Markov's ≤

- Motivation
- Intuition
- Formulation
- Proof
- Example
- Extensions

Probability Bounds

Often Want to bound probability of (typically bad) events

Excessive rain | Heavy traffic | Large loss | Disease outbreak

Now Markov's Inequality

Later Stronger bounds

Start Intuitive definition

Markov's Meerkats

Average meerkat height is 10”

Can half the meerkats be ≥ 40” tall?

No

If half of meerkats were ≥ 40” tall average would be ≥ ½ x 40” = 20”

F_{40} fraction of meerkats that are ≥ 40” tall

If $F_{40} \cdot 40 > 10$, average would be > 10

$F_{40} \cdot 40 \leq 10$ $F_{40} \leq 10/40 = ¼$

General μ $F_{4\cdot\mu} \cdot (4\cdot\mu) \leq \mu$ $F_{4\cdot\mu} \leq ¼$

Markov's ≤

X - nonnegative discrete or continuous r.v. with finite mean μ

Two forms

Intuitive, memorable

$$\forall\ \alpha \geq 1 \qquad P(X \geq \alpha\mu) \leq \frac{1}{\alpha}$$

A nonnegative r.v. is at least α times ≥ its mean with probability ≤ α

Direct proof, easier to apply, more common

$$a = \alpha\mu \qquad \forall\ a \geq \mu \qquad P(X \geq a) \leq \frac{\mu}{a}$$

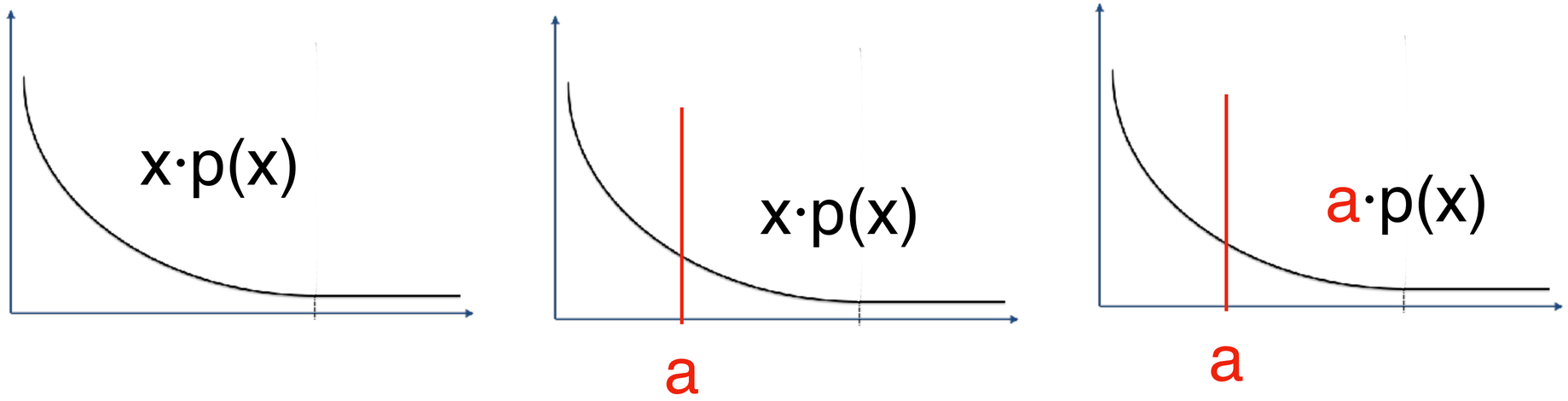
&= a\cdot P(X\ge a)

Proof

$$P(X \geq a) \leq \frac{\mu}{a}$$

Prove for discrete r.v.'s, same proof works for continuous, just $\sum \rightarrow \int$

$$\mu = \sum_x x \cdot p(x) \quad \geq \sum_{x \geq a} x \cdot p(x) \quad \geq \sum_{x \geq a} a \cdot p(x) \quad = a \cdot P(X \geq a)$$



Citation Counts

A journal paper is cited 8 times on average

Roughly right

Popular (multiple) hypothesis-testing paper

Cited ≥ 40,000 times

J. R. Statist. Soc. B (1995)
57, *No.* 1, *pp.* 289–300

Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing

By YOAV BENJAMINI† and YOSEF HOCHBERG

Bound probability that a paper gets cited ≥ 40,000 times

X: # paper citations

$X \geq 0$

$\mu = 8$

Markov

$$P(X \geq a) \leq \frac{\mu}{a}$$

$P(X \geq 40{,}000) \leq \mu / 40K = 8 / 40K = 0.02\%$

Generalize?

Can the Markov ≤ be: Generalized (conditions relaxed)?

Strengthened?

Generalization attempt: Can we remove non-negativity?

If X can be negative, then P(X ≥ a) can be close to 1 for any a

- large 0 a

$$p(x) = \begin{cases} 1-\epsilon & x = a \\ \epsilon & x = \frac{\mu-(1-\epsilon)a}{\epsilon} \end{cases}$$

$$EX = \mu \qquad p(X \geq a) = p(a) \approx 1$$ ❌

Strengthen?

Can we strengthen $P(X \geq a) \leq \frac{\mu}{a}$?

Can the ≤ hold with equality?

$$\mu = \sum_x x \cdot p(x) \quad \geq \sum_{x \geq a} x \cdot p(x) \quad \geq \sum_{x \geq a} a \cdot p(x) \quad = a \cdot P(X \geq a)$$

$$\boxed{\forall x \in (0, a), \quad p(x) = 0} \qquad \boxed{\forall x > a, \quad p(x) = 0}$$

$X \in \{0, a\}$

No sweeping improvements… continue looking…

Strengths

Weaknesses

Applies to all nonnegative random variables

Can always be used

Used to derive other inequalities: Chebyshev, Chernoff

Applies to all nonnegative random variables

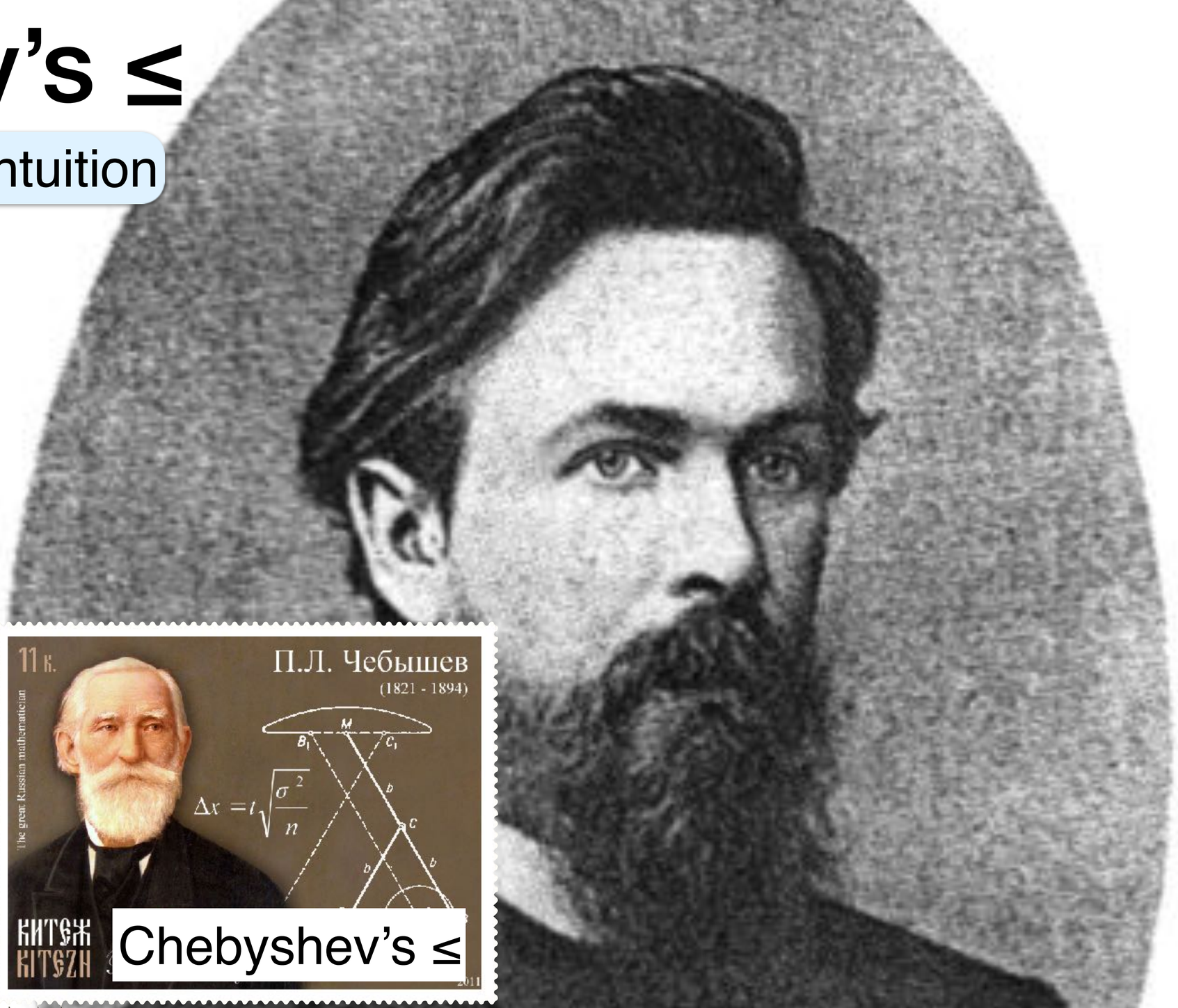
Limited to ≤ that hold for all distributions

Strengthen

Add assumptions

Markov's ≤

- Motivation, intuition
- Formulation
- Proof
- Example
- Extensions



Chebyshev's ≤

Chebyshev's

Motivation, intuition

Formulation

Proof

Example

Markov → Chebyshev

Markov

Probability that non-neg X is α times larger than its mean is $\leq 1/\alpha$

Chebyshev

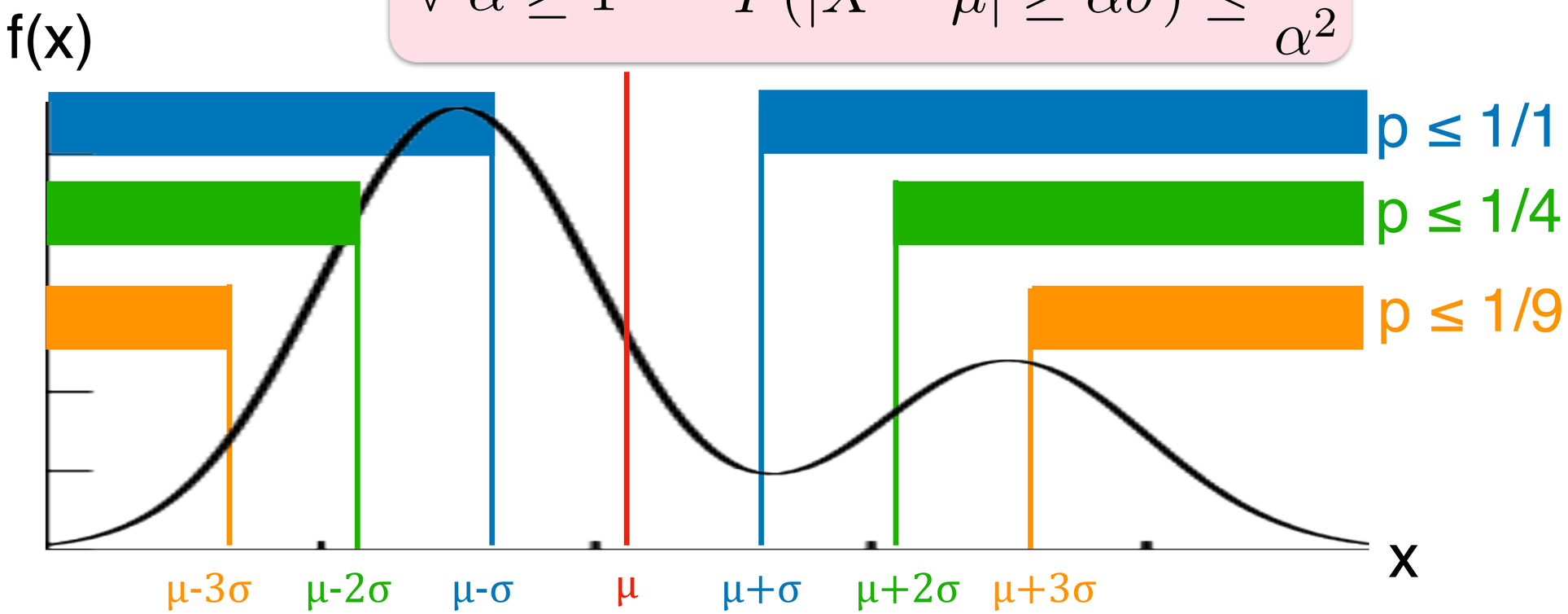
Probability that any X is more than $\alpha\sigma$ times further from μ is $\leq 1/\alpha^2$

Chebyshev's Inequality

Same two versions **1** Easier to illustrate, understand, remember

X is any discrete or continuous r.v. with finite mean μ and std σ

$$\forall\ \alpha \geq 1 \qquad P(|X-\mu| \geq \alpha\sigma) \leq \frac{1}{\alpha^2}$$



Two Formulations

X is any discrete or continuous r.v. with finite mean μ and std σ

1 Easier to visualize, understand, remember

$$\forall\ \alpha \geq 1 \qquad P(|X-\mu| \geq \alpha\sigma) \leq \frac{1}{\alpha^2}$$

2 Easier to prove, use $a = \alpha\sigma$

$$\forall\ a \geq \sigma \qquad P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Towards a Proof

Markov's Inequality

$$\forall\ a \geq \mu \qquad P(X \geq a) \leq \frac{\mu}{a}$$

Chebyshev's Inequality

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

$$P((X - \mu)^2 \geq a^2) \leq \frac{\sigma^2}{a^2}$$

Need: nonnegative r.v. with mean σ^2

Proof

$$\boxed{P(|X-\mu_X| \geq a) \leq \frac{\sigma_X^2}{a^2}}$$

Y Soon

X - any random variable $\quad \mu_X = EX \quad \sigma_X^2 = V(X) = E(X-\mu_X)^2$

$$Y = (X-\mu_X)^2 \qquad Y \geq 0 \qquad \mu_Y = E(X-\mu_X)^2 = \sigma_X^2$$

$$P(|X-\mu_X| \geq a) = P((X-\mu_X)^2 \geq a^2)$$

$$= P(Y \geq a^2)$$

Markov

$$\leq \frac{\mu_Y}{a^2} = \frac{\sigma_X^2}{a^2}$$

Citations

X - # paper citations

$\mu = 8$

Suppose $\sigma = 5$

$P(X \geq 28)$?

Markov

$P(X \geq 28) \leq 8/28 \approx 29\%$

$$P(X \geq a) \leq \frac{\mu}{a}$$

Chebyshev

$P(X \geq 28) = P(X - \mu \geq 20)$

$\leq P(|X - \mu| \geq 20)$

$\leq \left(\frac{\sigma}{20}\right)^2 = \left(\frac{5}{20}\right)^2 = \frac{1}{16} \approx 6.3\%$

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Markov: $\leq 0.02\%$

$P(X \geq 40{,}000) = P(X - \mu \geq 39{,}992)$

$\leq P(|X - \mu| \geq 39{,}992)$

$\leq \left(\frac{\sigma}{39{,}992}\right)^2 = \left(\frac{5}{39{,}992}\right)^2 = 1.6 \times 10^{-6}\ \%$

Survey Responses

Survey expected to result in $\mu = 1M$ responses with $\sigma = 50K$

Bound $P(0.8M < \#\text{ responses} < 1.2M)$

$0.8M = \mu - 4\sigma$ $\quad$ $1.2M = \mu + 4\sigma$

$P(\mu - 4\sigma < X < \mu + 4\sigma) = P(|X - \mu| < 4\sigma)$

$= 1 - P(|X - \mu| \geq 4\sigma)$

$\geq 1 - 1/16$

$= 15/16$

Mark x Che

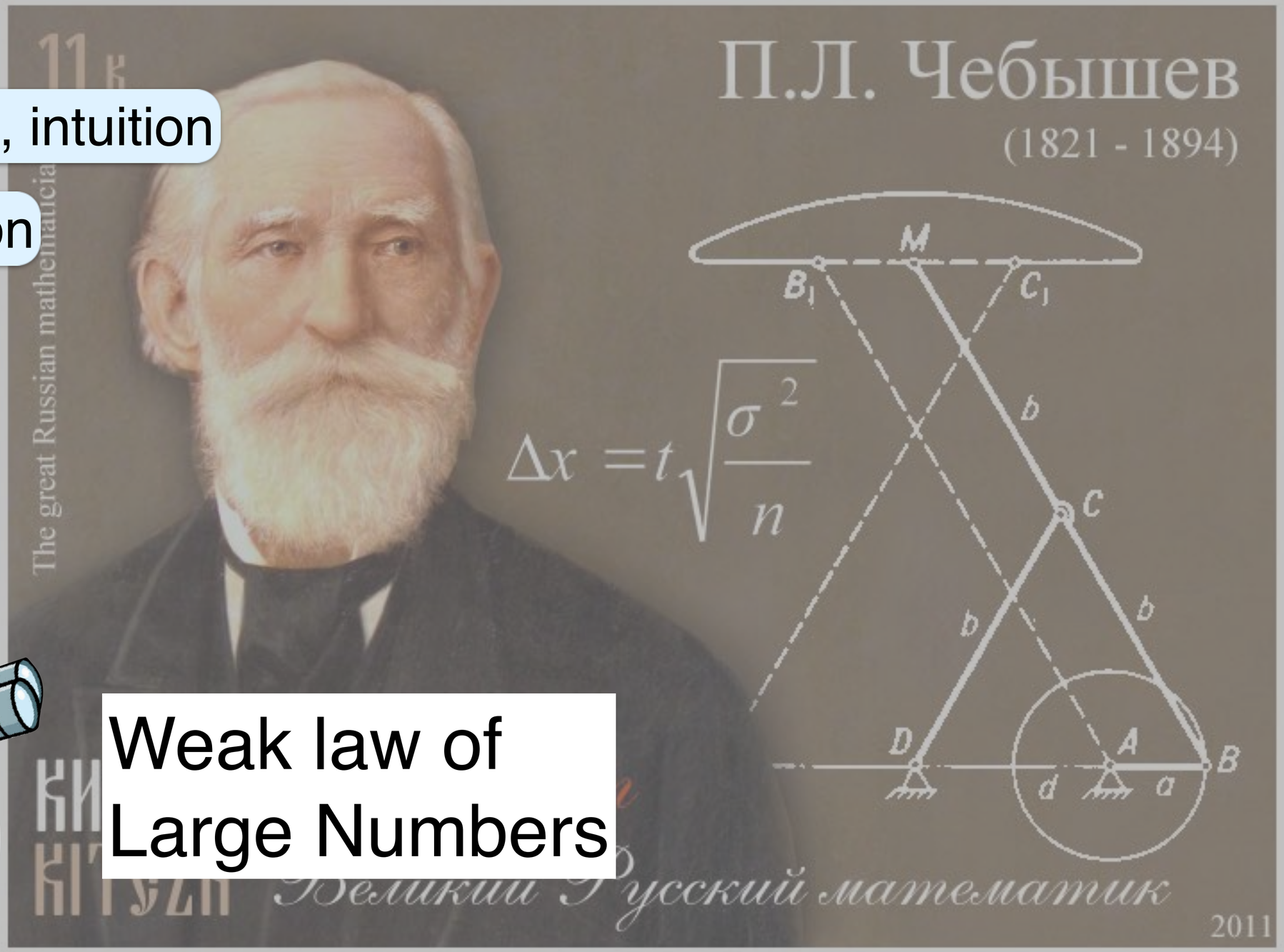
	Formula	Applies	Input	Range	Decreases
Markov	$P(X \geq a) \leq \frac{\mu}{a}$	X≥0	μ	a ≥ μ	Linearly
Chebyshev	$P(\lvert X - \mu \rvert \geq a) \leq \frac{\sigma^2}{a^2}$	Any X	μ and σ	a ≥ σ	Quadratically

Motivation, intuition

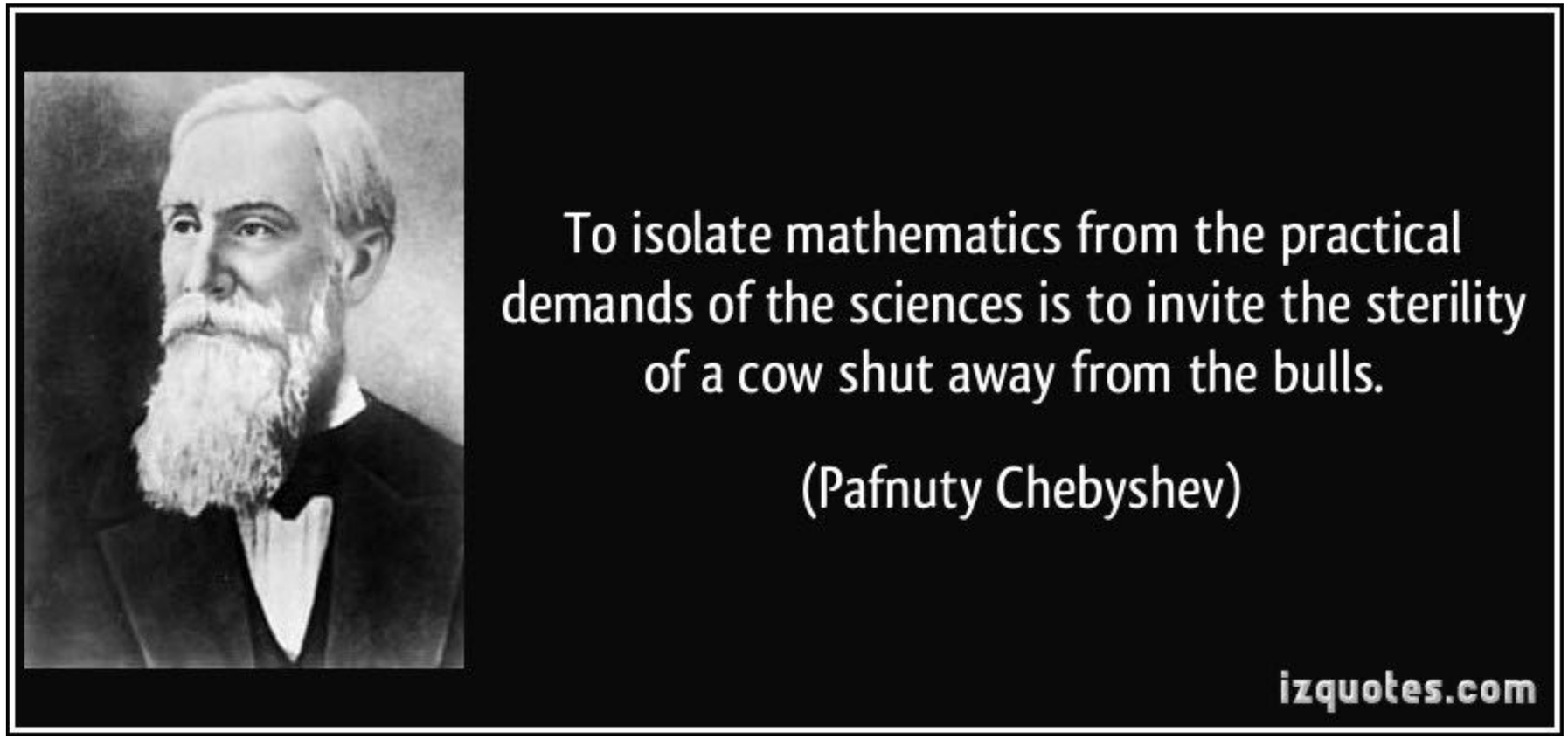
Formulation

Proof

Example

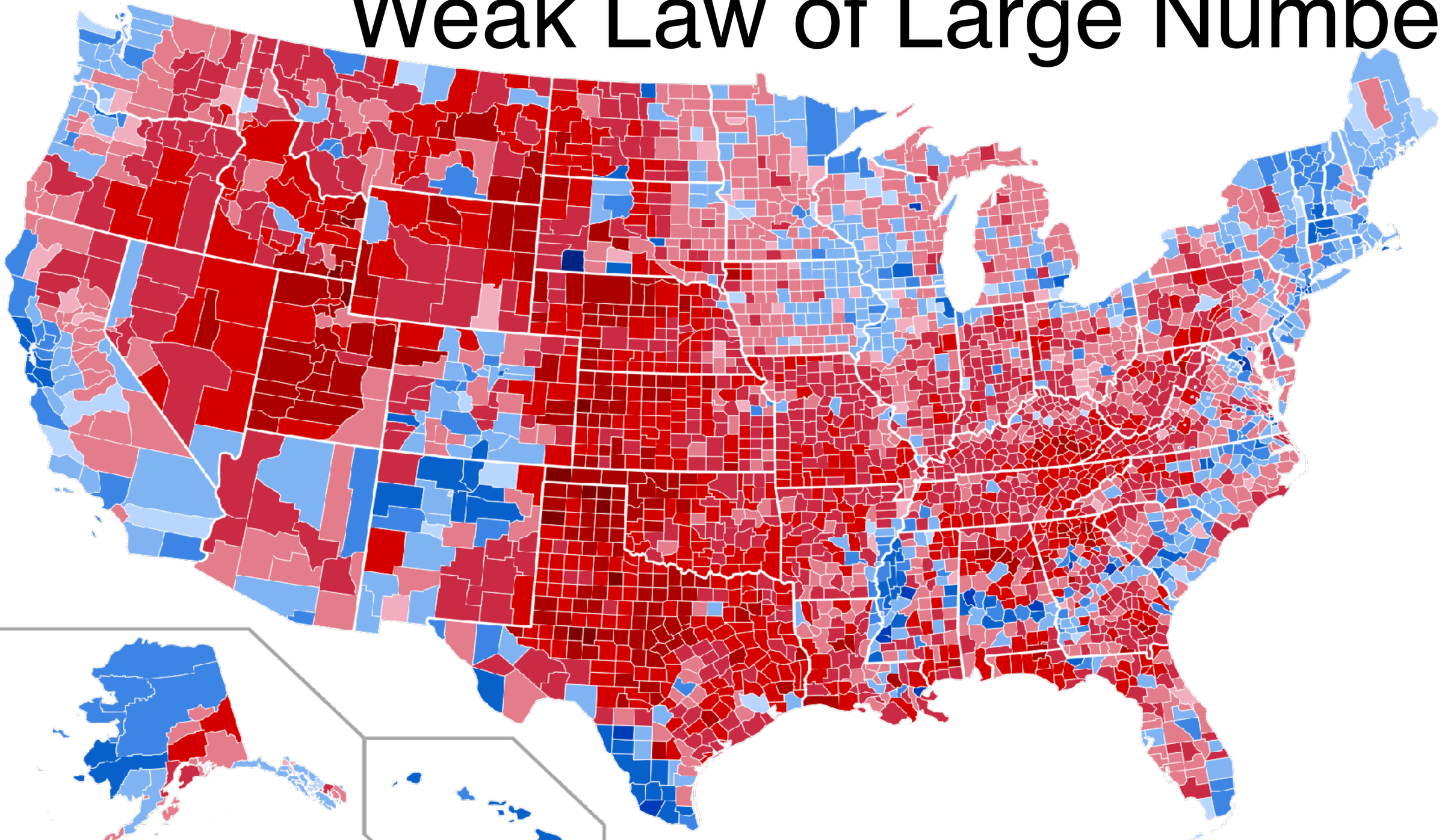


Weak law of Large Numbers



Alternative spellings: *Chebychev*, *Chebysheff*, Chebychov, *Chebyshov*;*Tchebychev*, *Tchebycheff,* *Tschebyschev*, *Tschebyschef*, *Tschebyscheff*

Weak Law of Large Numbers



Motivation

Probability theory based on sample averages converging to expectation

Flip many fair coins, fraction of heads converges to 1/2

Roll many fair dice, average value converges to 3.5

So far — Intuition

Now — Rigorous

Sample Mean

Sequence abbreviation $x^n \stackrel{\text{def}}{=} x_1, x_2, \ldots, x_n$

Mean $\overline{x^n} \stackrel{\text{def}}{=} \frac{x_1+x_2+\ldots+x_n}{n}$

$n = 4$ $x^4 \stackrel{\text{def}}{=} 3, 1, 4, 2$ $\overline{x^4} = \frac{3+1+4+2}{4} = 2.5$

n samples from a distribution $X^n = X_1, X_2, \ldots, X_n$

Sample mean $\overline{X^n} \stackrel{\text{def}}{=} \frac{X_1 + \ldots + X_n}{n}$

$\overline{X^n}$ is a random variable

Independent Samples

Independent random variables with the *same* distribution are Independent identically distributed (iid)

Independent $B_{0.3}$ r.v.'s are iid $B_{0.3}$, or iid

X_1, X_2, X_3 are iid $B_{0.3}$

Each X_i is $B_{0.3}$ selected $\perp\!\!\!\perp$ of all others

$P[\ (X_1 = 1, X_2 = 0, X_3 = 1) = 0.3 \cdot 0.7 \cdot 0.3 = 0.063$

Weak Law of Large Numbers

As # samples increases, the sample mean → distribution mean

$X^n = X_1,\ldots,X_n$ iid samples from distribution with finite mean μ and finite std σ

As n → ∞

$\overline{X^n}$ approaches μ

P(sample mean differs from μ by any given amount) ↘ 0 with n

$$P\left(|\overline{X^n} - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2} \cdot \frac{1}{n}$$

$\overline{X^n}$ "converges in probability" to μ

Polling Error

2016 Presidential elections

Poll 100,000 people

Assuming every person voted for Trump independently w. probability p

Bound the probability that off by more than 1%

WLLN

$$P\left(|\overline{X^n} - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2} \cdot \frac{1}{n}$$

$$\sigma^2 = p(1-p) \leq \tfrac{1}{4}$$

$$P(|\overline{X^{100,000}} - p| \geq 0.01) \leq \frac{1/4}{0.01^2 \cdot 100,000} = 2.5\%$$

Proof of WLLN $P\left(|\overline{X^n} - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n \cdot \epsilon^2}$

$X_1, X_2, \ldots,$ iid with finite μ and σ, sample mean $\overline{X^n} \stackrel{\text{def}}{=} \frac{1}{n} \sum X_i \qquad \sum = \sum_{i=1}^{n}$

Expectation

$$E\left(\overline{X^n}\right) = E\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n} \sum E(X_i) = \frac{1}{n} \sum \mu = \mu$$

Variance

$$V\left(\overline{X^n}\right) = V\left(\frac{1}{n} \sum X_i\right) = \frac{1}{n^2} V\left(\sum X_i\right) = \frac{1}{n^2} \sum V(X_i) = \frac{1}{n^2} \sum \sigma^2 = \frac{\sigma^2}{n}$$

Chebyshev

$$\forall \epsilon > 0 \qquad P\left(|\overline{X^n} - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{n \cdot \epsilon^2} \searrow 0$$

$$n \to \infty$$

Sensors

n sensors measure temperature t

Each reads $T_i = t + Z_i$ $\quad$ Z_i - noise with zero mean and variance ≤ 2

How many sensors needed to estimate t to ± ½ with probability ≥ 95%

$$P\left(|\overline{X^n} - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2} \cdot \frac{1}{n}$$

$$P(|\overline{T}^n - t| \geq 0.5) \leq \frac{2}{\frac{1}{4}n} \quad \leq 0.05$$

$$n \geq \frac{2}{\frac{1}{4} \cdot 0.05} \quad = : 2 \cdot 4 \cdot 20 = 160$$

Generalization

Same proof works when means μ_i and σ_i differ.

Just let $\mu \stackrel{\text{def}}{=} \frac{1}{n} \sum \mu_i$ and $\sigma^2 \stackrel{\text{def}}{=} \frac{1}{n} \sum \sigma_i^2$

$$P\left(|\overline{X^n} - \mu| \geq \epsilon\right) \leq \frac{\sigma^2}{\epsilon^2} \cdot \frac{1}{n}$$

Convergence in Probability

$X_1, X_2, \ldots$ infinite sequence of random variables

X_n converges in probability to a random variable Y

P(X_n differs from Y by any given fixed amount) ↘ 0 with n

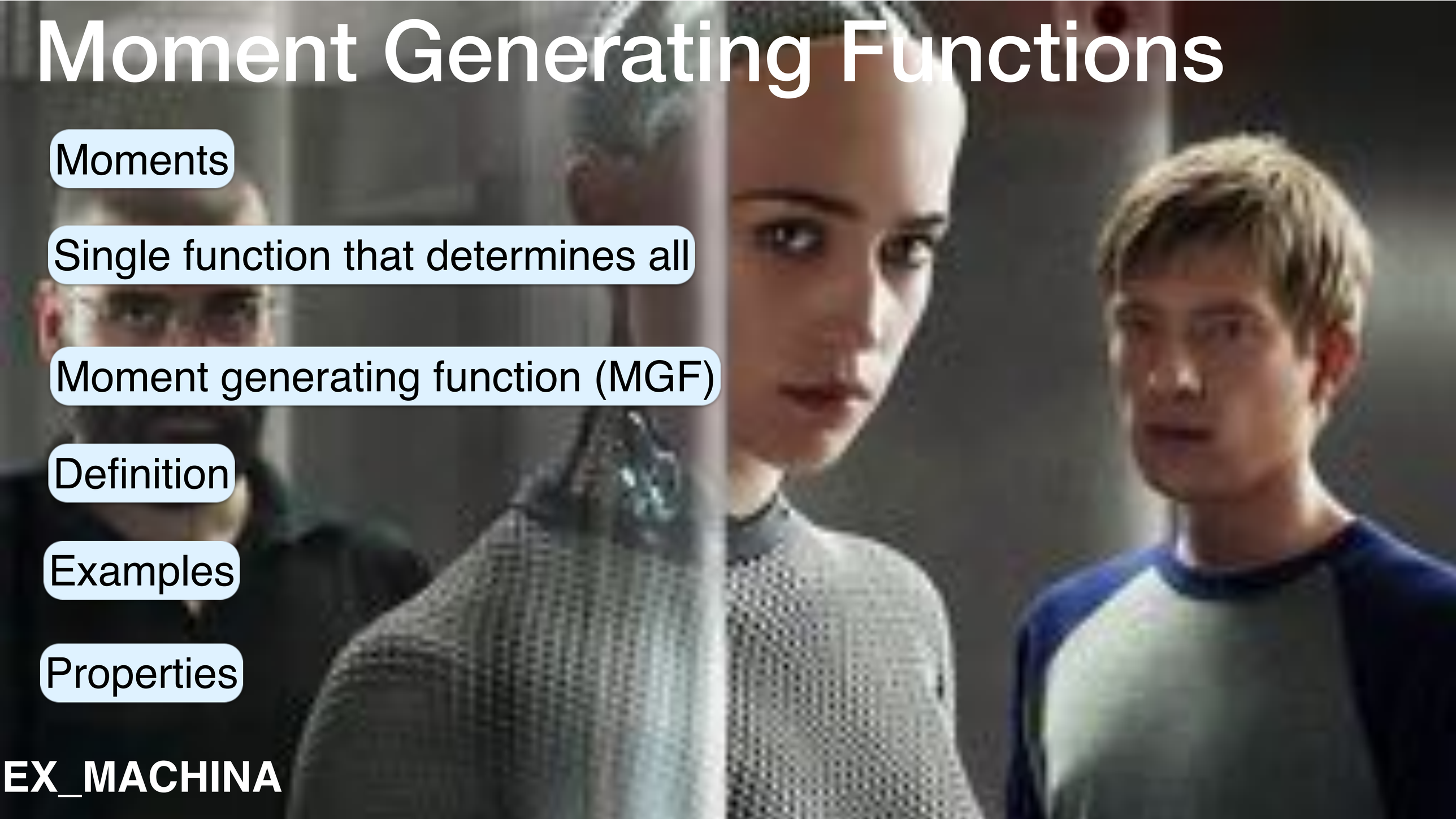
For every $\delta > 0$ P($|X_n - Y| \geq \delta$) ↘ 0 with n

For every $\delta>0$ and $\varepsilon>0$ there is an N s.t for all $n \geq N$

$P(|X_n - Y| \geq \delta) < \varepsilon$

WLLN: $\overline{X^n}$ converges in probability to μ

Moment Generating Functions



- Moments
- Single function that determines all
- Moment generating function (MGF)
- Definition
- Examples
- Properties

Moments

Expectations of powers of X are called Moments

$E(X)$ $E(X^2)$ $E(X^3)$

Sometimes called raw moments to distinguish from central moments

$E(X-\mu)^n$

Determine mean, variance, many other distribution properties

General method to determine all moments

Moment Generating Function (MGF)

Maps a random variable X to a real function M

M: $\mathbb{R} \rightarrow \mathbb{R}$

$$M(t) \stackrel{\text{def}}{=} M_X(t) \stackrel{\text{def}}{=} E[e^{tX}] = \begin{cases} \sum p(x)\, e^{tx} & \text{discrete} \\ \int f(x)\, e^{tx}\, dx & \text{continuous} \end{cases}$$

Determined by distribution p or f

X more convenient

Discrete

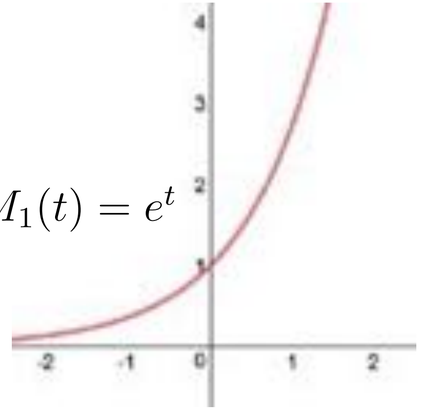
Continuous

Few Values

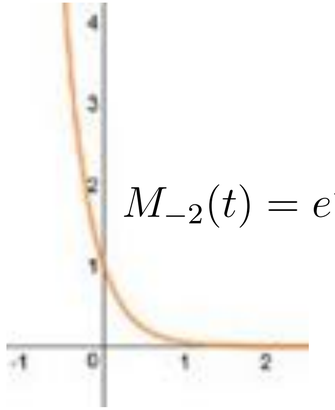
One value (constant)

$$X = c \qquad M_c(t) \;=\; E(e^{tX}) = \sum p(x)\, e^{tx} \;=\; e^{ct}$$

$M_1(t) = e^t$



$M_{-2}(t) = e^{-2t}$

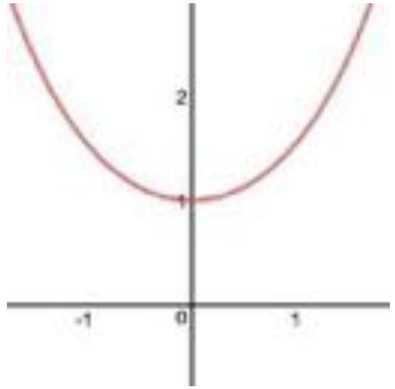


Two arbitrary values

$$p(c_1) = p_1 \quad p(c_2) = p_2 \qquad M(t) \;=\; E(e^{tX}) \;=\; p_1\, e^{c_1 t} + p_2\, e^{c_2 t}$$

$p_{-1} = p_1 = \frac{1}{2}$

$M(t) = \frac{1}{2}e^{-t} + \frac{1}{2}e^{t}$



0, 1 values (Bernoulli)

$X \sim B_p$ $\quad p_0 = 1 - p$ $\quad p_1 = p$

$$M(t) = \;(1-p)e^{t\cdot 0} + pe^{t\cdot 1} = \;(1-p) + pe^{t}$$

Properties

Positive

$$M(t) = E(e^{tX}) > 0$$

0 → 1

$$M(0) = E(e^{0X}) = E(e^0) = 1$$

Translation and Scaling

Translation X → X+b

$$M_{X+b}(t) = E(e^{t(X+b)}) = E(e^{tX} \cdot e^{tb}) = e^{tb} \cdot E(e^{tX}) = e^{tb} M_X(t)$$

Scaling X → aX

$$M_{aX}(t) = E(e^{t(aX)}) = E(e^{atX}) = M_X(at)$$

Translation and Scaling X → aX+b

$$M_{aX+b}(t) = e^{bt} \cdot M_{aX}(t) = e^{bt} \cdot M_X(at)$$

$$M_c(t) = e^{ct} \longrightarrow M_{a \cdot c+b} = e^{bt} \cdot M_c(at) = e^{bt} \cdot e^{cat} = e^{(ac+b) \cdot t}$$

Addition

Independent variables

The MGF of the sum is the product of the MGF's

Two variables

X ⫫ Y

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}] = E[e^{tX}] \cdot E[e^{tY}] = M_X(t) \cdot M_Y(t)$$

n variables

$X_1, X_2, \ldots, X_n \quad \perp\!\!\!\perp$

$$X \stackrel{\text{def}}{=} X_1 + X_2 + \ldots + X_n$$

$$M_X(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

Average

$X_1, X_2, \ldots, X_n \quad \perp\!\!\!\perp$

$$\overline{X} \stackrel{\text{def}}{=} \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$M_{\overline{X}}(t) = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right)$$

Moment Generation

$E(X^n)$ n^{th} "raw" moment of X

$M_x(t)$ determines ("generates") all moments of X

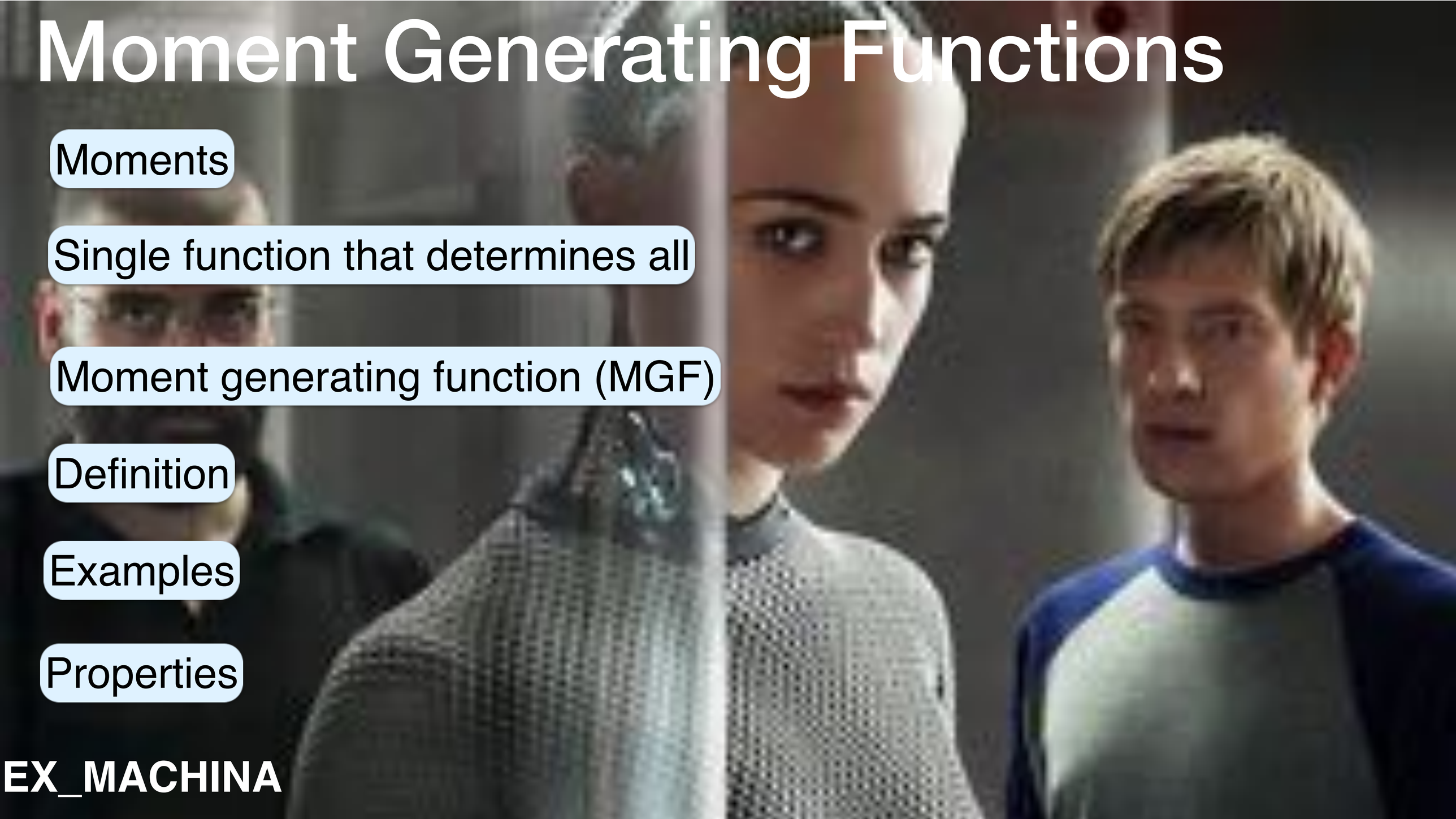
$$E(X) = M'_X(0) \qquad E(X^2) = M''_X(0) \qquad E(X^n) = M_X^{(n)}(0)$$

$$e^y = 1 + \frac{y}{1!} + \frac{y^2}{2!} + \frac{y^3}{3!} + \ldots$$

$$M_X(t) = E(e^{tX}) = E\left(1 + \frac{t}{1!}X + \frac{t^2}{2!}X^2 + \frac{t^3}{3!}X^3 + \ldots\right)$$

$$= 1 + \frac{t}{1!}E(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) + \ldots$$

Moment Generating Functions



- Moments
- Single function that determines all
- Moment generating function (MGF)
- Definition
- Examples
- Properties

Conveyor belt

Binomial

$B_{p,n} \quad 0 \le p \le 1, \; n \ge 0$

$p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$M(t) = \; [1 + p(e^t - 1)]^n$

$$M(t) = E(e^{tX}) = \sum_{k=0}^{n} p(k) \cdot e^{tk}$$

$$= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot e^{tk}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k}$$

$$= (pe^t + 1 - p)^n$$

Use….. Binomial Thm

$$= [1 + p(e^t - 1)]^n$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \, b^{n-k}$$

Moments

$E(X)$

$$M'(t) = n(pe^t + 1 - p)^{n-1} \cdot pe^t$$

$$M'(0) = n(pe^0 + 1 - p)^{n-1} \cdot pe^0 = np$$

Poisson

$P_\lambda, \ \lambda > 0$

$p(k) = e^{-\lambda} \ \frac{\lambda^k}{k!}$

$M(t) = \ e^{\lambda(e^t - 1)}$

$$M(t) \ = \ E(e^{tX}) \ = \ \sum_{k=0}^{\infty} p(k) \cdot e^{tk}$$

$$= \ \sum_{k=0}^{\infty} e^{-\lambda} \ \frac{\lambda^k}{k!} \cdot e^{tk}$$

$$= \ \sum_{k=0}^{\infty} e^{-\lambda} \ \frac{(\lambda e^t)^k}{k!}$$

$$= \ e^{\lambda e^t - \lambda} \cdot \sum_{k=0}^{\infty} e^{-\lambda e^t} \ \frac{(\lambda e^t)^k}{k!}$$

$$= \ e^{\lambda(e^t - 1)}$$

Standard Normal

$\mathcal{N}(0,1)$ $\quad f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\quad M(t) = e^{\frac{t^2}{2}}$

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx$$

$$\int_{-\infty}^{\infty} \rightarrow \int$$

$$= \int e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int \exp[-\frac{x^2-2tx}{2}]\, dx$$

$$= \frac{1}{\sqrt{2\pi}} \int \exp[-\frac{(x-t)^2}{2} + \frac{t^2}{2}]\, dx$$

$$= e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int \exp[-\frac{(x-t)^2}{2}]\, dx$$

$$= e^{\frac{t^2}{2}}$$

“Abnormal” Normal

Standard normal $Z \sim \mathcal{N}(0,1)$ $M_Z(t) = e^{\frac{t^2}{2}}$

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

$$X = \sigma Z + \mu$$

$$M_X(t) = M_{aZ+b}(t) = e^{bt} M_Z(at)$$

$$= e^{\mu t} \cdot e^{\frac{\sigma^2 t^2}{2}}$$

$$= e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Properties

Informally

If $M_X(t) = M_Y(t)$ then X and Y have the same distribution

Can invert $M_X(t)$ to obtain X

If $M_{X_n}(t) \rightarrow M_X(t)$ then the $f_{X_n} \rightarrow f_X$

Poisson and Binomial

Binomial

$$B_{p,n}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$M_{p,n}(t) = [1 + p(e^t - 1)]^n$$

Poisson

$$P_\lambda(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$M_\lambda(t) = e^{\lambda(e^t - 1)}$$

Convergence

$$B_{\frac{\lambda}{n},n} \to P_\lambda$$

$$M_{\frac{\lambda}{n},n}(t) \to M_\lambda(t) \ ?$$

$$(1 + \tfrac{1}{n})^n \to e$$

$$(1 + \tfrac{a}{n})^n = [(1 + \tfrac{a}{n})^{\frac{n}{a}}]^a \to e^a$$

$$M_{\frac{\lambda}{n},n}(t) = [1 + \tfrac{\lambda}{n}(e^t - 1)]^n \to e^{\lambda(e^t - 1)} = M_\lambda(t)$$

Moment Generating Functions

Moments

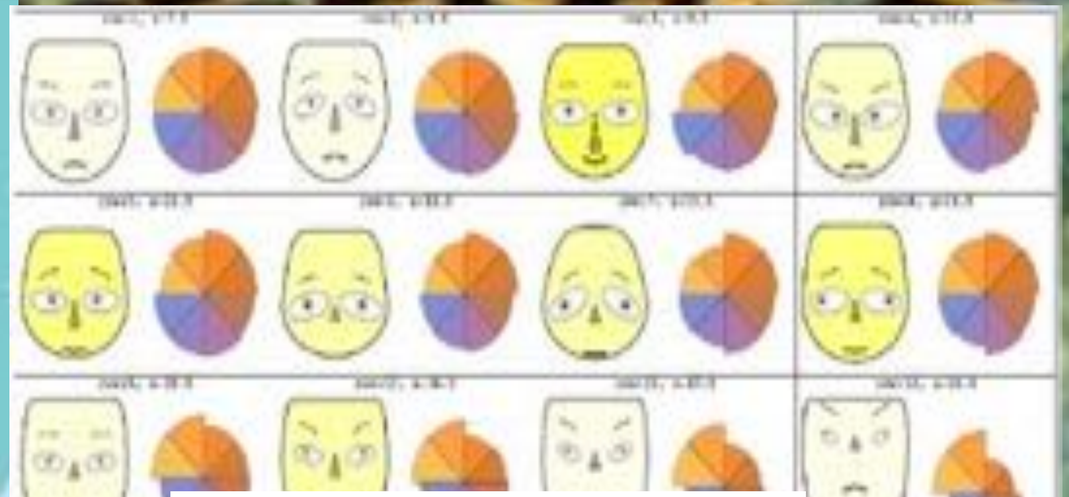
Single function that determines all

Moment generating function (MGF)

Definition

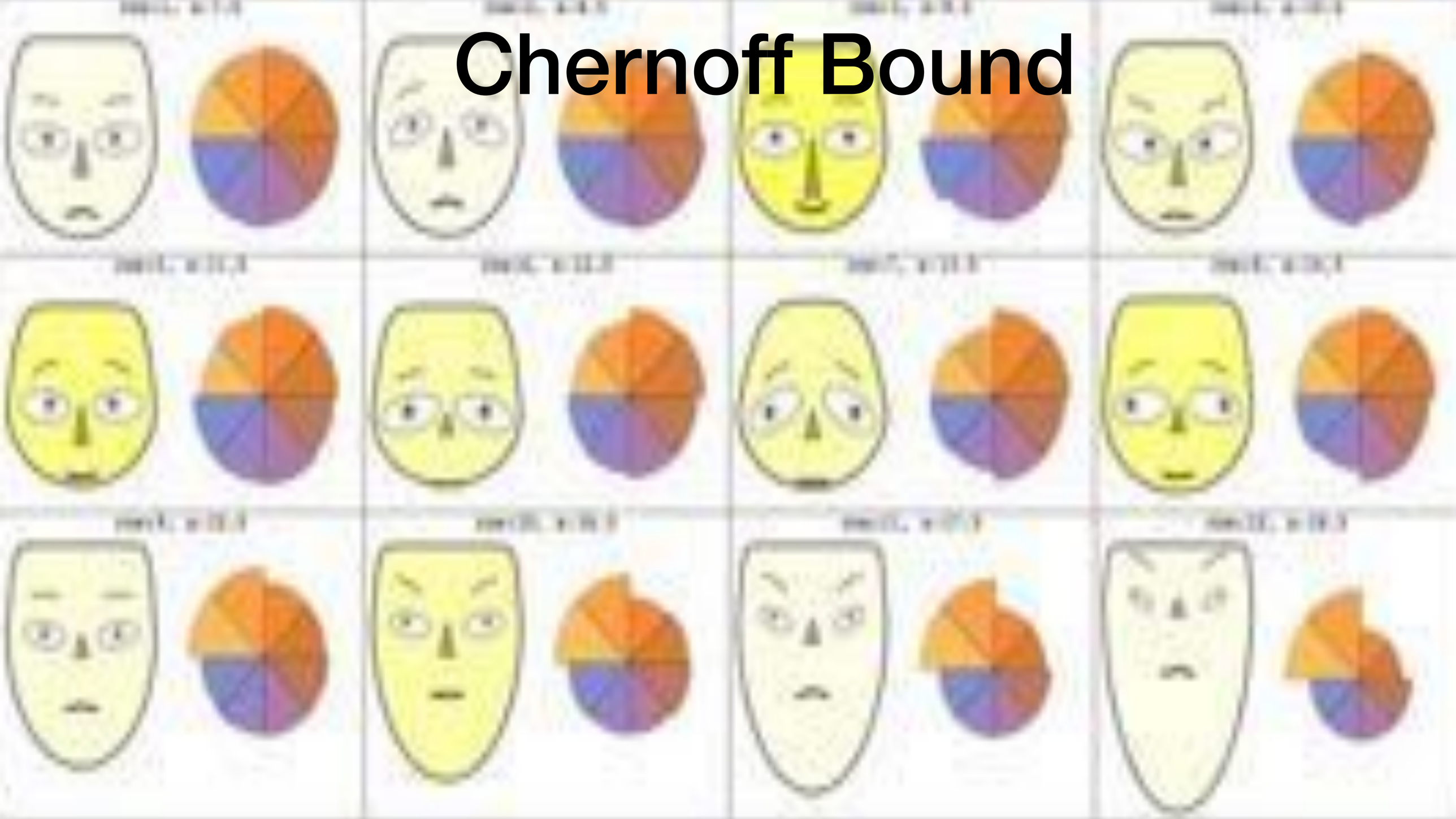
Examples

Properties



Chernoff Bound

Chernoff Bound



Herman Chernoff

1923 -

Statistician

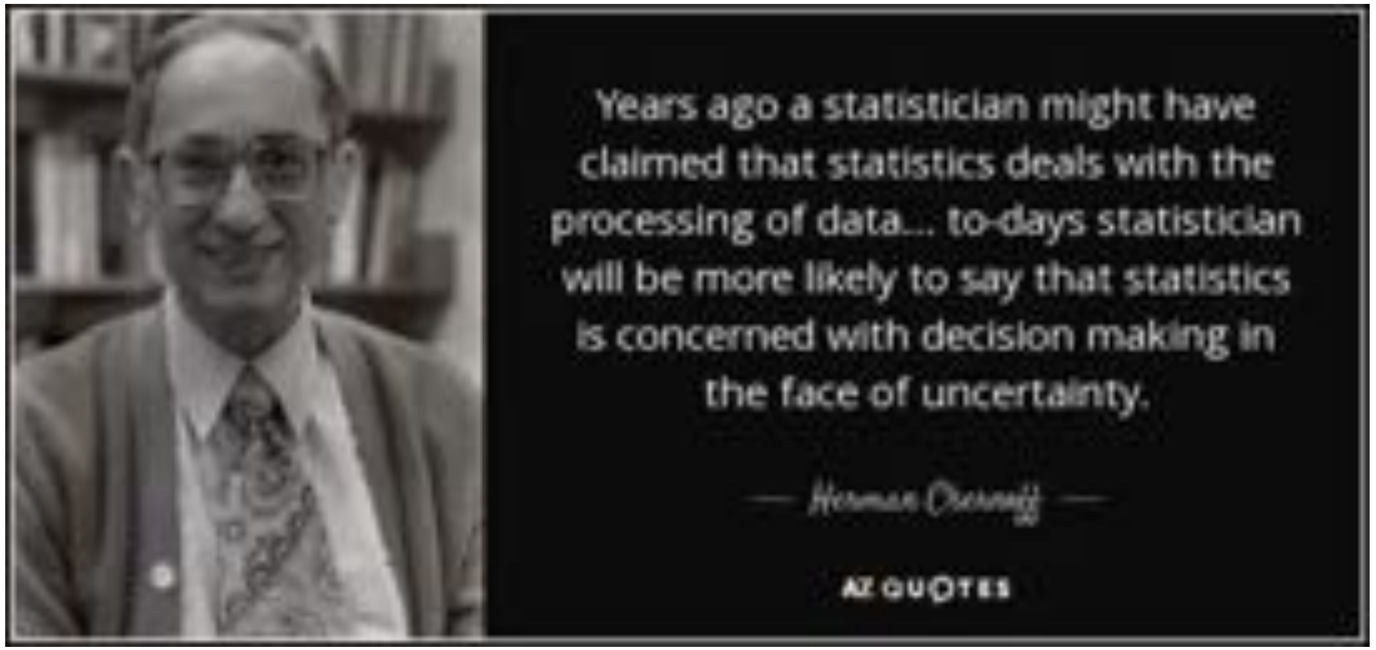
UIUC, Stanford, MIT, Harvard

Broad view of statistics

Chernoff bound

Markov → Chebyshev → Chernov

Chernoff faces



The New York Times
April 1, 2008

Chernoff Bound

$X \sim B_{p,n}$ $\mu = pn$ $P\left(X \geq (1+\delta)\mu\right)$

Decrease | Markov - constant | Chebyshev - linear | Chernoff - exponential

$\forall a \quad \forall t \geq 0$

$$X \geq a \quad \leftrightarrow \quad tX \geq ta \quad \leftrightarrow \quad e^{tX} \geq e^{ta}$$

Markov

$$P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}}$$

$E(e^{tX})$ | Evaluate | Bound | Incorporate | Simplify

Evaluate $E(e^{tX})$

Same as moment generating functions

Self-contained

$X \sim B_{p,n}$

$X = \sum_{i=1}^{n} X_i$

$X_i \sim B_p \perp\!\!\!\perp$

$$E\left(e^{tX}\right) = E(e^{t\Sigma X_i}) = E(e^{\Sigma t X_i}) = E\left(\prod_{i=1}^{n} e^{tX_i}\right) \overset{\perp\!\!\!\perp}{=} \prod_{i=1}^{n} E\left(e^{tX_i}\right) = [(1-p) + pe^t]^n$$

$$E(e^{tX_i}) = P(X_i = 0) \cdot e^{t \cdot 0} + P(X_i = 1) \cdot e^{t \cdot 1} = (1-p) + pe^t$$

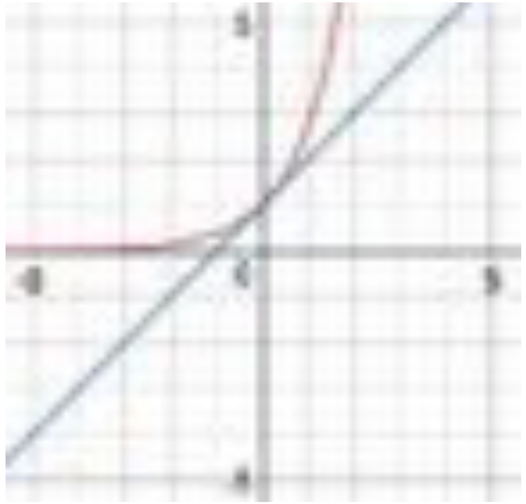
$$E\left(e^{tX}\right) = [(1-p) + pe^t]^n$$

Bound $E(e^{tX})$

$$E\left(e^{tX}\right) = [(1-p)+pe^t]^n \leq \left(e^{p(e^t-1)}\right)^n = e^{np(e^t-1)} = e^{\mu(e^t-1)}$$

$$(1-p)+pe^t = 1+p(e^t-1) \leq e^{p(e^t-1)}$$

$$1+x \leq e^x$$



$$e^x = 1+x+\frac{x^2}{2}+\ldots \geq 1+x$$

$$E\left(e^{tX}\right) \leq e^{\mu(e^t-1)}$$

Incorporate in Markov ≤

$X \sim B_{p,n}$ $\forall a \quad \forall t \geq 0$ $P(X \geq a) = P(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}}$

$E\left(e^{tX}\right) \leq e^{\mu(e^t - 1)}$

$P(X \geq a) \leq \frac{e^{\mu(e^t-1)}}{e^{ta}}$

$a = (1+\delta)\mu$ $\delta \geq 0$ $\forall t \geq 0$

$P(X \geq (1+\delta)\mu) \leq \frac{e^{\mu(e^t-1)}}{e^{\mu(1+\delta)t}} = e^{\mu((e^t-1)-t(1+\delta))}$

Optimize t

Optimize t

$\forall t \geq 0$

$$P\left(X \geq (1+\delta)\mu\right) \leq e^{\mu((e^t-1)-t(1+\delta))}$$

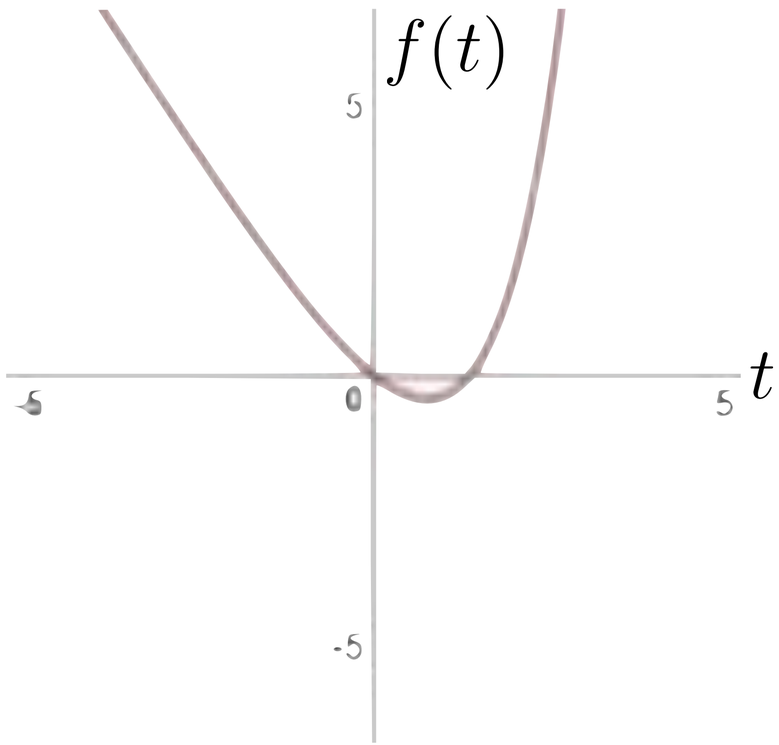
Find t minimizing

$$f(t) \stackrel{\text{def}}{=} (e^t - 1) - t(1+\delta)$$

$$f'(t) = e^t - (1+\delta) = 0$$

$$e^t = 1 + \delta \qquad t = \ln(1+\delta)$$

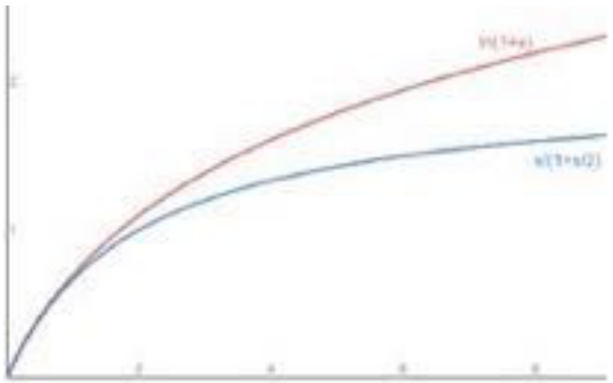
$$f''(t) = e^t \geq 0$$



$$P\left(X \geq (1+\delta)\mu\right) \leq e^{\mu((e^t-1)-t(1+\delta))} = e^{\mu[\delta-(1+\delta)\ln(1+\delta)]}$$

Final Simplification

$$P\left(X \geq (1+\delta)\mu\right) \leq e^{\mu[\delta-(1+\delta)\ln(1+\delta)]} \leq e^{-\frac{\delta^2}{2+\delta}\mu}$$



Inequality $\ln(1+x) \geq \frac{x}{1+\frac{x}{2}} \qquad \forall x \geq 0$

Define $f(x) = \ln(1+x) - \frac{x}{1+\frac{x}{2}}$

Show $f(x) \geq 0 \qquad \forall x \geq 0$

$$f(0) = \ln 1 - \frac{0}{1} = 0$$

$$f'(x) = \frac{1}{1+x} - \frac{(1+\frac{x}{2})-\frac{x}{2}}{(1+\frac{x}{2})^2} = \frac{1}{1+x} - \frac{1}{(1+\frac{x}{2})^2} = \frac{1}{1+x} - \frac{1}{1+x+\frac{x^2}{2}} \geq 0$$

$$f(x) \geq 0 \qquad \forall x \geq 0$$

$$\delta - (1+\delta)\ln(1+\delta) \leq \delta - (1+\delta)\frac{\delta}{1+\frac{\delta}{2}} = \delta\left[1 - \frac{1+\delta}{1+\frac{\delta}{2}}\right] = \delta\left[\frac{2+\delta-2-2\delta}{2+\delta}\right] = \frac{-\delta^2}{2+\delta}$$

Chernoff Bound

$X \sim B_{p,n}$

$\delta \geq 0$

$$P\left(X \geq (1+\delta)\mu\right) \leq e^{-\frac{\delta^2}{2+\delta}\mu}$$

$$P\left(X \leq (1-\delta)\mu\right) \leq e^{-\frac{\delta^2}{2}\mu}$$

Poll

47% vote D

6000 samples

X : # vote D

X ~ $B_{0.47,6000}$

Wrong:

X ≥ 0.5·6,000 = 3,000

$$\mu = np = 6000 \cdot 0.47 = 2820$$

$$3000 = (1+\delta)\mu$$

$$1+\delta = \frac{3000}{2820} = \frac{6000 \cdot 0.5}{6000 \cdot 0.47} = \frac{50}{47} \approx 1.06383$$

$$\delta \approx 0.0638$$

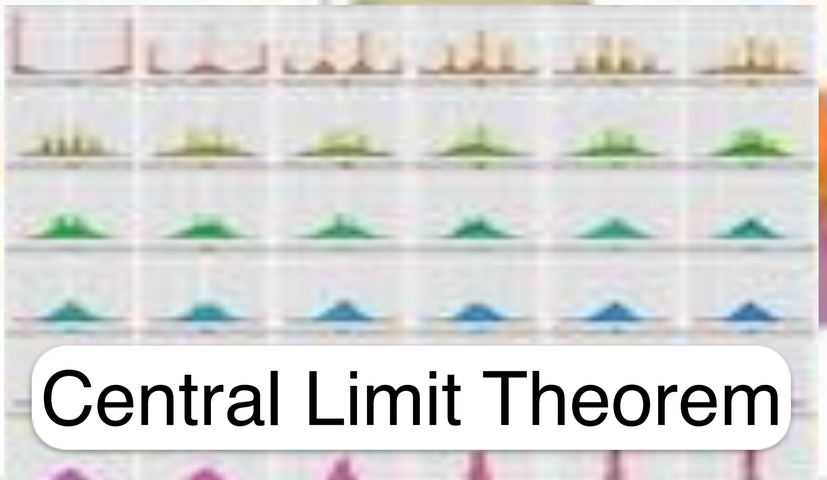
P(wrong)

$$P(X \geq 3000) = P(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{2+\delta}\mu} \approx 0.38\%$$

Chernoff Bound

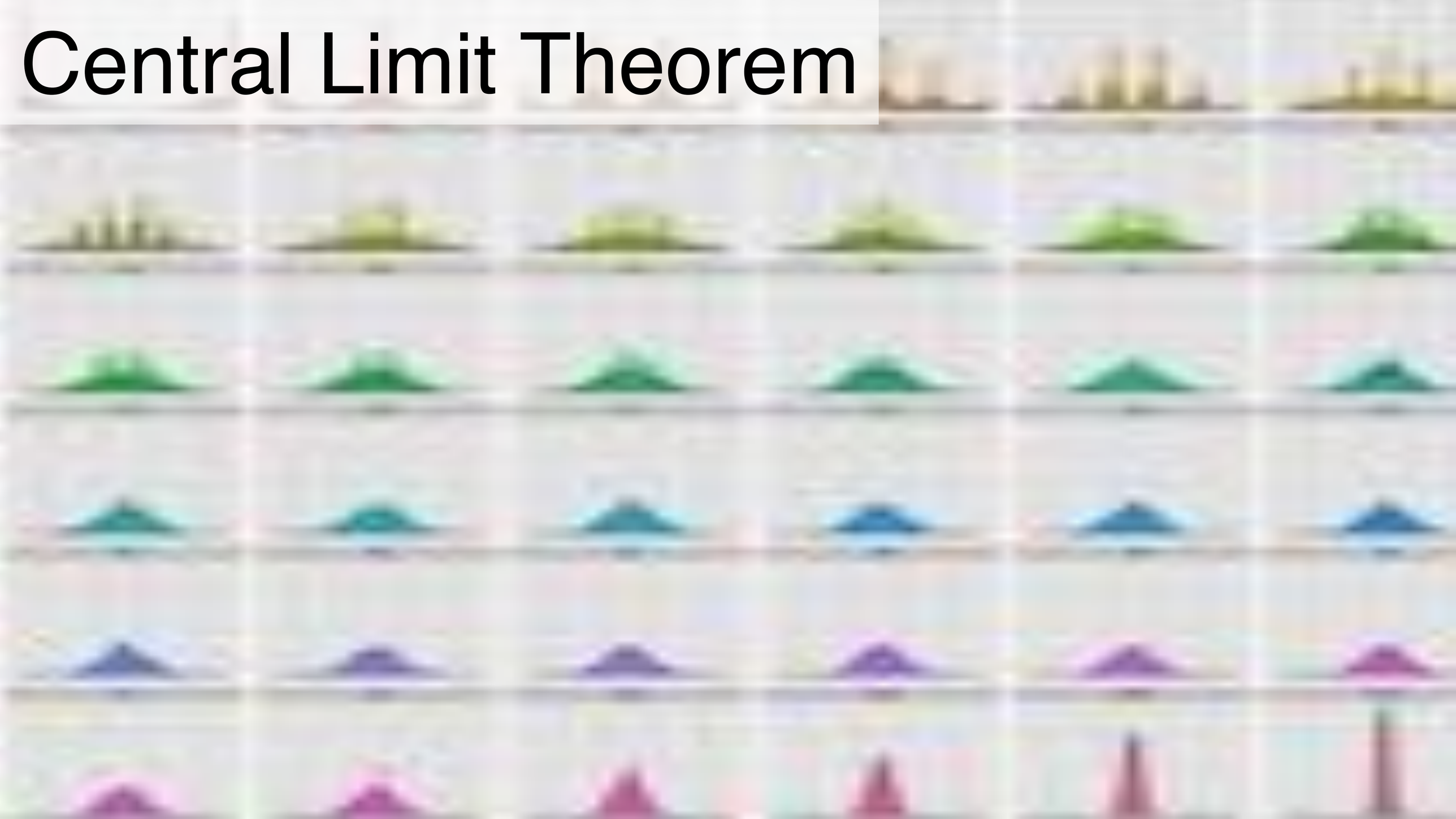
Exponential bound on exceeding mean by a constant factor

Next



Central Limit Theorem

Central Limit Theorem



Overview

Central Limit Theorem (CLT)

“Central" result in Statistics

Mild conditions

Normalized sum of random variables is roughly Normal

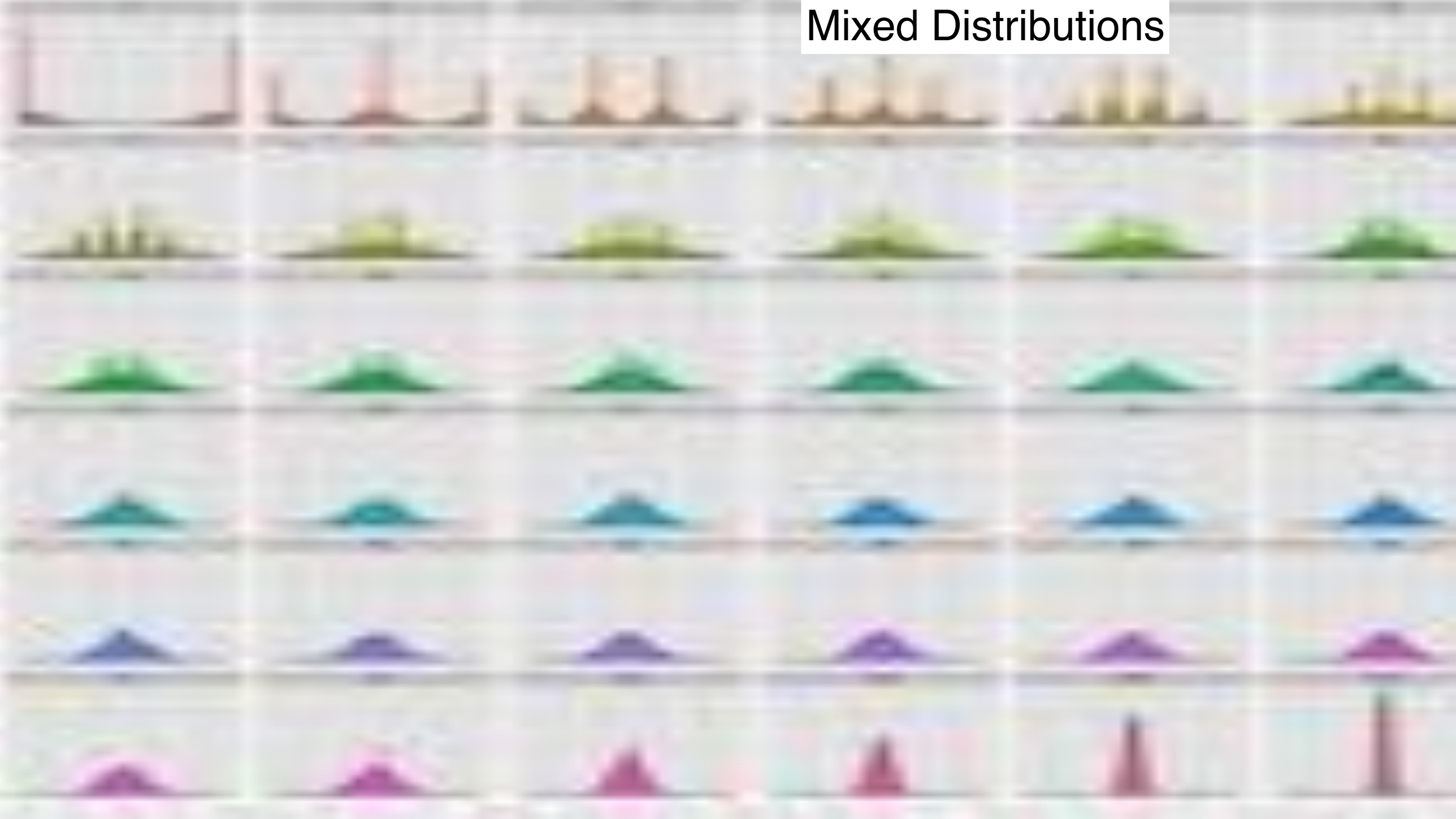
Discrete, continuous , mixed distributions

Explains why Bell Curve so popular

Allows for simple probability estimation

Seeing is believing

Mixed Distributions



Central Limit Theorem

iid - independent identically distributed

μ - mean

σ - standard deviation

Let $X_1, X_2, X_3, \ldots$ be iid with finite μ and σ.

As n→∞, the distribution of $\frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}}$ approaches $\mathcal{N}(0,1)$.

-nμ /σ Normalization

Without loss of generality, assume μ=0, σ=1

$X_1, X_2, X_3, \ldots$ iid with μ=0 and σ=1. As n→∞, the distribution of $\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}$ approaches $\mathcal{N}(0,1)$.

Greatly generalizes the Weak Law of Large Numbers

WLLN CLT

$X_1, X_2, X_3, \ldots$ iid with $\mu=0$ and $\sigma=1$

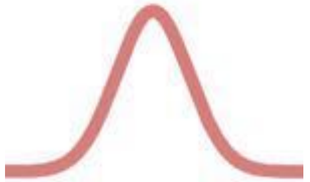
WLLN As $n \to \infty$, $\frac{X_1+X_2+\ldots+X_n}{n} \to 0$

$E = 0$ $\quad V(X_1 + \ldots + X_n) = n$ $\quad V\left(\frac{X_1+\ldots+X_n}{n}\right) = \frac{n}{n^2} = \frac{1}{n} \to 0$

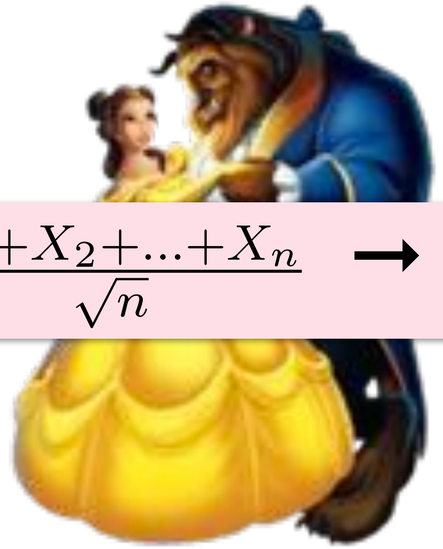
CLT As $n \to \infty$, the distribution of $\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}} \to \mathcal{N}(0,1)$ distribution

$E = 0$ $\quad V\left(\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}\right) = \frac{n}{n} = 1$

More importantly



CLT



As n→∞, the distribution of $\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}$ → $\mathcal{N}(0,1)$ distribution

$$V\left(\frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}\right) = \frac{n}{n} = 1$$

Normalize by $\gg \sqrt{n}$ V → 0 $\ll \sqrt{n}$ V → ∞ $\sqrt{n}$ V = 1

For any X_i

"Normal"

Converge to $\mathcal{N}(0,1)$

CLT

Does the extra resolution help?

ADD

Show example where following not useful

Markov (≤)

Chebyshev (≤ sigma)

WLLN (weak - in the wrong way)

Store Income

At a store

Customer's spending

$\mu = 80$

$\sigma = 40$

P(a customer spends ≤ 72)

10% below average

???

Care more about total revenue

100 Customers

P(average spending ≤ 72)

Can answer!

Thanks to CLT!

General CLT Application

$$X_1, \ldots, X_n \perp\!\!\!\perp, \sim \text{ any distribution with mean } \mu, \text{ and stdv } \sigma$$

$$\overline{X^n} \stackrel{\text{def}}{=} \frac{X_1 + \ldots + X_n}{n} \qquad P(\overline{X^n} \leq \alpha) = ?$$

$$Z_n \stackrel{\text{def}}{=} \frac{(X_1 + \ldots + X_n) - n\mu}{\sigma\sqrt{n}}$$

Typically ≥ 30

Roughly

CLT

For sufficiently large n

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

$$Z_n = \frac{n\overline{X^n} - n\mu}{\sigma\sqrt{n}} = \frac{n(\overline{X^n} - \mu)}{\sigma\sqrt{n}} = \frac{\overline{X^n} - \mu}{\sigma/\sqrt{n}}$$

$$P(\overline{X^n} \leq \alpha) = P\left(Z_n \leq \frac{\alpha - \mu}{\sigma/\sqrt{n}}\right) \approx \Phi\left(\frac{\alpha - \mu}{\sigma/\sqrt{n}}\right)$$

General formula for average of random variables

CDF

$X \sim N(0,1)$

$$\Phi(x) \triangleq F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-y^2}{2}} dy$$

No known formula

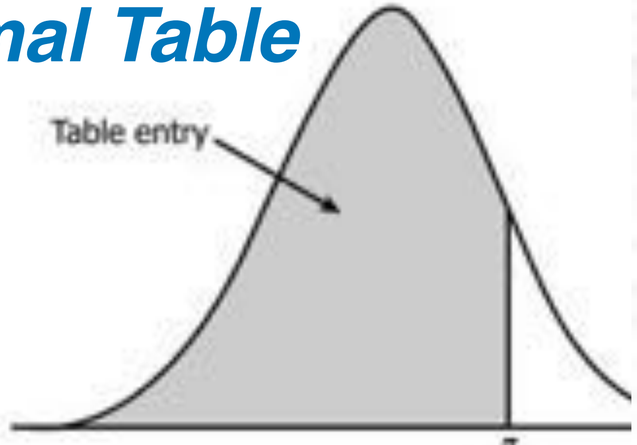
Instead use table or computer

Table for each μ, σ?

1 suffices!

Standard Normal Table

Z Table



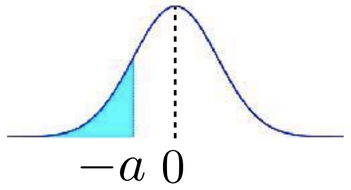
Store

Customer's spending X_i $\perp\!\!\!\perp$ $\mu = 80$ $\sigma = 40$ $n = 100$ Customers

P(average spending ≤ 72) 10% below average

$$P(\overline{X^n} \leq \alpha) \approx \Phi\left(\frac{\alpha-\mu}{\sigma/\sqrt{n}}\right)$$

$$P(\overline{X^{100}} \leq 72) \approx \Phi\left(\frac{72-80}{40/\sqrt{100}}\right) = \Phi(-\tfrac{8}{4}) = \Phi(-2)$$



$$P(X \leq -a) = \Phi(-a) = 1 - \Phi(a)$$

$$= 1 - \Phi(2) \approx 1 - 0.9772 = 0.0228 \approx 2.3\%$$

Change to insurance outlays so ≤ a

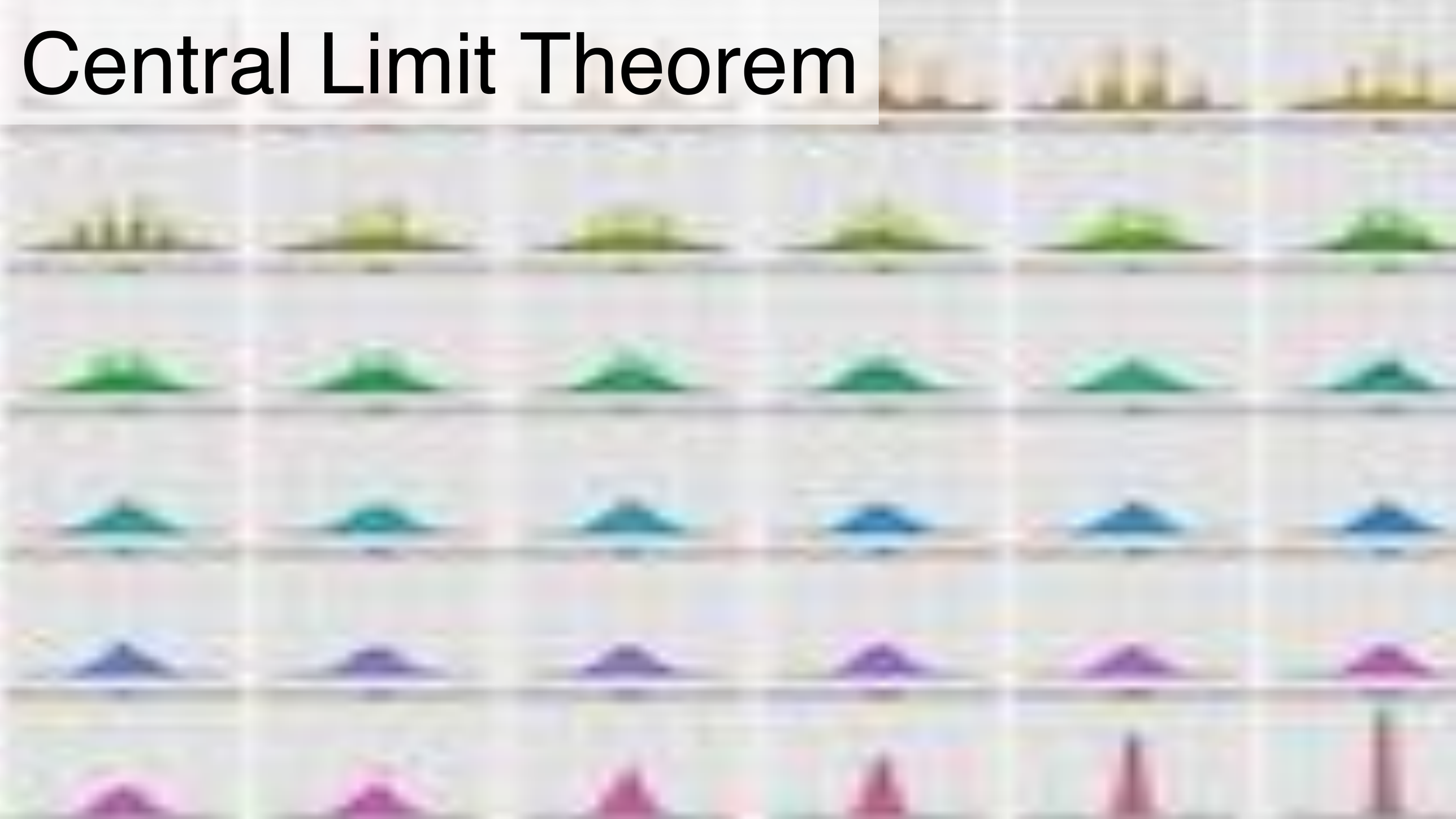
Do we need to artificially introduce multiple instances?

Not necessarily, some phenomena are MI by themselves

In communication, noise is gaussian

What about normalization by sqrt n - why would we do that

Central Limit Theorem



How the CLT is Made

Convergence

$X_1, X_2, X_3, \ldots$ iid

$\mu=0$

$\sigma=1$

$$Z_n \stackrel{\text{def}}{=} \frac{X_1+X_2+\ldots+X_n}{\sqrt{n}}$$

As $n \to \infty$, the distribution of Z_n approaches $\mathcal{N}(0,1)$.

Cumulative distribution function (CDF)

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

$$F_{Z_n}(x) = P(Z_n \leq x) = P\left(\frac{X_1+\ldots+X_n}{\sqrt{n}} \leq x\right)$$

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

$$F_{Z_n}(x) \underset{n\to\infty}{\longrightarrow} F_Z(x) \quad \forall x$$

CDF of standard normal distribution

Convergence “in distribution”

Moment Generating Functions

$$M_X(t) \stackrel{\text{def}}{=} E[e^{tX}]$$

Scaling and independent addition

$$M_{aX}(t) = E(e^{t \cdot aX}) = E(e^{at \cdot X}) = M_X(at)$$

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX} \cdot e^{tY}] = E[e^{tX}] \cdot E[e^{tY}] = M_X(t) \cdot M_Y(t)$$

Derivatives of M (at 0) are moment of X

$$M_X^{(n)}(0) = E[X^n]$$

$$M_X'(0) = E[X] \qquad M_X''(0) = E[X^2] \qquad M_X(0) = E[X^0] = 1$$

$$E[e^{0X}]$$

Standard Normal

$$Z \sim \mathcal{N}(0,1) \qquad M_Z(t) = e^{\frac{t^2}{2}}$$

Continuity

CDF - cumulative distribution fun.

MGF - moment generating fun.

Random Variables	CDF	MGF
$Z_1, Z_2, Z_3, \ldots$	$F_{Z_n}(t)$	$M_{Z_n}(t)$
Z	$F_Z(t)$	$M_Z(t)$

If $M_{Z_n}(t) \to M_Z(t) \quad \forall t$

then $F_{Z_n}(t) \to F_Z(t)$ wherever $F_Z(t)$ is continuous

Plan

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

$M_Z(t) = e^{\frac{t^2}{2}}$

Show $M_{Z_n}(t) \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}$

Imply $F_{Z_n}(t) \xrightarrow[n\to\infty]{} \Phi(t)$

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}}$$

$$M_{\frac{X_i}{\sqrt{n}}}(t) = M_{X_i}\left(\frac{t}{\sqrt{n}}\right) = M\left(\frac{t}{\sqrt{n}}\right)$$

$$M_{Z_n}(t) = \prod_{i=1}^{n} M_{\frac{X_i}{\sqrt{n}}}(t) = \prod_{i=1}^{n} M\left(\frac{t}{\sqrt{n}}\right) = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

Show $\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}$

L'Hôpital's Rule

Johann Bernoulli

Seek $\lim_{x \to a} \frac{f(x)}{g(x)}$ a - anything, finite or $\pm\infty$

$$\lim f(x) = 6 \quad \lim g(x) = 3 \longrightarrow \lim \frac{f(x)}{g(x)} = \frac{6}{3} = 2$$

$$\lim f(x) = \lim g(x) = 0 \quad \text{or} \quad \lim f(x),\ \lim g(x) = \pm\infty \longrightarrow \lim \frac{f(x)}{g(x)} \text{ unclear}$$

$$\text{If } \lim \frac{f'(x)}{g'(x)} \text{ exists} \qquad \text{then } \lim \frac{f(x)}{g(x)} = \lim \frac{f'(x)}{g'(x)}$$

Hospital → Hôpital → Ô

$$\lim_{n\to\infty}\left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{\frac{t^2}{2}} \quad \leftrightarrow \quad \lim_{n\to\infty} n\cdot\ln M\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2}$$

$$\lim_{n\to\infty} n\cdot\ln M\left(\frac{t}{\sqrt{n}}\right) = \lim_{u\to 0}\frac{\ln M(tu)}{u^2}$$

$$M(0) = E(X_i^0) = 1$$

$$M^{(i)}(0) = E(X^i)$$

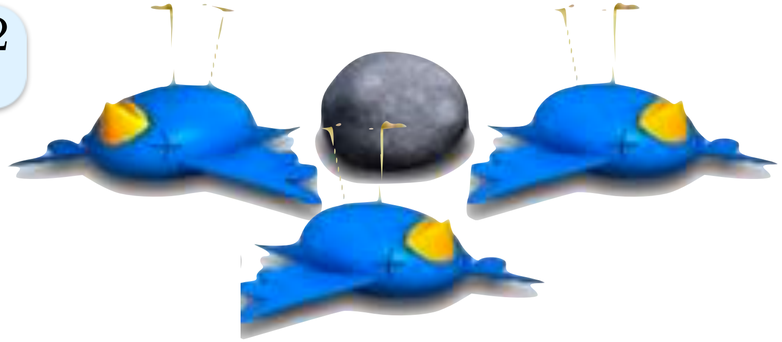
$$\infty\cdot 0 = ?$$

$$\boxed{u = \frac{1}{\sqrt{n}}}$$

$$\frac{0}{0} \rightarrow \hat{0}$$

$$n$$

$$1/\sqrt{n} \rightarrow u^2$$



$$\overset{\hat{0}}{=} \lim_{u\to 0}\frac{M'(tu)\cdot t}{M(tu)\cdot 2u}$$

t - constant

$$= \frac{t}{2}\lim_{u\to 0}\frac{M'(tu)}{u}$$

$$M'(0) = E(X_i^1) = \mu = 0$$

$$\overset{\hat{0}}{=} \frac{t}{2}\lim_{u\to 0}\frac{M''(tu)\cdot t}{1}$$

$$M''(0) = E(X_i^2) = \sigma^2 = 1$$

$$= \frac{t^2}{2}$$

Assembling

$$X_i \quad \text{iid}, \quad \mu = 0, \quad \sigma^2 = 1$$

$$Z_n = \frac{X_1 + \ldots + X_n}{\sqrt{n}}$$

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} \left[M\left(\frac{t}{\sqrt{n}}\right) \right]^n = e^{\frac{t^2}{2}} = M_Z(t)$$

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

$$F_{Z_n} \to F_Z = \Phi$$

Z_n approach Standard Normal distribution

General iid X_i just define $Z_n = \frac{X_1 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$

Generalizations

Showed for iid

Any original distribution, need just finite μ, σ, M

Discrete | Continuous | Mixed

More generally

Mild conditions

Independent | Weakly dependent | Multi-dimensional

Normal distribution very prevalent

How the CLT is Made