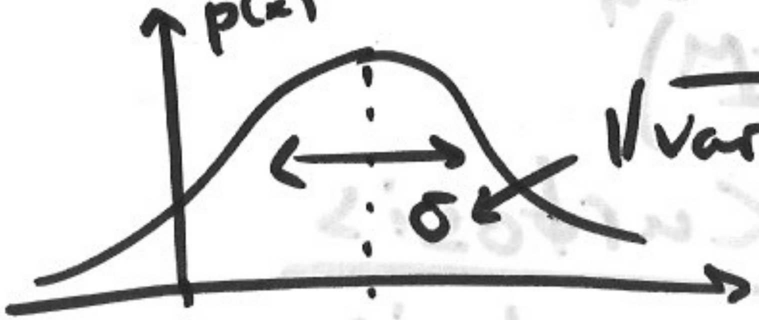
Moments of distributions (of random variables)

$$\mathbb{E}x = \int_x x \cdot p(x)\,dx$$ ← expectation (mean)

$$\underline{\mathbb{E}(x-\mathbb{E}x)^2} = \int (x-\mathbb{E}x)^2\, p(x)\,dx$$

variance

$$Std(x) = \sqrt{Var(x)}$$



$\sqrt{var\,x}$

$x \sim N(\mu, \sigma)$

$\mu = \mathbb{E}(x)$

$$\boxed{\mathbb{E}(x+y) = \mathbb{E}x + \mathbb{E}y}$$

linerity

$$\mathbb{E}(x-\mathbb{E}x)^2 = \mathbb{E}\left(x^2 - 2x\,\mathbb{E}x + (\mathbb{E}x)^2\right) =$$
$$= \mathbb{E}x^2 - 2(\mathbb{E}x)^2 + (\mathbb{E}x)^2 = \mathbb{E}x^2 - (\mathbb{E}x)^2$$

Idea! Let's study

$$\underbrace{\mathbb{E}x, \mathbb{E}x^2, \mathbb{E}x^3, \mathbb{E}x^4}_{\text{moments of r.v. } x}, \dots \quad \mathbb{E}x^n$$

1st 2nd 3rd

Standardized moments

$$\mathbb{E}\left(\frac{x-\mu}{\sigma}\right)^n \leftarrow \text{n-th standardized moment}$$

$\mu = \mathbb{E}x \quad \sigma = \sqrt{\mathbb{E}(x - \mathbb{E}x)^2}$

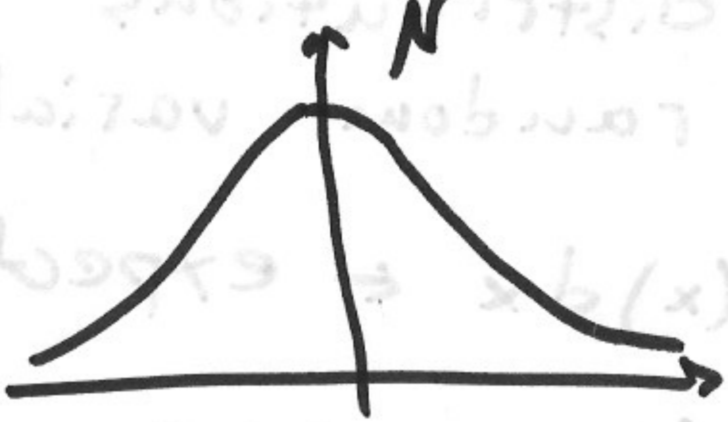
3-rd ~~st~~ s.m. $\mathbb{E}\left(\frac{x-\mu}{\sigma}\right)^3$

— skewness

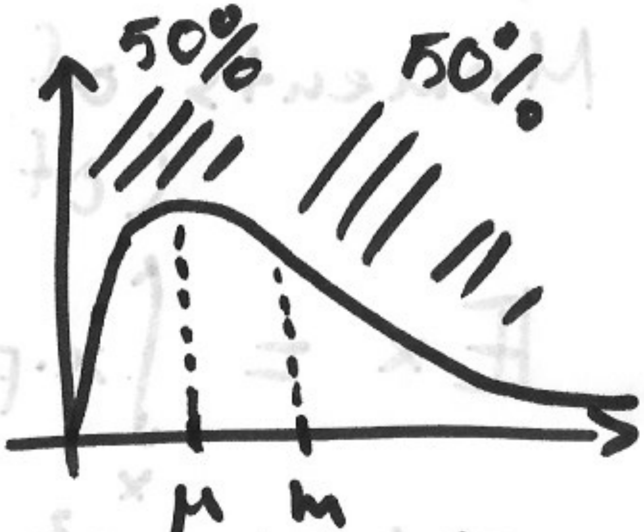
②

Skewness

negatively skewed



Skewness = 0



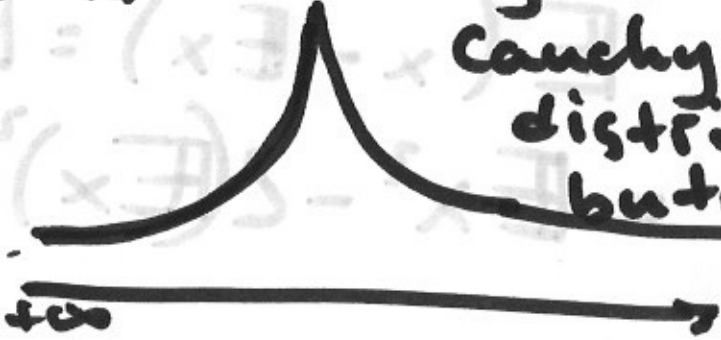
positively skewed

4-th moment, $\mathbb{E}\left(\frac{x-\mu}{\sigma}\right)^4$ – **Kurtosis**
(standardized)

measures how the tails of the distr. decay

$\sim e^{-x^2}$

$x \to +\infty$



Cauchy distribution

Idea!

$$\mathbb{E}(1) + \mathbb{E}(x) + \mathbb{E}\left(\frac{x^2}{2}\right) + \mathbb{E}\left(\frac{x^3}{3!}\right) + \dots + \mathbb{E}\left(\frac{x^n}{n!}\right) =$$

$$\frac{\int p(x)dx}{1} = \mathbb{E}\left(\underbrace{1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots}_{\to\infty}\right) = \mathbb{E}(e^x)$$

"e^x"

$x \to t \cdot x$

$$\mathbb{E}\left(1 + tx + \frac{(tx)^2}{2!} + \dots\right) = \mathbb{E}(e^{tx})$$

$$(e^{ax})'_x = a e^{ax}$$

$$\frac{d}{dt}\mathbb{E}(e^{tx}) = \mathbb{E}\left(\frac{d}{dt}e^{tx}\right) = \underline{\mathbb{E}(x \cdot e^{tx}) =}$$

set $t \to 0$

$$= \mathbb{E}(x)$$

4

so if $\cancel{xxx}$ $M_x(t) = \mathbb{E}_{p(x)}(e^{tx})$

$$\mathbb{E}(x^n) = \left(\frac{d^n}{dt^n} M_x(t)\right)\Bigg|_{t \to 0}$$

$M_x(t)$ - Moment generating function (m.g.f.)

$$M_x(t) = \int e^{tx} \cdot p(x)\, dx$$

← Laplace transform

$t \to it$

$i^2 = -1$

$$\underbrace{\int e^{itx} p(x) dx}_{\text{Fourier transform}} = \varphi(t)$$

$\varphi(t)$ – characteristic function

$$X \sim N(\mu, \sigma) \longrightarrow M_X(t) = e^{t\mu + \frac{1}{2}\sigma^2 t^2}$$

$$\varphi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$$

<u>Problem 2:</u>

$X \sim N(0, \sigma)$

$$\mathbb{E}(X^n) = \left(\frac{d}{dt}\right)^n e^{\frac{1}{2}\sigma^2 t^2}$$

$n = 2k$

$$\frac{d}{dt} e^{\cdots} = e^{\cdots} \cdot \cancel{\tfrac{1}{2}}\sigma^2 \cancel{2} t = 0$$

$t \to 0$

$$\left(\frac{d}{dt}\right)^2 e^{\cdots} = \cancel{e^{\cdots}\sigma^2 t} + e^{\cdots}\sigma^2 = \sigma^2$$

$t \to 0$

Hint: $\int u(x)\underbrace{v'(x)dx}_{dv} = u(x)v(x) - \int u'(x) \cdot v(x)dx$

$$\int u\,dv = uv - \int v\,du$$

$$\int x^n e^{-\frac{x^2}{2\sigma^2}} dx =$$

$$\Big(\underbrace{e^{-\frac{x^2}{2\sigma^2}}}_{v}\Big)' = e^{-\frac{x^2}{2\sigma^2}} \frac{-2x}{2\sigma^2} = -\frac{1}{\sigma^2} x e^{\cdots}$$

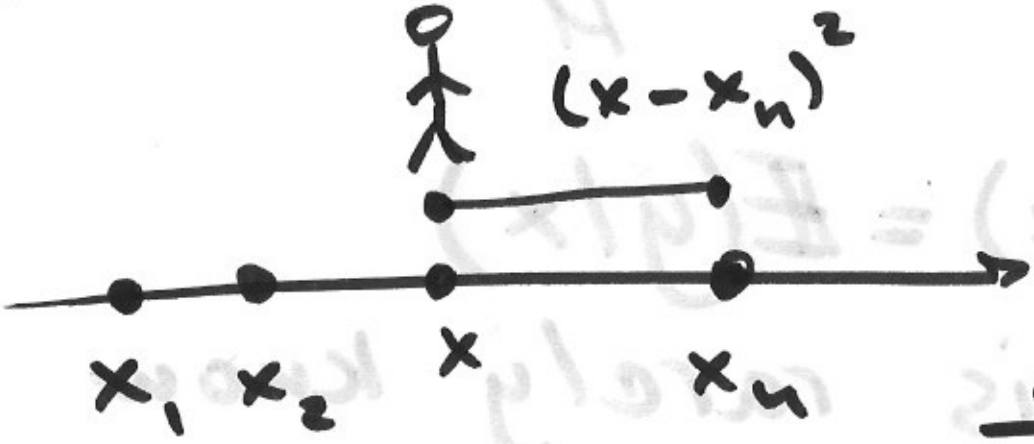
$$= -\sigma^2 \int \underbrace{x^{n-1}}_{u} \cdot \underbrace{\left(-\frac{1}{\sigma^2} x e^{-\frac{x^2}{2\sigma^2}}\right) dx}_{d e^{-\frac{x^2}{2\sigma^2}}} = \qquad \int_{-\infty}^{+\infty}$$

$$= -\sigma^2 \left(x^{n-1} \cdot e^{-\frac{x^2}{2\sigma^2}} - \int (n-1) x^{n-2} e^{-\frac{x^2}{2\sigma^2}} dx \right)$$

$$\int_{-\infty}^{+\infty} x^n \ldots dx = \sigma^2 (n-1) \cdot \int x^{n-2} \ldots dx$$

$$x^{n-1} e^{-\frac{x^2}{2\sigma^2}} \Big|_{-\infty}^{+\infty} = 0 - 0 = 0$$

Conditional expectation



$*\ (\ldots)^2 \to |\ldots|$

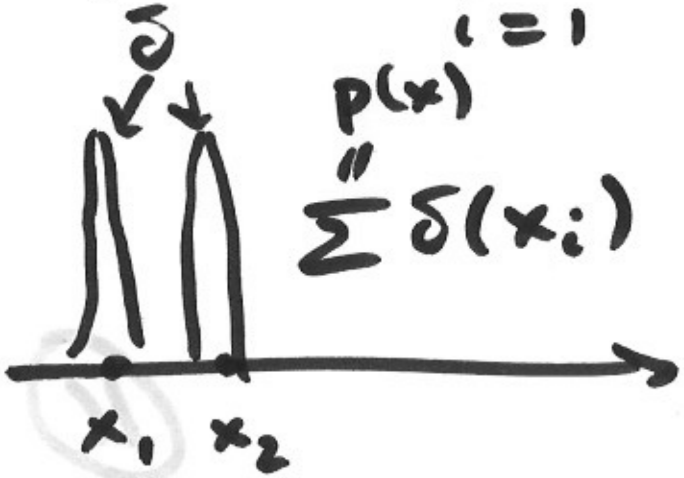
Laplace x|x

$$\tilde{x} = \arg\min_x \underbrace{\sum_{i=1}^{n} (x-x_i)^2}_{f(x)}$$

$$f'(x) = \sum_{i=1}^{n} 2(x-x_i) = 2\sum_i x - 2\sum x_i =$$

$$= 2\left(nx - \sum x_i\right) = 0$$

$$\tilde{x} = \frac{1}{n}\sum x_i$$



$p(x) = \sum \delta(x_i)$

$$\mathbb{E}_{p(x)}(x) = \arg\min_x \int (x-y)^2 p(y)\,dy$$

Conditional expectation

X - observation, $(x_1, x_2, \dots)$
$Y \leftarrow$ predict

Idea! predictor is just some function of x

$$\tilde{y} = H(x)$$

$$\mathbb{E}(y - H(x))^2 \to \min_H$$

~~argmin~~ $\tilde{H}(x) = \mathbb{E}(y|x)$

$p(x,y) \leftarrow$ is rarely know

Idea! restrict $H(x)$ to some class of functions:

For example, $H(x) = a + bx$

$$\mathbb{E}(y - H(x))^2 = f(a, b)$$

$$\tilde{b} = \frac{\mathbb{E}((x - \mathbb{E}x)(y - \mathbb{E}y))}{\operatorname{Var}(x)}$$

$\uparrow$

$\operatorname{Cov}(x, y)$

covariance

$$\tilde{a} = \mathbb{E}y - \tilde{b}\,\mathbb{E}(x)$$

$$\tilde{y} = \tilde{a} + \tilde{b} \cdot x$$

$$\mathbb{E}(y - \tilde{y})^2 = \operatorname{Var}(y) - \frac{(\operatorname{Cov}(x,y))^2}{\operatorname{Var}(x)}$$

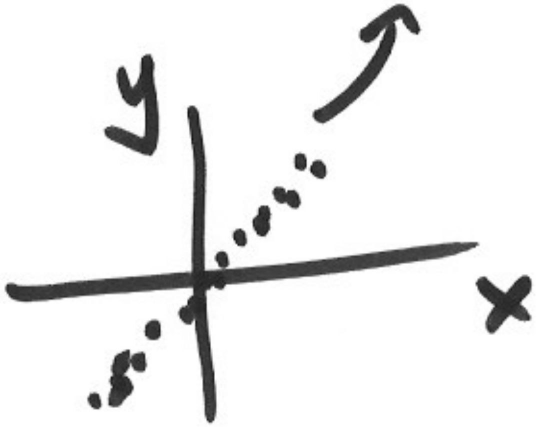
• if x, y - independent

$\leftarrow \mathbb{E}(x \cdot y) = (\mathbb{E}x) \cdot (\mathbb{E}y)$

$\rightarrow \mathrm{Cov}(x, y) = 0$

$\tilde{y} = \mathbb{E}(y) \qquad \mathbb{E}(y - \tilde{y})^2 = \mathrm{Var}(y)$

• $(\mathrm{Cov}(x, y))^2 = \mathrm{Var}(x)\mathrm{Var}(y)$



$\mathbb{E}(y - \tilde{y})^2 = \underline{\underline{0}}$