Math 180A: Introduction to Probability

Lecture B00 (Nemish)

math.ucsd.edu/~ynemish/teaching/180a

Lecture C00 (Au)

math.ucsd.edu/~bau/f20.180a

Today: ASV 3.5 (Gaussian distribution)

ASV 4.1 (Normal approximation)

Video: Prof. Todd Kemp, Fall 2019

Next: ASV 4.2, 4.3

Week 5: Quiz 3 (Wednesday, Nov 4)

Homework 4 (due Friday, Nov 6)

Regrades for Homework 3 (Nov 2-3)

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

Probability density

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

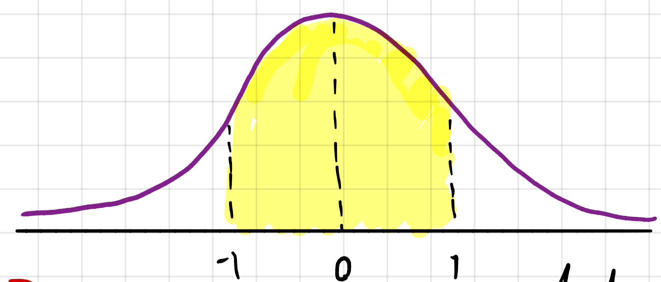
$$f(x)f(y) = \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

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

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

$$= \int_{-1}^{1} f(x)\,dx \quad \cancel{= \left(\int_{-1}^{1} f(x)\,dx \int_{-1}^{1} f(y)\,dy\right)^{1/2}} \quad \cancel{= \iint_{-1\,-1}^{1\;\;1} \frac{1}{2\pi} e^{-(x^2+y^2)/2}\, \overbrace{dx\,dy}^{r\,dr\,d\theta}}$$

in polar $e^{-r^2/2}$



CDF of the Gaussian

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

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

$$\left(\Phi(x) = \mathrm{Erf}(x/\sqrt{2})\right)$$

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

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

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

Normal Table of Values

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

This standard table (Appendix E in textbook) lists the values of

Eg. $\Phi(1.56) \doteq 0.9406$

Fact: $\Phi(-x) = 1 - \Phi(x)$

$$\Phi(-x) = \int_{-\infty}^{-x} f(t)\,dt \qquad s=-t,\ ds=-dt$$

$$= -\int_{\infty}^{x} f(-s)\,ds$$

$$= \int_{x}^{\infty} f(-s)\,ds$$

$$= \int_{-\infty}^{\infty} f(-s)\,ds - \int_{-\infty}^{x} f(-s)\,ds$$

$$= 1 - \int_{-\infty}^{x} f(s)\,ds = 1 - \Phi(x)$$

$\Phi(-1)$
$= 1 - \Phi(1)$
$= 1 - 0.8413$
$= 0.1587$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Mean and Variance

$X \sim N(0,1)$

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

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

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

$$= \lim_{r\to\infty} 0 = 0.$$

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

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

$$\int_{-\infty}^{\infty} \underbrace{t}_{u} \cdot \underbrace{te^{-t^2/2} dt}_{dv} = uv\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} v\, du = \underbrace{-te^{-t^2/2}\Big|_{-\infty}^{\infty}}_{0} - \underbrace{\int_{-\infty}^{\infty} -e^{-t^2/2} dt}_{\int_{-\infty}^{\infty} e^{-t^2/2} dt}$$

$$v = -e^{-t^2/2}$$

$$= \sqrt{2\pi}$$

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

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

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

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

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

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

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

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

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

Why should I care about normal distributions?

4.1

We've already seen one scaling limit: if $S_n \sim Bin(n,p)$

"Poisson Approx."

$p = \lambda/n$

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

This is for rare events.

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

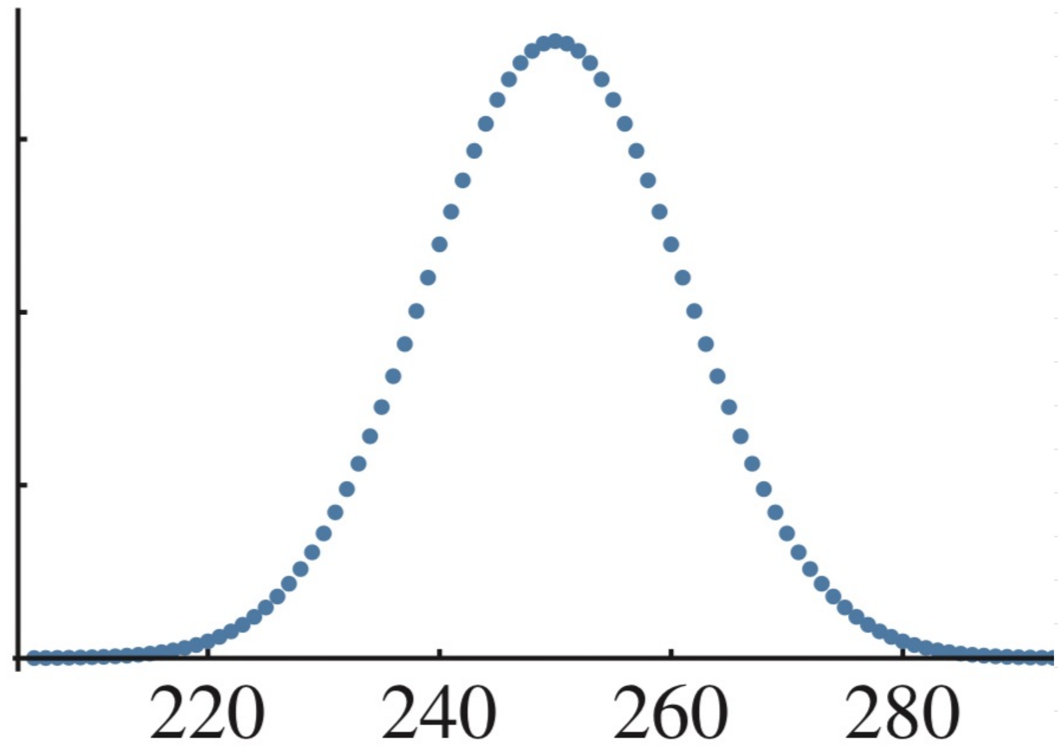
Eg. Toss a fair coin 500 times. What is the probability that the number of heads is between 240 and 260?

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

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

$\doteq 65.23\%$

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



It has a very distinct bell curve shape.

This is no accident: as $n \to \infty$, for fixed p, $\mathrm{Bin}(n,p)$ approximates a normal distribution!

Which one? Determined by mean and variance.

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

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

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

Vague Theorem

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

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

Binomial Central Limit Theorem

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

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

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

$$S \sim \text{Bin}(500, \tfrac{1}{2}) \qquad \begin{aligned} \mathbb{E}(S) &= 500 \cdot \tfrac{1}{2} = 250 \\ \text{Var}(S) &= 500 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2} = 125 \end{aligned}$$

$$\mathbb{P}(240 \le S \le 260)$$

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

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