3 - discrete random variables and their distributions

$\boxed{\text{distribution of a random variable}}$

random variable = function that depends on change

$$X = f(\omega)$$

domain: Ω sample space: $\mathbb{R}$, $\mathbb{N}$, $(0,\infty)$, $(0,1)$ ...

$f(\omega)$: is known when the experiment is completed

distribution of X: collection of all the probabilities

$$\boxed{P(x)} = P\{X = x\} \longrightarrow \text{probability mass function (pmf)}$$

cumulative distribution function (cdf): $\boxed{F(x)} = P\{X \le x\} = \sum_{y \le x} P(y)$

- support of the distribution = the set of possible values of X

★ for every outcome w, the variable X takes one and only one value x. this makes events $\{X = x\}$ disjoint and exhaustive, so:

$$\sum_x P(x) = \sum_x P(X = x) = 1$$

$\rightarrow$ cdf $F(x)$ is a non-decreasing function of x, always between 0 and 1,

$$\lim_{x \downarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \uparrow +\infty} F(x) = 1$$

example

tossing 3 coins, the number of heads:

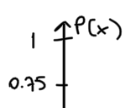
$$P\{X=0\} = P\{TTT\} = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{8}$$
$$P\{X=1\} = P\{HTT\} + P\{THT\} + P\{TTH\} = \tfrac{3}{8}$$
$$P\{X=2\} = P\{HHT\} + P\{HTH\} + P\{THH\} = \tfrac{3}{8}$$
$$P\{X=3\} = P\{HHH\} = \tfrac{1}{8}$$

$\hookrightarrow$

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

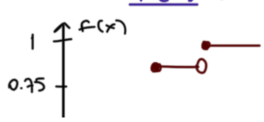
before we know the outcome w, we cannot know what X equals to. (but we can list possible values of X and their probabilities)

$f_X(x)$ $\underline{P(x)}$ $\rightarrow$ probability mass function



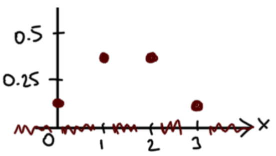
$x = 0 \quad P\{\frac{1}{8}\}$

$\underline{F(x)}$ $\rightarrow$ cumulative mass function

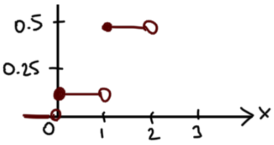


$x < 0 \rightarrow 0$

$x = 0 \rightarrow \frac{1}{8}$



$x=1 \quad P\{\frac{3}{8}\}$
$x=2 \quad P\{\frac{3}{8}\}$
$x=3 \quad P\{\frac{1}{8}\}$



$x=1 \rightarrow \frac{1}{8}+\frac{3}{8}=\frac{4}{8}$
$x=2 \rightarrow \frac{1}{8}+\frac{3}{8}+\frac{3}{8}=\frac{7}{8}$
$x=3 \rightarrow \frac{1}{8}+\frac{3}{8}+\frac{3}{8}+\frac{1}{8}=1$
$x>3 \rightarrow 1$

having 1.5 head — having number of heads less than 1.5

★ $P(1.5)=0$ but $F(1.5)=0.5$ $\quad F(1000)=1$

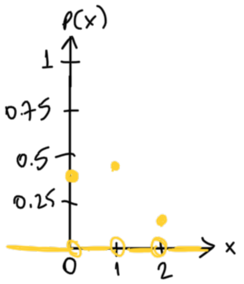
$F(x)$ is always on $[0,1]$ and it is non decreasing function

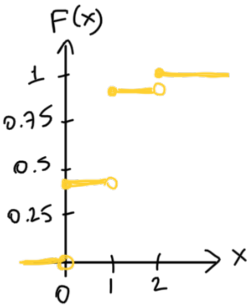
$$P\{a<X\leq b\}=F(b)-F(a)=P(b) \text{ if it is not continous}$$

exercise: there are independent two machines of a system. first will be broken with the probability 0.4, second with 0.3

mass funcion of break of the system:

$X: 0 \rightarrow$ no broken $\quad (0.6).(0.7)=0.42$
$X: 1 \rightarrow$ one broken $\quad (0.4)(0.7)+(0.6).(0.3)=0.46$
$X: 2 \rightarrow$ two broken $\quad (0.4)(0.3)=0.12$





cumulative distribution function:

$x<0 \rightarrow 0$
$F(0)= x=0 \rightarrow 0.42 = P(0)$
$0\leq x<1 \rightarrow 0.42$
$F(1)= x=1 \rightarrow 0.42+0.46=0.88 = F(0)+P(1)$
$1\leq x<2 \rightarrow 0.88$
$F(2)= x=2 \rightarrow 0.88+0.12=1 = F(1)+P(2)$
$x\geq 2 \rightarrow 1$

★ continous random variables ⇒ interval of values, uncountable, examples like time, weight, distance, temperature. we cannot use probability mass function, because the probability of any specific outcome among infinite sample space would be zero: $\frac{1}{\infty}$, so instead we use probability density function

↳ some are neither discrete nor continuous but mixed

distribution of a random vector

- when we deal with two factors to determine probability

☆ if X and Y are random variables (probility functions), the pair (X,Y) is a random vector

☆ its distribution joint distribution of X and Y

☆ marginal distribution = individual distribution of X and Y

- if $P(x,y)=P(x).P(y)$ then they are independent factor.
  if $P(x,y)\neq P(x).P(y)$, they are dependent.

$P_{(X,Y)}(x,y)$		y				$P_X(x)$
		0	1	2	3	
x	0	0.20	0.20	0.05	0.05	0.50
	1	0.20	0.10	0.10	0.10	0.50
$P_Y(y)$		0.40	0.30	0.15	0.15	1.00

explanation: x and y are not independent. when $P_{X,Y}(0,1) = 0.2$ $\neq$
but $P_X(0) = 0.50 \cdot P_Y(1) = 0.30 = 0.15$

☆ notice that the total of both $P_X(x)$ and $P_Y(y)$ is 1, on their own

$\boxed{\text{expectation and variance}}$

↳ the distribution of a random variable/vector can be summarized in a few vital characteristics: expectation, varience, standard deviation, covariance, and correlation.

$\boxed{\text{expectation}}$: weighted average (expected value or the mean), center of gravity

for discrete variable: $E(x) = \boxed{\mu} = \sum_x x f(x) = x_1 f(x_1) + x_2 f(x_2) + \dots = \boxed{\sum_x x P(x)}$

* if it is continuous random variable $\Rightarrow E(x) = \mu = \int_x x f(x).dx$

example = expected value of a single die

$$E(x) = 1.\underbrace{P(x=1)}_{\frac{1}{6}} + 2.\underbrace{P(x=2)}_{\frac{1}{6}} + 3.\underbrace{P(x=3)}_{\frac{1}{6}} + 4.\underbrace{P(x=4)}_{\frac{1}{6}} + 5.\underbrace{P(x=5)}_{\frac{1}{6}} + 6.\underbrace{(x=6)}_{\frac{1}{6}} = \frac{21}{6} = 3.5$$

example = $X = \begin{cases} 0 & \text{with probability } 0.75 \\ 1 & \text{with probability } 0.25 \end{cases}$ $\quad E(x) = 0.(0.75) + 1(0.25) = 0.25$

expected value of a function: $\sum_x g(x) f(x)$
↳ example: $g(x) = x^2 \rightarrow \sum_x x^2 f(x)$

linearity of expected values: $E(c_1 X_1 + c_2 X_2) = c_1 E(X_1) + c_2 E(X_2)$
or $E(aX + bY + c) = a E(X) + b E(Y) + c$

expected value of a product: if X and Y independent $E(XY) = E(X) \cdot E(Y)$

but in general $E(XY) = \int_y \int_x x \cdot y \cdot \underbrace{j(x,y)}_{\text{joint distribution}} \, dx \cdot dy$

$\boxed{\text{variance}}$: measurement of how much can a variable vary around its expectation
↳ expected squared deviation from the mean

for discrete variable: $Var(X) = \boxed{\sigma^2} = E\left[(X-\mu)^2\right] = E(X^2) - \mu^2$
$= E\,(X - EX)^2 = \boxed{\sum_x (x-\mu)^2 P(x)}$

☆ if the distance is not squared, the result is always $\mu - \mu = 0$
☆ variance is always non-negative
☆ variance $= 0$ only if all values of $x = \mu$

example: variance of a single die. ($\mu = 3.5$ from previous examples)

$= \underbrace{(1-3.5)^2}_{\frac{25}{4}} \cdot \underbrace{P(1)}_{\frac{1}{6}} + \underbrace{(2-3.5)^2}_{\frac{9}{4}} \cdot \underbrace{P(2)}_{\frac{1}{6}} + \underbrace{(3-3.5)^2}_{\frac{1}{4}} \underbrace{P(3)}_{\frac{1}{6}} + \underbrace{(4-3.5)^2}_{\frac{1}{4}} \underbrace{P(4)}_{\frac{1}{6}} + \underbrace{(5-3.5)^2}_{\frac{9}{4}} \cdot \underbrace{P(5)}_{\frac{1}{6}} + \underbrace{(6-3.5)}_{\frac{25}{4}} \cdot \underbrace{P}_{\frac{1}{6}}$

$= \frac{35}{12} \simeq 2.92$

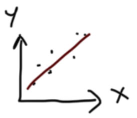
☆ if X and Y are independent $Var(X+Y) = Var(X) + Var(Y)$

$\boxed{\text{standard deviation}}$: square root of variance (because we want same units with measurements, since we take the square of them in variance)

$\boxed{\sigma} = Std(X) = \boxed{\sqrt{Var(X)}}$

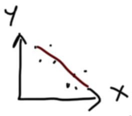
$\boxed{\text{covariance}}$: a measure of strength of a relationship between two random variables

Covariance $\boxed{\sigma_{XY}} = Cov(X,Y) = E\left[(X - \underset{\downarrow\atop E(X)}{\mu_X})(Y - \underset{\downarrow\atop E(Y)}{\mu_Y})\right] = E(XY) - E(X) \cdot E(Y)$

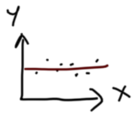


$Cov(X,Y) > 0$
positive

ex: height and weight of a person



$Cov(X,Y) < 0$
negative

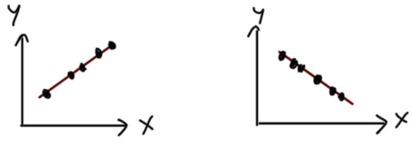


$Cov(X,Y) = 0$
zero covariance

☆ if X and Y are independent $Cov(X,Y) = 0$ (reverse is not always true)

correlation (coefficient) : the covariance standardized to the range of [-1, 1] (normalized)

$$Corr(X,Y) = \boxed{\rho_{XY}} = \frac{Cov(X,Y)}{\sqrt{Var(X) . Var(Y)}} = \frac{Cov(X,Y)}{Std(X) . Std(Y)}$$



perfect correlation $\rho = \pm 1$ ☆ $\rho = 0$ → weak or no correlation

☆ adding a constant does not affect the variables' variance or covariance

☆ multiplying by a constant does not change the correlation coefficient