MATH 180A: INTRO TO PROBABILITY
(FOR DATA SCIENCE)

www.math.ucsd.edu/~tkemp/180A

Today: § 3.4 - 3.5

Next: § 4.1 - 4.2

Midterm 1: solutions posted

Provisional grades posted tonight

Variance

3.4

Definition: The variance of a random variable X is

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

I.e. first compute $\mu := \mathbb{E}(X)$. Then apply the function $g(x) = (x-\mu)^2$ to X, and compute $\mathbb{E}(g(X))$.

* If X is discrete, $\mathrm{Var}(X) =$

* If X is continuous, $\mathrm{Var}(X) =$

In any case, $\mathrm{Var}(X) \geq 0$. Its square root is

$\sigma(X) = \sqrt{\mathrm{Var}(X)}$ is called standard deviation.

Eg. If $X \sim \text{Ber}(p)$, $\mathbb{E}(X) = p$, $\therefore$

Eg. If $U \sim \text{Unif}([a,b])$, $\mathbb{E}(U) = \frac{a+b}{2}$, $\therefore$

Variance is a measure of how "spread out from the mean" the distribution is. For example:

Theorem: Let X be a random variable with finite expectation $\mathbb{E}(X) = \mu$. Then

$$\mathrm{Var}(X) = 0 \quad \text{iff} \quad \mathbb{P}(X = \mu) = 1.$$

(i.e. $X \equiv \mu$.)

Pf. ($\Longleftarrow$)

($\Longrightarrow$) [Here we will assume X is discrete, for now.]

Chebyshev's Inequality

If X has finite mean $\mathbb{E}(X) = \mu$ and finite variance $Var(X) = \sigma^2$, then

$$\mathbb{P}(|X-\mu| \geq k\sigma)$$

Two ideas in the proof:

(1) Indicators. $\mathbb{1}_A = \begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{if } A \text{ does not} \end{cases}$

$\mathbb{1}_A$ is a 0-1 valued r.v. It is a $Ber(p)$ with $p =$

$\therefore \mathbb{E}(\mathbb{1}_A) =$

(2) Monotonicity. FACT: if $X \leq Y$ then

(Intuitive; for proof, we must wait until week 8.)

Proof of Chebyshev:

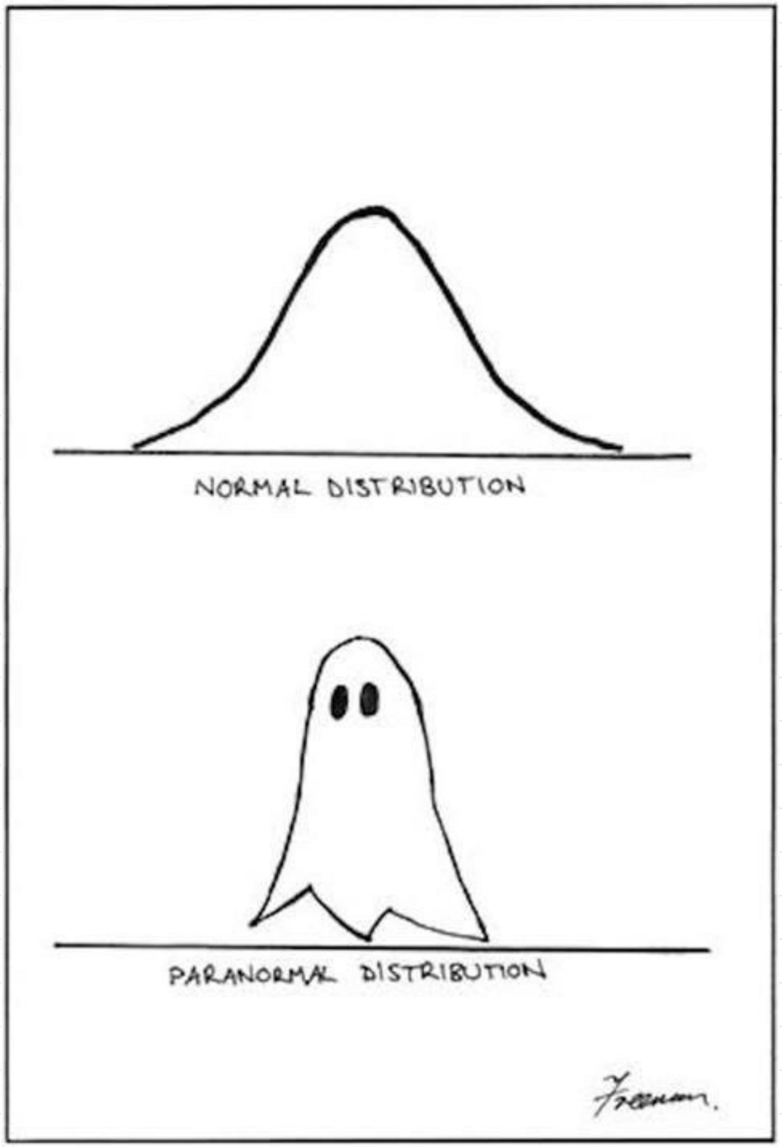
Normal (Gaussian) Distribution

3.5

The standard normal distribution $\mathcal{N}(0,1)$ is given by the density

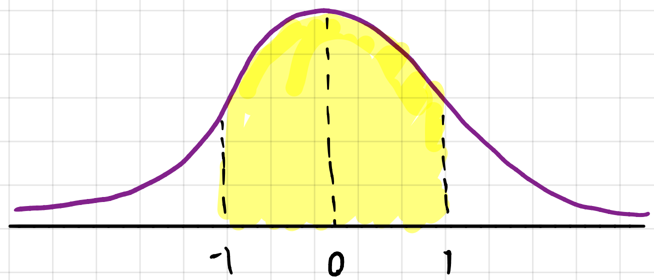
$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

Check:



CDF of $\mathcal{N}(0,1)$

Suppose $X \sim \mathcal{N}(0,1)$. What is $\mathbb{P}(|X| \leq 1)$?



$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$$

Mean and Variance $\quad X \sim \mathcal{N}(0,1)$

$\mathbb{E}(X) =$

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

General Normal $\mathcal{N}(\mu, \sigma^2)$

Let $X \sim \mathcal{N}(0,1)$. For $\sigma > 0$, $\mu \in \mathbb{R}$, let $Y = \sigma X + \mu$.

$$\mathbb{P}(Y \leq y) = \mathbb{P}(\sigma X + \mu \leq y)$$

$$\therefore \quad f_Y(y) = \frac{d}{dy} \mathbb{P}(Y \leq y)$$

Fact: If $a, b \in \mathbb{R}$, $\mathbb{E}(aX+b) = a\mathbb{E}(X)+b$ & $Var(aX+b) = a^2 Var(X)$

$\mathbb{E}(\sigma X + \mu) = \sigma \mathbb{E}(X) + \mu = 0 + \mu = \mu$. $\quad Var(\sigma X + \mu) = \sigma^2 Var(X) = \sigma^2$

$$\therefore \quad \mathbb{P}(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{(Chebyshev)}$$