}, & 0\le x<1.\\ \frac{4}{8}, & 1\le x<2\\ \frac{7}{8}, & 2\le x<3\\ 1, & x\ge 3.\end{cases}$$



$0\le F(x)\le 1$,

$\lim_{x\to-\infty}F(x)=0$, $\lim_{x\to+\infty}F(x)=1$.

From $F(x)$ to determine $p(x)$?

$P(X=0)=F(0)-F(0-)=\frac{1}{8}-0=\frac{1}{8}$.

$P(X=1)=F(1)-F(1-)=\frac{4}{8}-\frac{1}{8}=\frac{3}{8}$

$P(X=2)=F(2)-F(2-)=\frac{7}{8}-\frac{4}{8}=\frac{3}{8}$

$P(X=3)=F(3)-F(3-)=1-\frac{7}{8}=\frac{1}{8}$.

r.v.s X and Y are independent: $X\sim p(x_i)$, $i=1,2,\cdots$

$Y\sim p(y_j)$, $j=1,2,\cdots$

if $P(X=x_i \text{ and } Y=y_j)=P(X=x_i)P(Y=y_j)$, $\forall\, i,j=1,2,\cdots$

§ 2.1.1. Bernoulli Random Variables

X	0	1
$p(x)$	$1-p$	p

$P(X=0)=p(0)=1-p$
$P(X=1)=p(1)=p.$

A Bernoulli r.v. X takes on only two values: 0 and 1.

useful representation: $P(X=x)=p(x)=\begin{cases} p^x\cdot(1-p)^{1-x}, & \text{if } x=0 \text{ or } 1 \\ 0, & \text{otherwise}\end{cases}$

e.g. 1. If A is an event, define indicator r.v. I_A:

$I_A(\omega)=\begin{cases}1, & \text{if } \omega\in A\\ 0, & \text{if } \omega\notin A.\end{cases}$ then I_A is a Bernoulli r.v.

$P(I_A=1)=P(A)=:p.$
$P(I_A=0)=P(A^c)=1-p,$

I_A	0	1
$p(x)$	$1-p$	p

.

Binomial distribution.

Suppose n independent experiments, or trials, are performed, where n is fixed. and each trial results in a "success" with probability p and a "failure" with probability $1-p$. The total number of successes, X is a binomial r.v. with parameters n and p. e.g. toss a coin 10 times; toss a die 20 times,

$P(X=k)=?$ p.m.f.

$$P(X=k)=p(k)=\underbrace{\binom{n}{k}}_{\text{total number of such sequences}}\underbrace{p^k\cdot(1-p)^{n-k}}_{\text{any particular sequence of } k \text{ successes}},\quad k=0,1,2,\cdots,n.$$

$X\sim B(n,p)$.

In R, $p(k)=dbinom(k,n,p)$

Expl. A. Tay-Sachs disease is a rare but fatal disease of genetic origin occurring chiefly in infants and children, especially those of Jewish or eastern European extraction. If a couple are both carriers of Tay-Sachs disease, a child of theirs has prob. 0.25 of being born with the disease. If such a couple has four children, what is the p.m.f for the number of children who will have the disease?

Ans: Let X denote the number of children with the disease. Then $X\sim B(4, 0.25)$.

$P(X=0)=\binom{4}{0}(0.25)^0(0.75)^4=0.316.$ from $P(X=k)=\binom{4}{k}(0.25)^k(1-0.25)^{4-k}$

$P(X=1)=\binom{4}{1}(0.25)^1(0.75)^3=0.422.$

$p(k) = dbinom(k, 4, 0.25)$

p.m.f.

X	0	1	2	3	4
$p(x)$	0.316	0.422	0.211	0.047	0.004

prob. 13. A multiple-choice test consists of 20 items, each with four choices. A student is able to eliminate one of the choices on each question as incorrect and chooses randomly from the remaining three choices. A passing grade is 12 items or more correct.

a. What's prob. the student passes?

b. Answer (a) again, assuming the student can eliminate two of the choices on each question.

Ans: a. Let X be the number of items the student answers correctly.

then $X \sim B(20, \frac{1}{3})$, $P(X=k)=\binom{20}{k}(\frac{1}{3})^k(1-\frac{1}{3})^{20-k}$.

Let B = "student passes"

then $B=\{X\geq 12\}$, $P(B)=P(X\geq 12)=\sum_{k=12}^{20}\binom{20}{k}(\frac{1}{3})^k(1-\frac{1}{3})^{20-k}$.

$\left(=\sum_{k=12}^{20}P(X=k)\right)$

using R, $P(B)=1-P(X\leq 11)$ In R: $F(11)=P(X\leq 11)$ $= pbinom(11, 20, 1/3)$.

$= 1-F(11) = 1-pbinom(11, 20, 1/3)$

$= 0.012973.$

b. Let Y be the number of items that student answers correctly.

then $Y \sim B(20, \frac{1}{2})$.

$P(B)=P(Y\geq 12)=1-P(Y\leq 11)=1-F(11)$

$=1-pbinom(11, 20, \frac{1}{2})$

$=0.251722.$