Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev’s enduring inequality, the magisterial law of large numbers

Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.

Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

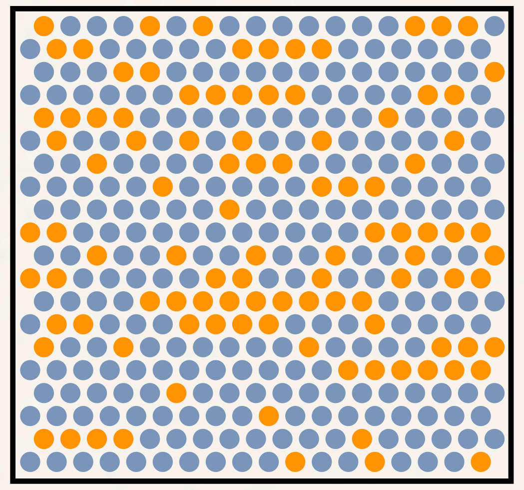
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll



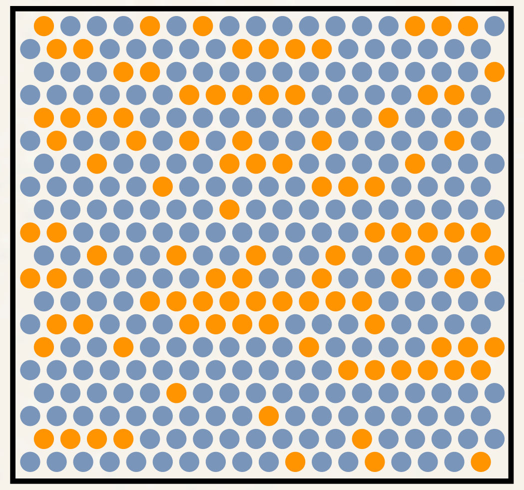
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll



Repeated independent trials: $X_1, X_2, \dots, X_n, \dots \sim \text{Bernoulli}(p)$

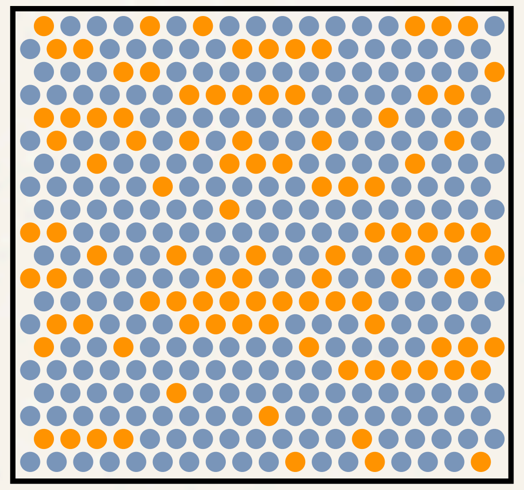
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll



Repeated independent trials: $X_1, X_2, \ldots, X_n, \cdots \sim \text{Bernoulli}(p)$

Accumulated successes: $S_n = X_1 + X_2 + \cdots + X_n \sim \text{Binomial}(n, p)$

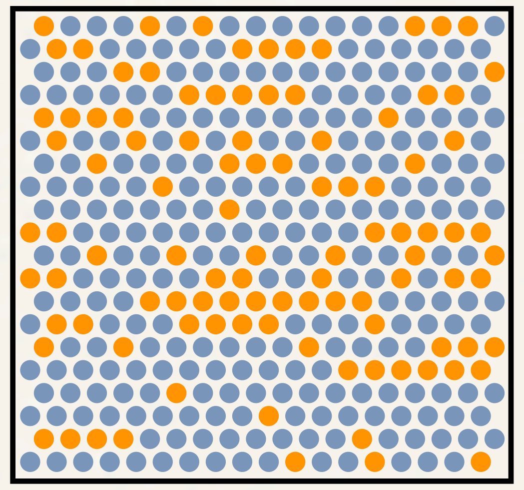
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll



Repeated independent trials: $X_1, X_2, \dots, X_n, \dots \sim \text{Bernoulli}(p)$

Accumulated successes: $S_n = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$

Estimate the unknown population proportion p by the relative frequency of successes S_n/n

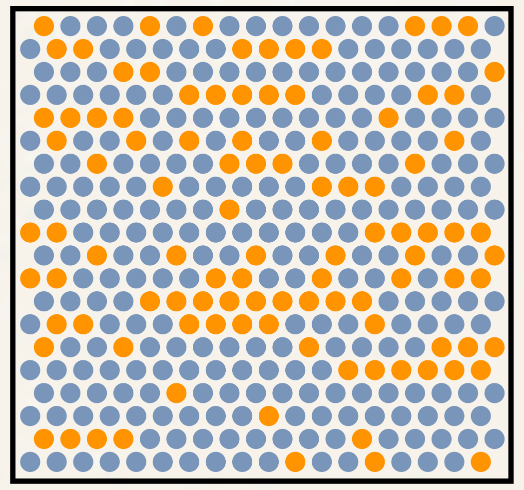
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll

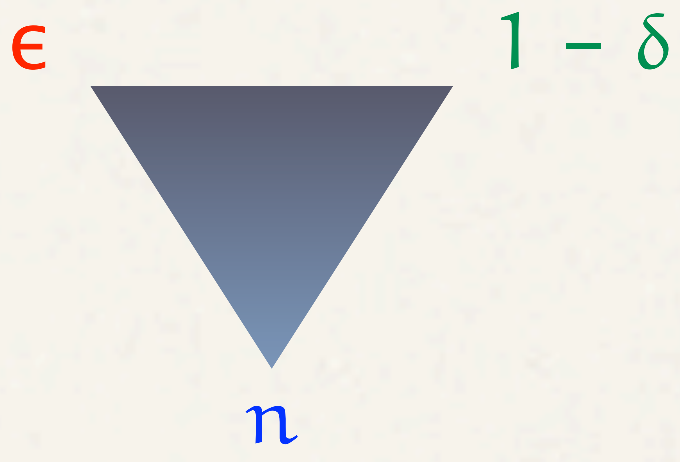


Repeated independent trials: $X_1, X_2, \dots, X_n, \dots \sim \text{Bernoulli}(p)$

Accumulated successes: $S_n = X_1 + X_2 + \cdots + X_n \sim \text{Binomial}(n, p)$

Estimate the unknown population proportion p by the relative frequency of successes S_n/n

The dance of error, confidence, and sample size



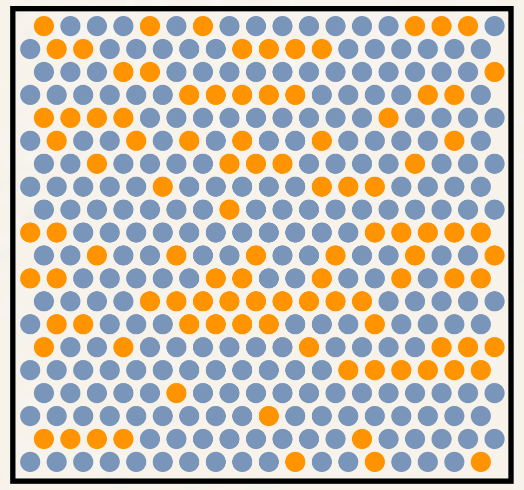
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll

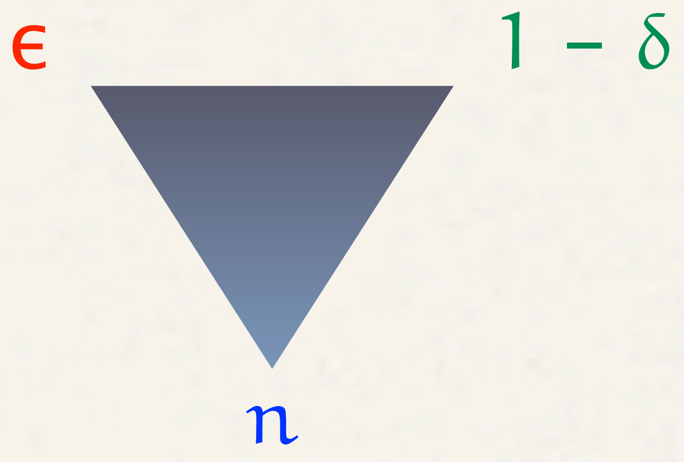


Repeated independent trials: $X_1, X_2, \dots, X_n, \dots \sim \text{Bernoulli}(p)$

Accumulated successes: $S_n = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$

Estimate the unknown population proportion p by the relative frequency of successes S_n/n

The dance of error, confidence, and sample size



We want:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \delta$$

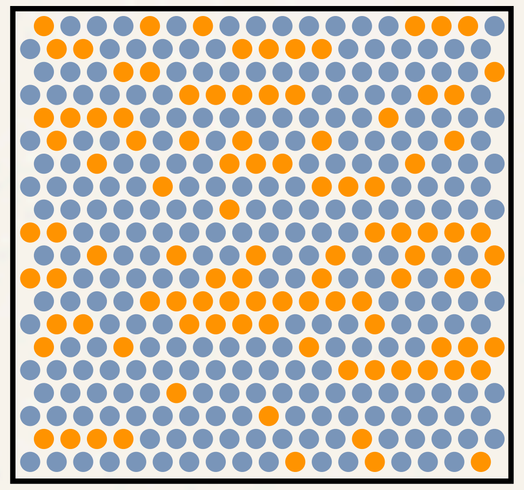
Summary of Tableau 11, Part 1

The fabulous limit laws
Chebyshev's enduring inequality, the magisterial law of large numbers

Random sample

- Independence.
- The subtlety of bias.

A model of a poll

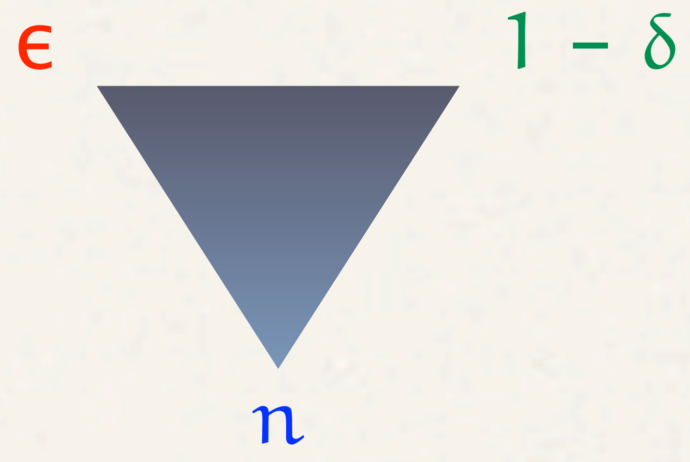


Repeated independent trials: $X_1, X_2, \dots, X_n, \dots \sim \text{Bernoulli}(p)$

Accumulated successes: $S_n = X_1 + X_2 + \cdots + X_n \sim \text{Binomial}(n, p)$

Estimate the unknown population proportion p by the relative frequency of successes S_n/n

The dance of error, confidence, and sample size



We want:

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \le \delta$$

Chebyshev's inequality, the law of large numbers

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \le \frac{1}{4n\epsilon^2} \to 0 \qquad (n \to \infty)$$

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1-\delta$.

Error ϵ	Confidence $1-\delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

Applications

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Applications

- Why polls work:

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Applications

- Why polls work:
  - Estimate p by S_n/n.

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Applications

- Why polls work:
  - Estimate p by S_n/n.
  - A sample of size 5,556 achieves error of no more than 3% with confidence 95%.

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

Applications

- Why polls work:
  - Estimate p by S_n/n.
  - A sample of size 5,556 achieves error of no more than 3% with confidence 95%.
- Why drug testing works:

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

Applications

- Why polls work:
  - Estimate p by S_n/n.
  - A sample of size 5,556 achieves error of no more than 3% with confidence 95%.
- Why drug testing works:
  - Randomised double sample (drug, placebo): estimate the chances of the observed discrepancy if the drug were ineffective (or has a dangerous side effect).

If $n \geq 1/(4\epsilon^2\delta)$ then the estimate S_n/n of p has an error of no more than ϵ with confidence at least $1 - \delta$.

Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

Applications

- Why polls work:
  - Estimate p by S_n/n.
  - A sample of size 5,556 achieves error of no more than 3% with confidence 95%.
- Why drug testing works:
  - Randomised double sample (drug, placebo): estimate the chances of the observed discrepancy if the drug were ineffective (or has a dangerous side effect).
  - A doubling of sample size for a given confidence.