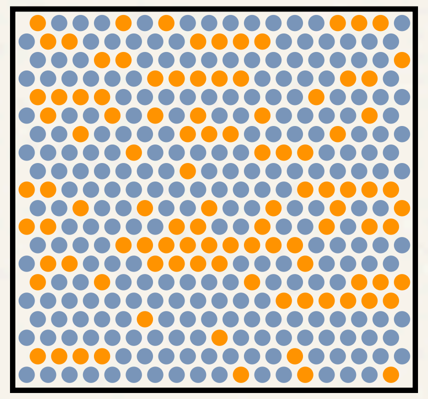
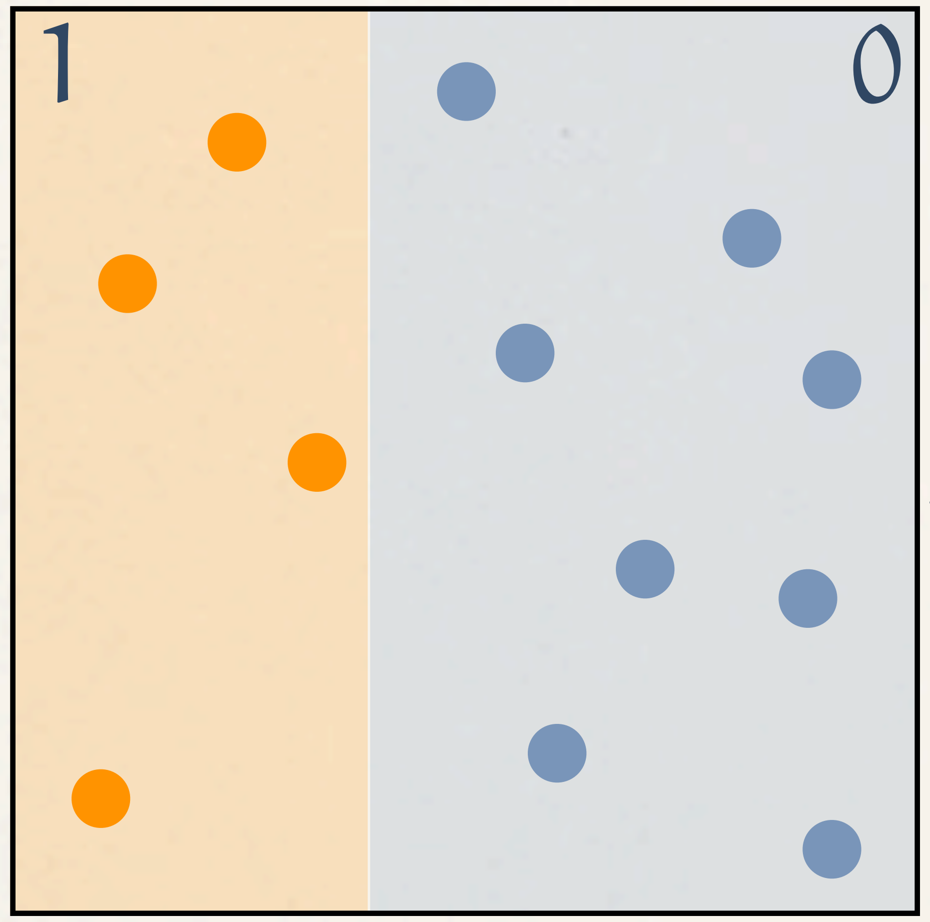
Why polls *really* work

A model for a poll



Fixed but unknown proportion p



Fixed but unknown proportion $q = 1 - p$

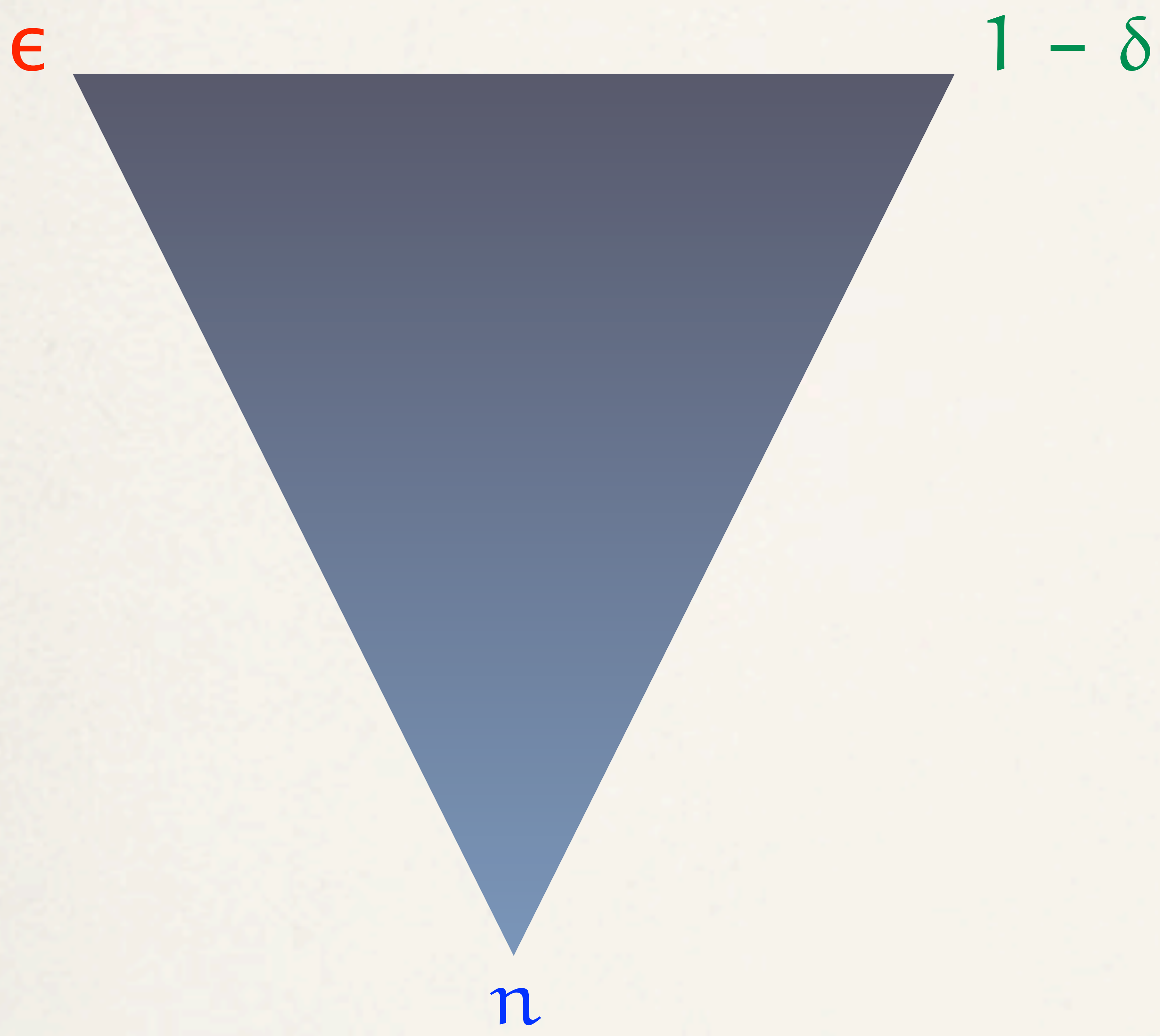
X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}
0	1	1	0	0	0	0	1	0	0	1	0

Bernoulli(p) trials: $X_1, X_2, \dots, X_n = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q. \end{cases}$

Accumulated successes: $S_n = X_1 + X_2 + \cdots + X_n$

Estimate the fixed, but unknown, bias p by the chance-driven sample mean S_n/n.

How are the error ϵ, the confidence $1 - \delta$, and the sample size n related?



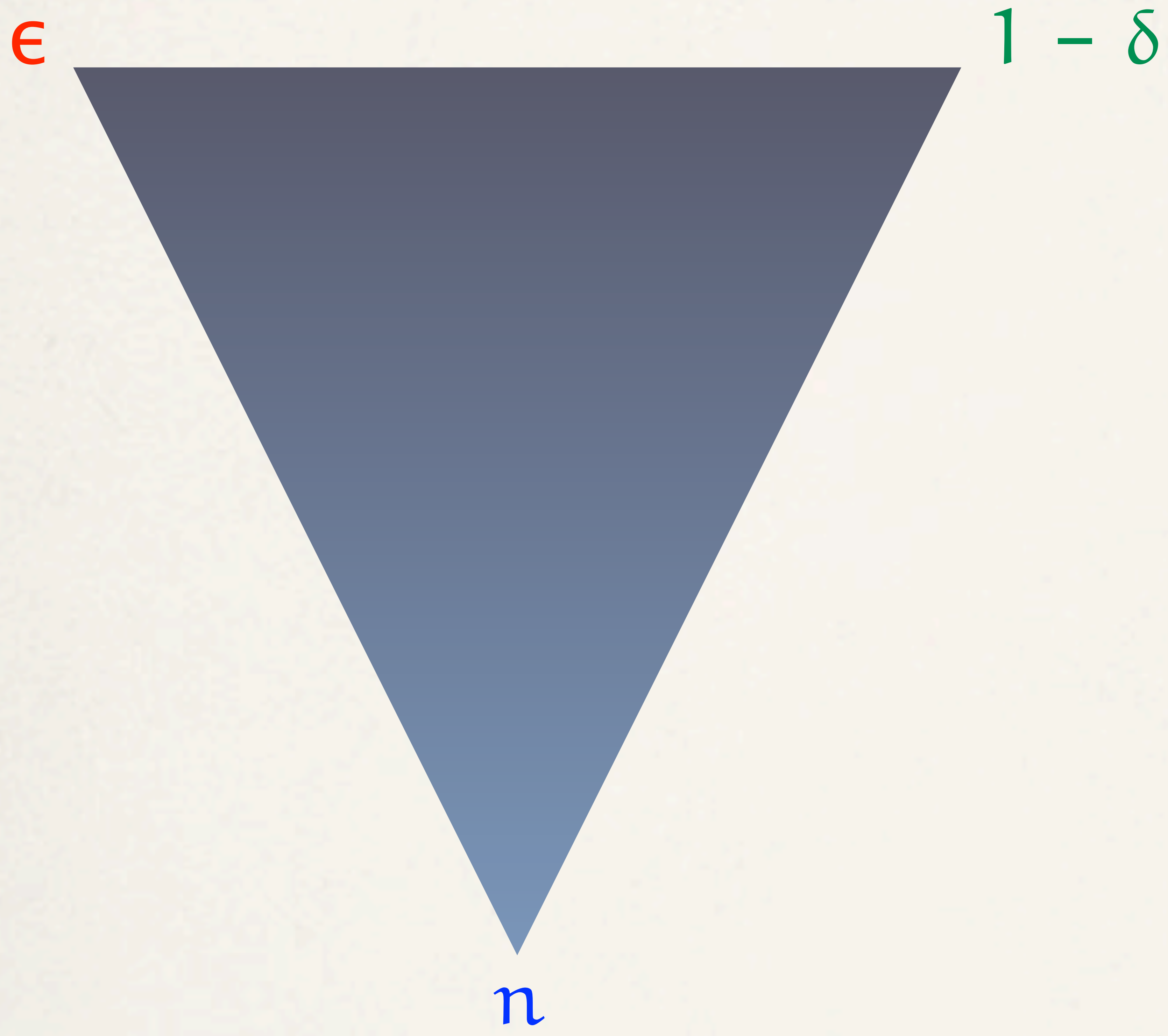
Error ϵ	Confidence $1 - \delta$	Sample size n
0.10	0.90	250
0.05	0.95	2000
0.03	0.95	5556

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \frac{1}{4n\epsilon^2} \leq \delta$$

Chebyshev!

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

How are the error ϵ, the confidence $1 - \delta$, and the sample size n related?



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

$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \underbrace{\leq \frac{1}{4n\epsilon^2}}_{\text{Chebyshev!}} \leq \delta$$

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