DISCRETE RANDOM VARIABLES

DEFINITION : A *discrete random variable* is a *function* $X(s)$ from a *finite* or *countably infinite* sample space $\mathcal{S}$ to the real numbers :

$$X(\cdot) \quad : \quad \mathcal{S} \quad \to \quad \mathbb{R} \ .$$

EXAMPLE : Toss a coin 3 times in sequence. The sample space is

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

and examples of random variables are

- $X(s) \ = \$ the number of Heads in the sequence ; *e.g.*, $X(HTH) \ = \ 2$,

- $Y(s) \ = \$ The index of the first H ; *e.g.*, $Y(TTH) \ = \ 3$,

  0 if the sequence has no H , *i.e.*, $Y(TTT) \ = \ 0$.

NOTE : In this example $X(s)$ and $Y(s)$ are actually *integers* .

Value-ranges of a random variable correspond to *events* in $\mathcal{S}$.

EXAMPLE : For the sample space

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

with

$$X(s) \;=\; \text{the number of Heads}\,,$$

the value

$X(s) \;=\; 2\,,$ corresponds to the event $\{HHT\,,\,HTH\,,\,THH\}\,,$

and the values

$1 < X(s) \leq 3\,,$ correspond to $\{HHH\,,\,HHT\,,\,HTH\,,\,THH\}\,.$

NOTATION : If it is clear what $\mathcal{S}$ is then we often just write

$$X \qquad \text{instead of} \qquad X(s)\,.$$

Value-ranges of a random variable correspond to *events* in $\mathcal{S}$,

and

$$\textit{events} \text{ in } \mathcal{S} \text{ have a } \textit{probability} \, .$$

Thus

$$\textit{Value-ranges} \text{ of a random variable have a } \textit{probability} \, .$$

EXAMPLE : For the sample space

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

with

$$X(s) = \text{the number of Heads} \, ,$$

we have

$$P(0 < X \leq 2) = \frac{6}{8} \, .$$

QUESTION : What are the values of

$$P(X \leq -1) \, , \; P(X \leq 0) \, , \; P(X \leq 1) \, , \; P(X \leq 2) \, , \; P(X \leq 3) \, , \; P(X \leq 4) \, ?$$

NOTATION : We will also write $p_X(x)$ to denote $P(X = x)$.

EXAMPLE : For the sample space

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

with

$$X(s) = \text{the number of Heads}\,,$$

we have

$$p_X(0) \equiv P(\ \{TTT\}\ ) = \tfrac{1}{8}$$

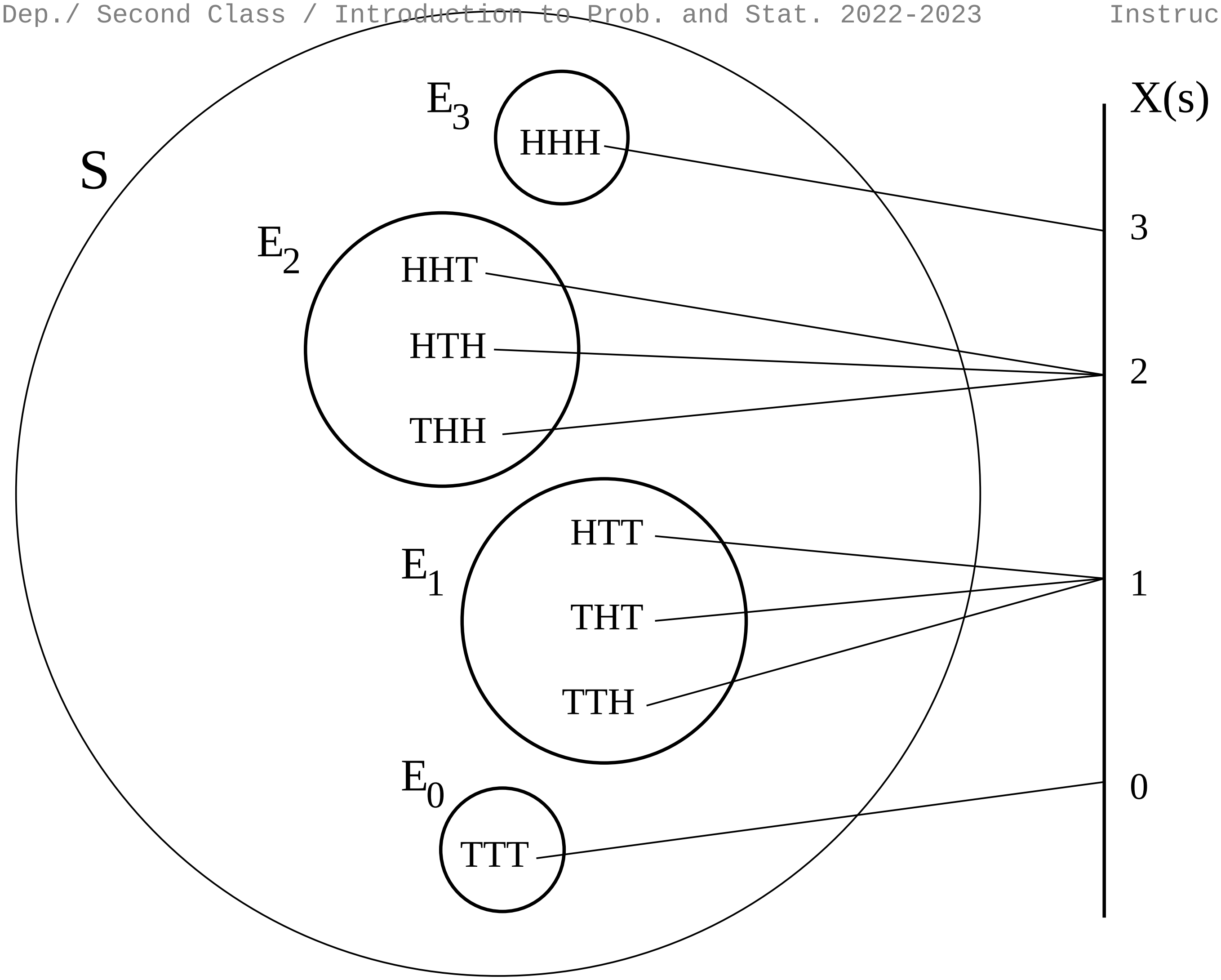
$$p_X(1) \equiv P(\ \{HTT\,,\, THT\,,\, TTH\}\ ) = \tfrac{3}{8}$$

$$p_X(2) \equiv P(\ \{HHT\,,\, HTH\,,\, THH\}\ ) = \tfrac{3}{8}$$

$$p_X(3) \equiv P(\ \{HHH\}\ ) = \tfrac{1}{8}$$

where

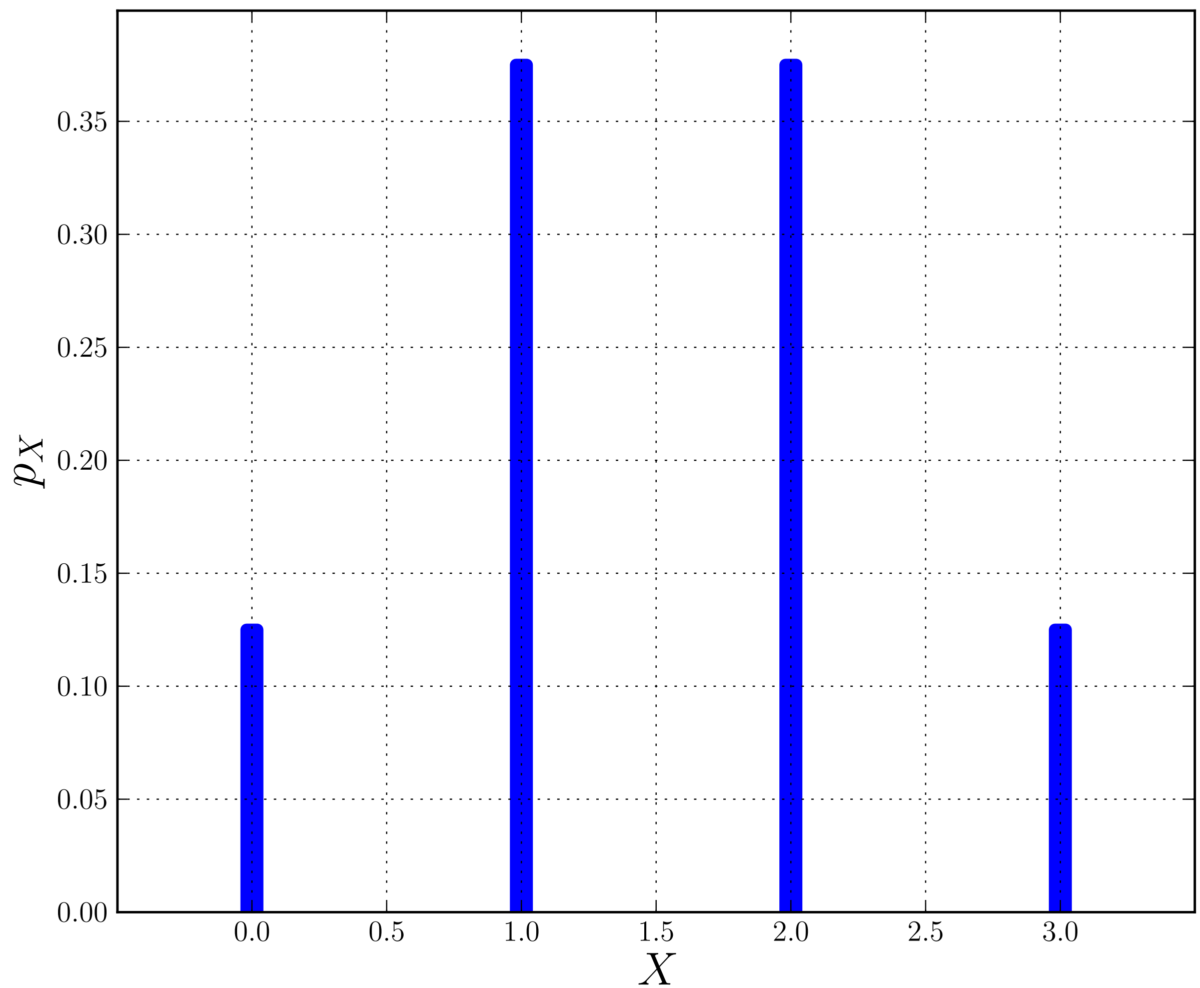
$$p_X(0) + p_X(1) + p_X(2) + p_X(3) = 1\,. \qquad (\text{ Why ? })$$



Graphical representation of X .

The *events* E_0, E_1, E_2, E_3 are *disjoint* since $X(s)$ is a *function* !

$(X : S \to \mathbb{R}$ must be defined for *all* $s \in S$ and must be *single-valued*.)



The graph of p_X .

DEFINITION :

$$p_X(x) \equiv P(X = x) ,$$

is called the *probability mass function* .

DEFINITION :

$$F_X(x) \equiv P(X \leq x) ,$$

is called the (*cumulative*) *probability distribution function* .

PROPERTIES :

- $F_X(x)$ is a *non-decreasing* function of x . (Why ?)
- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$. (Why ?)
- $P(a < X \leq b) = F_X(b) - F_X(a)$. (Why ?)

NOTATION : When it is clear what X is then we also write

$$p(x) \text{ for } p_X(x) \quad \text{and} \quad F(x) \text{ for } F_X(x) .$$

EXAMPLE : With $X(s) =$ the number of Heads , and

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

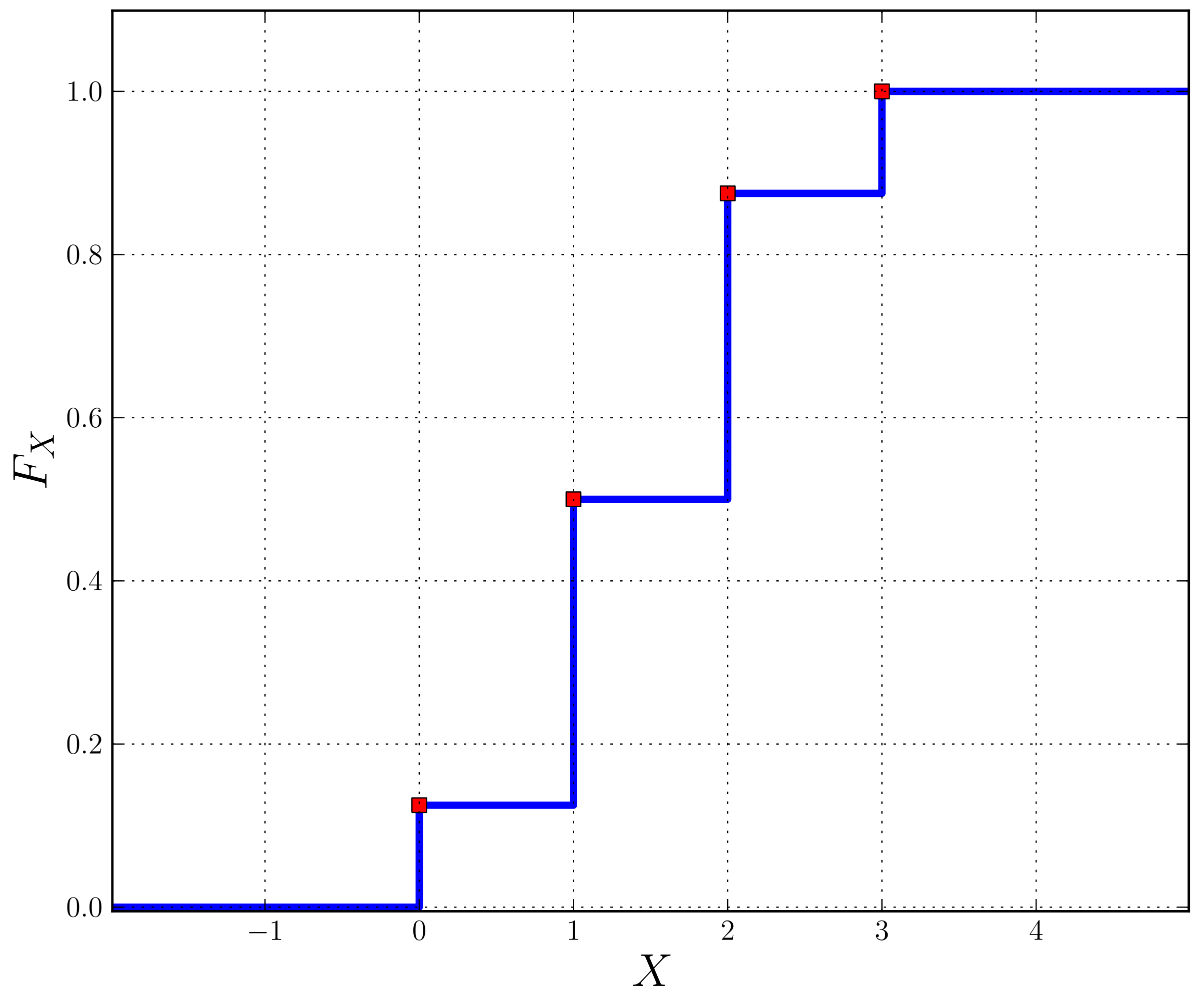
$$p(0) = \tfrac{1}{8} \quad , \quad p(1) = \tfrac{3}{8} \quad , \quad p(2) = \tfrac{3}{8} \quad , \quad p(3) = \tfrac{1}{8}\,,$$

we have the *probability distribution function*

$$\begin{aligned}
F(-1) &\equiv P(X \leq -1) = 0 \\
F(\,0) &\equiv P(X \leq \,0) = \mathbf{\tfrac{1}{8}} \\
F(\,1) &\equiv P(X \leq \,1) = \tfrac{4}{8} \\
F(\,2) &\equiv P(X \leq \,2) = \mathbf{\tfrac{7}{8}} \\
F(\,3) &\equiv P(X \leq \,3) = 1 \\
F(\,4) &\equiv P(X \leq \,4) = 1
\end{aligned}$$

We see, for example, that

$$\begin{aligned}
P(0 < X \leq 2) &= P(X = 1) + P(X = 2) \\
&= F(2) - F(0) = \tfrac{7}{8} - \tfrac{1}{8} = \tfrac{6}{8}\,.
\end{aligned}$$



The graph of the *probability distribution function* F_X .

EXAMPLE : Toss a coin until "Heads" occurs.

Then the sample space is *countably infinite* , namely,

$$\mathcal{S} = \{H \ , \ TH \ , \ TTH \ , \ TTTH \ , \ \cdots \} .$$

The *random variable* X is the *number of rolls* until "Heads" occurs :

$$X(H) = 1 \quad , \quad X(TH) = 2 \quad , \quad X(TTH) = 3 \quad , \quad \cdots$$

Then

$$p(1) = \tfrac{1}{2} \quad , \quad p(2) = \tfrac{1}{4} \quad , \quad p(3) = \tfrac{1}{8} \quad , \quad \cdots \qquad (\text{ Why ? })$$

and

$$F(n) = P(X \leq n) = \sum_{k=1}^{n} p(k) = \sum_{k=1}^{n} \frac{1}{2^k} = 1 - \frac{1}{2^n} ,$$

and, as should be the case,

$$\sum_{k=1}^{\infty} p(k) = \lim_{n\to\infty} \sum_{k=1}^{n} p(k) = \lim_{n\to\infty} (1 - \frac{1}{2^n}) = 1 .$$

NOTE : The outcomes in $\mathcal{S}$ *do not have equal probability* !

EXERCISE : Draw the *probability mass* and *distribution functions*.

$X(s)$ is the *number of tosses* until "Heads" occurs $\cdots$

REMARK : We can also take $\mathcal{S} \equiv \mathcal{S}_n$ as *all ordered outcomes of length* n. For example, for $n = 4$,

$$\begin{aligned}\mathcal{S}_4 \quad = \quad \{ \quad & \tilde{H}HHH \ , \quad \tilde{H}HHT \ , \quad \tilde{H}HTH \ , \quad \tilde{H}HTT \ , \\ & \tilde{H}THH \ , \quad \tilde{H}THT \ , \quad \tilde{H}TTH \ , \quad \tilde{H}TTT \ , \\ & T\tilde{H}HH \ , \quad T\tilde{H}HT \ , \quad T\tilde{H}TH \ , \quad T\tilde{H}TT \ , \\ & TT\tilde{H}H \ , \quad TT\tilde{H}T \ , \quad TTT\tilde{H} \ , \quad TTTT \quad \} \ .\end{aligned}$$

where for each outcome the first "Heads" is marked as $\tilde{H}$.

Each outcome in $\mathcal{S}_4$ has *equal probability* 2^{-n} (here $2^{-4} = \frac{1}{16}$) , and

$$p_X(1) = \tfrac{1}{2} \quad , \quad p_X(2) = \tfrac{1}{4} \quad , \quad p_X(3) = \tfrac{1}{8} \quad , \quad p_X(4) = \tfrac{1}{16} \quad \cdots ,$$

independent of n .

Joint distributions

The *probability mass function* and the *probability distribution function* can also be functions of *more than one variable*.

EXAMPLE : Toss a coin 3 times in sequence. For the sample space

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

we let

$$X(s) \;=\; \#\ \text{Heads} \quad , \quad Y(s) \;=\; \text{index of the first } H \quad (0 \text{ for } TTT)\,.$$

Then we have the *joint probability mass function*

$$p_{X,Y}(x,y) \;=\; P(X=x\,,\ Y=y)\,.$$

For example,

$$\begin{aligned} p_{X,Y}(2,1) &= P(X=2\,,\ Y=1) \\ &= P(\ 2\ \text{Heads}\ ,\ 1^{\text{st}}\ \text{toss is Heads}) \\ &= \tfrac{2}{8} \;=\; \tfrac{1}{4}\,. \end{aligned}$$

EXAMPLE : (continued $\cdots$) For

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

$X(s) =$ number of Heads, and $Y(s) =$ index of the first H ,

we can list the values of $p_{X,Y}(x,y)$:

Joint probability mass function $p_{X,Y}(x,y)$

	$y=0$	$\mathbf{y=1}$	$y=2$	$y=3$	$p_X(x)$
$x=0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x=1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$\mathbf{x=2}$	0	$\mathbf{\frac{2}{8}}$	$\frac{1}{8}$	0	$\frac{3}{8}$
$x=3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

NOTE :

- The *marginal probability* p_X is the probability mass function of X.
- The *marginal probability* p_Y is the probability mass function of Y.

EXAMPLE : (continued $\cdots$)

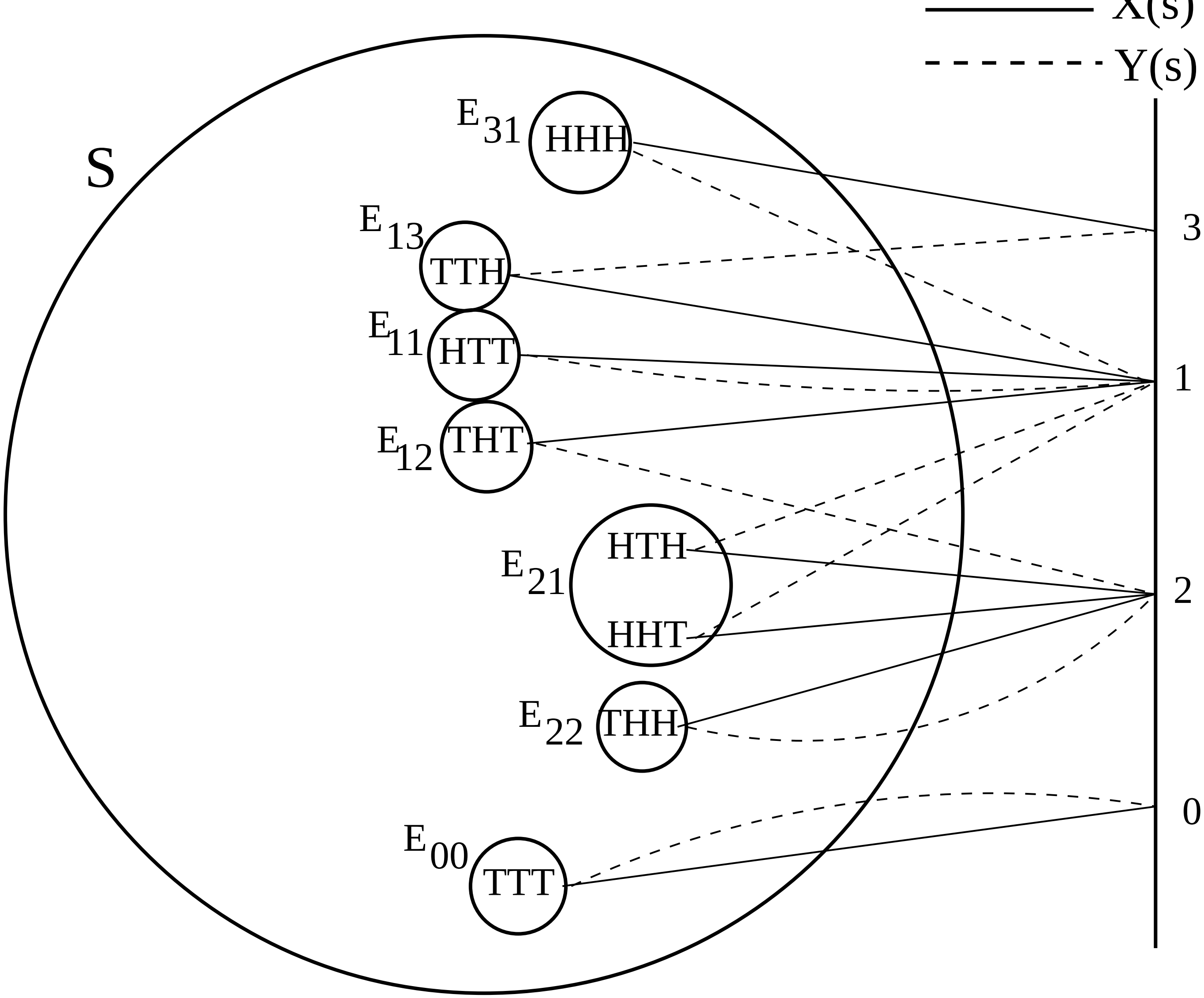
$X(s)$ = number of Heads, and $Y(s)$ = index of the first H .

	$y=0$	$\mathbf{y=1}$	$y=2$	$y=3$	$p_X(x)$
$x=0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x=1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$\mathbf{x=2}$	0	$\mathbf{\frac{2}{8}}$	$\frac{1}{8}$	0	$\mathbf{\frac{3}{8}}$
$x=3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\mathbf{\frac{4}{8}}$	$\frac{2}{8}$	$\frac{1}{8}$	1

For example,

- $X=2$ corresponds to the *event* $\{HHT\ ,\ HTH\ ,\ THH\}$.
- $Y=1$ corresponds to the *event* $\{HHH\ ,\ HHT\ ,\ HTH\ ,\ HTT\}$.
- $(X=2$ *and* $Y=1)$ corresponds to the *event* $\{HHT\ ,\ HTH\}$.

QUESTION : Are the events $X=2$ and $Y=1$ *independent* ?



The *events* $E_{i,j} \equiv \{ s \in S \ : \ X(s) = i \ , \ Y(s) = j \}$ are *disjoint* .

QUESTION : Are the events $X = 2$ and $Y = 1$ *independent* ?

DEFINITION :

$$p_{X,Y}(x, y) \quad \equiv \quad P(X = x \ , \ Y = y) \ ,$$

is called the *joint probability mass function* .

DEFINITION :

$$F_{X,Y}(x, y) \quad \equiv \quad P(X \leq x \ , \ Y \leq y) \ ,$$

is called the *joint* (*cumulative*) *probability distribution function* .

NOTATION : When it is clear what X and Y are then we also write

$$p(x, y) \quad \text{for} \quad p_{X,Y}(x, y) \ ,$$

and

$$F(x, y) \quad \text{for} \quad F_{X,Y}(x, y) \ .$$

EXAMPLE : *Three* tosses : $X(s) = \#$ Heads, $Y(s) =$ index 1^{st} H .

Joint probability mass function $p_{X,Y}(x,y)$

	$y=0$	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x=1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x=2$	0	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
$x=3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

Joint distribution function $F_{X,Y}(x,y) \equiv P(X \leq x, Y \leq y)$

	$y=0$	$y=1$	$y=2$	$y=3$	$F_X(\cdot)$
$x=0$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$x=1$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{4}{8}$	$\frac{4}{8}$
$x=2$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{6}{8}$	$\frac{7}{8}$	$\frac{7}{8}$
$x=3$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{7}{8}$	1	1
$F_Y(\cdot)$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{7}{8}$	1	1

Note that the distribution function F_X is a *copy* of the 4th column, and the distribution function F_Y is a *copy* of the 4th row. (Why ?)

In the preceding example :

Joint probability mass function $p_{X,Y}(x,y)$

	$y=0$	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x=1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x=2$	0	$\frac{2}{8}$	$\mathbf{\frac{1}{8}}$	$\mathbf{0}$	$\frac{3}{8}$
$x=3$	0	$\frac{1}{8}$	$\mathbf{0}$	$\mathbf{0}$	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

Joint distribution function $F_{X,Y}(x,y) \equiv P(X \leq x, Y \leq y)$

	$y=0$	$y=1$	$y=2$	$y=3$	$F_X(\cdot)$
$x=0$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$x=1$	$\frac{1}{8}$	$\mathbf{\frac{2}{8}}$	$\frac{3}{8}$	$\mathbf{\frac{4}{8}}$	$\frac{4}{8}$
$x=2$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{6}{8}$	$\frac{7}{8}$	$\frac{7}{8}$
$x=3$	$\frac{1}{8}$	$\mathbf{\frac{5}{8}}$	$\frac{7}{8}$	$\mathbf{1}$	1
$F_Y(\cdot)$	$\frac{1}{8}$	$\frac{5}{8}$	$\frac{7}{8}$	1	1

QUESTION : Why is

$$P(1 < X \leq 3\,, 1 < Y \leq 3) \;=\; F(3,3) - F(1,3) - F(3,1) + F(1,1)\ ?$$

EXERCISE :

Roll a *four-sided die* (tetrahedron) *two* times.

(The sides are marked 1 , 2 , 3 , 4 .)

Suppose each of the four sides is equally likely to end facing down.

Suppose the *outcome* of a *single roll* is the side that faces *down* (!).

Define the random variables X and Y as

X = result of the *first roll* , Y = *sum* of the two rolls.

- What is a good choice of the *sample space* $\mathcal{S}$?
- How many outcomes are there in $\mathcal{S}$?
- List the values of the *joint probability mass function* $p_{X,Y}(x,y)$.
- List the values of the *joint cumulative distribution function* $F_{X,Y}(x,y)$.

EXERCISE :

Three balls are selected at random from a bag containing

2 *red* , 3 *green* , 4 *blue* balls .

Define the *random variables*

$$R(s) \; = \; \text{the number of } \textit{red} \text{ balls drawn,}$$

and

$$G(s) \; = \; \text{the number of } \textit{green} \text{ balls drawn .}$$

List the values of

- the *joint probability mass function* $p_{R,G}(r, g)$.
- the *marginal probability mass functions* $p_R(r)$ and $p_G(g)$.
- the *joint distribution function* $F_{R,G}(r, g)$.
- the *marginal distribution functions* $F_R(r)$ and $F_G(g)$.

Independent random variables

Two discrete random variables $X(s)$ and $Y(s)$ are *independent* if

$$P(X = x\ ,\ Y = y) \ = \ P(X = x) \cdot P(Y = y)\ , \quad \text{for } \textit{all } x \text{ and } y\ ,$$

or, equivalently, if their *probability mass functions* satisfy

$$p_{X,Y}(x, y) \ = \ p_X(x) \ \cdot \ p_Y(y)\ , \quad \text{for } \textit{all } x \text{ and } y\ ,$$

or, equivalently, if the *events*

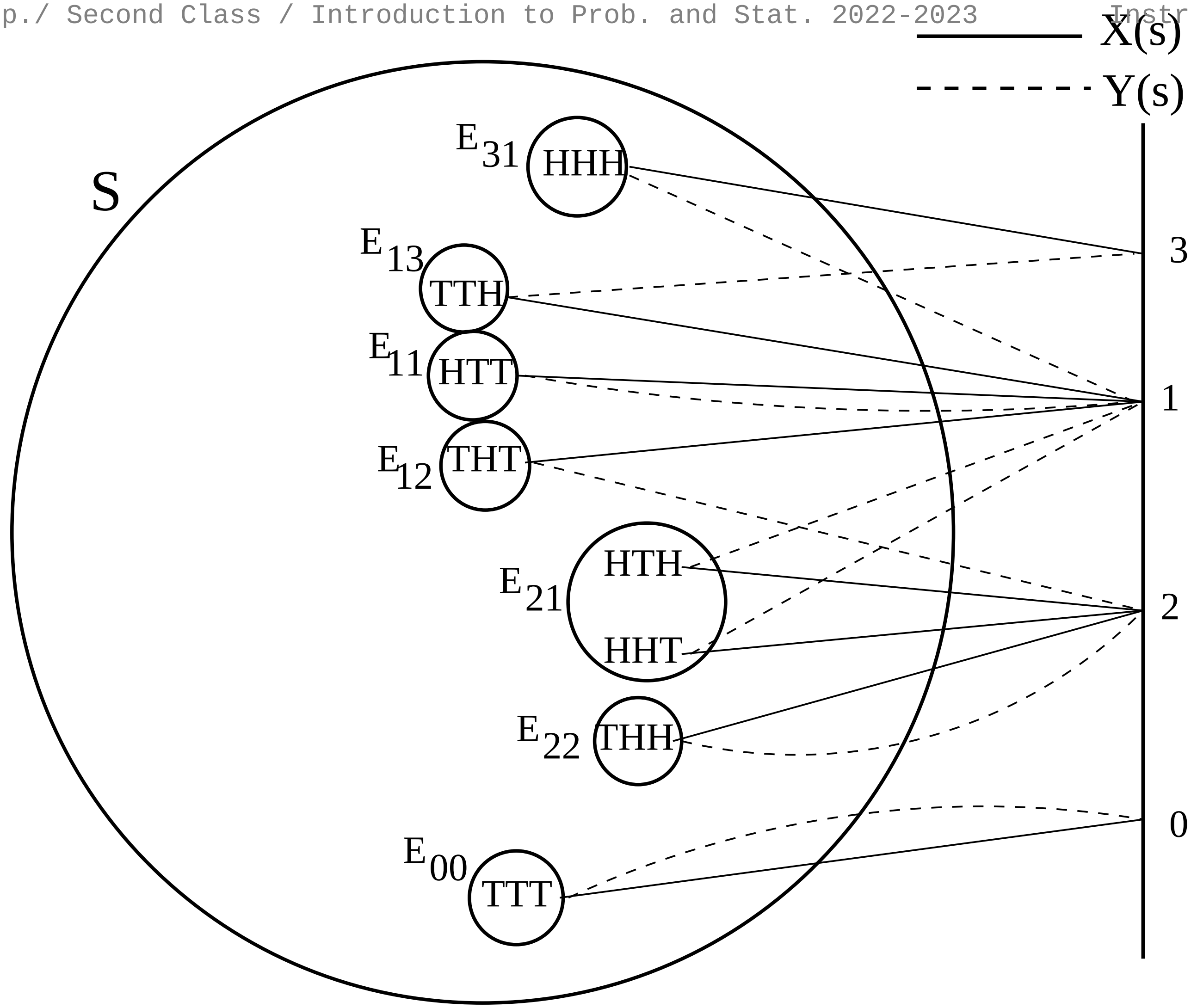
$$E_x \ \equiv \ X^{-1}(\{x\}) \quad \text{and} \quad E_y \ \equiv \ Y^{-1}(\{y\})\ ,$$

are independent *in the sample space* $\mathcal{S}$, *i.e.*,

$$P(E_x E_y) \ = \ P(E_x) \cdot P(E_y)\ , \quad \text{for } \textit{all } x \text{ and } y\ .$$

NOTE :

- In the current *discrete* case, x and y are typically *integers* .
- $X^{-1}(\{x\}) \ \equiv \ \{\ s \in \mathcal{S} \ : \ X(s) = x\ \}$.



$\textit{Three}$ tosses : $X(s) = \#$ Heads, $Y(s) =$ index 1^{st} H .

- What are the values of $p_X(2)$, $p_Y(1)$, $p_{X,Y}(2,1)$?
- Are X and Y *independent* ?

RECALL :

$X(s)$ and $Y(s)$ are *independent* if for all x and y :

$$p_{X,Y}(x,y) \;=\; p_X(x) \;\cdot\; p_Y(y)\ .$$

EXERCISE :

Roll a die two times in a row.

Let

$$X \text{ be the result of the } 1^{\text{st}} \text{ roll },$$

and

$$Y \text{ the result of the } 2^{\text{nd}} \text{ roll }.$$

Are X and Y *independent* , *i.e.*, is

$$p_{X,Y}(k,\ell) \;=\; p_X(k) \cdot p_Y(\ell), \qquad \text{for all } \; 1 \;\leq\; k,\ell \;\leq\; 6 \;\;?$$

EXERCISE :

Are these random variables X and Y *independent* ?

Joint probability mass function $p_{X,Y}(x, y)$

	$y = 0$	$y = 1$	$y = 2$	$y = 3$	$p_X(x)$
$x = 0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x = 1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x = 2$	0	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
$x = 3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

EXERCISE : Are these random variables X and Y *independent* ?

Joint probability mass function $p_{X,Y}(x,y)$

	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=1$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
$x=2$	$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
$x=3$	$\frac{1}{9}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$p_Y(y)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Joint distribution function $F_{X,Y}(x,y) \equiv P(X \leq x, Y \leq y)$

	$y=1$	$y=2$	$y=3$	$F_X(x)$
$x=1$	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{2}$
$x=2$	$\frac{5}{9}$	$\frac{25}{36}$	$\frac{5}{6}$	$\frac{5}{6}$
$x=3$	$\frac{2}{3}$	$\frac{5}{6}$	1	1
$F_Y(y)$	$\frac{2}{3}$	$\frac{5}{6}$	1	1

QUESTION : Is $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$?

PROPERTY :

The *joint distribution function* of *independent* random variables X and Y satisfies

$$F_{X,Y}(x,y) \quad = \quad F_X(x)\cdot F_Y(y)\ , \qquad \text{for all } x,y\ .$$

PROOF :

$$\begin{aligned}
F_{X,Y}(x_k,y_\ell) \quad &= \quad P(X \leq x_k\ ,\ Y \leq y_\ell) \\
&= \quad \textstyle\sum_{i\leq k}\ \sum_{j\leq \ell}\ p_{X,Y}(x_i,y_j) \\
&= \quad \textstyle\sum_{i\leq k}\ \sum_{j\leq \ell}\ p_X(x_i)\cdot p_Y(y_j) \qquad \text{(by independence)} \\
&= \quad \textstyle\sum_{i\leq k}\ \{\ p_X(x_i)\cdot \sum_{j\leq \ell}\ p_Y(y_j)\ \} \\
&= \quad \{\textstyle\sum_{i\leq k}\ p_X(x_i)\ \}\ \cdot\ \{\sum_{j\leq \ell}\ p_Y(y_j)\ \} \\
&= \quad F_X(x_k)\cdot F_Y(y_\ell)\ .
\end{aligned}$$

Conditional distributions

Let X and Y be discrete random variables with *joint probability mass function*

$$p_{X,Y}(x, y) \; .$$

For given x and y , let

$$E_x \; = \; X^{-1}(\{x\}) \qquad \text{and} \qquad E_y \; = \; Y^{-1}(\{y\}) \; ,$$

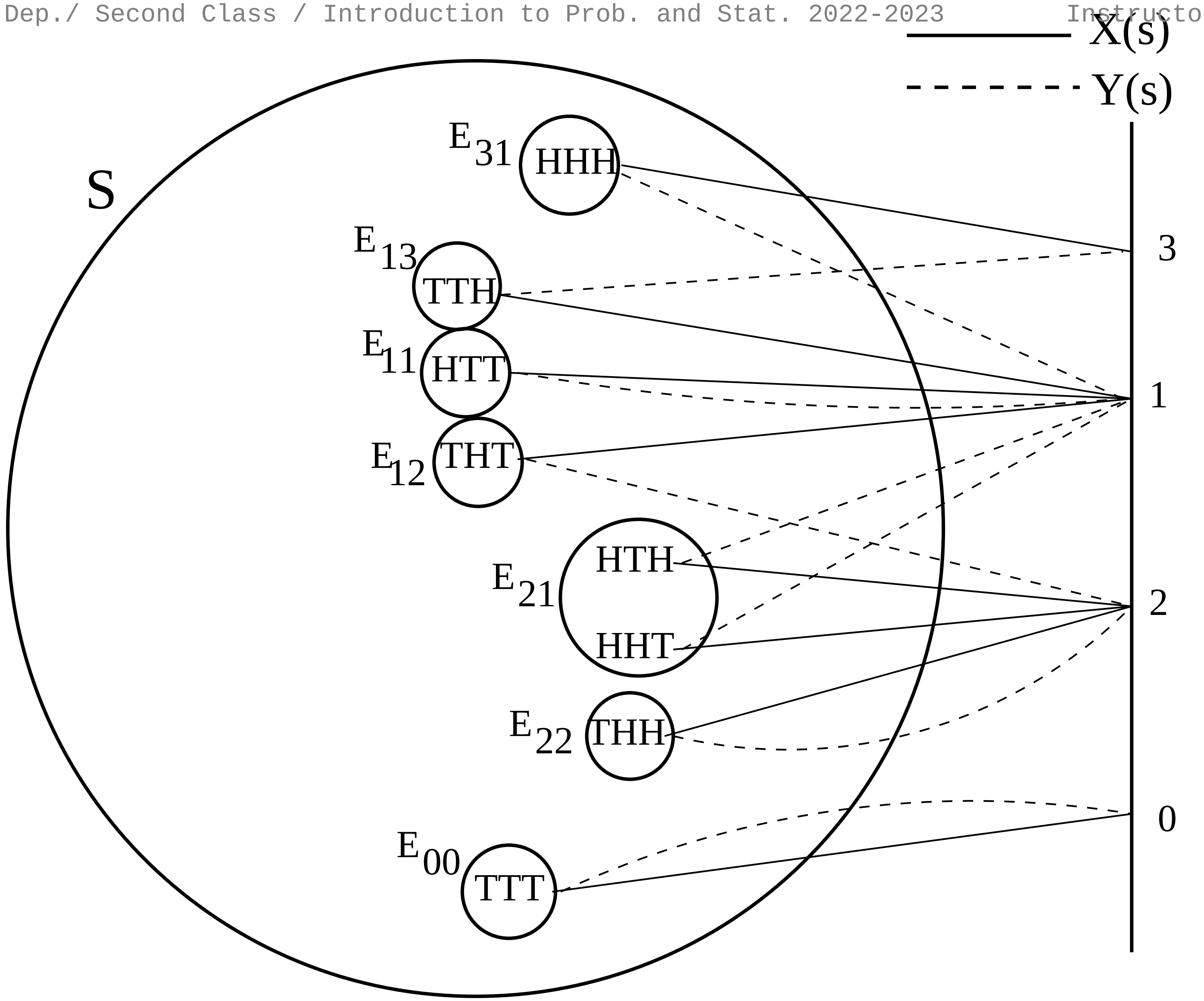
be their corresponding *events* in the sample space $\mathcal{S}$.

Then

$$P(E_x|E_y) \;\equiv\; \frac{P(E_xE_y)}{P(E_y)} \;=\; \frac{p_{X,Y}(x,y)}{p_Y(y)} \; .$$

Thus it is natural to define the *conditional probability mass function*

$$p_{X|Y}(x|y) \;\equiv\; P(X = x \mid Y = y) \;=\; \frac{p_{X,Y}(x,y)}{p_Y(y)} \; .$$



Three tosses : $X(s) = \#$ Heads, $Y(s) =$ index 1^{st} H .

- What are the values of $P(X = 2 \mid Y = 1)$ and $P(Y = 1 \mid X = 2)$?

EXAMPLE : (3 tosses : $X(s) = \#$ Heads, $Y(s) =$ index 1^{st} H.)

Joint probability mass function $p_{X,Y}(x, y)$

	$y = 0$	$y = 1$	$y = 2$	$y = 3$	$p_X(x)$
$x = 0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x = 1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x = 2$	0	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
$x = 3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

Conditional probability mass function $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$.

	$y = 0$	$y = 1$	$y = 2$	$y = 3$
$x = 0$	1	0	0	0
$x = 1$	0	$\frac{2}{8}$	$\frac{4}{8}$	1
$x = 2$	0	$\frac{4}{8}$	$\frac{4}{8}$	0
$x = 3$	0	$\frac{2}{8}$	0	0
	1	1	1	1

EXERCISE : Also construct the Table for $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$.

EXAMPLE :

Joint probability mass function $p_{X,Y}(x,y)$

	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=1$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
$x=2$	$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
$x=3$	$\frac{1}{9}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$p_Y(y)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Conditional probability mass function $p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$.

	$y=1$	$y=2$	$y=3$
$x=1$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
$x=2$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$x=3$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
	1	1	1

QUESTION : What does the last Table tell us?

EXERCISE : Also construct the Table for $P(Y=y|X=x)$.

Expectation

The *expected value* of a discrete random variable X is

$$E[X] \equiv \sum_k x_k \cdot P(X = x_k) = \sum_k x_k \cdot p_X(x_k) .$$

Thus $E[X]$ represents the *weighted average value* of X .

($E[X]$ is also called the *mean* of X .)

EXAMPLE : The *expected value* of *rolling a die* is

$$E[X] = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + \cdots + 6 \cdot \tfrac{1}{6} = \tfrac{1}{6} \cdot \textstyle\sum_{k=1}^{6} k = \tfrac{7}{2} .$$

EXERCISE : Prove the following :

- $E[aX] = a\, E[X]$,
- $E[aX + b] = a\, E[X] + b$.

EXAMPLE : Toss a coin until "Heads" occurs. Then

$$\mathcal{S} = \{H \ , \ TH \ , \ TTH \ , \ TTTH \ , \ \cdots \} \ .$$

The *random variable* X is the *number of tosses* until "Heads" occurs :

$$X(H) = 1 \quad , \quad X(TH) = 2 \quad , \quad X(TTH) = 3 \ .$$

Then

$$E[X] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + \cdots = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{2^k} = 2 \ .$$

n	$\sum_{k=1}^{n} k/2^k$
1	0.50000000
2	1.00000000
3	1.37500000
10	1.98828125
40	1.99999999

REMARK :

Perhaps using $\mathcal{S}_n = \{$all sequences of n tosses$\}$ is better $\cdots$

The expected value of a *function of a random variable* is

$$E[g(X)] \equiv \sum_k g(x_k)\, p(x_k) \ .$$

EXAMPLE :

The *pay-off* of rolling a die is $\$k^2$, where k is the side facing up.

What should the *entry fee* be for the betting to break even?

SOLUTION : Here $g(X) = X^2$, and

$$E[g(X)] = \sum_{k=1}^{6} k^2 \frac{1}{6} = \frac{1}{6} \frac{6(6+1)(2 \cdot 6 + 1)}{6} = \frac{91}{6} \cong \$15.17 \ .$$

The expected value of a function of *two* random variables is

$$E[g(X,Y)] \equiv \sum_k \sum_\ell g(x_k, y_\ell)\, p(x_k, y_\ell) .$$

EXAMPLE :

	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=1$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
$x=2$	$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
$x=3$	$\frac{1}{9}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$p_Y(y)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$E[X] = 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{3} + 3 \cdot \tfrac{1}{6} = \mathbf{\tfrac{5}{3}} ,$$

$$E[Y] = 1 \cdot \tfrac{2}{3} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} = \mathbf{\tfrac{3}{2}} ,$$

$$\begin{aligned} E[XY] &= 1 \cdot \tfrac{1}{3} + 2 \cdot \tfrac{1}{12} + 3 \cdot \tfrac{1}{12} \\ &+ 2 \cdot \tfrac{2}{9} + 4 \cdot \tfrac{1}{18} + 6 \cdot \tfrac{1}{18} \\ &+ 3 \cdot \tfrac{1}{9} + 6 \cdot \tfrac{1}{36} + 9 \cdot \tfrac{1}{36} = \mathbf{\tfrac{5}{2}} . \end{aligned}$$

(So ?)

PROPERTY :

- If X and Y are *independent* then $E[XY] = E[X]\,E[Y]$.

PROOF :

$$
\begin{aligned}
E[XY] &= \sum_k \sum_\ell x_k\, y_\ell\, p_{X,Y}(x_k, y_\ell) \\
&= \sum_k \sum_\ell x_k\, y_\ell\, p_X(x_k)\, p_Y(y_\ell) \qquad \text{(by independence)} \\
&= \sum_k \{ x_k\, p_X(x_k) \sum_\ell y_\ell\, p_Y(y_\ell)\} \\
&= \{\sum_k x_k\, p_X(x_k)\} \cdot \{\sum_\ell y_\ell\, p_Y(y_\ell)\} \\
&= E[X] \cdot E[Y] \,.
\end{aligned}
$$

EXAMPLE : See the preceding example !

PROPERTY : $E[X+Y] \;=\; E[X] \;+\; E[Y]\,.$ (**Always** !)

PROOF :

$$
\begin{aligned}
E[X+Y] &= \textstyle\sum_k \sum_\ell (x_k + y_\ell)\, p_{X,Y}(x_k, y_\ell) \\
&= \textstyle\sum_k \sum_\ell x_k\, p_{X,Y}(x_k, y_\ell) \;+\; \sum_k \sum_\ell y_\ell\, p_{X,Y}(x_k, y_\ell) \\
&= \textstyle\sum_k \sum_\ell x_k\, p_{X,Y}(x_k, y_\ell) \;+\; \sum_\ell \sum_k y_\ell\, p_{X,Y}(x_k, y_\ell) \\
&= \textstyle\sum_k \{x_k \sum_\ell p_{X,Y}(x_k, y_\ell)\} \;+\; \sum_\ell \{ y_\ell \sum_k p_{X,Y}(x_k, y_\ell)\} \\
&= \textstyle\sum_k \{x_k\, p_X(x_k)\} \;+\; \sum_\ell \{y_\ell\, p_Y(y_\ell)\} \\
&= E[X] \;+\; E[Y]\,.
\end{aligned}
$$

NOTE : X and Y need not be independent !

EXERCISE :

Probability mass function $p_{X,Y}(x,y)$

	$y=6$	$y=8$	$y=10$	$p_X(x)$
$x=1$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
$x=2$	0	$\frac{1}{5}$	0	$\frac{1}{5}$
$x=3$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
$p_Y(y)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	1

Show that

- $E[X] = 2 \quad , \quad E[Y] = 8 \quad , \quad E[XY] = 16$
- X and Y are *not* independent

Thus if

$$E[XY] = E[X]\,E[Y] \, ,$$

then it does not necessarily follow that X and Y are independent !

Variance and Standard Deviation

Let X have *mean*

$$\mu = E[X] .$$

Then the *variance* of X is

$$Var(X) \equiv E[\,(X-\mu)^2] \equiv \sum_k (x_k - \mu)^2\, p(x_k) ,$$

which is the average weighted *square distance* from the mean.

We have

$$\begin{aligned} Var(X) &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 . \end{aligned}$$

The *standard deviation* of X is

$$\sigma(X) \equiv \sqrt{Var(X)} = \sqrt{E[\,(X-\mu)^2]} = \sqrt{E[X^2] - \mu^2} \,.$$

which is the average weighted *distance* from the mean.

EXAMPLE : The *variance* of *rolling a die* is

$$Var(X) = \sum_{k=1}^{6} [k^2 \cdot \frac{1}{6}] - \mu^2$$

$$= \frac{1}{6} \frac{6(6+1)(2\cdot 6+1)}{6} - (\frac{7}{2})^2 = \frac{35}{12} \,.$$

The *standard deviation* is

$$\sigma = \sqrt{\frac{35}{12}} \cong 1.70 \,.$$

Covariance

Let X and Y be random variables with *mean*

$$E[X] = \mu_X \quad , \quad E[Y] = \mu_Y \ .$$

Then the *covariance* of X and Y is defined as

$$Cov(X,Y) \equiv E[\,(X-\mu_X)\,(Y-\mu_Y)\,] = \sum_{k,\ell} (x_k-\mu_X)\,(y_\ell-\mu_Y)\,p(x_k,y_\ell)\ .$$

We have

$$\begin{aligned} Cov(X,Y) &= E[\,(X-\mu_X)\,(Y-\mu_Y)\,] \\ &= E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y] \\ &= E[XY] - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y \\ &= E[XY] - E[X]\,E[Y]\ . \end{aligned}$$

We defined

$$\begin{aligned} Cov(X,Y) &\equiv E[\,(X-\mu_X)\,(Y-\mu_Y)\,] \\ &= \sum_{k,\ell} (x_k-\mu_X)\,(y_\ell-\mu_Y)\,p(x_k,y_\ell) \\ &= E[XY] \;-\; E[X]\,E[Y]\,. \end{aligned}$$

NOTE :

$Cov(X,Y)$ measures "*concordance*" or "*coherence*" of X and Y :

- If $X > \mu_X$ when $Y > \mu_Y$ and $X < \mu_X$ when $Y < \mu_Y$ then

$$Cov(X,Y) \;>\; 0\,.$$

- If $X > \mu_X$ when $Y < \mu_Y$ and $X < \mu_X$ when $Y > \mu_Y$ then

$$Cov(X,Y) \;<\; 0\,.$$

EXERCISE : Prove the following :

- $Var(aX+b) \;=\; a^2\, Var(X)$,
- $Cov(X,Y) \;=\; Cov(Y,X)$,
- $Cov(cX,Y) \;=\; c\, Cov(X,Y)$,
- $Cov(X,cY) \;=\; c\, Cov(X,Y)$,
- $Cov(X+Y,Z) \;=\; Cov(X,Z) \;+\; Cov(Y,Z)$,
- $Var(X+Y) \;=\; Var(X) \;+\; Var(Y) \;+\; 2\, Cov(X,Y)$.

PROPERTY :

If X and Y are *independent* then $Cov(X,Y) = 0$.

PROOF :

We have already shown (with $\mu_X \equiv E[X]$ and $\mu_Y \equiv E[Y]$) that

$$Cov(X,Y) \equiv E[\,(X-\mu_X)\,(Y-\mu_Y)\,] = E[XY] - E[X]\,E[Y] ,$$

and that if X and Y are *independent* then

$$E[XY] = E[X]\,E[Y] .$$

from which the result follows.

EXERCISE : (already used earlier $\cdots$)

Probability mass function $p_{X,Y}(x,y)$

	$y=6$	$y=8$	$y=10$	$p_X(x)$
$x=1$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
$x=2$	0	$\frac{1}{5}$	0	$\frac{1}{5}$
$x=3$	$\frac{1}{5}$	0	$\frac{1}{5}$	$\frac{2}{5}$
$p_Y(y)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	1

Show that

- $E[X] \;=\; 2 \quad , \quad E[Y] \;=\; 8 \quad , \quad E[XY] \;=\; 16$
- $Cov(X,Y) \;=\; E[XY] \;-\; E[X]\,E[Y] \;=\; 0$
- X and Y are *not* independent

Thus if

$$Cov(X,Y) \;=\; 0\ ,$$

then it does not necessarily follow that X and Y are independent !

PROPERTY :

If X and Y are *independent* then

$$Var(X+Y) \ = \ Var(X) \ + \ Var(Y) \ .$$

PROOF :

We have already shown (in an exercise !) that

$$Var(X+Y) \ = \ Var(X) \ + \ Var(Y) \ + \ 2\, Cov(X,Y) \ ,$$

and that if X and Y are *independent* then

$$Cov(X,Y) \ = \ 0 \ ,$$

from which the result follows.

EXERCISE :

Compute

$$E[X] \;,\; E[Y] \;,\; E[X^2] \;,\; E[Y^2]$$

$$E[XY] \;,\; Var(X) \;,\; Var(Y)$$

$$Cov(X,Y)$$

for

Joint probability mass function $p_{X,Y}(x,y)$

	$y=0$	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=0$	$\frac{1}{8}$	0	0	0	$\frac{1}{8}$
$x=1$	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{8}$
$x=2$	0	$\frac{2}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$
$x=3$	0	$\frac{1}{8}$	0	0	$\frac{1}{8}$
$p_Y(y)$	$\frac{1}{8}$	$\frac{4}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	1

EXERCISE :

Compute

$$E[X] \ , \ E[Y] \ , \ E[X^2] \ , \ E[Y^2]$$

$$E[XY] \ , \ Var(X) \ , \ Var(Y)$$

$$Cov(X,Y)$$

for

Joint probability mass function $p_{X,Y}(x,y)$

	$y=1$	$y=2$	$y=3$	$p_X(x)$
$x=1$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{2}$
$x=2$	$\frac{2}{9}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
$x=3$	$\frac{1}{9}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
$p_Y(y)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	1