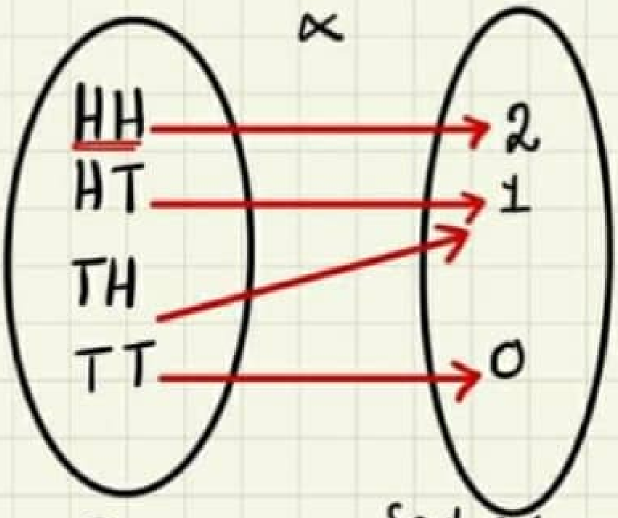
Statistics, lecture 12 :- Chapter 4

*Random variable:- (r.v.)
A function from the sample space ($\Omega$) to a set of real numbers "x, y"

e.g) when tossing a fair coin 2 times, let the (r.v.) X be the no. of heads obtained

sol) $\Omega = \{HH, HT, TT, TH\}$

$\propto$




$\Omega$ $\{0, 1, 2\} \Rightarrow$ support.

$X(HH) = 2$
$X(HT) = 1$
$X(TH) = 1$
$X(TT) = 0$

حسب المتغيّر العشوائي الذي يحدد السؤال نعرف الارقام.

*Support for (r.v.)*

Countable set $\{0, 1, 2\}$ "Discrete"

Interval $0 < X < 1$ "Continuous"



*Discrete (r.v.)* ع نركّز عليه

(r.v.): X represents a value associated with each outcome of a Probability experiment.

يعني بيربط كل نتيجة من نتائج التجربة برقم

Example 1, book, p 213:

1. Discrete $\Rightarrow \{0, 1, 2, 3 \ldots 500\}$
2. Continuous $\Rightarrow 0 < X < 21$

*Probability distribution*

| $X$ | $x_1$ | $x_2$ | ... |
|---|---|---|---|
| $P(X)$ | $P(X=x_1)$ | $P(X=x_2)$ | ..... |

$r.f = \frac{f}{\sum f}$

e.g) In the previous example,

| $X$ | 0 | 1 | 2 |
|---|---|---|---|
| $P(X)$ | $\frac{1}{4}$ | $\frac{2}{4}$ | $\frac{1}{4}$ |

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

Probability distribution اسم الجدول

Note: $\sum_{x} P(X=x) = 1$

اذاً .. لمّا يعطيك جدول، كيف تتأكد انه

Probability distribution ؟

الجواب: لازم يحقق شرطين معاً

① $0 \le P(X=x) \le 1 \Rightarrow$

② $\sum P(X=x) = 1$

e.g) For the following Probability distribution, find k:

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.1 | 0.3 | k | 0.2 |

Sol)

$0.1 + 0.3 + k + 0.2 = 1$

$k = 0.4$

ملاحظة: ممكن يعطيك Frequency distribution ويطلب تحوله الى Probability distribution انظر للمثال التالي:

e.g)

| $x$ | 1 | 2 | 3 | 4 | sum |
|---|---|---|---|---|---|
| f | 4 | 3 | 2 | 1 | 10 |
| r.f | $\frac{4}{10}$ | $\frac{3}{10}$ | $\frac{2}{10}$ | $\frac{1}{10}$ | |

sol)



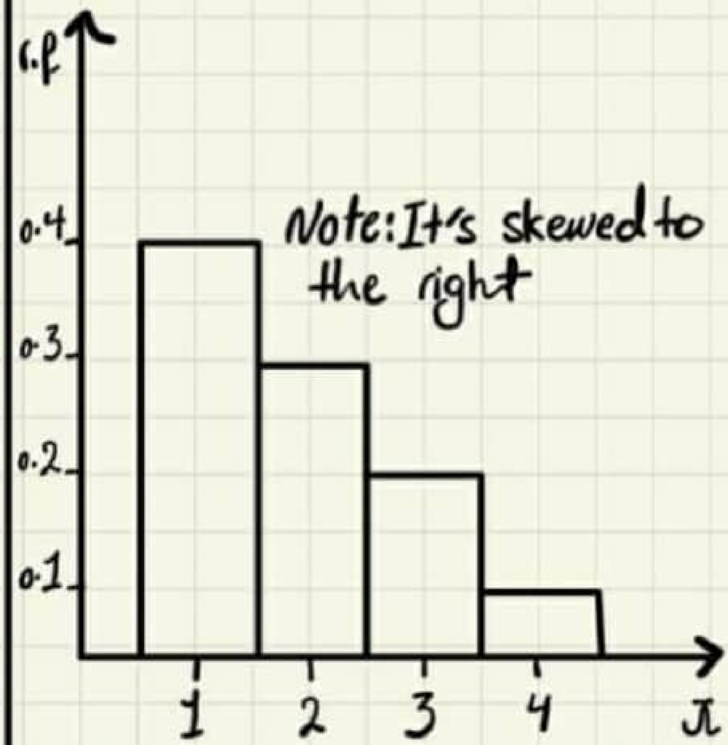
| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

لاحظ: الجواب عبارة عن قيم $x$ ونسبة كل وحدة فيهم

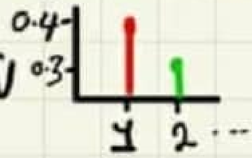
*Try to solve examples 2,3,4 book, P 214+215

*How to represent Probability distribution ؟*

*Draw a Histogram.




ملاحظة: يوجد طريقة اخرى وهي تكون برسم خط فقط من القيمة ل r.f الخاص بها




لكن كتابنا معتمد Histogram

e.g)

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

Find:

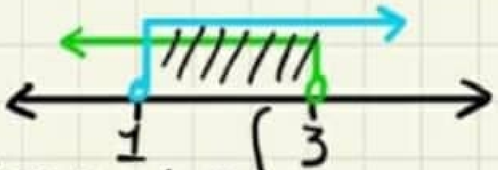
i) $P(1<X<3) = P(X=2) = 0.3$

ii) $P(X>3) = P(X=1) + (P X=2) = 0.7$

iii) $P(X>1 \mid X<3) =$

$$\frac{P(X>1 \cap X<3)}{P(X<3)} =$$

* كيف اوجد التقاطع في هاي الحالة؟
عن طريق خط الاعداد .




المنطقة المشتركة
بين السهمين.

$$\frac{P(X>1 \cap X<3)}{P(X<3)} = \frac{P(1<X<3)}{P(X<3)}$$

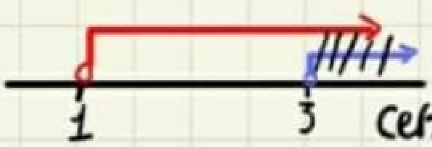
$$= \frac{0.3}{0.7}$$

$$= \frac{3}{7}$$

ملاحظة: من الآن فصاعداً اي جدول بيعطيك اياه وظيفتك تتأكد اذا Probability distribution أو لا .

iv) $P(X>1 \mid X>3) =$

$$\frac{P(X>3)}{P(X>3)} = 1$$




certain event

5

ليش certain؟ لانه بالتحليل: اذا انا متأكد انه $X>3$ ، شو احتمالية تكون $X>1$ ؟ أكيد ^^ .

v) $P(X<2 \mid X<3) = \frac{P(X<2)}{P(X<3)}$

$$= \frac{4}{7}$$


vi) $P(X<2 \mid X>3) = \frac{P(\emptyset)}{P(X>3)} = 0$


لا أنا ضامن انه $X>3$ ، شو احتمال تكون اقل من 2 مستحيلة ^^

vii) $P(X=2.5) = 0$

قاعدة: اي رقم مش موجود بالجدول احتماليته صفر .

viii) $P(X=5) = 0$

*** Probability density function ***

$f(x) = P(X=x)$ is called a p.d.f if:

i) $P(X=x) \geq 0$ for all $x$

ii) $\sum P(X=x) = 1$

6

e.g) If $P(X=x) = k.x$ , $x \Rightarrow \{1, 2, 3, 4\}$ is a P.d.f, then find k:

Sol) نقوم بتحويله الى جدول ثم نكمل الحل

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | k | 2k | 3k | 4k |

$k + 2k + 3k + 4k = 1$

$10k = 1 \Rightarrow k = 0.1$

e.g) If $P(X=x) = k.x^2$, $x \in \{1, 2, 3\}$ is a P.d.f, then find k:

| $x$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(X=x)$ | k | 4k | 9k |

$k + 4k + 9k = 1$

$14k = 1 \Rightarrow k = \frac{1}{14}$

***The commulative distribution function***

$F(a) = P(X \leq a)$ "c.d.f"

e.g)

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

Find:

i) $F(1) = P(X \leq 1) = P(1) = 0.4$

ii) $F(2) = P(X \leq 2) = P(1) + P(2) = 0.7$

iii) $F(3) = P(X \leq 3) = P(1) + P(2) + P(3) = 0.9$

iv) $F(4) = P(X \leq 4) = 1$

لاحظ :- $F(X)$ .. وكأننا اخذنا احتمالية $X$ و جمعناها لاحتمالية ما قبلها ، يعني نفس فكرة "c.f"

v) $F(5) = P(X \leq 5) = 1$

vi) $F(7) = 1$

لاحظ: اذا كانت $X$ أكبر من أكبر قيمة فإن $F(X) = 1$

vii) $F(0) = P(X \leq 0) = 0$

ملاحظة: اذا كانت $X$ اصغر من اصغر قيمة فان $F(X) = 0$

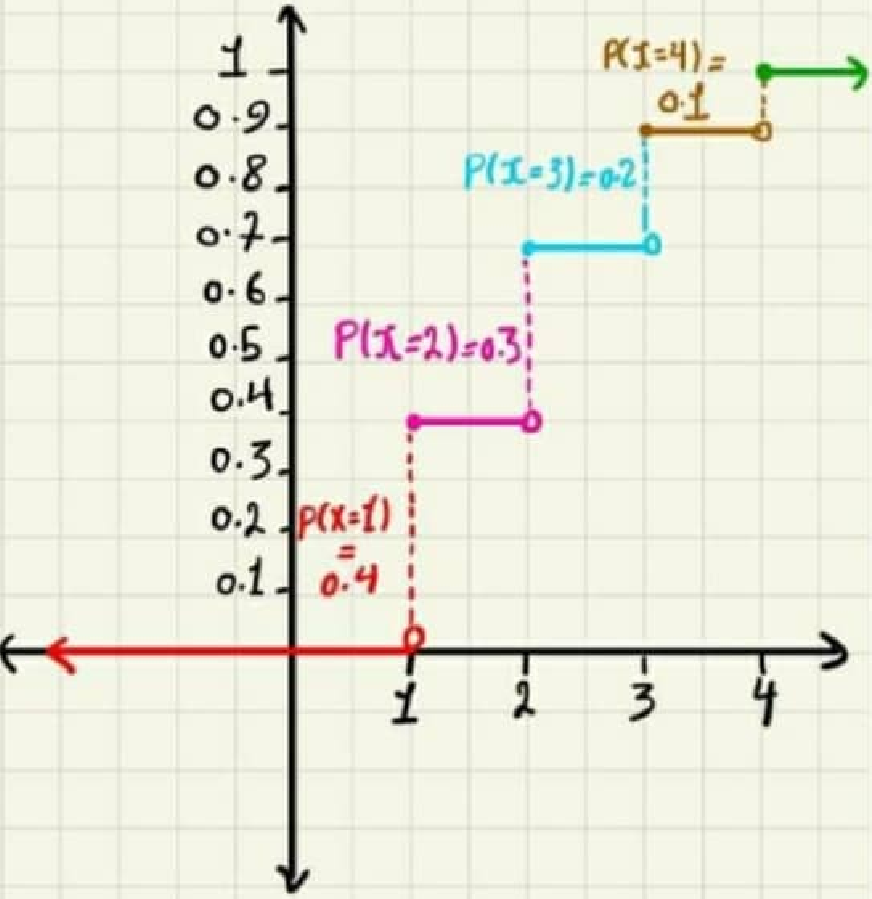
viii) $F(2.1) = P(X \leq 2.1) = P(X \leq 2)$

$F(2.5) = P(X \leq 2.5) = P(X \leq 2)$

$F(2.9) = P(X \leq 2.9) = \underbrace{P(X \leq 2)}_{= F(2)}$

لاحظ :- احتمالية اي رقم بين 2 و 3 مثلاً هي صفر لكن "c.d.f" يجمع الاحتمالية والاصفر منها بالتالي قيمة لاي رقم بين 2 و 3 هي $F(2)$ وقس ذلك على بقية الارقام

## * How to represent a "c.d.f"?




"step function" or "non-decreasing function"

* يمكن التعبير عنه من خلال الاقتران المتشعب ايضًا

$$F(x) = \begin{cases} 0 & , \ x < 1 \\ 0.4 & , \ 1 \le x < 2 \\ 0.7 & , \ 2 \le x < 3 \\ 0.9 & , \ 3 \le x < 4 \\ 1 & , \ 4 \le x \end{cases}$$

ممكن يعطيك اياه جاهز ويطلب "p.d.f"

⇒ اعمل جدول



| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

يعني رجعنا نحكي بالفكرة ^^

## * The expectation value *

Mean

related to the population

$$\mu = E(X) = \sum_{x} x \cdot P(X=x)$$

بضرب كل حالة باحتمالها وبعدين بجمعهم

e.g)

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

Find $\mu = E(X)$

sol)

$\mu = E(X) = 0.4 + 2 \times 0.3 + 3 \times 0.2 + 4 \times 0.1 = 2$

* ملاحظة: لو حسبنا الـ الاوساط {1,2,3,4} رح يطلع 2.5 ... طيب هو ليش طلع هنا 2 ؟ لانه هنا اسمه weighted Mean وكأنه بشكل ميزان بكفتيه [0.4, 0.3} , {0.1, 0.2] اكيد رح ترجح الاوزان الاثقل بالتالي رح يكون اقرب لل 2 .

* طيب متى بكون 2.5 بالضبط ؟ لما يكون symmetric

e.g)

| $x$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(X=x)$ | a | b | a |

Find $\mu$ :

sol)
we can't find $\mu$ by calculations
لكن لو انتبهنا فهو symmetric distribution
لانه الاحتمالية باليمين مساوية للاحتمالية باليسار
بالتالي $\underline{\mu = 2}$
قيمة $\mu$ هي العدد الي بالنص
اذا كان symmetric

e.g)

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | a | 0.3 | b | 0.1 |

If $\mu = 2$, then find a, b :

sol) $2 = a + 2\times0.3 + 3b + 0.1\times4$
$a + 3b = 1$ ---- ①

$a + 0.3 + b + 0.1 = 1$
$a + b = 0.6$ ---- ②
$a = 0.4$ , $b = 0.2$

* Solve example 5, book, P: 216
* " " 7, " , P: 218



* Properties for $E(X)$:-
i) $E(a) = a$ , a: constant
ii) $E(aX) = aE(X)$ وكأنك ضربت كل خانة بـ a
iii) $E(X \mp Y) = E(X) \mp E(Y)$
بيتوزع عالجمع والطرح لكن لا يوزع عالضرب
iv) $E(g(X)) = \sum_x g(x) \cdot P(X=x)$

e.g) If $\mu = E(X) = 10$, then find:
i) $E(\mu) = E(10) = 10$
ii) $E(2X) = 2 \times E(X) = 20$
iii) $E(3-2X) = E(3) - E(2X)$
$3 - 20 = -17$

e.g)

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

Find:
i) $E(X) = 0.4 + 0.6 + 0.6 + 0.1 = 2$
ii) $E(X^2) = 5$
نقوم بتربيع كل $x$ و نضربها في احتمالها و نجمع
iii) $E(\frac{1}{X}) = 0.4 + \frac{1}{2}\times0.3 + \frac{1}{3}\times0.2 + \frac{1}{4}\times0.1$

*The variance ($\sigma^2$)*
$\sigma^2 = var(X) = E(X-\mu)^2$
or $\sigma^2 = E(X^2) - (E(X))^2$

standard deviation $\rightarrow sd(X) = \sigma = \sqrt{var(X)}$

e.g)

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

Find $\sigma^2$:

sol) $E(X^2), E(X)$ دايماً ابدأ الحل بايجاد

$E(X)=2 \Rightarrow$ حليناه سابقاً

$E(X^2)=5 \Rightarrow$ " "

$var(X)=E(X^2)-(E(X))^2$

$5-(2)^2=1$

$S.D(X)=\sqrt{1}=1$

or

| $x-\mu$ | $-1$ | 0 | 1 | 2 |
|---|---|---|---|---|
| $(x-\mu)^2$ | 1 | 0 | 1 | 4 |
| $P(X=x)$ | 0.4 | 0.3 | 0.2 | 0.1 |

$E(X-\mu)^2 = 0.4+0+0.2+0.4$

$1$

shortcut formula is easier for calculations

*******

e.g) If $P(X=x)=k\cdot\left(\frac{2}{3}\right)^x,\ x\in\{0,1,2,\dots\}$ is a P.d.f then find:

i) The value of $k$

ii) $P(X\leq 9)$



e.g)

| $x$ | 0 | 1 | 2 | 3 | ... |
|---|---|---|---|---|---|
| $P(X=x)$ | $k$ | $k(\frac{2}{3})$ | $k(\frac{4}{9})$ | $k(\frac{8}{27})$ | ... |

$k+k(\frac{2}{3})+k(\frac{4}{9})+k(\frac{8}{27})+\dots=1$

↳ Geometric series (G.P)

$a+ar+ar^2+ar^3\dots$

i) $S_n=\dfrac{a(1-r^n)}{1-r}$

↳ sum of the first n terms

ii) $S_\infty=\dfrac{a}{1-r}$ ; $|r|<1$

نعود الى مثالنا بعد ان تذكرنا المتسلسلة الحسابية

$a=k \quad r=\frac{2}{3}$

$S_\infty=1$

$\dfrac{a}{1-r}=1 \Rightarrow \dfrac{k}{1-\frac{2}{3}}=1$

$1=\frac{1}{3}$

ii) $P(X\leq 9)$

$P(X=0)+P(X=1)+\dots P(X=9)$

$\underbrace{k+k(\tfrac{2}{3})+k(\tfrac{4}{9})+\dots+k(\tfrac{2}{3})^9}_{10\ \text{terms}}$

$S_{10}=\dfrac{\frac{1}{3}\left(1-(\frac{2}{3})^{10}\right)}{1-(\frac{2}{3})}$

$1-(\frac{2}{3})^{10}$

e.g) When throwing a fair die 2 times, S to be the sum for the 2 numbers obtained. Find the Prob. distribution of S

sol) $\Omega = \{(1,1),(1,2) \dots (6,6)\}$

| S | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P(S=s)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

symmetric