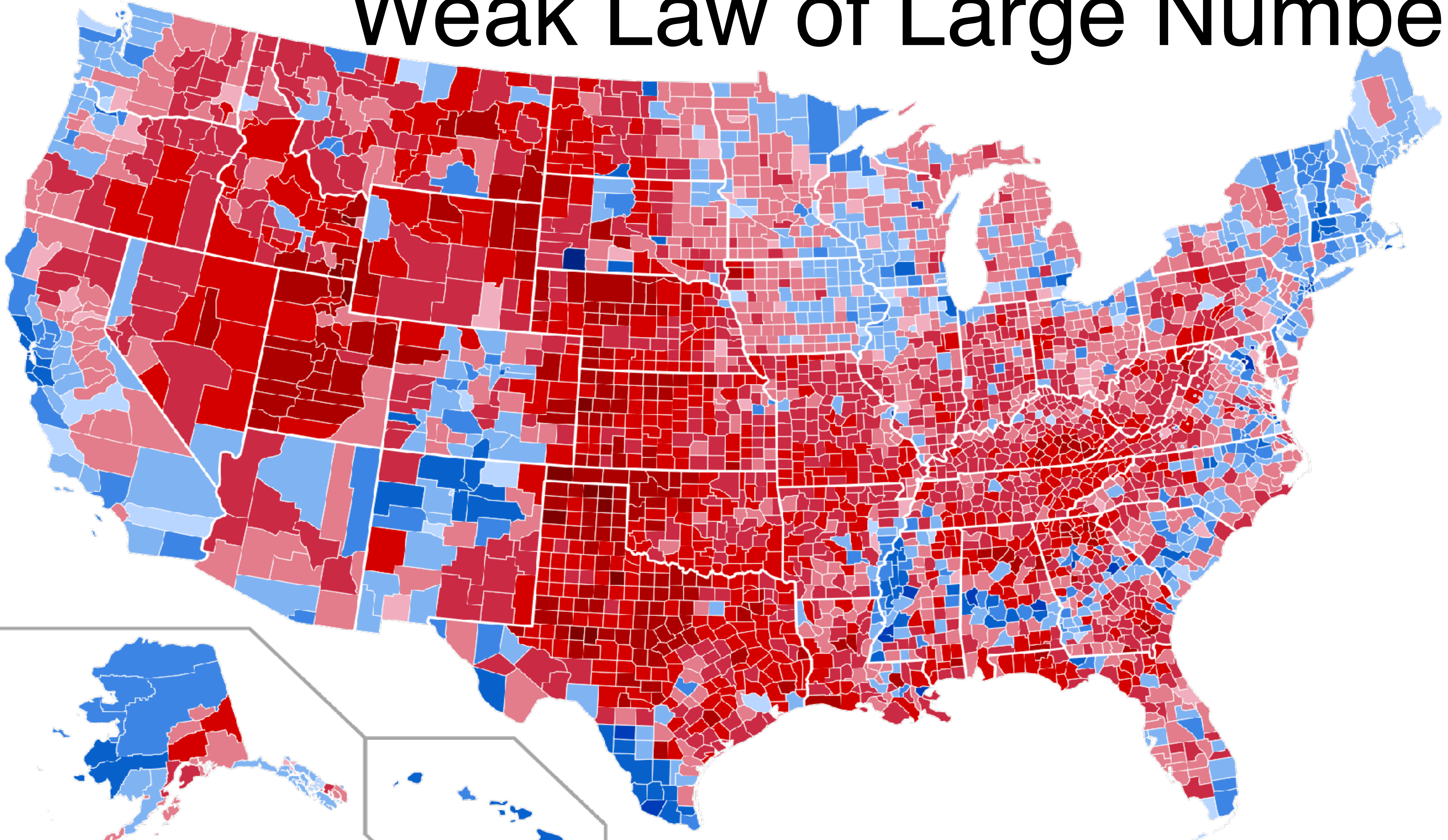
Weak Law of Large Numbers



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 $\overline{X^n} \stackrel{\text{def}}{=} \frac{1}{n} \sum X_i$ $\quad \sum = \sum_{i=1}^{n}$

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

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

$$n \to \infty$$

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$) \searrow 0$ with n

For every $\delta > 0$ $\quad P(|X_n - Y| \geq \delta) \searrow 0$ 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

Weak Law of Large Numbers

Next

Stronger bounds via
The Chernoff Bound and
Moment Generating Functions