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

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

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

Markov's ≤

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

Two forms

Intuitive, memorable

$$\forall\ \alpha \ge 1 \qquad P(X \ge \alpha\mu) \le \frac{1}{\alpha}$$

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

Direct proof, easier to apply, more common

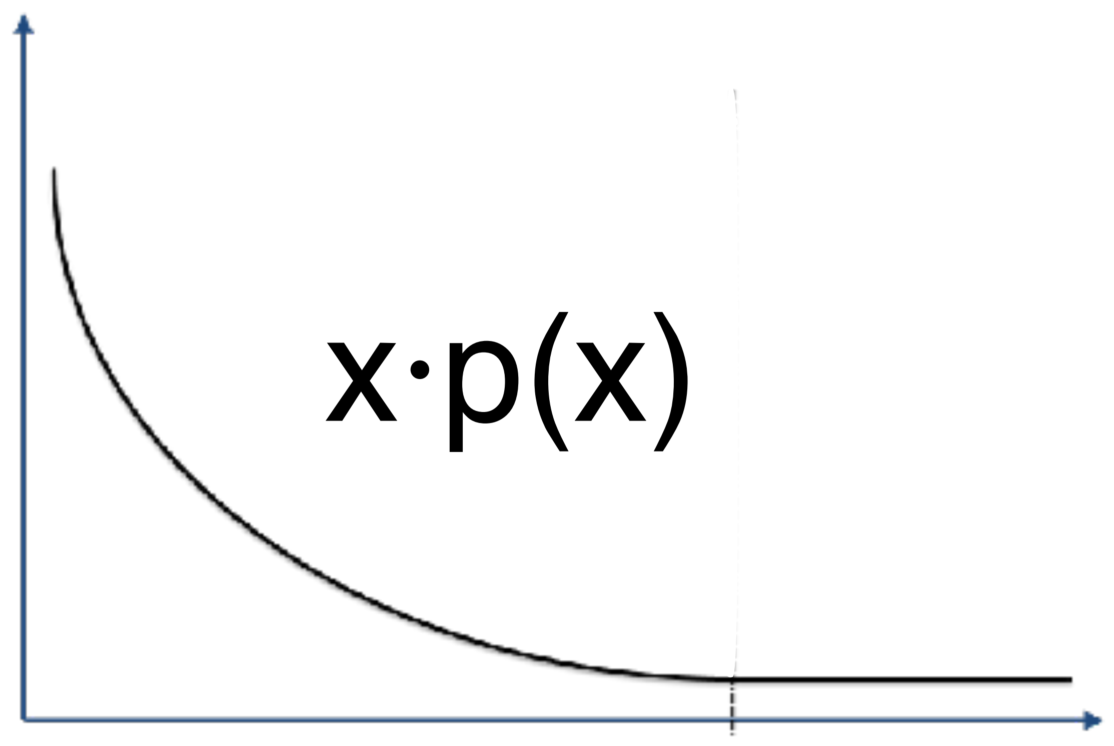
$$a = \alpha\mu \qquad \forall\ a \ge \mu \qquad P(X \ge a) \le \frac{\mu}{a}$$

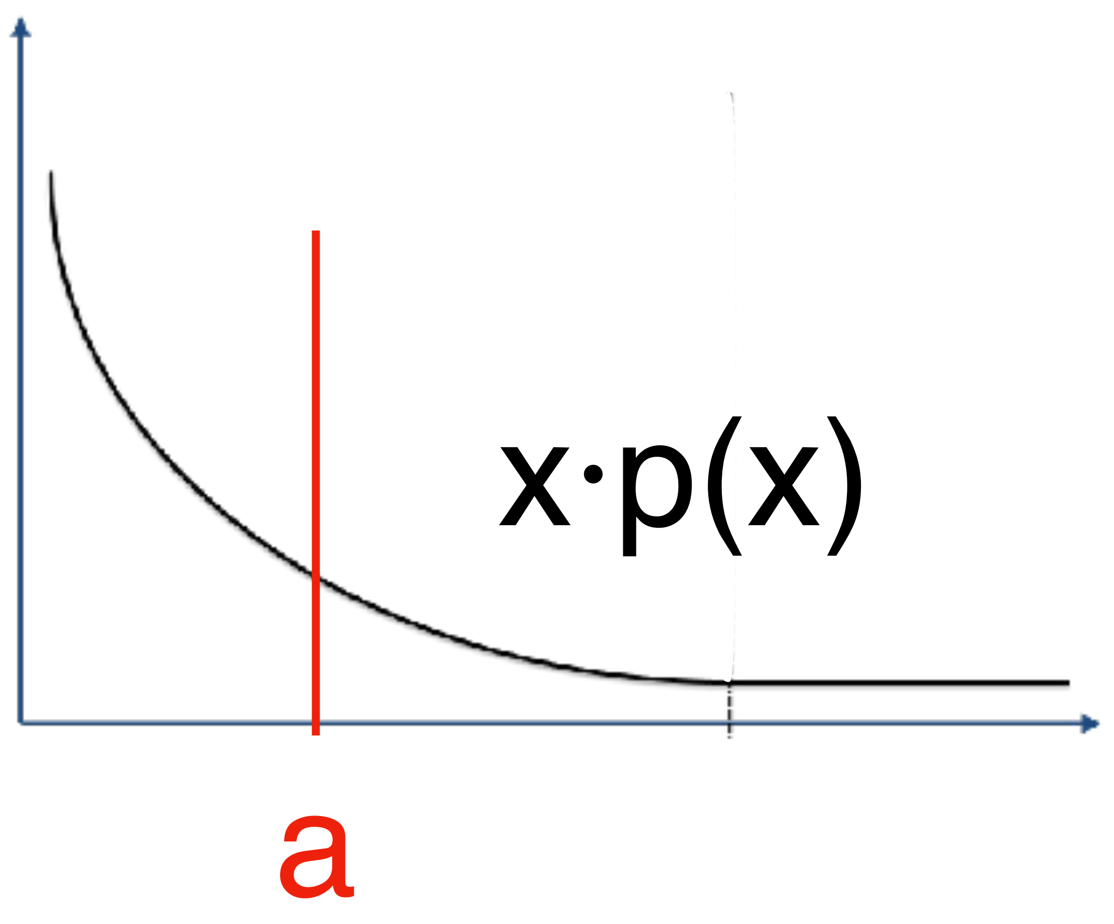
Proof

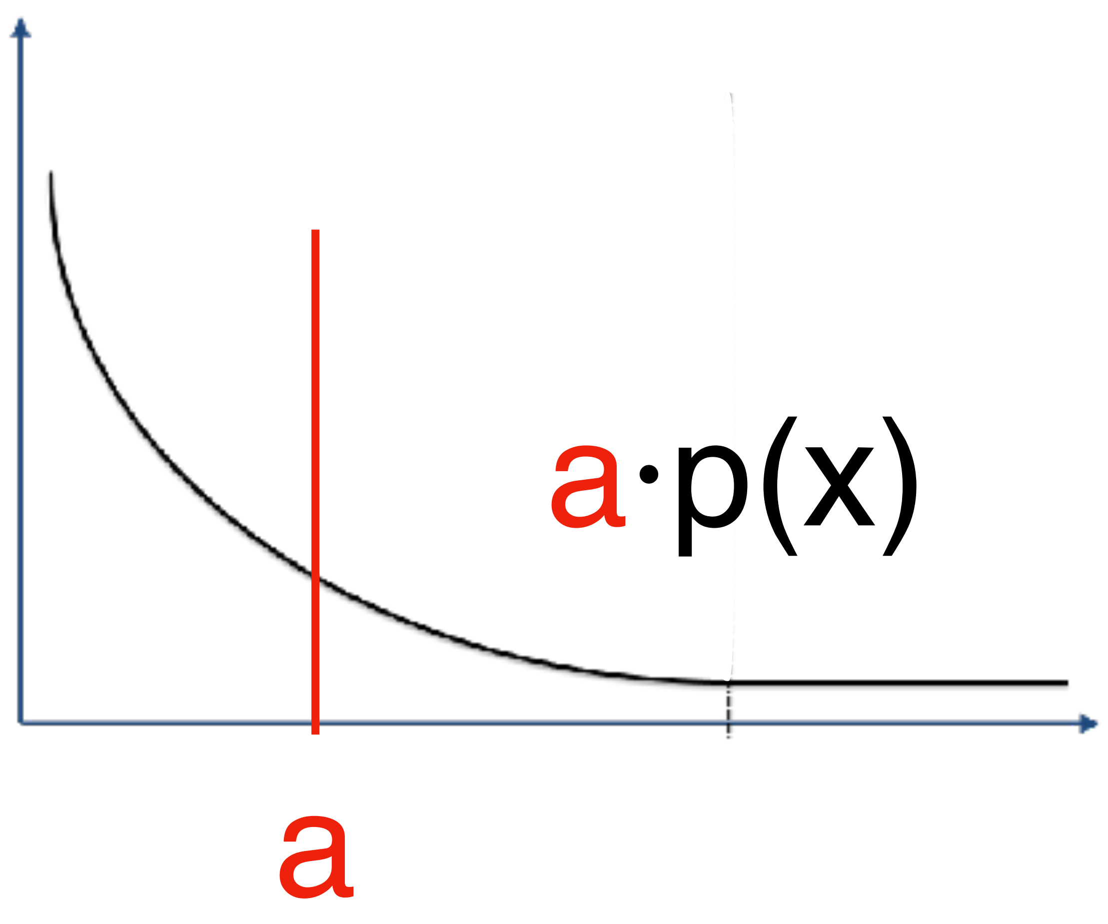
$$P(X \geq a) \leq \frac{\mu}{a}$$

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

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







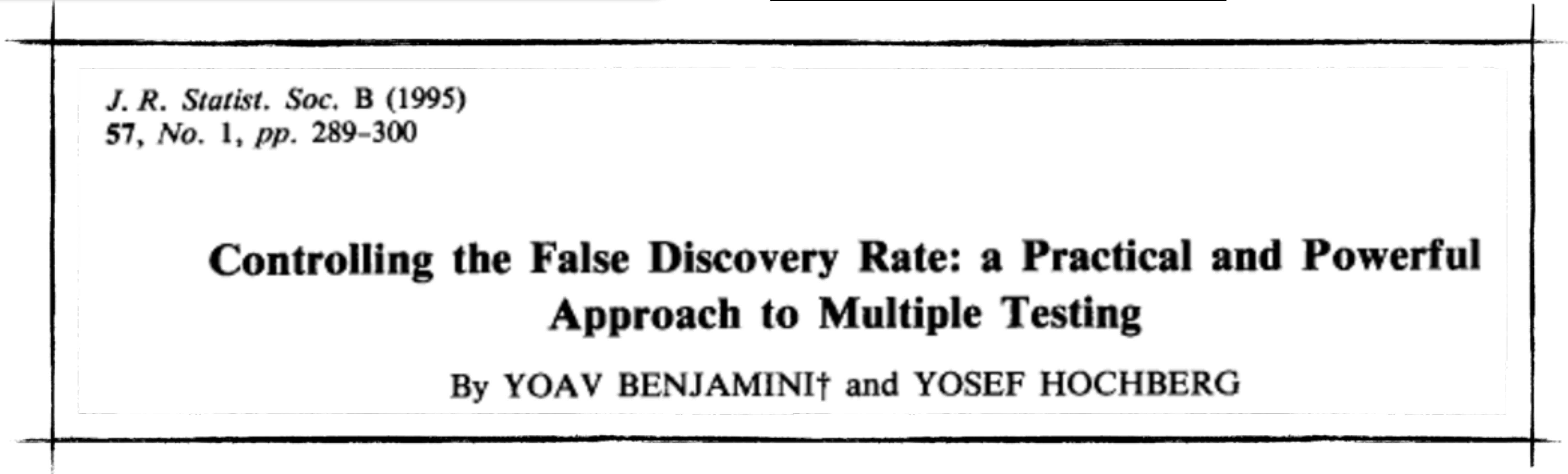
Citation Counts

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

Can the ≤ hold with equality?

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

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

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

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

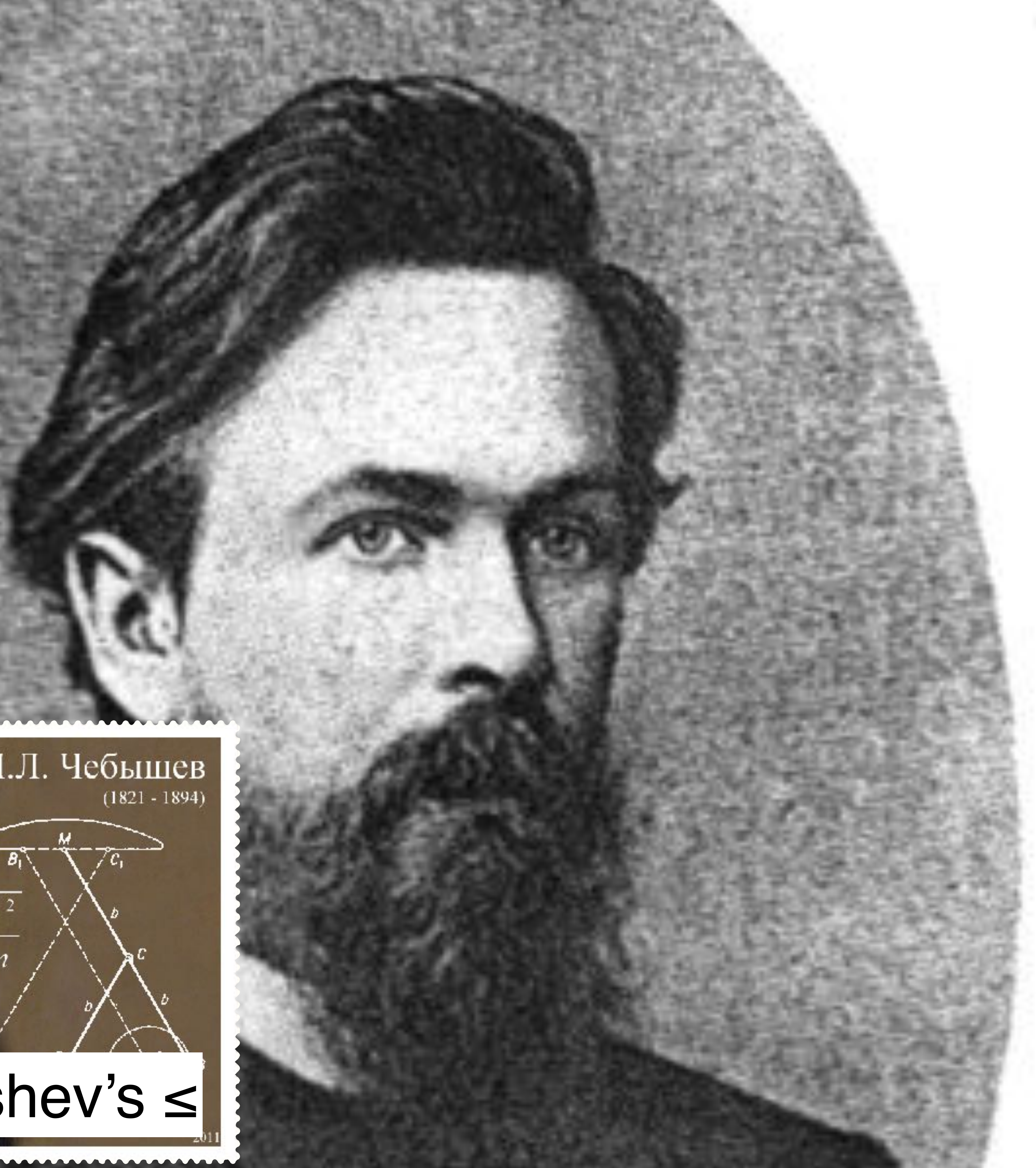
Limited to ≤ that hold for all distributions

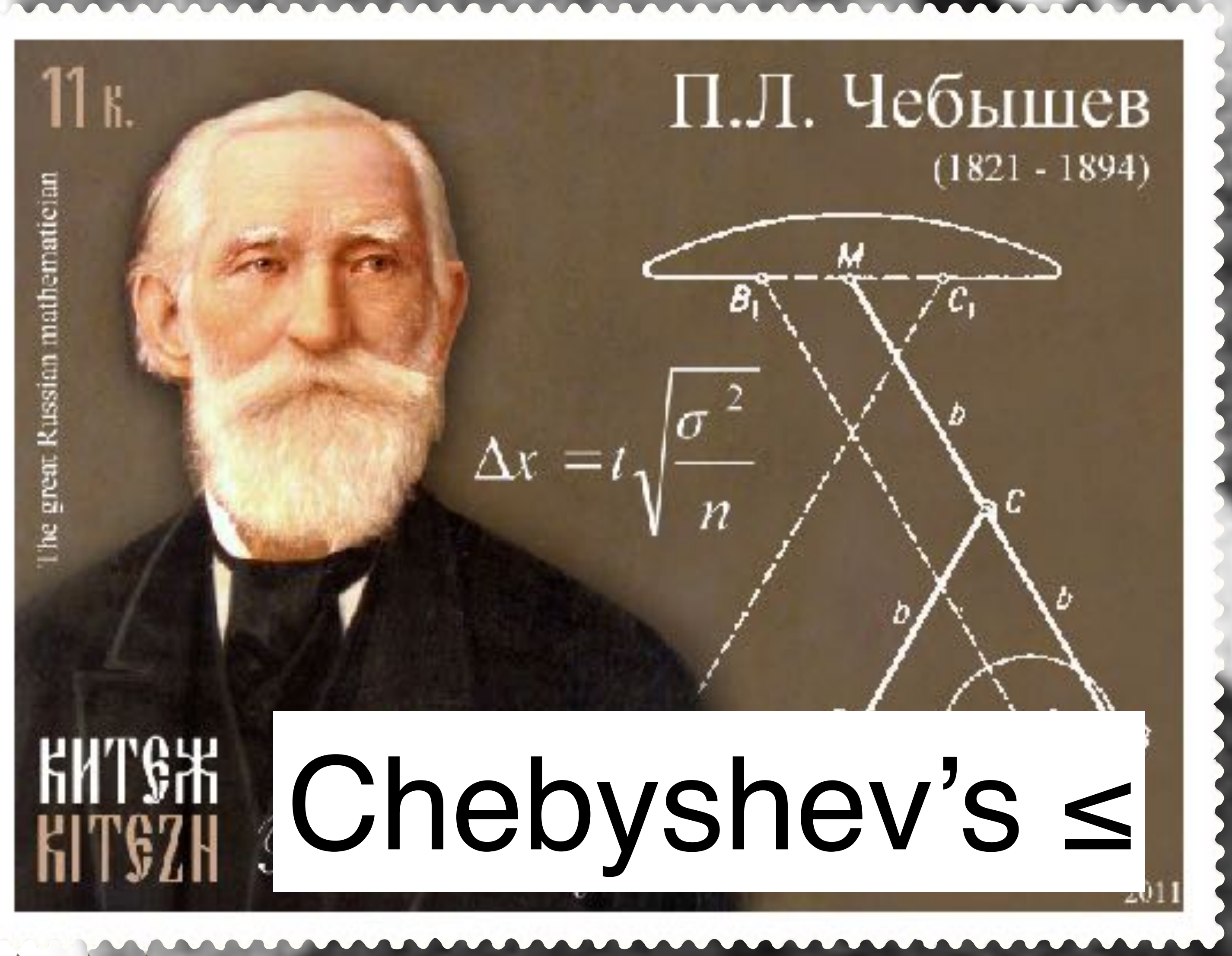
Strengthen

Add assumptions

Markov's ≤

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





Chebyshev's ≤

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

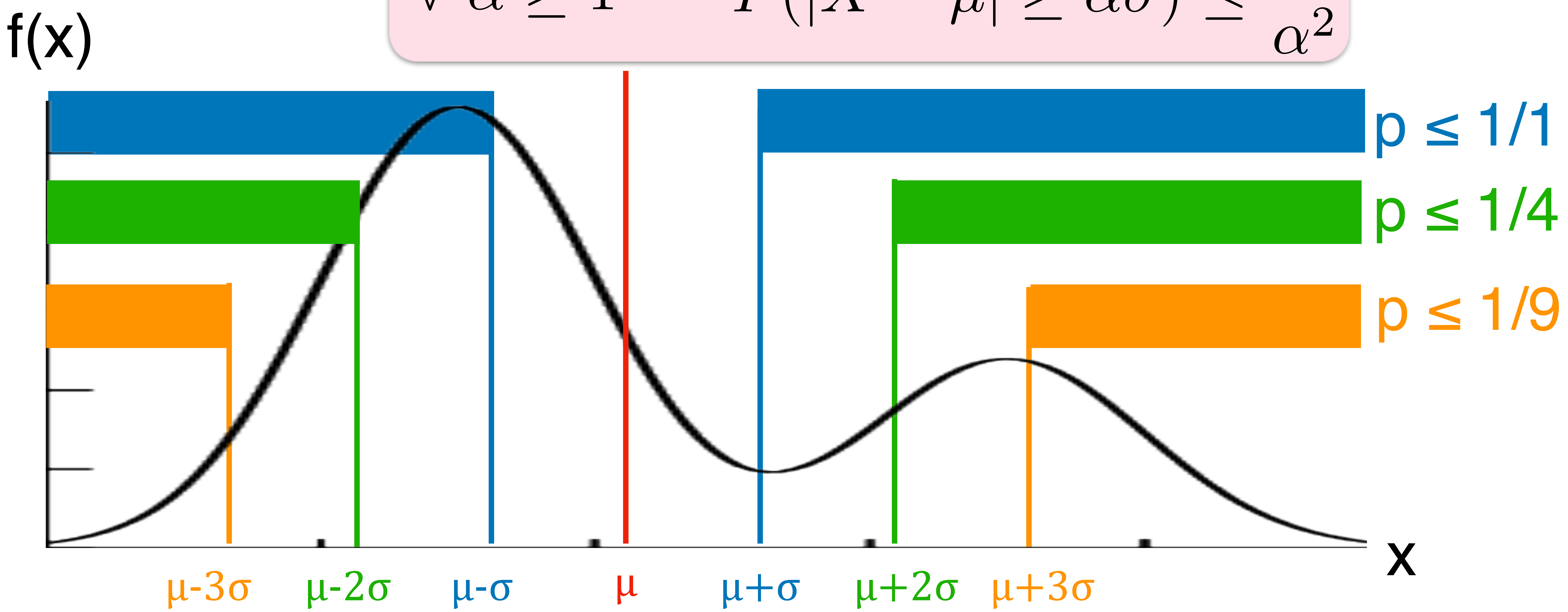
Probability that any X is more than $\alpha\sigma$ times further from μ is $\leq 1/\alpha^2$

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

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

Y Soon

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

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

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

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

Markov

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

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 μ = 1M responses with σ = 50K

Bound P(0.8M < # responses < 1.2M)

$0.8M = \mu - 4\sigma$

$1.2M = \mu + 4\sigma$

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

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

$$\geq 1 - 1/16$$

$$= 15/16$$

Mark x Che

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

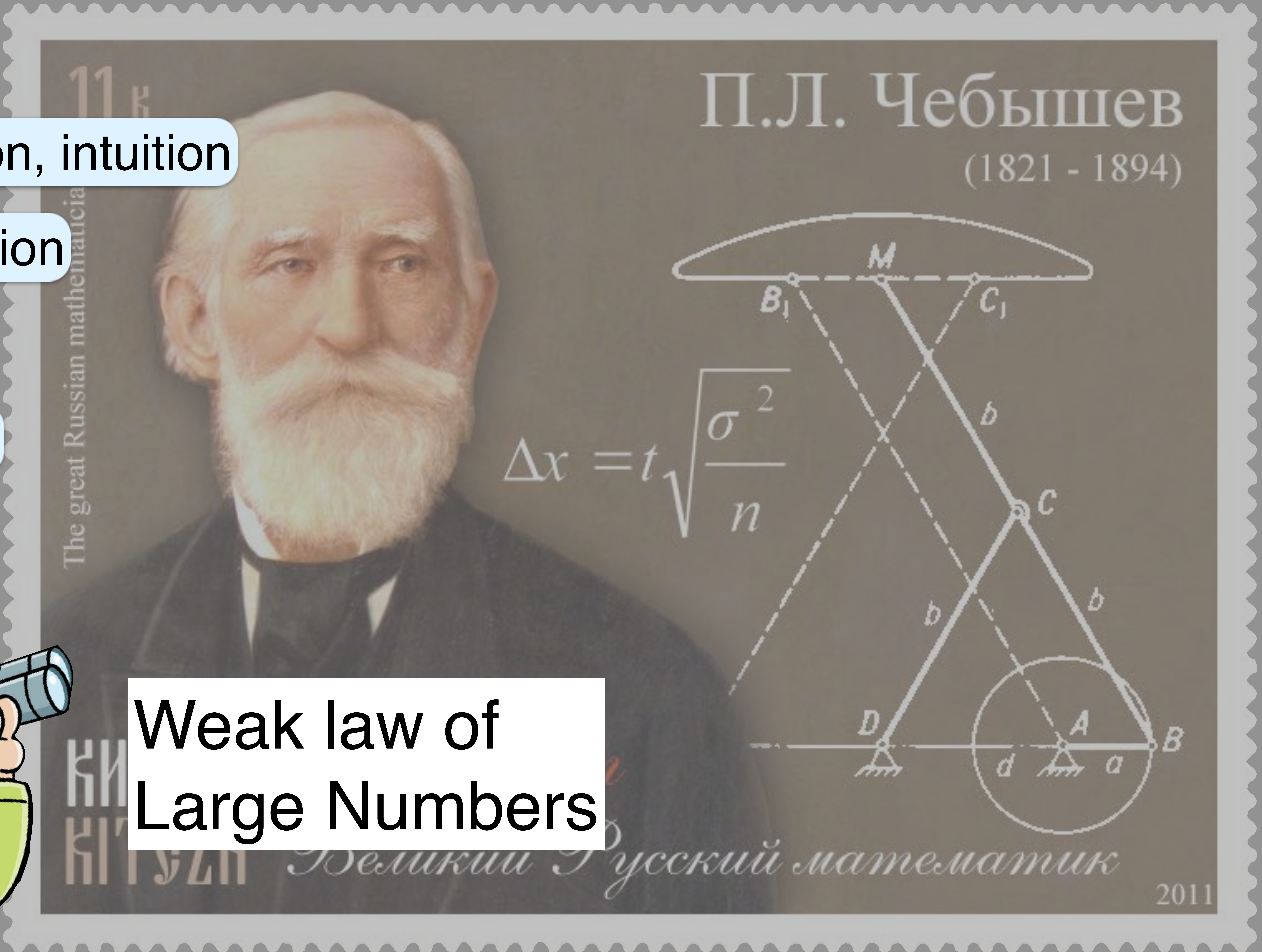
Weak law of Large Numbers

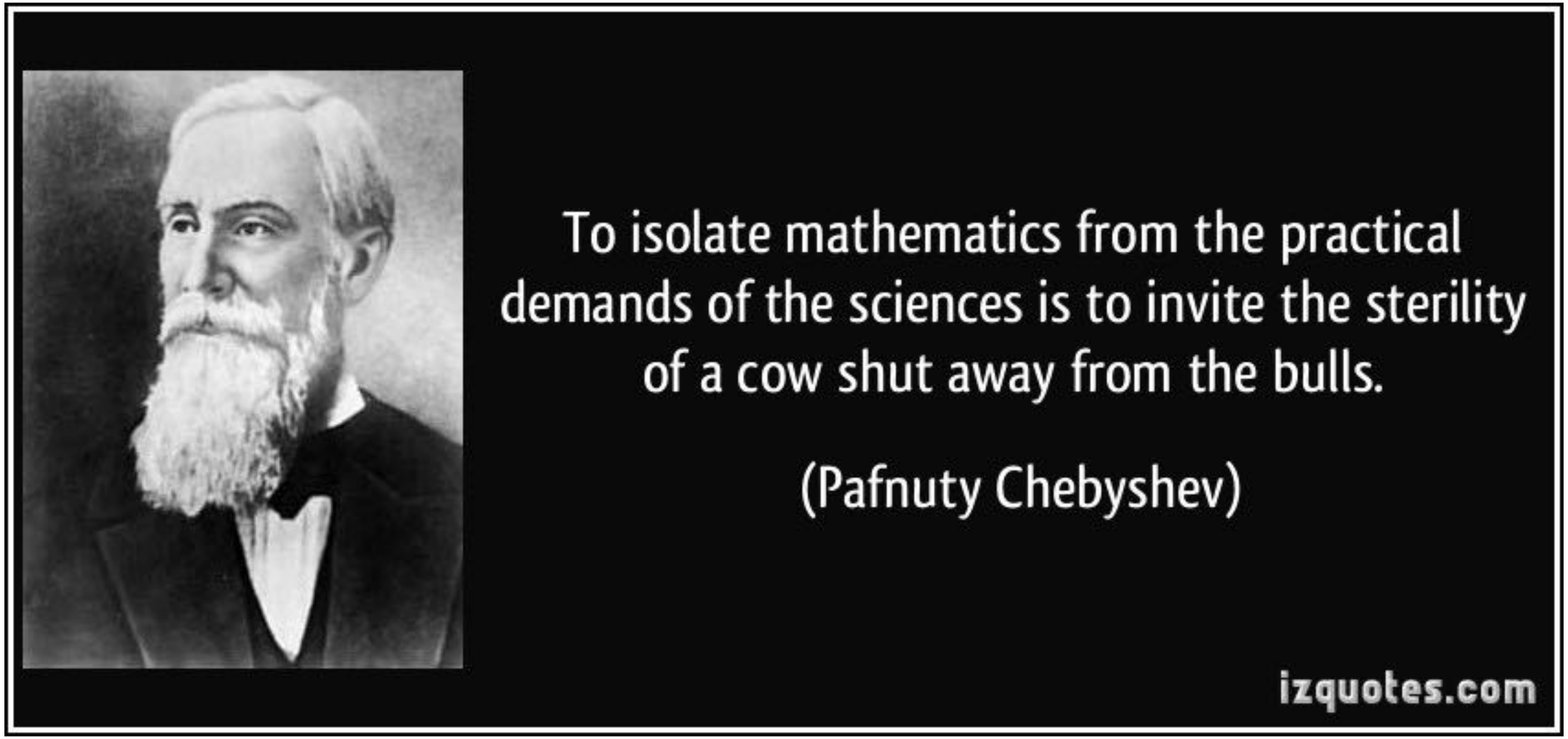
Motivation, intuition

Formulation

Proof

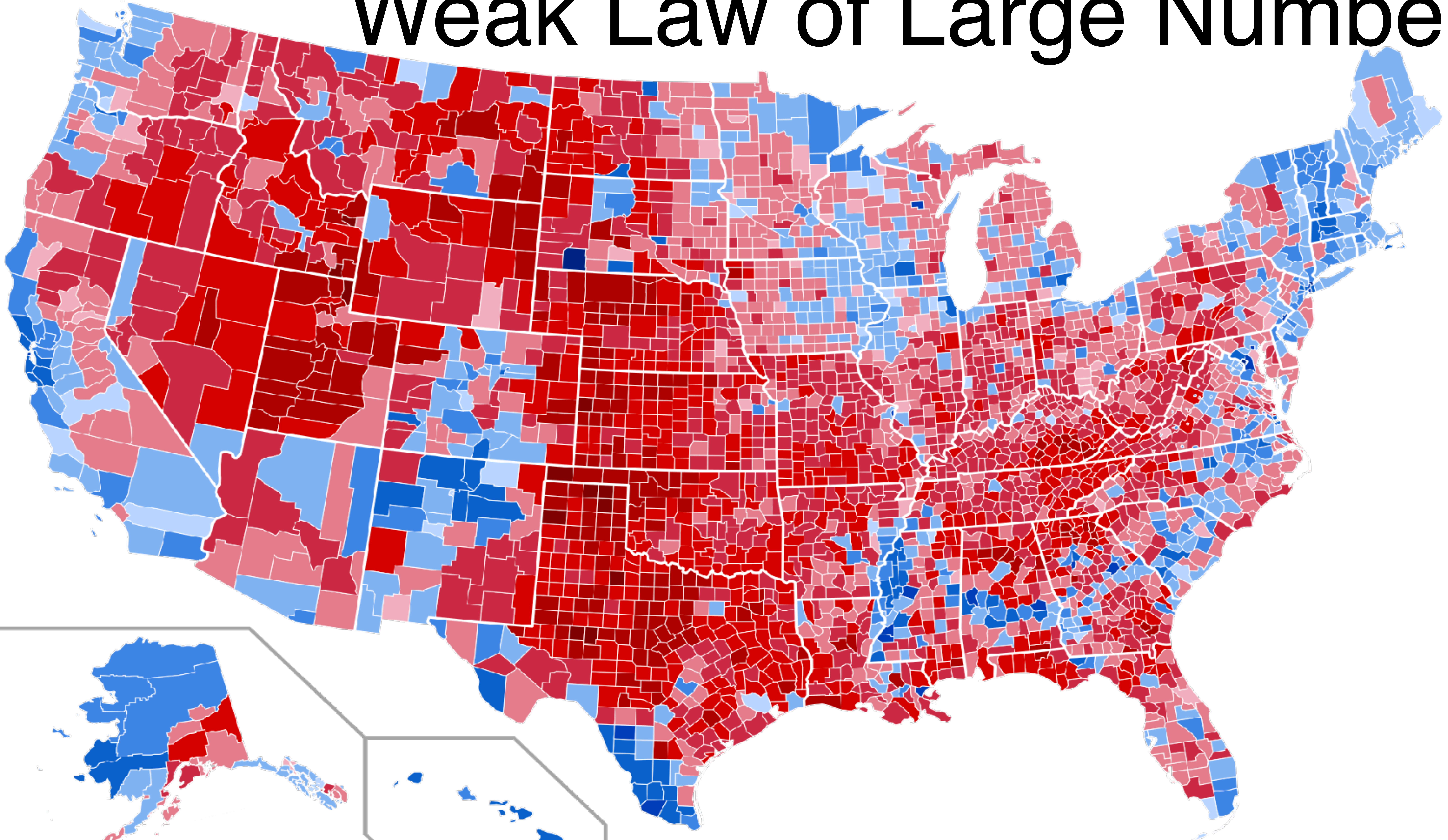
Example





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 $n \to \infty$

$\overline{X^n}$ 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

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

Variance

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

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 $T_i = t+Z_i$ 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