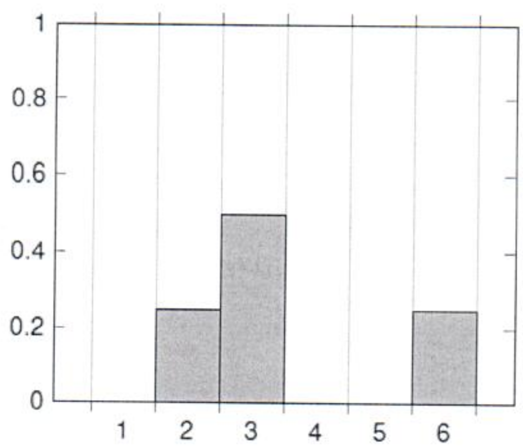
Example 4.2.1 *Let X and Y be discrete random variables with the following distributions*

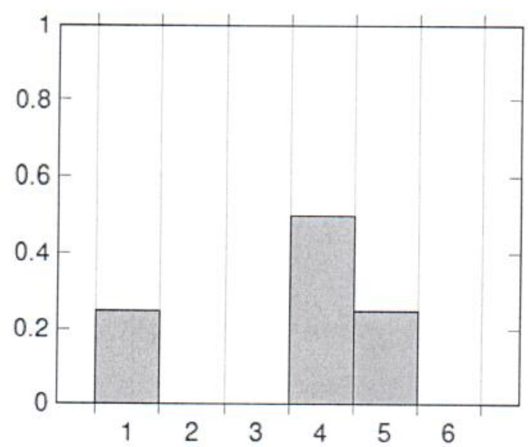
x	$P(X = x)$	y	$P(Y = y)$
1	0	1	1/4
2	1/4	2	0
3	1/2	3	0
4	0	4	1/2
5	0	5	1/4
6	1/4	6	0

Show that these distributions have the same mean and variance.

X



Y



For X: $0 + \frac{1}{2} + \frac{3}{2} + 0 + 0 + \frac{3}{2}$

$$\mu_x = E(X) = 1 \cdot 0 + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 4 \cdot 0 + 5 \cdot 0 + 6 \cdot \frac{1}{4} = \frac{7}{2} = 3.5$$

$$\mathrm{var}(X) = \sigma_x^2 = E((X-\mu)^2) = \cancel{\left(1-\frac{7}{2}\right)^2 \cdot 0} + \left(2-\frac{7}{2}\right)^2 \cdot \frac{1}{4}$$
$$+ \left(3-\frac{7}{2}\right)^2 \cdot \frac{1}{2} + \cancel{\left(4-\frac{7}{2}\right)^2 \cdot 0} + \cancel{\left(5-\frac{7}{2}\right)^2 \cdot 0}$$
$$+ \left(6-\frac{7}{2}\right)^2 \cdot \frac{1}{4}$$
$$= (-1.5)^2 \cdot \frac{1}{4} + (-0.5)^2 \cdot \frac{1}{2} + (2.5)^2 \cdot \frac{1}{4}$$
$$= \frac{9}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} + \frac{25}{4} \cdot \frac{1}{4} = \frac{9}{16} + \frac{1}{8} + \frac{25}{16} = \frac{36}{16}$$
$$= 2.25$$

For Y:
$$\mu_Y = E(Y) = 1 \cdot \frac{1}{4} + 2 \cdot 0 + 3 \cdot 0 + 4 \cdot \frac{1}{2} + 5 \cdot \frac{1}{4} + 6 \cdot 0$$
$$= \frac{1}{4} + 2 + \frac{5}{4} = 3.5$$
$$\mathrm{var}(Y) = \sigma_Y^2 = E((Y-\mu_Y)^2) = (-2.5)^2 \cdot \frac{1}{4} + (-1.5)^2 \cdot 0$$
$$+ \dots = 2.25$$

Let us now compute the 3rd moment about the mean.

For X:

$$\mu_3 = E((X-\mu)^3) = \cancel{\left(1-\frac{7}{2}\right)^3 \cdot 0} + \left(2-\frac{7}{2}\right)^3 \cdot \frac{1}{4}$$

$$+ \left(3-\frac{7}{2}\right)^3 \cdot \frac{1}{2} + \cancel{\left(4-\frac{7}{2}\right)^3 \cdot 0}$$

$$+ \cancel{\left(5-\frac{7}{2}\right)^3 \cdot 0} + \left(6-\frac{7}{2}\right)^3 \cdot \frac{1}{4}$$

$$= \left(\frac{-3}{2}\right)^3 \cdot \frac{1}{4} + \left(-\frac{1}{2}\right)^3 \cdot \frac{1}{2} + \left(\frac{5}{2}\right)^3 \cdot \frac{1}{4}$$

$$\ldots = 3$$

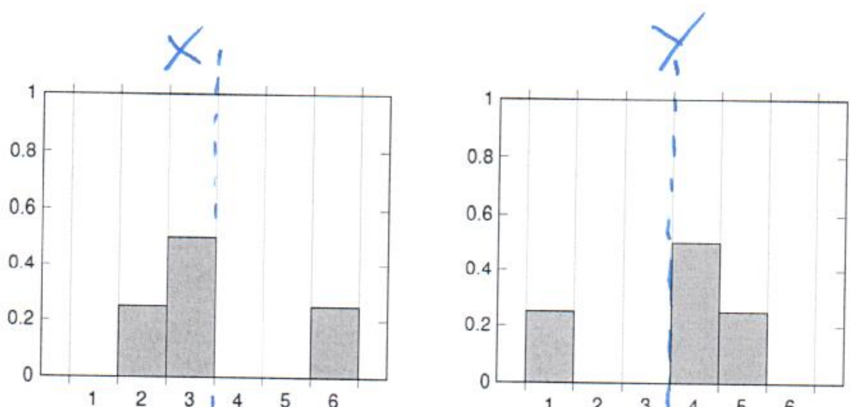
we found on p. 133 $\mu_Y = 3.5$

For Y: $\mu_3 = E((Y-\mu)^3) = (1-\frac{7}{2})^3 \cdot \frac{1}{4} + (2-\frac{7}{2})^3 \cdot 0$

$$+ (3-\tfrac{7}{2})^3 \cdot 0 + (4-\tfrac{7}{2})^3 \cdot \tfrac{1}{2}$$

$$+ (5-\tfrac{7}{2})^3 \cdot \tfrac{1}{4} + (6-\tfrac{7}{2})^3 \cdot 0$$

$$= \cdots = -3$$



The 3rd moment about the mean describes the symmetry of the graph about the mean.

Notice that the distribution on the left has a higher proportion of its probabilities to the left of the mean $\mu = \frac{7}{2}$. [redacted]

The opposite is true for the distribution on the right. [redacted]

Example 4.2.2 *Let random variable X be the number of points on a regular 6-sided die. Compute mean and variance of X.*

The mean is

$$\mu = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= \frac{21}{6} = 3.5$$

The variance is

$$\sigma^2 = E((X-\mu)^2) = (1-3.5)^2 \cdot \frac{1}{6} + (2-3.5)^2 \cdot \frac{1}{6} + (3-3.5)^2 \cdot \frac{1}{6}$$

$$+ (4-3.5)^2 \cdot \frac{1}{6} + (5-3.5)^2 \cdot \frac{1}{6} + (6-3.5)^2 \cdot \frac{1}{6}$$

$$= \frac{1}{6}\left[(-2.5)^2 + (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 + (2.5)^2\right]$$

$$= \frac{1}{6}\left[6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25\right]$$

$$= \frac{1}{6} \cdot (17.5) = \frac{1}{6} \cdot \frac{35}{2} = \frac{35}{12} \approx 2.9167$$

$$(X-\mu)^2 = X^2 - 2\mu X + \mu^2$$

Theorem on p. 120

Consider the variance. Using properties of expected values we have

var$(X) = \sigma^2$ — 2nd moment about the mean of X

$$\begin{aligned}
\mathrm{var}(X) = \sigma^2 = E((X-\mu)^2) &= \underline{E(X^2 - 2\mu X + \mu^2)} = E(X^2) - E(2\mu X) + E(\mu^2) \\
&= E(X^2) - 2E(\mu X) + E(\mu^2) \\
&= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2 \\
&= E(X^2) - \underbrace{2\mu \cdot \mu}_{-2\mu^2} + \mu^2 \\
&= \underbrace{E(X^2)}_{=\mu_2'} - \mu^2
\end{aligned}$$

We summarize this in a theorem.

Theorem 4.2.3

$$\text{variance} — \sigma^2 = \mu_2' - \mu^2$$

Example 4.2.4 *Redo the previous die rolling problem with theorem.*

We first need to find the mean μ:

$$\mu = E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} = 3.5$$

Next we find $\mu_2' = E(X^2)$:

$$\mu_2' = E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdot 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$$

$$= \frac{1}{6} + \frac{4}{6} + \frac{9}{6} + \frac{16}{6} + \frac{25}{6} + \frac{36}{6} = \frac{91}{6}$$

Then using the theorem, the variance is

$$\sigma^2 = \underset{E(X^2)}{\underset{\shortparallel}{\mu_2'}} - \mu^2 = \frac{91}{6} - (3.5)^2 \approx 2.9167.$$

Theorem 4.2.5 *If X has variance σ^2, then for constants a and b*

$$var(\underbrace{aX+b}_{Y}) = a^2\underbrace{\sigma^2}_{var(X)}.$$

Proof: Let $Y = aX + b$, and let $\mu = E(X)$.

Then

$$\underbrace{E(Y)}_{\text{mean of } Y} = E(aX+b) = a\underbrace{E(X)}_{\mu} + b = \underbrace{a\mu + b}_{\text{mean of } Y}.$$

For the variance of Y, we have

$$\text{var}(Y) = E((Y - (a\mu + b))^2)$$

$$= E((aX \cancel{+b} - a\mu \cancel{-b})^2) = E((aX - a\mu)^2)$$

$$= E(a^2X^2 - 2a^2\mu X + a^2\mu^2)$$

$$= E((aX - a\mu)^2)$$

$$= E(a^2X^2 - 2a^2X\mu + a^2\mu^2)$$

$$= E(a^2X^2) - E(2a^2\mu X) + E(a^2\mu^2)$$

$$= a^2E(X^2) - 2a^2\mu E(X) + a^2\mu^2 \quad = a^2E(X^2) - 2a^2\mu E(X) + a^2\mu^2$$

(by Theorem 4.1.10)

$$a^2E(X^2) - 2a^2\mu E(X) + a^2\mu^2$$

$$a^2(E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2)$$

$$a^2(E(X^2) - 2\mu^2 + \mu^2)$$

$$a^2(\underbrace{E(X^2) - \mu^2}_{var(X) = \sigma^2})$$

$$= a^2(E(X^2) - 2\mu^2 + \mu^2)$$

$$= a^2(E(X^2) - \mu^2)$$

$$= a^2\sigma^2 \quad \text{(by Theorem 4.2.3)}$$

σ^2 : variance
σ : standard deviation

4.2.1 Chebyshev's Theorem

standard deviation

The next important theorem shows how σ describes the spread of the probability distribution.

Theorem 4.2.6 (Chebyshev's Theorem) *Let X be a random variable with mean μ and standard deviation σ. Then for any $k > 0$,*

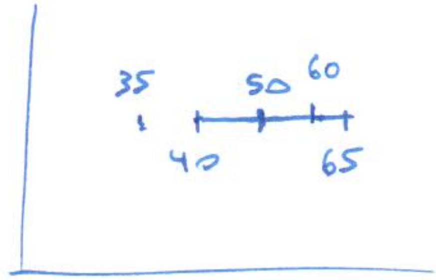
$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

In words, the probability that values for X lie within k standard deviations of the mean is at least $1 - \frac{1}{k^2}$.

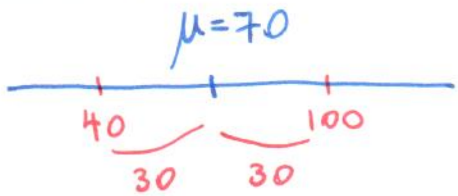
(The proof of this theorem can be found in the lecture slides.)

Chebyshev's Theorem:

$$P(|X-\underbrace{\mu}_{\text{mean}}| < k\underbrace{\sigma}_{\text{standard deviation}}) \geq 1 - \frac{1}{k^2}$$



Example 4.2.7 *The mean score of an exam is 70, with a standard deviation of 5. At least what percentage of the data set lies between 40 and 100?*



$\mu = 70$
$\sigma = 5$

$|100-70| = |40-70| = 30$

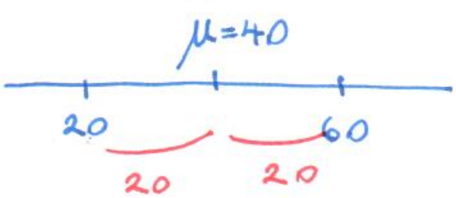
$30 = k\sigma \Rightarrow k = 6$ ($\sigma = 5$)

Consider $k = 6$ in the theorem.

$$P(|X-\overset{\mu}{70}| < \overset{k}{6}\cdot\overset{\sigma}{5}) \geq 1 - \frac{1}{6^2} = 1 - \frac{1}{36} = \frac{35}{36}$$

By Chebyshev's Theorem, at least $\frac{35}{36} \approx 97\%$ of the data lies between 40 and 100.

Example 4.2.8 *The mean age of a flight attendant is 40, with a standard deviation of 8. At least what percent of the data set lies between 20 and 60?*



$\mu = 40$
$\sigma = 8$

$|60-40| = |20-40| = 20$

$20 = k\cdot\sigma \Rightarrow k = 2.5$ ($\sigma = 8$)

So, consider $k = 2.5$ in the theorem

Then, by Chebyshev's Theorem:

$$P(|X-\underbrace{40}_{\mu}| < \underbrace{2.5}_{k}\cdot\underbrace{8}_{\sigma}) \geq 1 - \frac{1}{(2.5)^2} = 1 - \frac{1}{6.25} = 1 - \frac{4}{25} = \frac{21}{25} = 0.84$$

So, by Chebyshev's Theorem, at least 84% of the data lies between 20 and 60.

For what comes next, we need to know the **Maclaurin Series for** e^x,

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \cdots = \sum_{i=0}^{\infty} \frac{1}{i!}x^i,$$

and term-by-term differentiation

$$\frac{d}{dx}\left(\sum_{i=0}^{\infty} f_i(x)\right) = \sum_{i=0}^{\infty} \frac{d}{dx}\left(f_i(x)\right).$$

Now we can talk about **moment generating functions**.

The **moment generating function of a random variable** X, where it exists, is given by

$$\text{discrete case:} \quad M_X(t) = E(e^{tX}) = \sum_x e^{tx} \cdot f(x)$$

$$\text{continuous case:} \quad M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx$$

where $f(x)$ is the probability distribution/density of X.

We will see why this name is appropriate.

Maclaurin series:

$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots$

(discrete case)

Expanding the expression for $M_X(t)$,

$e^{tx} = 1 + tx + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \dots$

$M_X(t)$

$$= \sum_x e^{tx} \cdot f(x)$$

$$= \sum_x \left(1 + (tx) + \frac{1}{2!}(tx)^2 + \frac{1}{3!}(tx)^3 + \dots\right) \cdot f(x)$$

$$= \sum_x f(x) + (tx)f(x) + \frac{(tx)^2}{2!}f(x) + \frac{(tx)^3}{3!}f(x) + \dots$$

$$= \sum_x f(x) + \sum_x tx f(x) + \sum_x \frac{(tx)^2}{2!} f(x) + \sum_x \frac{(tx)^3}{3!} f(x) + \dots$$

$$= \underbrace{\sum_x f(x)}_{\text{0th moment about the origin}} + t\underbrace{\sum_x xf(x)}_{E(X)' \text{ 1st moment about the origin}} + \frac{t^2}{2!}\underbrace{\sum_x x^2 f(x)}_{E(X^2)' \text{ 2nd moment about the origin}} + \frac{t^3}{3!}\underbrace{\sum_x x^3 f(x)}_{E(X^3)'} + \dots$$

We see the rth moments about the origin appearing in the terms of the series.

Moment generating function of X

$$M_X(t) = \sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \dots$$

To extract the ith moment, we take the ith derivative with respect to t, and evaluate at $t = 0$.

For example, to get the 2nd moment: $\frac{d^2}{dt^2}M_X(t)\Big|_{t=0} =$

2nd derv. with resp. to t

$$\frac{d^2}{dt^2}\left(\sum_x f(x) + t\sum_x xf(x) + \frac{t^2}{2!}\sum_x x^2 f(x) + \frac{t^3}{3!}\sum_x x^3 f(x) + \dots\Big|_{t=0}\right)$$

$$\frac{d}{dt}M_X(t) = 0 + \sum_x xf(x) + \frac{2t}{2!}\sum_x x^2 f(x) + \frac{3t^2}{3!}\sum_x x^3 f(x) + \dots$$

(1st derivative)

Take 2nd derivative of each term with respect to t,

$$= \left(0 + 0 + \sum_x x^2 f(x) + \cancel{\overset{0}{t}\sum_x x^3 f(x) + \frac{t^2}{2}\sum_x x^4 f(x) + \dots}\Big|_{t=0}\right)$$

Evaluating at $t=0$, makes all terms beyond $\sum_x x^2 f(x)$ equal to 0.

Letting $t = 0$ gives $\frac{d^2}{dt^2}M_X(t)\Big|_{t=0} = \sum_x x^2 f(x) = E(X^2)$.

2nd moment about the origin.

Example 4.2.9 *Let X be a discrete random variable with distribution $f(x) = \frac{1}{8}\binom{3}{x}$ for $x = 0, 1, 2, 3$.*

"3 choose x"

The moment generating function for X is $M_X(t) = \sum_{x=0}^{3} e^{tx} \cdot \left(\frac{1}{8}\binom{3}{x}\right)$

$$M_X(t) = e^{t\cdot 0}\cdot\left(\frac{1}{8}\cdot\binom{3}{0}\right) + e^{t\cdot 1}\left(\frac{1}{8}\cdot\binom{3}{1}\right) + e^{t\cdot 2}\cdot\left(\frac{1}{8}\cdot\binom{3}{2}\right) + e^{t\cdot 3}\left(\frac{1}{8}\cdot\binom{3}{3}\right)$$

$$= \frac{1}{8}\left(e^0\underbrace{\binom{3}{0}}_{=1} + e^t\underbrace{\binom{3}{1}}_{=3} + e^{2t}\underbrace{\binom{3}{2}}_{=3} + e^{3t}\underbrace{\binom{3}{3}}_{=1}\right) = \frac{1}{8}(1+e^t)^3$$

expand the cubic expression to verify the identity

To find the mean (1st moment about the origin): $E(X)$

$$\left.\frac{d}{dt}M_X(t)\right|_{t=0} = \left.\frac{d}{dt}\frac{1}{8}(1+e^t)^3\right|_{t=0}$$

CHAIN RULE!

$$= 3\cdot\frac{1}{8}(1+e^t)^2\cdot e^t$$

$$= \left.\frac{3}{8}(1+e^t)^2 e^t\right|_{t=0} = \frac{3}{8}2^2 = \frac{3}{2} \quad \frac{3}{8}\cdot 4 = \frac{12}{8} = \frac{3}{2}$$

$$= \frac{3}{8}(1+\underbrace{e^0}_{=1})^2 e^0$$