Note 8

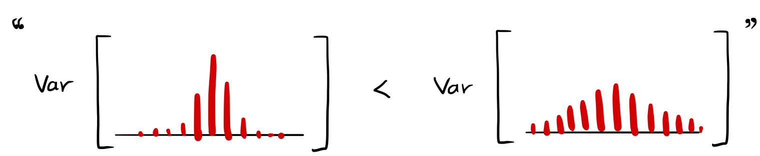
2.3. Special Mathematical Expectations

Mean

- The expected value $E(X)$ is often called the **mean** of X (or of its distribution), reflecting the fact that $E(X) = \sum_{x \in S} xp(x)$ may be regarded as a weighted mean.
- The mean $E(X)$ of X is often denoted by the Greek letter μ (mu).

Variance and Standard Deviation

- The expected value $E[(X-\mu)^2] = E[(X-E(X))^2]$ is called the **variance** of X and denoted by $\mathrm{Var}(X)$. This is a measure of how much the distribution is dispersed.



- Here are some properties of variance:

Proposition

Let X be a random variable of the discrete type such that $E(X^2)$ exists. Then

(1) $\mathrm{Var}(X)$ is always non-negative.

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

(3) $\mathrm{Var}(X) = 0$ implies that X is a constant random variable, i.e., $P(X = c) = 1$ for some constant c.

Proof. Write $\mu = E(X)$. Then (1) is an immediate consequence of the fact that $(X-\mu)^2 \geq 0$. Indeed, if S is the space of X and p is the PMF of X, then

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_{x \in S} \underbrace{(x-\mu)^2 p(x)}_{\geq 0} \geq 0.$$

Next, (2) is a consequence of the following computation:

$$E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu \underbrace{E[X]}_{=\mu} + \underbrace{E[\mu^2]}_{=\mu^2} = E[X^2] - \mu^2.$$

Finally, we prove (3). Assume that $\mathrm{Var}(X) = 0$. Since each $(x-\mu)^2 p(x)$ is non-negative and their sum is zero, it follows that $(x-\mu)^2 p(x) = 0$ for each $x \in S$. In particular, if $x \neq \mu$ then $p(x) = 0$. Since the sum of values of $p(x)$ must be 1, this leaves only one possible scenario that $p(\mu) = 1$, i.e., $P(X = \mu) = 1$. □

- $\sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of X and denoted by the Greek letter σ (sigma). Consequently, $\mathrm{Var}(X) = \sigma^2$.

Example (Uniform distribution)

Let X have a uniform distribution on $S = \{1, \cdots, m\}$, where m is a positive integer. That is, $p(x) = \frac{1}{m}$ for each $x \in S$. Then

- The mean of X is

$$\mu = E(X) = \sum_{x \in S} x p(x) = \frac{1}{m} \sum_{x=1}^{m} x = \frac{1}{m} \cdot \frac{m(m+1)}{2} = \frac{m+1}{2}.$$

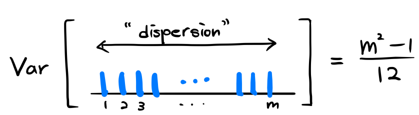
- To compute the variance of X, we first evaluate $E(X^2)$:

$$E(X^2) = \sum_{x \in S} x^2 p(x) = \frac{1}{m} \sum_{x=1}^{m} x^2 = \frac{1}{m} \cdot \frac{m(m+1)(2m+1)}{6} = \frac{(m+1)(2m+1)}{6}.$$

Then

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = \frac{(m+1)(2m+1)}{6} - \left(\frac{m+1}{2}\right)^2 = \frac{m^2-1}{12}.$$

Note that $\mathrm{Var}(X)$ increases in m, conforming to the intuition that $\mathrm{Var}(X)$ is a measure of how much a distribution is spread out:



- The following proposition discusses how both the mean and the variable changes under a linear function:

Proposition

Let a and b be constants. Then

$$E(aX+b) = aE(X) + b \qquad \text{and} \qquad \mathrm{Var}(aX+b) = a^2\,\mathrm{Var}(X).$$

In particular, if $Y = aX + b$, then $\mu_Y = a\mu_X + b$ and $\sigma_Y = |a|\sigma_X$.

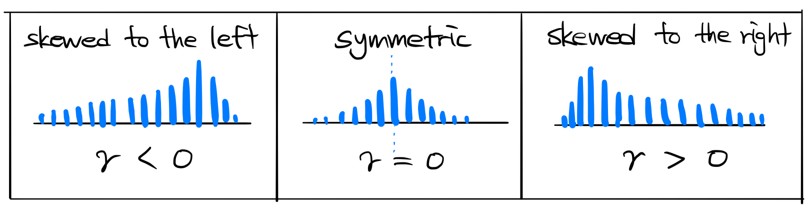
Proof. The first part is an easy consequence of the linearity of $E(\cdot)$ and $E(b) = b$. For the second part,

$$\begin{aligned}\operatorname{Var}(aX+b) &= E[((aX+b) - E(aX+b))^2] \\ &= E[(aX + \not{b} - aE(X) - \not{b})^2] \\ &= a^2 E[(X - E(X))^2] \\ &= a^2 \operatorname{Var}(X).\end{aligned}$$

Finally, writing $Y = aX + b$ and taking square to both sides of $\operatorname{Var}(Y) = a^2 \operatorname{Var}(X)$ gives $\sigma_Y = |a|\sigma_X$ as desired. □

Index of skewness

- The quantity $\dfrac{E[(X-\mu)^3]}{\sigma^3}$ is called the **index of skewness** of X, often denoted by the Greek letter γ (gamma).
- In the "unimodal case", i.e., if the probability histogram has a single peak, then γ roughly describes the tendency of how the distribution is skewed:



Moment generating function

- The function

$$M(t) := E(e^{tX})$$

  is called the **moment generating function (MGF)** of X, provided it exists near $t = 0$.

- If X is a random variable of the discrete type with PMF $p(x)$ and the space $S = \{b_1, b_2, \cdots\}$, then

$$M(t) = E(e^{tX}) = \sum_{x \in S} e^{tx} p(x) = e^{tb_1} p(b_1) + e^{tb_2} p(b_2) + \cdots.$$

  If this exists near $t = 0$, then we can indeed read out the "coefficient" of each e^{tx} in the right-hand side and hence can completely determine the PMF $p(x)$ out of $M(t)$. The consequence of this observation is two-fold:

(1) Theoretically, $M(t)$ uniquely determines the distribution of X, if exists. It is like the DNA of distribution, as it endoces all the information on the distribution.

(2) Practically, we can extract various information on the distribution of X from $M(t)$.

Example

Assume that X has the MGF

$$M(t) = \frac{3e^t + 2e^{2t} + e^{3t}}{6}.$$

- By reading out the exponents, we find that the space of X is $S = \{1, 2, 3\}$.
- For each $x \in S$, reading out the "coefficient" of e^{tx} gives:

$$M(t) = \underbrace{\boxed{\frac{3}{6}}}_{=p(1)} e^t + \underbrace{\boxed{\frac{2}{6}}}_{=p(2)} e^{2t} + \underbrace{\boxed{\frac{1}{6}}}_{=p(3)} e^{3t}$$

Example (Poisson distribution)

Let $\lambda > 0$ and assume that X has the MGF

$$M(t) = e^{\lambda(e^t - 1)}.$$

The right-hand side may not look like a "series in e^t". However, using the famous Taylor series $e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n$, we get

$$M(t) = e^{\lambda e^t} e^{-\lambda} = \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} e^{-\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} e^{nt}.$$

- By reading out the exponents, the space of X is $S = \{0, 1, 2, \cdots\}$.
- For each $x \in S$, the "coefficient" of e^{xt} is $\dfrac{\lambda^x e^{-\lambda}}{x!}$, which is the value of $p(x)$.

- Although any information on the distribution can be extracted from the MGF, it is particularly easier to compute the moments of X with the MGF.

Proposition

Suppose that the MGF $M(t)$ of X exists near $t = 0$. Then, for each positive integer r, we have

$$E(X^r) = M^{(r)}(0).$$

This is called the ***r*-th moment** of X.

Sketch of Proof. If X is of the discrete type with PMF $p(x)$, then

$$\frac{d^r}{dt^r}M(t) = \frac{d^r}{dt^r}\sum_{x\in S} e^{tx}p(x) = \sum_{x\in S}\frac{d^r}{dt^r}e^{tx}p(x) = \sum_{x\in S} x^r e^{tx}p(x) = E(X^r e^{tX}).$$

(In the second step, interchanging the order of summation and differentiation can be justified by the knowledge from the branch of mathematics called analysis.) Plugging $t = 0$ then proves the desired equality. □

Example (Geometric distribution)

If X has a geometric distribution with parameter $p \in (0, 1)$, i.e., with the PMF

$$p(x) = \begin{cases} q^{x-1}p, & \text{for } x = 1, 2, 3, \cdots; \\ 0, & \text{otherwise,} \end{cases}$$

then the MGF is

$$M(t) = E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx}q^{x-1}p = \frac{pe^t}{1-qe^t},$$

provided $qe^t < 1$ (or equivalently, $t < -\ln q$). Then

$$E(X) = M'(X) = \left.\frac{pe^t}{(1-qe^t)^2}\right|_{t=0} = \frac{1}{p}$$

and

$$E(X^2) = M''(0) = \left.\frac{pe^t(1+qe^t)}{(1-qe^t)^3}\right|_{t=0} = \frac{1+q}{p^2}.$$

So it follows that

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{q}{p^2}.$$