Math 3501 - Probability and Statistics I

2.3 - Special Mathematical Expectations

Mean

In what follows, suppose that X is a discrete random variable with space S and pmf $f(x)$.

Definition

The mean of the random variable X (or of its distribution) is

$$\boxed{\mu = E(X) = \sum_{x \in S} x f(x) \ .}$$

mean is a measure of location for the distribution of X

Notes:

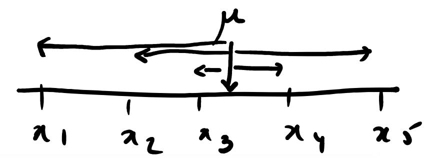
1) the mean μ is also referred to as the *first moment of X about the origin*.

2) the first moment (1st X) about the mean is always zero:

$$E[(X - \mu)] = E(X) - E(\mu) = \mu - \mu = 0 \ .$$

linearity

Variance and standard deviation



Definition

The *variance of the random variable* X (or of its distribution), denoted $\mathrm{Var}(X)$, is the second moment of X about the mean:

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_{x \in S} (x-\mu)^2 f(x)$$

The positive square root of the variance is called the *standard deviation of* X and is denoted by the Greek letter σ (sigma)

$$\sigma = \sqrt{\mathrm{Var}(X)}\ .$$

Note: Variance may also be denoted by σ^2, since

$$\sigma^2 = \left(\sqrt{\mathrm{Var}(X)}\right)^2 = \mathrm{Var}(X)\ .$$

Property

Variance may be computed in another way:

$$\mathrm{Var}(X) = E\left(X^2\right) - \left(E(X)\right)^2.$$

Either use the definition to compute $\mathrm{Var}(X) = E\left[(X-\mu)^2\right]$ or observe that:

$$\begin{aligned}
\mathrm{Var}(X) = E\left[(X-\mu)^2\right] &= E\left[X^2 - 2\mu X + \mu^2\right] \\
\text{linearity } &= E\left[X^2\right] - 2\mu \underbrace{E[X]}_{\mu} + \mu^2 \\
&= E\left[X^2\right] - 2\mu^2 + \mu^2 = E\left[X^2\right] - \mu^2
\end{aligned}$$

$$\boxed{\mathrm{Var}(X) = E\left[X^2\right] - \left(E[X]\right)^2}$$

often easier than using the definition!

Interpretation

The mean $\mu = E(X)$ provides a measure of location:

- it gives the middle or center of the distribution of X relative to the pmf $f(x)$

Variance and the standard deviation are measures of dispersion or spread:

- they indicate how much the points in S spread out around the mean μ relative to the pmf $f(x)$

$Var(X) < Var(Y)$

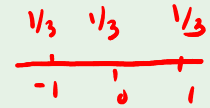
Note:

$$\begin{cases} \text{units of } \mu = \text{units } \sigma = \text{units of } X \\ \text{units of } \sigma^2 = (\text{units of } X)^2 \end{cases}$$

Example

Let the random variable X have the pmf

$$f(x) = \frac{1}{3}, \quad x \in \overbrace{\{-1, 0, 1\}}^{S}.$$



Find its the variance σ_X^2 and standard deviation σ_X.

capital x

X

small cap x: values that X takes

$$E[X] = \sum_{x \in S} x \cdot f(x) = \sum_{x=-1}^{1} x \cdot \frac{1}{3} = 0$$

$$\sigma_X^2 = \mathrm{Var}(X) = E[X^2] - (\underbrace{E[X]}_{0})^2 = E[X^2] = \sum_{x \in S} x^2 \cdot f(x) = \sum_{x=-1}^{1} x^2 \cdot \frac{1}{3} = \frac{2}{3}$$

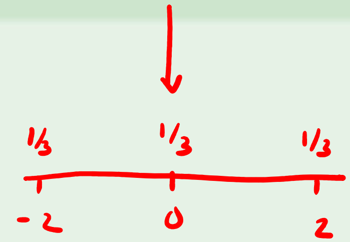
r.v notation: capital latin letter

$$\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{2}{3}} = \frac{\sqrt{6}}{3}$$

Example

Let the random variable Y have the pmf

$$f(x) = \frac{1}{3}\,, \quad x \in \overbrace{\{-2, 0, 2\}}^{S_Y}\,.$$



Find its the variance σ_Y^2 and standard deviation σ_Y.

Compare with the previous example.

$$E[Y] = \sum_{y \in S_Y} y \cdot f(y) = (-2)\cdot f(-2) + 0 \cdot f(0) + 2 \cdot f(2) = (-2)\cdot\frac{1}{3} + 0\cdot\frac{1}{3} + 2\cdot\frac{1}{3} = 0$$

same as $E[X]$

$$E[X] = E[Y]$$

$$\sigma_Y^2 = Var(Y) = E[Y^2] - \big(\underbrace{E[Y]}_{=0}\big)^2 = E[Y^2] = \sum_{y \in S_Y} y^2 \cdot f(y) = (-2)^2\cdot\frac{1}{3} + 0^2\cdot\frac{1}{3} + 2^2\cdot\frac{1}{3} = \frac{8}{3}$$

$$\sigma_Y = \sqrt{Var(Y)} = \sqrt{\frac{8}{3}} = \frac{2\sqrt{6}}{3}$$

as expected

$$\frac{2}{3} = \sigma_X^2 < \sigma_Y^2 = \frac{8}{3}$$

Example

Let X have a uniform distribution on the first m positive integers.

Find its mean and variance.

X takes values on $\boxed{S_X = \{1, 2, \dots, m\}}$

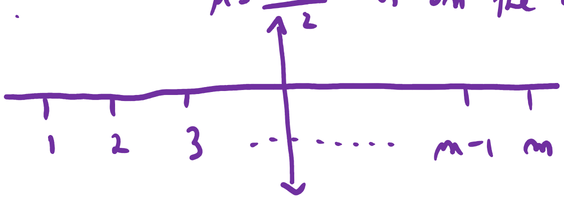
X uniform on $\{1, 2, \dots, m\}$ means that the pmf of X is

$$f(x) = \frac{1}{m}, \quad x = 1, 2, \dots, m$$

Key example: fair 6-faced die
↳ $m = 6$

Mean:
$$\mu = E[X] = \sum_{x \in S_X} x \cdot f(x) = \sum_{x=1}^{m} x \cdot \frac{1}{m} = \frac{1}{m} \underbrace{\sum_{x=1}^{m} x}_{\text{sum of first } m \text{ integers}} = \frac{1}{m} \frac{m(m+1)}{2} = \frac{m+1}{2}$$

$\mu = \frac{m+1}{2}$ is on the exactly middle of the set S_X



EXAMPLE:
For a fair die, the average score is $\frac{7}{2} = 3.5$

NOTE: $\sum_{x=1}^{N} x = \frac{N(N+1)}{2}$

$$\sigma_X^2 = \mathrm{Var}(X) = E[X^2] - (E[X])^2$$

we already compute $E[X] = \frac{m+1}{2}$

$$E[X^2] = \sum_{x \in S_X} x^2 \cdot \underbrace{f(x)}_{1/m} = \sum_{x=1}^{m} x^2 \cdot \frac{1}{m} = \frac{1}{m} \sum_{x=1}^{m} x^2 = \frac{1}{m} \cdot \frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}$$

Recall:

$$\sum_{x=1}^{m} x^2 = \frac{m(m+1)(2m+1)}{6}$$

Then, we obtain that

$$\sigma_X^2 = E[X^2] - (\underbrace{E[X]}_{\frac{m+1}{2}})^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2$$

$$= (m+1)\left(\frac{2m+1}{6} - \frac{m+1}{4}\right) =$$

$$= (m+1)\left(\frac{4m+2-3m-3}{12}\right) = \frac{(m+1)(m-1)}{12} = \frac{m^2-1}{12}$$

$$\sigma_X^2 = \mathrm{Var}(X) = (m^2-1)/12$$

Properties

SUMMARY:

→ (1) $E[aX+b] = aE[X]+b$

(2) $Var(aX+b) = a^2 Var(X)$

(3) $\sigma_{aX+b} = |a|\sigma_X$

Let X be a random variable with mean μ_X and variance σ_X^2.

Define a new random variable as $\boxed{Y = aX + b,}$ where a and b are constants.

$\mu_Y = ?$ $\quad \sigma_Y^2 = ?$

Then:

1) the mean of Y is

linearity $\quad E[b]=b$

$$\mu_Y = E(Y) = E(aX + b) = a\underbrace{E(X)}_{\mu_X} + b = a\mu_X + b \leftarrow \mu_Y = a\mu_X + b$$

2) the variance of Y is

$$a^2 E[(X-\mu_X)^2]$$

$$\sigma_Y^2 = E\left[(Y - \mu_Y)^2\right] = E\left[(aX + b - a\mu_X - b)^2\right] = E\left[a^2(X - \mu_X)^2\right] = a^2\sigma_X^2 *$$

$$(aX - a\mu_X)^2 = a^2(X-\mu_X)^2$$

3) the standard deviation of Y is

$$\sigma_Y = |a|\sigma_X \ .$$

Example

Let X be a random variable with mean $\underbrace{\mu = 1}_{E[X]=1}$ and variance $\underbrace{\sigma^2 = 2}_{Var(X)=2}$.

Find:

$E(X+2) = E[X] + E[2] = 1 + 2 = 3$

$E(3X-2) = 3E[X] - 2 = 3.(1) - 2 = 1$

$\mathrm{Var}(X-1) = Var(X) = 2$

$\mathrm{Var}(2X) = 2^2 Var(X) = 2^2 . 2 = 8$

$\mathrm{Var}(2X+5) = Var(2X) = 2^2 . Var(X) = 8$

Higher order moments

Definition

Let r be a positive integer.

The *rth moment of X (or of its distribution) about the origin* is given by

$$E(X^r) = \sum_{x \in S} x^r f(x) ,$$

mean $\mu = E[X]$ is the 1st moment about origin

provided the sum is absolutely convergent.

Given $b \in \mathbb{R}$, the *rth moment of X (or of its distribution) about b* is given by

$$E[(X-b)^r] = \sum_{x \in S} (x-b)^r f(x) ,$$

Variance $\sigma^2 = E[(X-\mu)^2]$ is the 2nd moment of X about the mean

provided the sum is absolutely convergent.

Another, less used, type of higher order moment:

Let r be a positive integer. The *rth factorial moment*, denoted $E[(X)_r]$, is defined as

$$E[(X)_r] = E[X(X-1)(X-2)\cdots(X-r+1)] \ ,$$

provided the mathematical expectation on the right hand side exists.

Note: the second factorial moment is equal to the difference of the second and first moments about the origin:

$$\underbrace{E[X(X-1)]}_{E[(X)_2]} = E\left(X^2\right) - E(X) \ .$$

$$E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X]$$

Index of skewness

Definition

The *index of skewness of the random variable X* (or of its distribution), which we will denote by γ, equals the third moment of X about the mean normalized by σ^3:

$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3}\,.$$

← gives a sense of symmetry of the distribution or the lack of it

Notes:

1) γ is a unitless (scale-free) quantity

2) If the distribution of X is symmetric about the mean μ, then $\boxed{\gamma = 0}$

- The converse is not true: there are non-symmetric distributions with $\gamma = 0$

X symmetric $\Longrightarrow$ $\gamma = 0$

But $\gamma = 0$ $\not\Longrightarrow$ X is symmetric

Exercise:

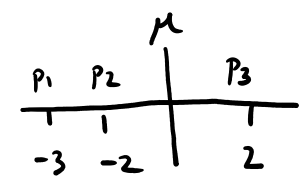
Pick p_1, p_2, p_3 so that

$p_1, p_2, p_3 > 0$

$p_1 + p_2 + p_3 = 1$

$E[X] = 0$

$\gamma = 0$



If the histogram of a random variable is such that:

- it has a single peak;
- more total probability lies to the right of the peak than to the left;
- the heights of the bars in the probability histogram decrease roughly monotonically away from the peak and more slowly to its right than to its left.

Then:

1) the index of skewness γ is positive $\gamma > 0$

2) we say that the *distribution is skewed to the right*.

right tail is heavier than left tail

If the histogram of a random variable is such that:

- it has a single peak;
- more total probability lies to the left of the peak than to the right;
- the heights of the bars in the probability histogram decrease roughly monotonically away from the peak and more slowly to its left than to its right.

Then:

$\gamma < 0$

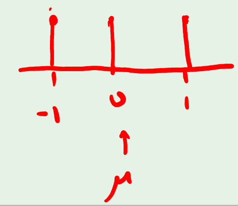
1) the index of skewness γ is negative

2) we say that the *distribution is skewed to the left*.

Example

Let X be the random variable with pmf

$$f(x) = 1/3, \quad x \in \{-1, 0, 1\} .$$



Give the index of skewness of X.

We know that $E[X] = 0$ (previous example)

Since the distribution of X is symmetric about the mean $\mu = 0$ then $\gamma = 0$

$$\boxed{f(\mu - x) = f(\mu + x) \quad \text{for all } x \in \mathbb{R}}$$

symmetry $\Rightarrow \gamma = 0$

Property

The index of skewness may also be computed using: /

$$E\left[(X-\mu)^3\right] = E\left(X^3\right) - 3\mu E\left(X^2\right) + 2\mu^3$$
$$= E\left(X^3\right) - 3\mu\sigma^2 - \mu^3 \; .$$

analogue to the property

$Var(X) = E[X^2] - (E[X])^2$

Proof. $E[(X-\mu)^3] \underbrace{=}_{\text{binomial expansion}} E[X^3 - 3\mu X^2 + 3\mu^2 X - \mu^3] \underbrace{=}_{\text{linearity of exp. value}} E[X^3] - 3\mu E[X^2] + \underbrace{3\mu^2 \overbrace{E[X]}^{\mu}}_{3\mu^3} - \mu^3$

$\sigma^2 = E[X^2] - (\overbrace{E[X]}^{\mu})^2$

$E[X^2] = \sigma^2 + \mu^2$

$= E[X^3] - 3\mu \underbrace{E[X^2]}_{\sigma^2+\mu^2} + 2\mu^3$

$= E[X^3] - 3\mu(\sigma^2 + \mu^2) + 2\mu^3$

$= E[X^3] - 3\mu\sigma^2 - \mu^3$

$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i \cdot b^{n-i}$ $\qquad (a+b)^3 = a^3 + 3a^2 b + 3ab^2 + b^3$

Example

Let X be the random variable with pmf

$$f(x) = \frac{4-x}{6}, \quad x \in \{1, 2, 3\}.$$

$$f(x) = \begin{cases} \frac{3}{6}, & x=1 \\ \frac{2}{6}, & x=2 \\ \frac{1}{6}, & x=3 \end{cases}$$

Determine the index of skewness of X.

We want to determine $\gamma = \dfrac{E[(X-\mu)^3]}{\sigma^3}$

✓ $\mu = E[X] = \sum_{x=1}^{3} x \cdot f(x) = 1 \cdot \frac{3}{6} + 2 \cdot \frac{2}{6} + 3 \cdot \frac{1}{6} = \frac{10}{6}$ ✓

$E[X^2] = \sum_{x=1}^{3} x^2 \cdot f(x) = (1)^2 \cdot \frac{3}{6} + (2)^2 \cdot \frac{2}{6} + (3)^2 \cdot \frac{1}{6} = \frac{20}{6}$ ✓

✓ $\sigma^2 = Var(X) = E[X^2] - (E[X])^2 = \frac{20}{6} - \left(\frac{10}{6}\right)^2 = \frac{20}{6} - \frac{100}{36} = \frac{20}{36}$ ←

$\sigma = \sqrt{Var(X)} = \frac{\sqrt{20}}{6} = \frac{2\sqrt{5}}{6} = \frac{\sqrt{5}}{3} \implies \sigma^3 = (\sigma)^3 = \left(\frac{\sqrt{5}}{3}\right)^3$

$$E[X^3] = \sum_{x=1}^{3} x^3 \cdot f(x) = (1)^3 \cdot \frac{3}{6} + (2)^3 \cdot \frac{2}{6} + (3)^3 \cdot \frac{1}{6} =$$

$$= \frac{3}{6} + \frac{16}{6} + \frac{27}{6} = \frac{46}{6}$$

$$E[(X-\mu)^3] = E[X^3] - 3\mu\sigma^2 - \mu^3 = \frac{46}{6} - 3 \cdot \left(\frac{10}{6}\right) \cdot \frac{20}{36} - \left(\frac{10}{6}\right)^3$$

2 slides before

$\mu = 10/6$

$\frac{20}{36}$

$$= \frac{46 \times 36 - 600 - 1000}{216}$$

$$= \frac{56}{216}$$

$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{56/216}{\left(\sqrt{5}/3\right)^3} = \ldots \text{ calculator}$$