

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

[www.math.ucsd.edu/~tkemp/180A](http://www.math.ucsd.edu/~tkemp/180A)

Today: § 3.5, 4.1

Next: § 4.2-4.3

HW3 grades released; regrade requests Tue 10/29 8am-11pm

Midterm grades released; regrade requests Wed 10/30 8am-11pm

Lab 3 grades released; regrade requests Thu 10/31 8am-11pm

HW4 now posted; due next Friday (Nov 1) by 11:59pm.

# Standard Normal / Gaussian $\mathcal{N}(0, 1)$

Probability density

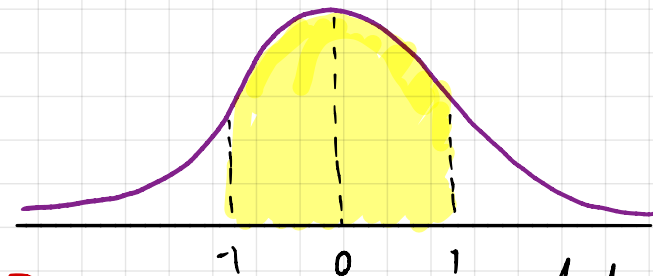
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$f(x)f(y) = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

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

$$\mathbb{P}(|X| \leq 1) = \mathbb{P}(-1 \leq X \leq 1)$$

$$= \int_{-1}^1 f(x) dx = \int_{-1}^1 \int_{-1}^1 f(x)f(y) dx dy$$



CDF  
of the Gaussian

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

$$\text{"Erf}(x) = \int_{-\infty}^x \frac{1}{\sqrt{\pi}} e^{-t^2} dt = \Phi(\sqrt{2}x)$$

$(\Phi(x) = \text{Erf}(x/\sqrt{2}))$

$$\mathbb{P}(|X| \leq 1) = \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1$$

$$= 2(0.8413) - 1 = 1.6826 - 1 = 68.26\%$$



# Mean and Variance

$$X \sim \mathcal{N}(0, 1)$$

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} t f_X(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-t^2/2} dt$$

$$\frac{d}{dt} (-e^{-t^2/2}) = t e^{-t^2/2}$$

$$= \lim_{r \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-r}^r t e^{-t^2/2} dt = \lim_{r \rightarrow \infty} \frac{1}{\sqrt{2\pi}} (-e^{-t^2/2}) \Big|_{-r}^r$$

$$= \lim_{r \rightarrow \infty} 0 = 0$$

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = 1$$

$$= \mathbb{E}(X^2) = \left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt$$

$$\int_{-\infty}^{\infty} t \cdot t e^{-t^2/2} dt = \underbrace{u}_{t} \underbrace{dv}_{t e^{-t^2/2}} = uv \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v du = \underbrace{-te^{-t^2/2}}_0 \Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} -e^{-t^2/2} dt = \int_{-\infty}^{\infty} e^{-t^2/2} dt = \sqrt{2\pi}$$

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

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

$$\mathbb{P}(Y \leq t) = \mathbb{P}(\sigma X + \mu \leq t) = \mathbb{P}(\sigma X \leq t - \mu) = \mathbb{P}(X \leq (t - \mu)/\sigma)$$

$$\begin{aligned} \therefore f_Y(t) &= \frac{d}{dt} \mathbb{P}(Y \leq t) = \Phi' \left( \frac{t - \mu}{\sigma} \right) \cdot \frac{1}{\sigma} \\ &= \frac{1}{\sigma} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(t - \mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t - \mu)^2}{2\sigma^2}} \end{aligned}$$

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

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

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

$$\begin{aligned} \mathbb{P}\left(\left|\frac{Y - \mu}{\sigma}\right| \geq k\right) &= \mathbb{P}(|X| \geq k) = \int_{-k}^{-1} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt + \int_1^k \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &= 2 \int_k^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \end{aligned}$$

$\leq \frac{1}{k} e^{-k^2/2}$

Why should I care about normal distributions?

4.1

We've already seen one scaling limit: if  $S_n \sim \text{Bin}(n, p)$   
 $p = \lambda/n$

"Poisson Approx."

$$\Rightarrow \lim_{n \rightarrow \infty} P(S_n = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

This is for rare events.

But what if we are sampling trials where success is not so rare?

E.g. Toss a fair coin 500 times. What is the probability that the number of heads is between 240 and 260?

# Heads =  $S \sim \text{Bin}(500, \frac{1}{2})$

$$P(240 \leq S \leq 260) = \sum_{k=240}^{260} \binom{500}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{500-k} = 65.23\%$$

Here is a plot of the probability mass function of the  $\text{Bin}(500, \frac{1}{2})$  distribution.

It has a very distinct **bell curve** shape.

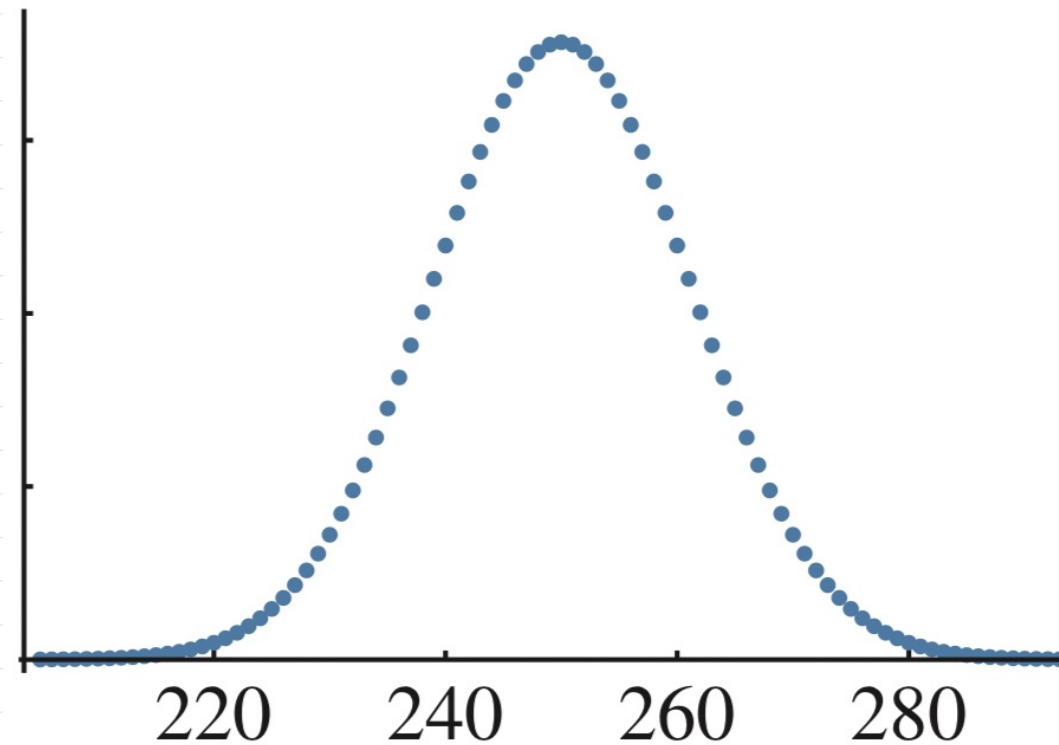
This is no accident:  
as  $n \rightarrow \infty$ , for fixed  $p$ ,

$\text{Bin}(n, p)$  approximates  
a normal distribution!

$$S_n \sim \text{Bin}(n, p)$$

$$\hookrightarrow \mathbb{E}(S_n) = np.$$

$$\hookrightarrow \text{Var}(S_n) = np(1-p)$$



Which one? Determined by  
mean and variance.

### Vague Theorem

For  $n$  large and  $p$  not close to  
0 or 1,

$$\text{Bin}(n, p) \approx \mathcal{N}(np, np(1-p))$$

# Binomial Central Limit Theorem

Fix  $p \in (0, 1)$ . For each  $n$ , let  $S_n \sim \text{Bin}(n, p)$ .

For any fixed  $a \leq b$ ,

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Eg. Toss a fair coin 500 times. What is the probability that the number of heads is between 240 and 260?

$$S \sim \text{Bin}(500, \frac{1}{2})$$

$$\mathbb{E}(S) = 500 \cdot \frac{1}{2} = 250$$

$$\text{Var}(S) = 500 \cdot \frac{1}{2} \cdot \frac{1}{2} = 125$$

$$P(240 \leq S \leq 260)$$

$$= P(-10 \leq S - 250 \leq 10) = P\left(\frac{-10}{\sqrt{125}} \leq \frac{S - 250}{\sqrt{125}} \leq \frac{10}{\sqrt{125}}\right)$$

$$\approx \Phi\left(\frac{10}{\sqrt{125}}\right) - \Phi\left(\frac{-10}{\sqrt{125}}\right) = 62.89\%$$