Example 3.1

y	0	1	2	3	4	5	6
$p_Y(y)$	0.10	0.15	0.20	0.25	0.20	0.06	0.04

① $\sum p_Y(y) = 1$

② $0 \leq p_Y(y) \leq 1$

Q1 $p_Y(y=2) = P(Y=2) = 0.20$ Exactly 2 lines

Q2. At most two lines?

~~$P_Y(y$~~ $p_Y(y \leq 2) = P(Y \leq 2)$

$= P(Y=0) + P(Y=1) + P(Y=2)$

$= 0.10 + 0.15 + 0.20 = 0.45$.

$P(Y \leq y)$ is called CDF

cumulative distribution function.

$F_Y(y) = P(Y \leq y)$

↓

The visualization of pmf and CDF

[Go back to notes of L3].

$\boxed{\text{Expected value}}$

Let Y be a discrete random variable with pmf $p_Y(y)$, the expected value of Y is given by $\mu = E(Y) = \sum_{\text{all } y} y \cdot p_Y(y)$

μ is a weighted average of the possible values of Y. Each value y is weighted by its probability.

In statistical applications, $\mu = E(Y)$ is called the <u>population mean</u>.

Example 3.1 revisited.

The expected value of Y is

$$\mu = E(Y) = \sum y \cdot p_Y(y)$$

$$= 0 \times 0.10 + 1 \times 0.15 + 2 \times 0.2 + 3 \times 0.25$$

$$+ 4 \times 0.20 + 5 \times 0.06 + 6 \times 0.04 = 2.64.$$

Interpretation

on average, we would expect 2.64 calls at a given time.

① why Not sum up / n

② Don't divide by n.

Result.

Expected value of Y

function of

Let Y be a discrete random variable with pmf $p_Y(y)$. Suppose that g is a real valued function. Then $g(Y)$ is a random variable

and $$E[g(Y)] = \sum_{\text{all } y} g(\cancel{y}) \cancel{p} \, p_Y(y)$$

$$= \sum_{\text{all } y} g(y) p_Y(y)$$

[~~g(Y)~~ i function of R.V. is also R.V]

[Can find $g(Y)$]

properties.

Let Y be a discrete random variable with pmf $p_Y(y)$. Suppose that $g_1 \dots g_k$ are real valued functions, and let c be any real constant. Expectations satisfy the following properties.

(a) $E(c) = c$

(b) $E[cg(Y)] = c\,E[g(Y)]$

(c) $E\left(\sum_{j=1}^{k} g_j(Y)\right) = \sum_{j=1}^{k} E[g_j(Y)]$

[These rules also apply to continuous distributions as well].

Example 3.2.

In an one-hour period, # gallons of a certain toxic chemical that is produced at a local plant has the following p.m.f.

y	0	1	2	3
$p(y)$	0.2	0.3	0.3	0.2

(a) Compute the expected number of gallons produced during a one-hour period.

Solution:

$$\mu = E(Y) = \sum_{\text{all } y} y\, p_Y(y) = 0 \times 0.2 + 1 \times 0.3 + 2 \times 0.3 + 3 \times 0.2 = 1.5$$

Hence, we would expect 1.5 gallons of toxic chemical to be produced per hour on average.

(b) The cost (in 100\$) to produce Y gallons of this chemical per hour is

$$g(Y) = 3 + 12Y + 2Y^2.$$

What is the expected cost in a one-hour period?

The expected value of $g(Y)$, where

$$g(Y) = 3 + 12Y + 2Y^2.$$

using property 3.

$$E[g(Y)] = E(3 + 12Y + 2Y^2)$$

$$= 3 + 12E(Y) + 2E(Y^2)$$

$$= 3 + 12 \times 1.5 + 2E(Y^2) \quad (1)$$

$$E(Y^2) = \sum_{\text{all } y} g(y) p_Y(y)$$

$$= \sum_{\text{all } y} y^2 p_Y(y)$$

$$= 0^2 \times 0.2 + 1^2 \times 0.3 + 2^2 \times 0.3$$

$$+ 3^2 \times 0.2 = 3.3$$

plugging back to (1) and we get

$$E[g(Y)] = 3 + 12 \times 1.5 + 2 \times 3.3 = 27.6.$$

The expected cost is 2,760.00 on average.

$\boxed{\text{Variance}}$ of Y

Let Y be a discrete random variable with pmf $p_Y(y)$ and expected value $E(Y) = \mu$.

The variance of Y is given by

$$\sigma^2 = \operatorname{var}(Y) = E[(Y-\mu)^2]$$

[expectation of a function of Y]

$$= \sum_{\text{all } y} (y-\mu)^2 \cdot p_Y(y)$$

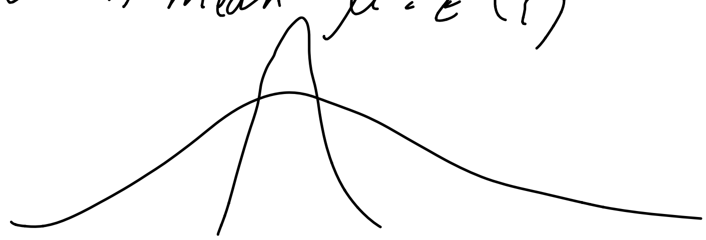
The standard deviation of Y is the positive square root of the variance.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\operatorname{var}(Y)}$$

Facts: The variance σ^2 satisfies the following.

(a) $\sigma^2 \geq 0$. $\sigma^2 = 0$ if and only if the random variable Y has a degenerate distribution, i.e. all the probability mass is located at one support point

(b). The large σ^2 is, the more spread in the possible values of Y about the population mean $\mu = E(Y)$



(c) σ is measured in original units and σ^2 is measured in $(\text{units})^2$.