PROBABILITY DISTRIBUTIONS

DISCRETE RANDOM VARIABLES

We've already seen examples where the possible outcomes of an experiment form a DISCRETE SET $\{x_1, x_2, \ldots, x_M\}$:

E.g. flipping a coin $\{H, T\}$
\# of heads in n coin tosses $\{0, 1, \ldots, n\}$
\# of correct lottery numbers $\{0, 1, \ldots 6\}$
roll of a die $\{1, 2, \ldots, 6\}$
\# of dice rolls needed to get a SIX $\{1, 2, 3, \ldots\}$
Pip count on 2 dice $\{2, 3, \ldots 12\}$
\# α-particles emitted by source in 1 sec $\{0, 1, \ldots, N_{at.}\}$

[In many cases the outcomes are integers, though they don't <u>have</u> to be!]. The probabilities P_i of the various outcomes x_i form a PROBABILITY DISTRIBUTION over the SAMPLESPACE.

EXAMPLE 1: # of heads in 3 coin tosses

$$P_0 = P(TTT) = \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{8}$$

$$P_1 = P_2 = \tfrac{3}{8} \quad \text{[SEE L2]}$$

$$P_3 = \tfrac{1}{8}.$$

EXPECTATION VALUE

... is just another word for the average that would be obtained in an experiment of many trials $[N \gg 1]$. If x_i occurs N_i times,

$$\boxed{\begin{aligned} E(X) \equiv \langle X \rangle &= \frac{1}{N} \sum_i N_i x_i \\ &= \sum_i P_i x_i \end{aligned}}$$

[WARNING: also written as $\bar{X}$]

EXAMPLE 2: If X [the random variable] is the # of heads in 3 coin tosses,

$$\langle X \rangle = \sum_{i=0}^{3} i P_i$$

$$= 0 \times \tfrac{1}{8} + 1 \times \tfrac{3}{8} + 2 \times \tfrac{3}{8} + 3 \times \tfrac{1}{8} = \tfrac{3}{2}$$

VARIANCE & STANDARD DEVIATION

... measure the SPREAD of a distribution, i.e., how far X typically DEVIATES from $\langle X \rangle$. It would be awkward to use $\langle |X - \langle X \rangle| \rangle$ as this measure. Instead use the VARIANCE

$$\boxed{\mathrm{var}(X) \equiv \sigma^2 = \langle (X - \langle X \rangle)^2 \rangle}$$

$\sigma \equiv$ "STANDARD DEVIATION"
$\equiv$ "ROOT-MEAN-SQUARE [R.M.S.] DEVIATION"

[VARIANCE & S.D., CONT$^{\underline{D}}$]

To compute the variance, it's often convenient to use

$$
\begin{aligned}
\mathrm{var}(x) &= \langle (x - \langle x \rangle)^2 \rangle \\
&= \langle x^2 - 2x\langle x \rangle + \langle x \rangle^2 \rangle \\
&= \sum_i P_i \left(x_i^2 - 2x_i \langle x \rangle + \langle x \rangle^2 \right) \\
&= \sum_i P_i x_i^2 - 2\langle x \rangle \sum_i P_i x_i + \langle x \rangle^2 \overbrace{\sum_i P_i}^{1} \\
&= \langle x^2 \rangle - 2\langle x \rangle \langle x \rangle + \langle x \rangle^2
\end{aligned}
$$

or

$$\boxed{\mathrm{var}(x) = \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2}$$

EXAMPLE 3 : For $x = [\#\text{ heads in 3 coin tosses}]$,

$$
\begin{aligned}
\langle x^2 \rangle &= \sum_{i=0}^{3} i^2 P_i \\
&= 0^2 \times \tfrac{1}{8} + 1^2 \times \tfrac{3}{8} + 2^2 \times \tfrac{3}{8} + 3^2 \times \tfrac{1}{8} = 3
\end{aligned}
$$

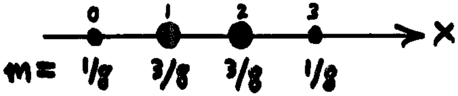
$\langle x \rangle = \frac{3}{2}$ [from EXAMPLE 2]

$$
\begin{aligned}
\Rightarrow \sigma^2 &= \langle x^2 \rangle - \langle x \rangle^2 \\
&= 3 - \left(\tfrac{3}{2}\right)^2 = \tfrac{3}{4}
\end{aligned}
$$

$$\Rightarrow \sigma = \frac{\sqrt{3}}{2} \simeq 0.87$$

MECHANICAL ANALOGY [DIGRESSION]

You could think of the P_i as MASSES at positions x_i on the x-axis. For EXAMPLES 1-3,



Expectation value

$$\langle x \rangle = \sum P_i x_i = \frac{\sum P_i x_i}{\sum P_i}$$

$\equiv$ CENTRE OF MASS

Variance

$$\langle (x - \langle x \rangle)^2 \rangle = \sum P_i (x_i - \langle x \rangle)^2$$

$\equiv$ MOMENT OF INERTIA ABOUT CENTRE OF MASS

$$
\begin{aligned}
\langle x^2 \rangle = \sum P_i x_i^2 &\equiv \text{MOMENT OF INERTIA ABOUT } x = 0 \\
&= \sigma^2 + \langle x \rangle^2 \\
&= \sigma^2 + \left(\sum P_i\right) \langle x \rangle^2
\end{aligned}
$$

$\equiv$ PARALLEL AXIS THEOREM

EXAMPLE 4: The GEOMETRIC DISTRIBUTION.

of dice rolls needed to get a six:

$P_1 = \frac{1}{6}$ [six on 1st attempt]

$P_2 = \frac{5}{6} \times \frac{1}{6}$ [six on 2nd attempt]

$P_3 = \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6}$, etc.

More generally, if the chance of success on one attempt is p and chance of failure is q [$= 1-p$], then

$$\left.\begin{aligned} P_1 &= p \\ P_2 &= qp \\ P_3 &= q^2 p \\ &\vdots \\ P_n &= q^{n-1} p \end{aligned}\right\}$$

GEOMETRIC DISTN [the P_n form a geometric sequence with common ratio q]

Always sensible to <u>check</u> that the probabilities add up to 1:

$$\begin{aligned} \sum_{n=1}^{\infty} P_n &= p + pq + pq^2 + \ldots \\ &= p + q(p + pq + pq^2 + \ldots) \\ &= p + q \sum_{n=1}^{\infty} P_n \end{aligned}$$

$$\Rightarrow \underline{\sum_{n=1}^{\infty} P_n = \frac{p}{1-q} = 1.}$$

[EXAMPLE 4, CONTD]

Expected # of attempts:

$$\langle n \rangle = \sum_{n=1}^{\infty} n P_n = \sum_{n=1}^{\infty} n \cdot pq^{n-1} = p \sum_{n=1}^{\infty} nq^{n-1}$$

Don't need to remember formula for sum of arithmetic-geometric series! Just notice that $nq^{n-1} = \frac{d}{dq}(q^n)$

$$\begin{aligned} \Rightarrow \sum_{1}^{\infty} nq^{n-1} &= \frac{d}{dq}\left(\sum_{1}^{\infty} q^n\right) \\ &= \frac{d}{dq}\left(\frac{q}{1-q}\right) \quad \text{[by same method as on last page]} \\ &= \frac{1}{1-q} + \frac{q}{(1-q)^2} = \frac{1}{(1-q)^2} \quad \text{(R)} \end{aligned}$$

Hence $\underline{\langle n \rangle = p/(1-q)^2 = 1/p}$.

Differentiating (R) again helps with calculating the variance...

[See next sheet!]

[VARIANCE OF GEOMETRIC DISTRIBUTION]

Use $\mathrm{var}(n) = \langle n^2 \rangle - \langle n \rangle^2$

$$= \langle n(n-1) + n \rangle - \langle n \rangle^2$$

$$= \langle n(n-1) \rangle + \langle n \rangle - \langle n \rangle^2 \qquad (S)$$

Now, $\langle n(n-1) \rangle = \sum_{n=1}^{\infty} n(n-1) P_n$

$$= \sum_{n=1}^{\infty} n(n-1)\, p q^{n-1}$$

$$= (pq) \left\{ \sum_{n=1}^{\infty} n(n-1)\, q^{n-2} \right\}$$

But the quantity $\{\cdots\}$ is $\frac{d}{dq}$ of left-hand side of (R) [previous sheet]:

$$\Rightarrow \langle n(n-1) \rangle = (pq) \frac{d}{dq} \left(\sum_{n=1}^{\infty} n q^{n-1} \right)$$

$$= (pq) \frac{d}{dq} \left(\frac{1}{(1-q)^2} \right) \quad \text{[using (R)]}$$

$$= (pq) \times \frac{2}{(1-q)^3}$$

$$= \frac{2pq}{p^3} = \frac{2q}{p^2}.$$

Now use (S):

$$\mathrm{var}(n) = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2} \quad \text{[using } p+q=1\text{]}$$

$\Rightarrow$ standard deviation $\sigma = \sqrt{\mathrm{var}(n)} = \dfrac{\sqrt{q}}{p}$.