CSE 312

Foundations of Computing II

17: Normal Distribution & Central Limit Theorem

Anonymous questions: www.slido.com/6995617

Review Continuous RVs

Probability Density Function (PDF).

$f: \mathbb{R} \to \mathbb{R}$ s.t.

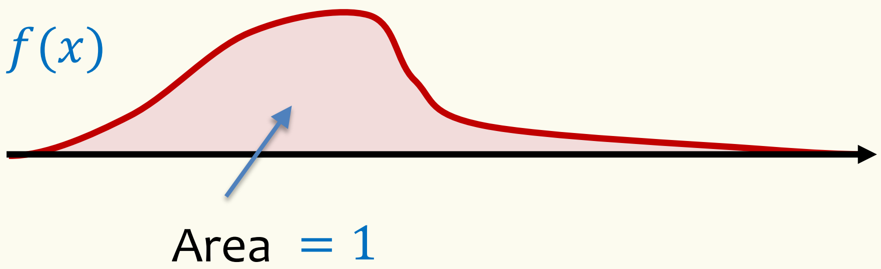
- $f(x) \geq 0$ for all $x \in \mathbb{R}$
- $\int_{-\infty}^{+\infty} f(x)\, \mathrm{d}x = 1$

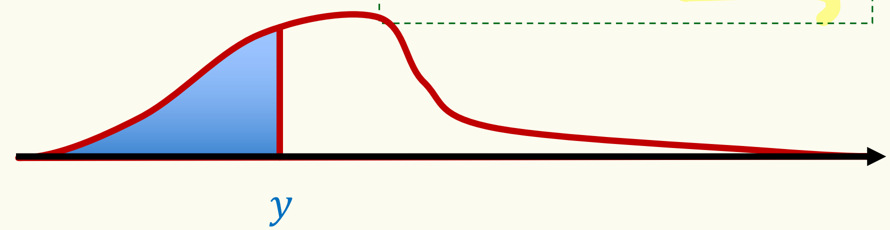
Cumulative Distribution Function (CDF).

$P(X \leq y) = F(y) = \int_{-\infty}^{y} f(x)\, \mathrm{d}x$

$(< y)$

Theorem. $f(x) = \frac{dF(x)}{dx}$

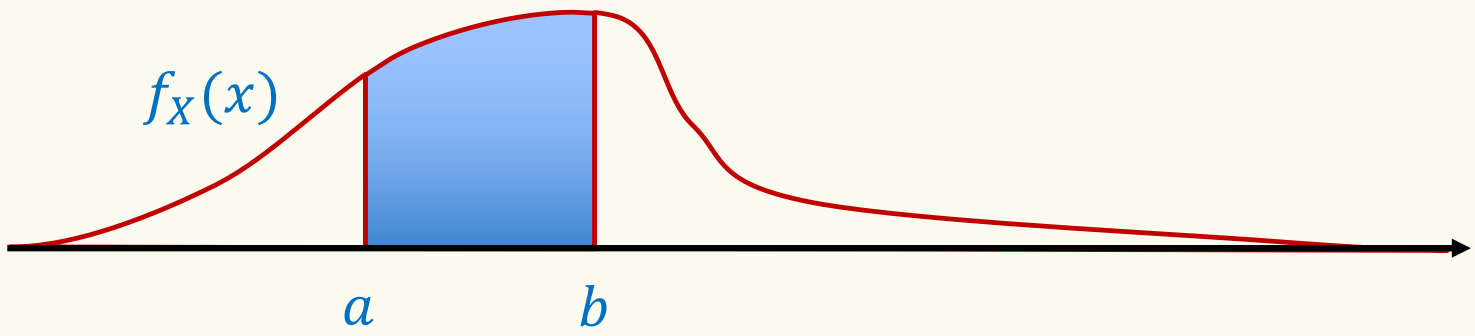




Density ≠ Probability !

$F_X(y) = P(X \leq y)$

Review Continuous RVs



$$P(X \in [a,b]) = \int_a^b f_X(x)\mathrm{d}x = F_X(b) - F_X(a)$$

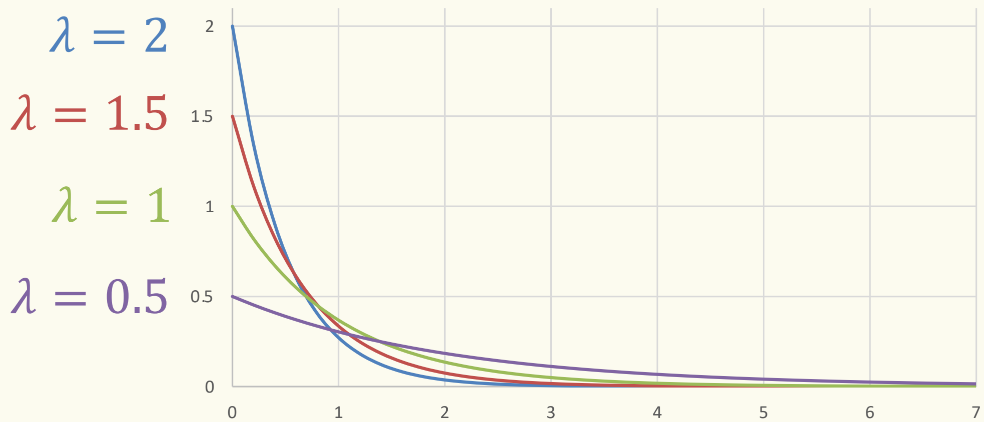
Review Exponential Distribution

Definition. An **exponential random variable** X with parameter $\lambda \geq 0$ is follows the exponential density

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

We write $X \sim \text{Exp}(\lambda)$ and say X that follows the exponential distribution.

CDF: For $y \geq 0$,
$F_X(y) = 1 - e^{-\lambda y}$



Agenda

- Normal Distribution
- Practice with Normals
- Central Limit Theorem (CLT)

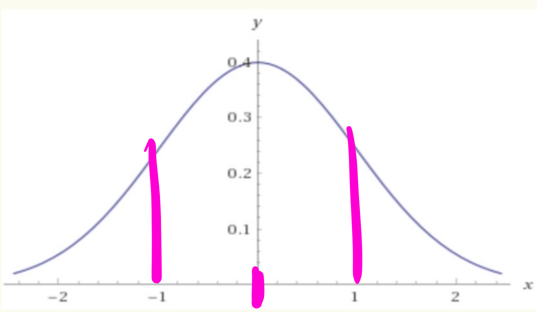
The Normal Distribution

Carl Friedrich Gauss

Definition. A **Gaussian (or <u>normal</u>) random variable** with parameters $\mu \in \mathbb{R}$ and $\sigma \geq 0$ has density

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

We say that X follows the Normal Distribution, and write $X \sim \mathcal{N}(\mu, \sigma^2)$.



$\mathcal{N}(0, 1)$.

No closed form expression for CDF…

$$F(x) = \int_{-\infty}^{x} f_X(y)\, dy$$

The Normal Distribution. $X \sim \mathcal{N}(\mu, \sigma^2)$

Carl Friedrich Gauss

Definition. A **Gaussian (or <u>normal</u>) random variable** X with parameters $\mu \in \mathbb{R}$ and $\sigma \geq 0$ has density

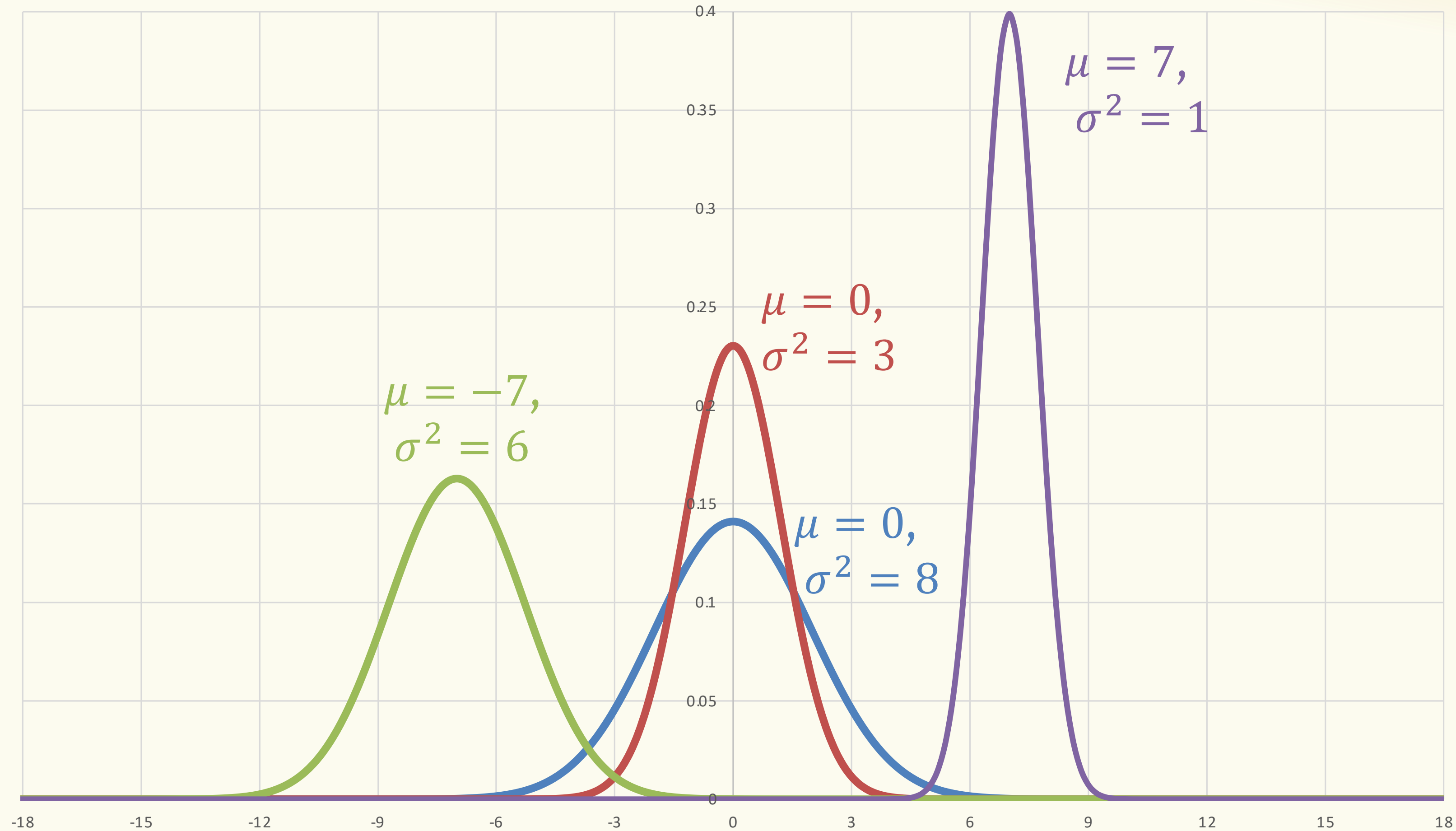
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Fact. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathbb{E}[X] = \mu$, and $\mathrm{Var}(X) = \sigma^2$

Proof of expectation is easy because density curve is symmetric around μ, $f_X(\mu - x) = f_X(\mu + x)$, but proof for variance requires integration of $e^{-x^2/2}$

The Normal Distribution

Aka a “Bell Curve” (imprecise name)



Standard normal distribution

Standard (unit) normal = $\mathcal{N}(0, 1)$

CDF. $\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} \mathrm{d}x$ for $Z \sim \mathcal{N}(0, 1)$

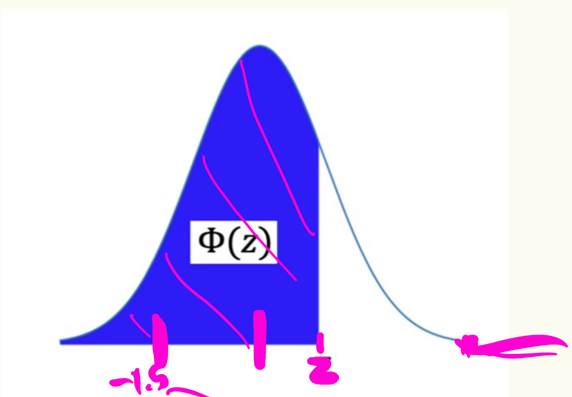
Note: $\Phi(z)$ has no closed form – generally given via tables

z=0.00
3.09

Table of Standard Cumulative Normal Density $\mathcal{N}(0,1)$

$P(Z \le 0.98) = \Phi(0.98) \approx 0.8365$

$P(Z \le 1) = \Phi(1.00) \approx 0.84134$



$\Phi(z) = P(Z \le z)$

N(0,1)

Φ Table: $\mathbb{P}(Z \le z)$ when $Z \sim \mathcal{N}(0,1)$

0.42

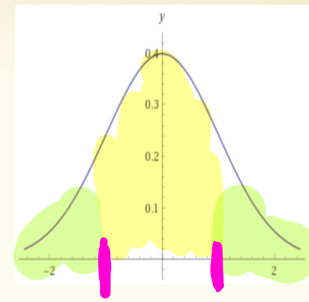
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

$\sigma^2 = 1$

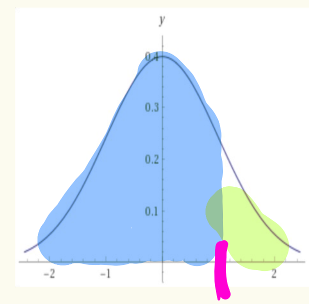
$\sigma = 1$

The Standard Normal CDF

$P(Z \leq 1) = \Phi(1.00) \approx 0.84$



What is the probability that a standard Normal is within one standard deviation of its mean?



$P(-1 \leq Z \leq 1) = P(Z \leq 1) - P(Z \leq -1)$

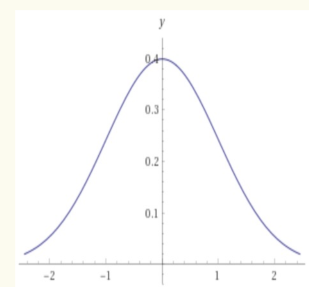
$= P(Z \leq 1) - P(Z > 1)$

$= P(Z \leq 1) - (1 - P(Z \leq 1))$

$= 2P(Z \leq 1) - 1$

$= 2\Phi(1) - 1 = 2 \cdot 0.84 - 1$

$= 0.68$



What happens when the mean/variance are not 0/1?

Turns out that by shifting and scaling, we can ``convert'' any normal distribution into a standard normal (for purposes of calculation and lookup in the tables).

a, b constants.

For a moment, suppose X is any r.v. with mean μ and variance σ^2 and $Y = aX + b$.

What are the mean and variance of Y?

Mean of Y: $E(Y) = E(aX+b) = aE(X) + b = a\mu + b$

Variance of Y: $Var(Y) = Var(aX+b) = Var(aX) = a^2 Var(X)$

$\frac{X - 3.15}{5}$ $\frac{X - \mu_X}{\sigma_X}$

Closure of normal distribution – Under Shifting and Scaling

Fact. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = aX + b$ also normally distributed.

$Y \sim N(a\mu + b, a^2\sigma^2)$

Mean of Y: $a\mu + b$

Variance of Y: $a^2\sigma^2$

In particular: $\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$

$\sigma^2 > 0$

Closure of normal distribution – Under Shifting and Scaling

Fact. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Mean and variance immediate. The fact that Y is still normal is not obvious, but can show with algebra that the PDF of $Y = aX + b$ is still normal.

Useful for computing $F_X(z) = P(X \le z)$

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

$$P(X \le z) = P\left(\frac{X-\mu}{\sigma} \le \frac{z-\mu}{\sigma}\right)$$

$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$

$$= P\left(Z \le \frac{z-\mu}{\sigma}\right) = \Phi\left(\frac{z-\mu}{\sigma}\right)$$

$N(0,1)$

Closure of normal distribution – Under Shifting and Scaling

Fact. If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Y = aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

Mean and variance immediate. The fact that Y is still normal is not obvious, but can show with algebra that the PDF of $Y = aX + b$ is still normal.

Useful because

If $X \sim \mathcal{N}(\mu, \sigma^2)$, we compute $F_X(z) = P(X \leq z)$ as follows:

$$F_X(z) = P(X \leq z) = P\left(\frac{X - \mu}{\sigma} \leq \frac{z - \mu}{\sigma}\right) = \Phi\left(\frac{z - \mu}{\sigma}\right)$$

Agenda

- Normal Distribution
- Practice with Normals
- Central Limit Theorem (CLT)

Key observation

standardizing

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$

Therefore,

$$F_X(z) = P(X \leq z) = P\left(\frac{X-\mu}{\sigma} \leq \frac{z-\mu}{\sigma}\right) = \Phi\left(\frac{z-\mu}{\sigma}\right)$$

And can look up the value in the standard normal table.

Example

Let $X \sim \mathcal{N}(0.4, 4 = 2^2)$.

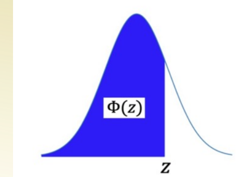
$P(X \leq 1.2) = P\left(\frac{X - 0.4}{2} \leq \frac{1.2 - 0.4}{2}\right)$

$\frac{X-0.4}{2} \sim N(0,1)$, $\quad \frac{1.2-0.4}{2} = 0.4$

$= P(Z \leq 0.4) = \Phi(0.4)$

$= 0.65542$

Table of Standard Cumulative Normal Density



Φ Table: $\mathbb{P}(Z \leq z)$ when $Z \sim \mathcal{N}(0,1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

Example

Let $X \sim \mathcal{N}(0.4, 4 = 2^2)$.

$$P(X \leq 1.2) = P\left(\frac{X - 0.4}{2} \leq \frac{1.2 - 0.4}{2}\right)$$

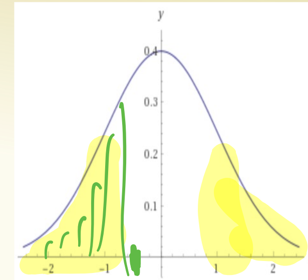
$$= P\left(\frac{X - 0.4}{2} \leq 0.4\right) = \Phi(0.4) \approx 0.6554$$

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

0.1	0.5398	0.5438
0.2	0.5793	0.5832
0.3	0.6179	0.6217
0.4	0.6554	0.6591
0.5	0.6915	0.6950
0.6	0.7257	0.7291
0.7	0.7580	0.7611

Example

Let $X \sim \mathcal{N}(3, 16)$.



$P(2 < X < 5) =$ $\Pr\left(\frac{2-3}{4} \le \frac{X-3}{4} \le \frac{5-3}{4}\right)$

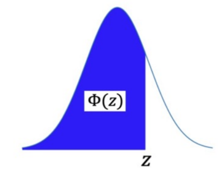
$\frac{2-3}{4} \to -\frac{1}{4}$, $\quad \frac{X-3}{4} \sim N(0,1)$, $\quad \frac{5-3}{4} = \frac{1}{2}$

$= P\left(-\frac{1}{4} \le Z \le \frac{1}{2}\right) = \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{4}\right)$

$= \Phi\left(\frac{1}{2}\right) - \left(1 - \Phi\left(\frac{1}{4}\right)\right)$

$= \Phi\left(\frac{1}{2}\right) + \Phi\left(\frac{1}{4}\right) - 1$

Table of Standard Cumulative Normal Density



Φ Table: $\mathbb{P}(Z \leq z)$ when $Z \sim \mathcal{N}(0,1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

Example

Let $X \sim \mathcal{N}(3, 16)$.

$$P(2 < X < 5) = P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{5-3}{4}\right)$$

$$= P\left(-\frac{1}{4} < Z < \frac{1}{2}\right)$$

$$= \Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{1}{4}\right)$$

$$= \Phi\left(\frac{1}{2}\right) - \left(1 - \Phi\left(\frac{1}{4}\right)\right) \approx 0.29017$$

Example – How Many Standard Deviations Away?

Let $X \sim \mathcal{N}(\mu, \sigma^2)$.

$$P(|X - \mu| < k\sigma) = P\left(\frac{|X - \mu|}{\sigma} < k\right) =$$

$$= P\left(-k < \frac{X - \mu}{\sigma} < k\right) = \Phi(k) - \Phi(-k)$$

e.g. $k = 1$: 68%
$k = 2$: 95%
$k = 3$: 99%

Summary so far

- Normal distributions stay normal under shifting and scaling.
- To "standardize" a normal random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, you subtract the mean and divide by the standard deviation, i.e.,

$$\frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

- This allows you to use the standard normal tables (showing $\Phi(z) = P(Z \leq z)$ for $Z \sim \mathcal{N}(0, 1)$) to do calculations for any normal distribution. (Also must use symmetry of normal distributions.)

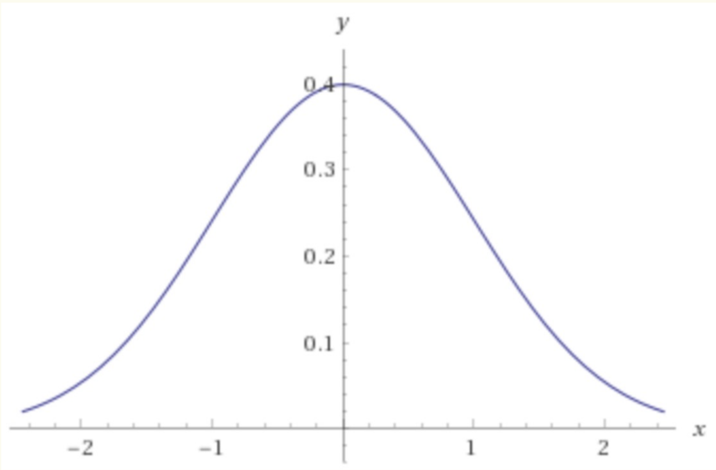
Closure of the normal -- under addition

Fact. If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ (both independent normal RV) then $aX + bY + c \sim \mathcal{N}(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$

Note: The special thing is that **the sum of normal RVs is still a normal RV.** The values of the expectation and variance are **not** surprising.

Why not surprising?

- Linearity of expectation (always true)
- When X and Y are independent, $\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$



Normal Distribution

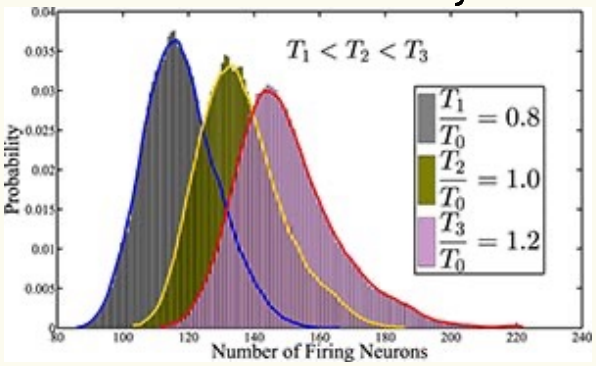
Paranormal Distribution

Agenda

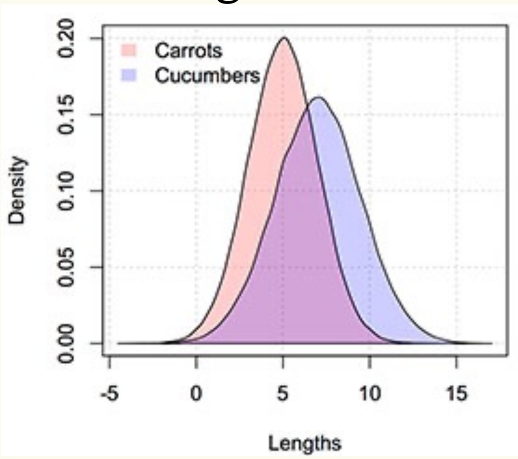
- Normal Distribution
- Practice with Normals
- Central Limit Theorem (CLT)

Normal Distributions EVERYWHERE – why?

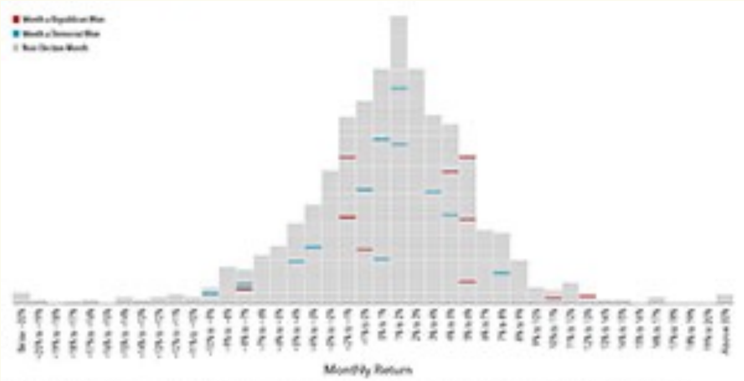
Neuron Activity



Vegetables



S&P 500 Returns after Elections



Examples from:
https://galtonboard.com/probabilityexamplesinlife

Sums of i.i.d. RVs look normal!

i.i.d. = independent and identically distributed

$X_1, \dots, X_n$ i.i.d. with expectation μ and variance σ^2

Define $S_n = X_1 + \cdots + X_n$

Empirical observation:

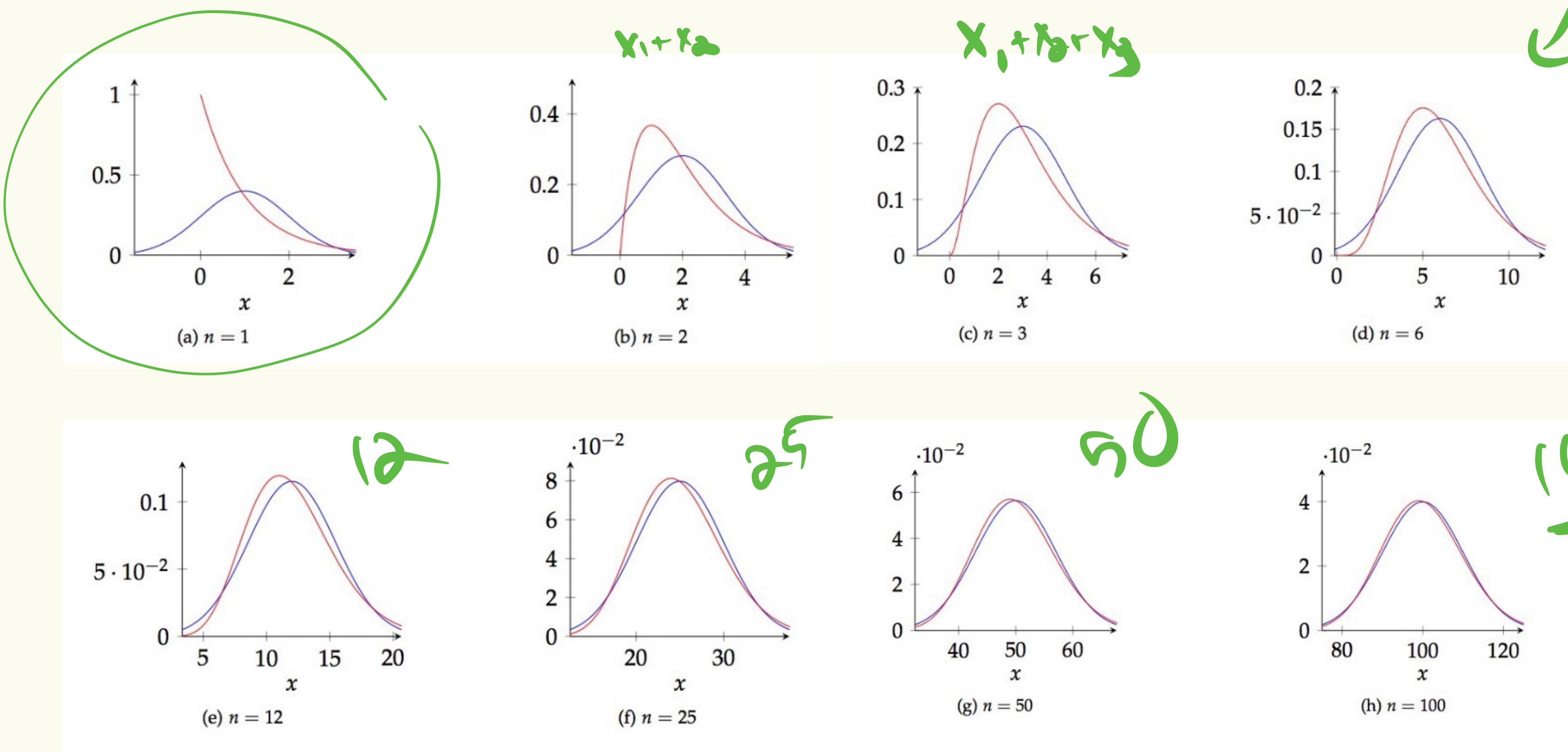
S_n looks like a normal RV as n grows.

$\mathbb{E}[S_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = n\mu$

$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n) = n\sigma^2$

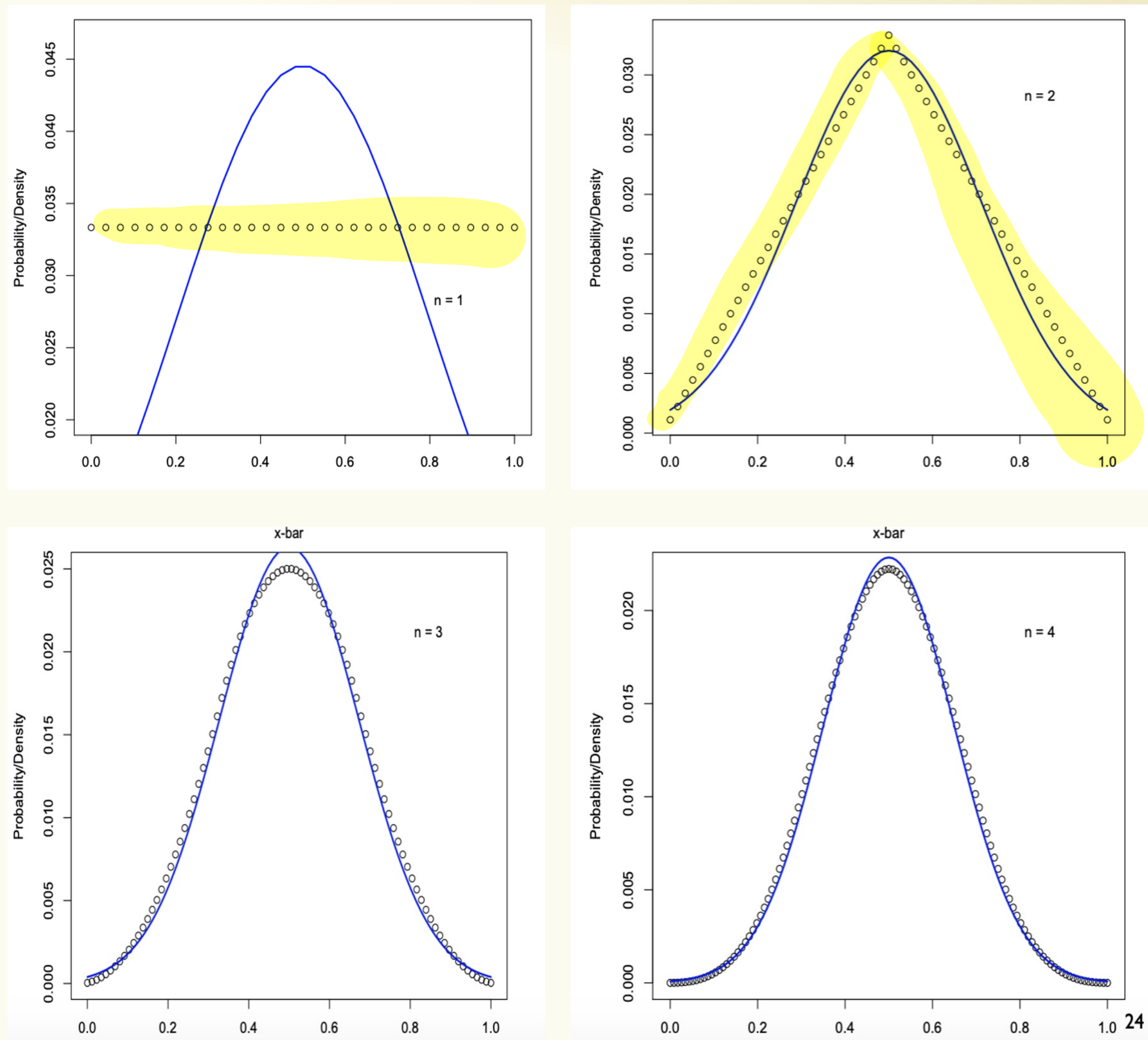
S_n look like $N(n\mu, n\sigma^2)$

Example: Sum of n i.i.d. $\mathrm{Exp}(1)$ random variables

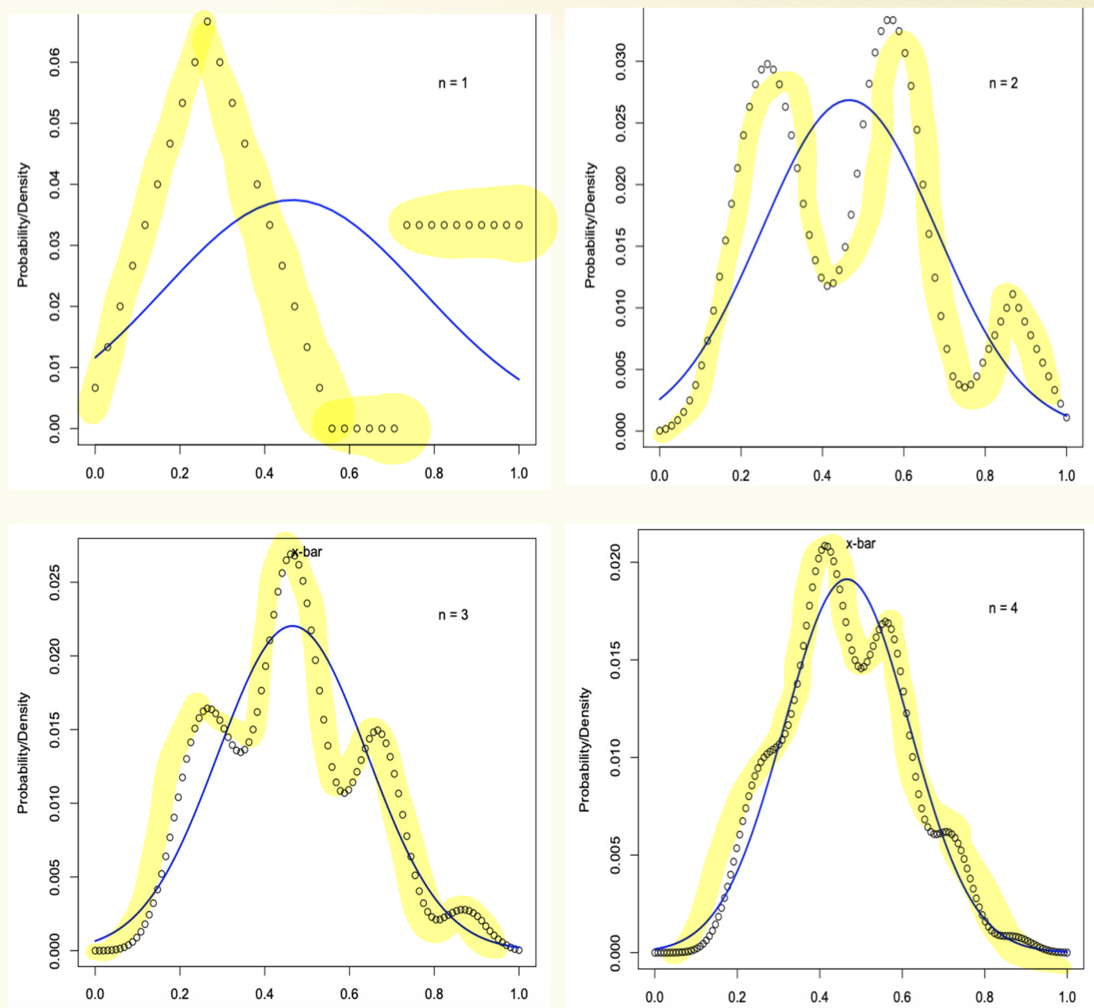


Example: sum of uniform r.v.s

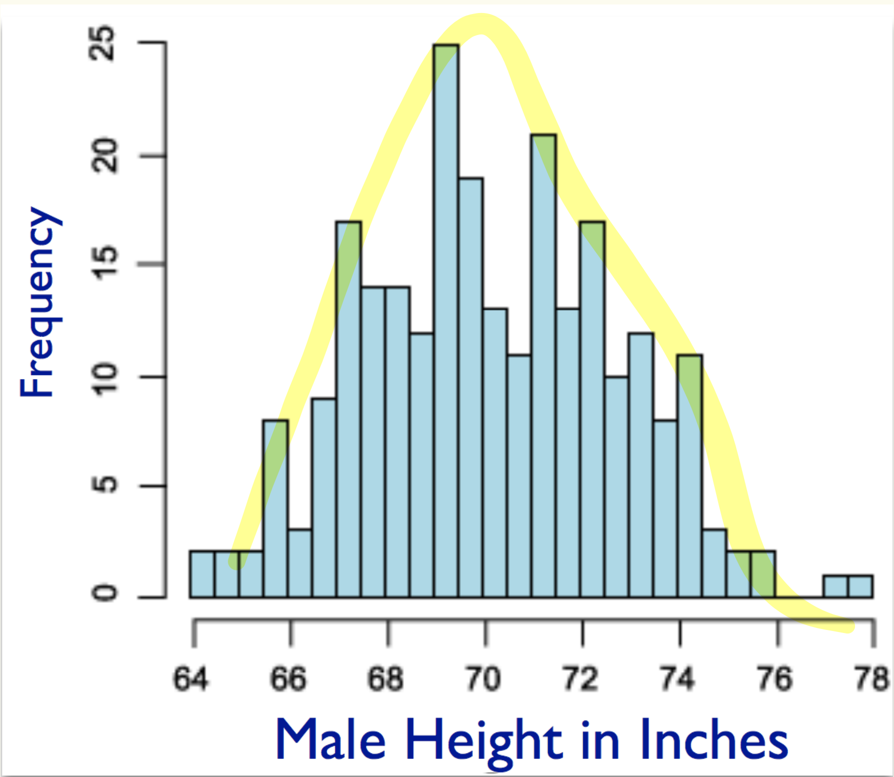
$$\frac{X_1+X_2+\dots+X_n}{n}$$



CLT : Sum of some other weird i.i.d. r.v.s



Suppose that what we see in nature results from combining (summing) many independent random observations…



Then distribution might look normal. e.g. Height distribution resembles Gaussian.

R.A.Fisher (1918) observed that the height is likely the outcome of the sum of many independent random parameters, i.e., can written as

$$X = X_1 + \cdots + X_n$$

Central Limit Theorem

$X_1, \dots, X_n$ i.i.d., each with expectation μ and variance σ^2

Define $S_n = X_1 + \cdots + X_n$,

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$E(S_n) = n\mu$

$Var(S_n) = n\sigma^2$

st dev of $S_n = \sqrt{n}\,\sigma$

Then distribution of $Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges to that of a normal distribution with mean 0 and variance 1 as $n \to \infty$.

Central Limit Theorem

$X_1, \ldots, X_n$ i.i.d., each with expectation μ and variance σ^2

Define $S_n = X_1 + \cdots + X_n$ and

$$Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$$\mathbb{E}[Y_n] = \frac{1}{\sigma\sqrt{n}}(\mathbb{E}[S_n] - n\mu) = \frac{1}{\sigma\sqrt{n}}(n\mu - n\mu) = 0$$

$$\text{Var}(Y_n) = \frac{1}{\sigma^2 n}\left(\text{Var}(S_n - n\mu)\right) = \frac{\text{Var}(S_n)}{\sigma^2 n} = \frac{\sigma^2 n}{\sigma^2 n} = 1$$

Central Limit Theorem

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Theorem. (Central Limit Theorem) The CDF of Y_n converges to the CDF of the standard normal $\mathcal{N}(0,1)$, i.e.,

$$\lim_{n\to\infty} P(Y_n \leq y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} \mathrm{d}x$$

$$= \Phi(y)$$

Central Limit Theorem

$$Y_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Theorem. (Central Limit Theorem) The CDF of Y_n converges to the CDF of the standard normal $\mathcal{N}(0,1)$, i.e.,

$$\lim_{n\to\infty} P(Y_n \leq y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-x^2/2} \mathrm{d}x$$

Also stated as:

- $\lim_{n\to\infty} Y_n \to \mathcal{N}(0,1)$
- $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i \to \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$ for $\mu = \mathbb{E}[X_i]$ and $\sigma^2 = \mathrm{Var}(X_i)$

CLT application

- Transmitting a signal. 100 sources independently add noise to signal, each Unif(-1,1). If |total noise| > 10, signal is corrupted.
- Use the CLT to estimate the approximate probability that the signal is not corrupted.

$E(X_i)=0$ $\quad Var(X_i)=\frac{1}{3}$

X_i: noise from i^{th} source. $\quad X_i \sim Unif(-1,1)$

$S_n = X_1 + X_2 + \dots + X_{100}$ $\quad E(S_n) = 100 \cdot 0 = 0$

$Var(S_n) =$

Agenda

- Polling

Magic Mushrooms

In Fall 2020, Oregonians voted on whether to legalize the therapeutic use of “magic mushrooms”.

Poll to determine the fraction p of the population expected to vote in favor.

- Call up a random sample of n people to ask their opinion
- Report the empirical fraction

Questions

- Is this a good estimate?
- How to choose n?

Polling Accuracy

Often see claims that say

"Our poll found 80% support. This poll is accurate to within 5% with 98% probability"*

Will unpack what this and how they sample enough people to know this is true.

* When it is 95% this is sometimes written as "19 times out of 20"

Formalizing Polls

Population size N, true fraction of voting in favor p, sample size n.

Problem: We don't know p, want to estimate it

Polling Procedure

for $i = 1, \ldots, n$:

1. Pick uniformly random person to call (prob: $1/N$)
2. Ask them how they will vote

$$X_i = \begin{cases} 1, & \text{voting in favor} \\ 0, & \text{otherwise} \end{cases}$$

Report our estimate of p: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Formalizing Polls

What type of r.v. is X_i?

Poll: www.slido.com/6995617

	Type	$\mathbb{E}[X_i]$	$\mathrm{Var}(X_i)$
a.	Bernoulli	p	$p(1-p)$
b.	Bernoulli	p	p^2
c.	Geometric	p	$\frac{1-p}{p^2}$
d.	Binomial	np	$np(1-p)$

Population size N, true fraction of voting in favor p, sample size n.

Problem: We don't know p

Polling Procedure

for $i = 1, \dots, n$:

1. Pick uniformly random person to call (prob: $1/N$)
2. Ask them how they will vote

$$X_i = \begin{cases} 1, & \text{voting in favor} \\ 0, & \text{otherwise} \end{cases}$$

Report our estimate of p: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

Random Variables

What type of r.v. is X_i?

	Type	$\mathbb{E}[X_i]$	$\mathrm{Var}(X_i)$
a.	Bernoulli	p	$p(1-p)$
b.	Bernoulli	p	p^2
c.	Geometric	p	$\frac{1-p}{p^2}$
d.	Binomial	np	$np(1-p)$

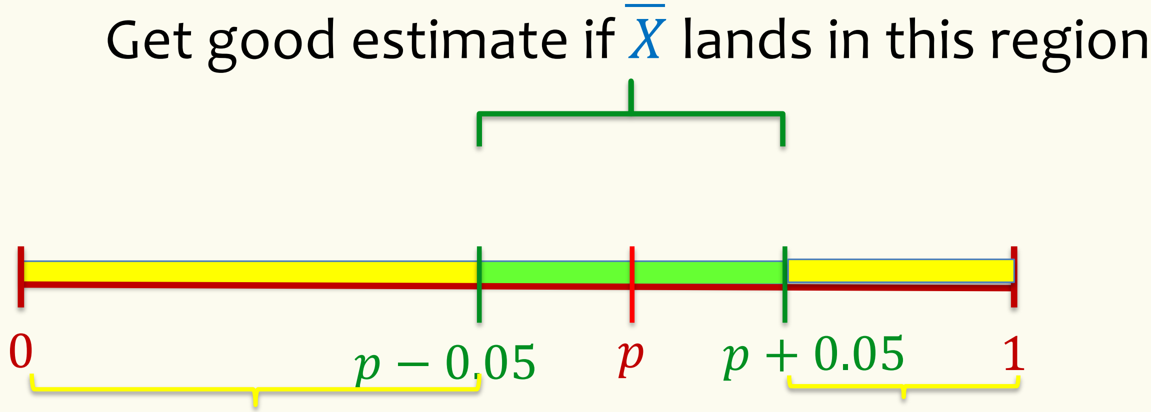
What about $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$?

Poll: www.slido.com/6995617

	$\mathbb{E}[\bar{X}]$	$\mathrm{Var}(\bar{X})$
a.	np	$np(1-p)$
b.	p	$p(1-p)$
c.	p	$p(1-p)/n$
d.	p/n	$p(1-p)/n$

Roadmap: Bounding Error

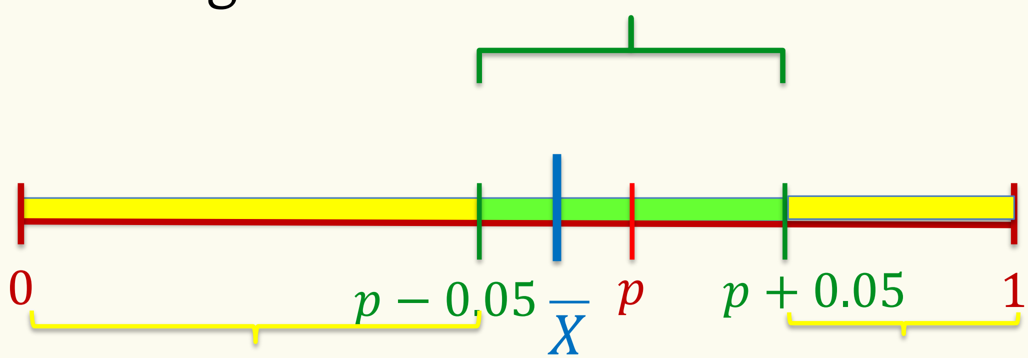
Goal: Find the value of n such that 98% of the time, the estimate $\bar{X}$ is within 5% of the true p



Roadmap: Bounding Error

Goal: Find the value of n such that 98% of the time, the estimate $\bar{X}$ is within 5% of the true p

Get good estimate if $\bar{X}$ lands in this region



Want $P\left(\left|\bar{X} - p\right| > 0.05\right) \leq 0.02$

Central Limit Theorem

www.slido.com/6995617

Poll: In the limit $\overline{X}$ is…?

a. $\mathcal{N}(0,1)$
b. $\mathcal{N}(p, p(1-p))$
c. $\mathcal{N}(p, p(1-p)/n)$
d. I don't know

With i.i.d random variables $X_1, X_2, \dots, X_n$ where $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$

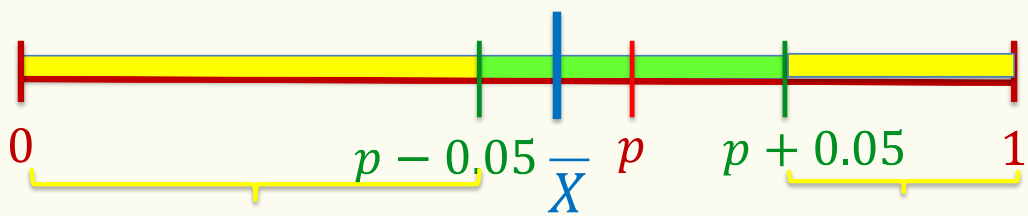
As $n \to \infty$,

$$\frac{X_1 + X_2 + \cdots X_n \; - n\mu}{\sigma\sqrt{n}} \to \mathcal{N}(0,1)$$

Restated: As $n \to \infty$,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Roadmap: Bounding Error



Want $P\left(\left|\overline{X} - p\right| > 0.05\right) \leq 0.02$