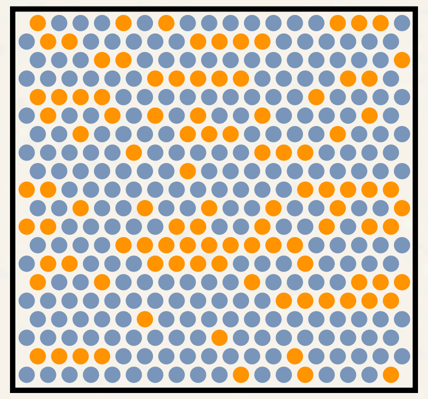
Why polls work

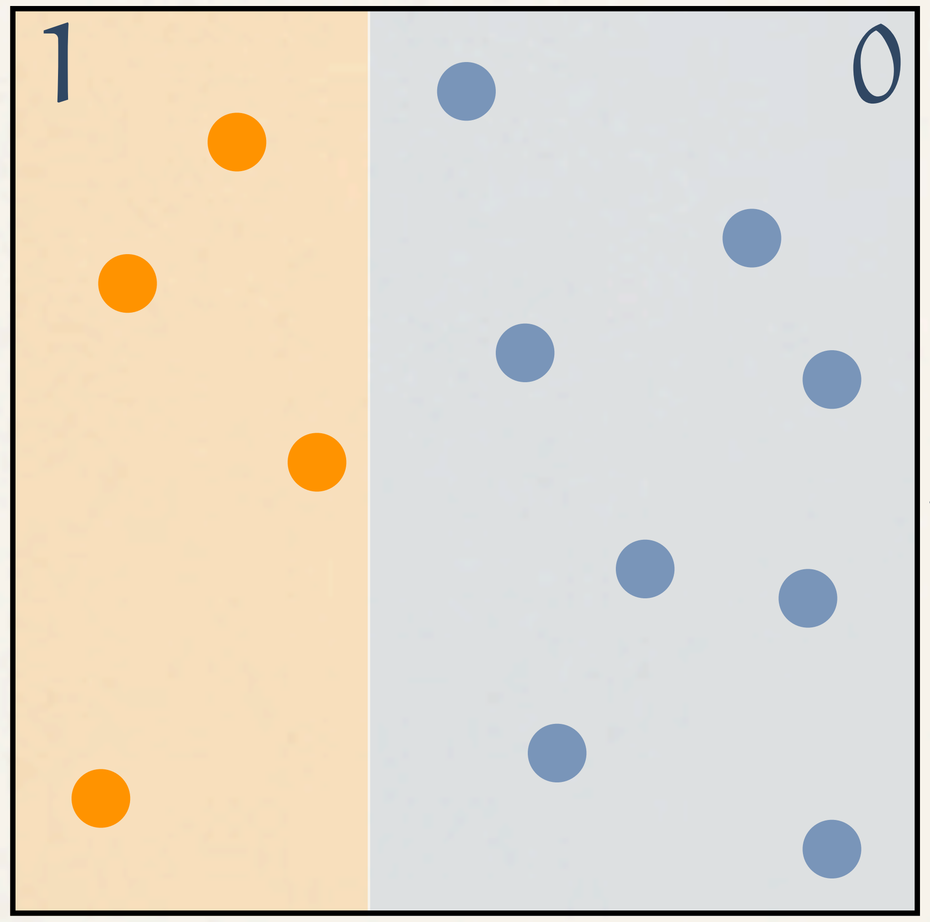
$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \leq \frac{1}{4n\epsilon^2}$$

A model for a poll



Random sample: repeated independent trials

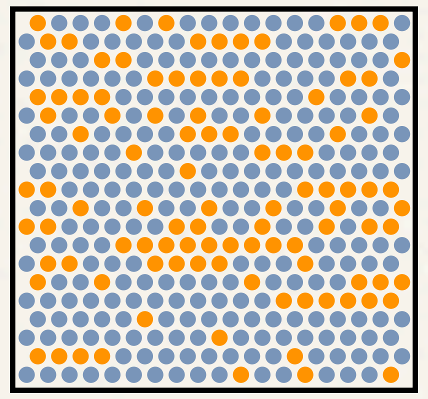
Fixed but unknown proportion p



Fixed but unknown proportion $q = 1 - p$

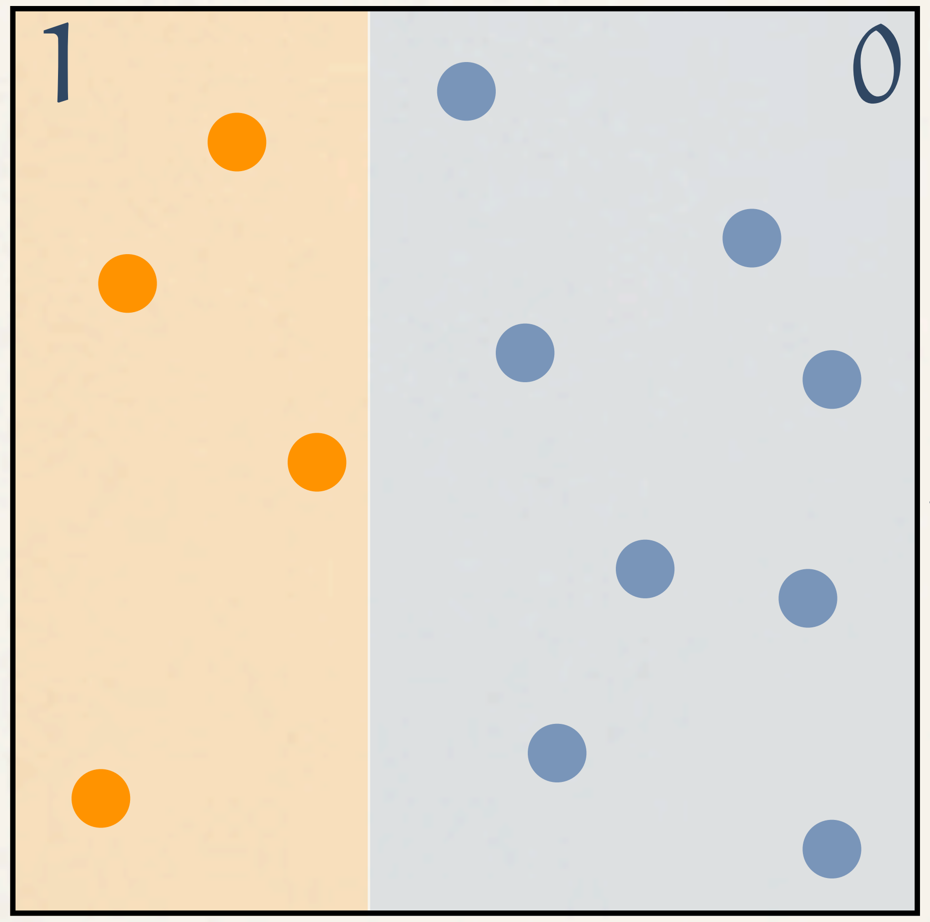
Bernoulli(p) trials: $X_1, X_2, \dots, X_n = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } q. \end{cases}$

A model for a poll



Random sample: repeated independent trials

Fixed but unknown proportion p

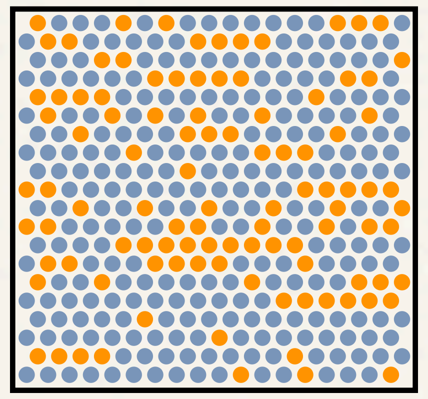


Fixed but unknown proportion $q = 1 - p$

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

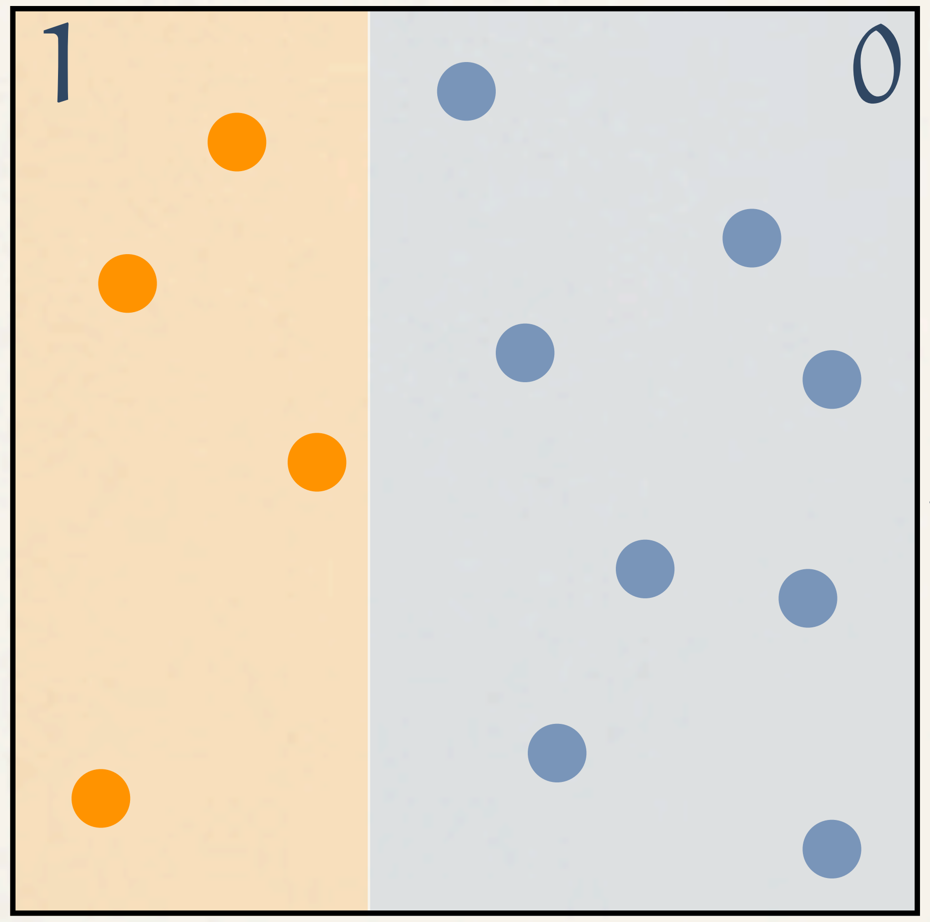
Binomial(n, p): $S_n = X_1 + X_2 + \cdots + X_n$

A model for a poll



Random sample: repeated independent trials

Fixed but unknown proportion p



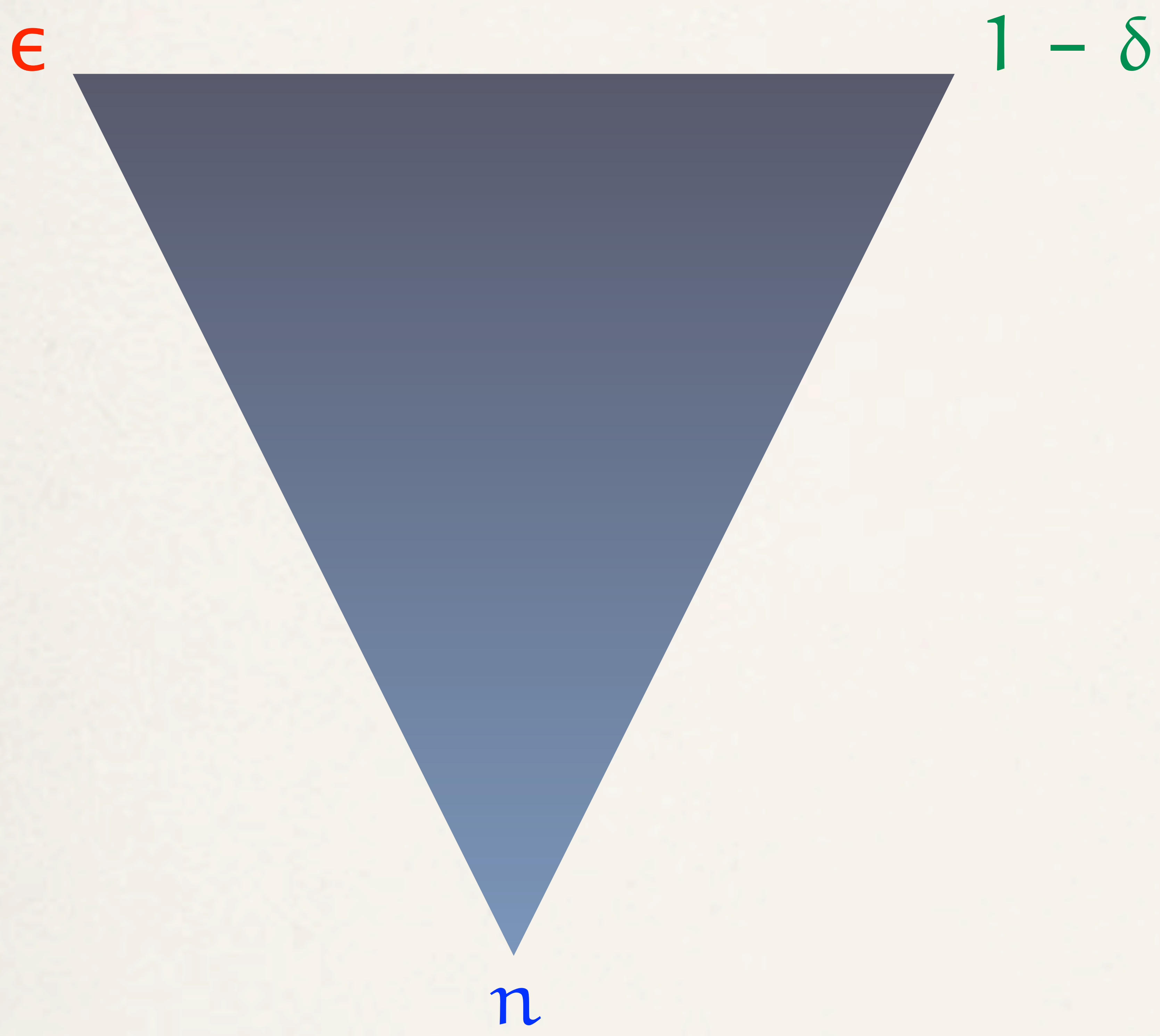
Fixed but unknown proportion $q = 1 - p$

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

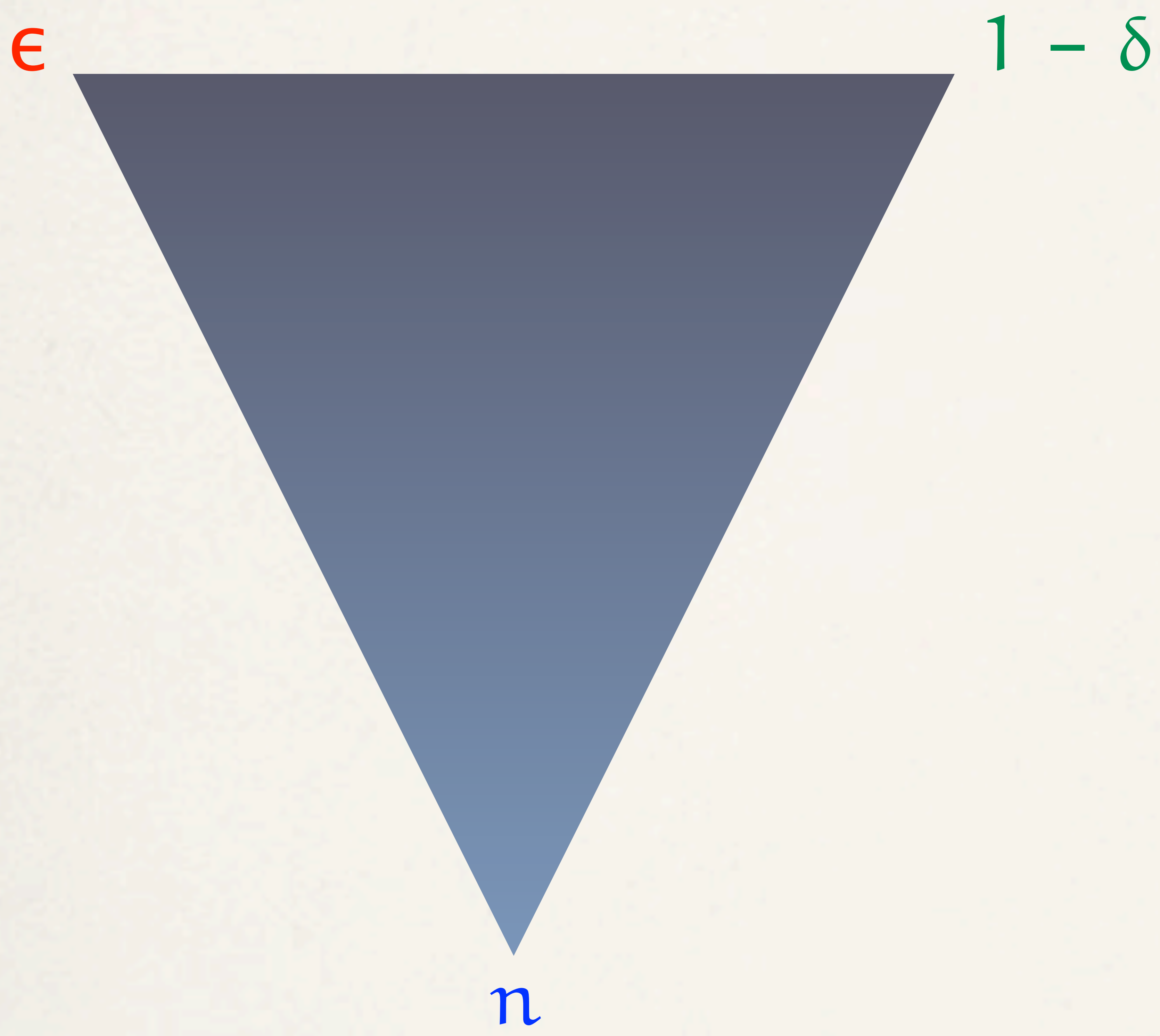
Binomial(n, p): $S_n = X_1 + X_2 + \cdots + X_n$

Estimate the fixed but unknown population proportion p by the relative frequency S_n/n of accumulated successes in the sample.

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

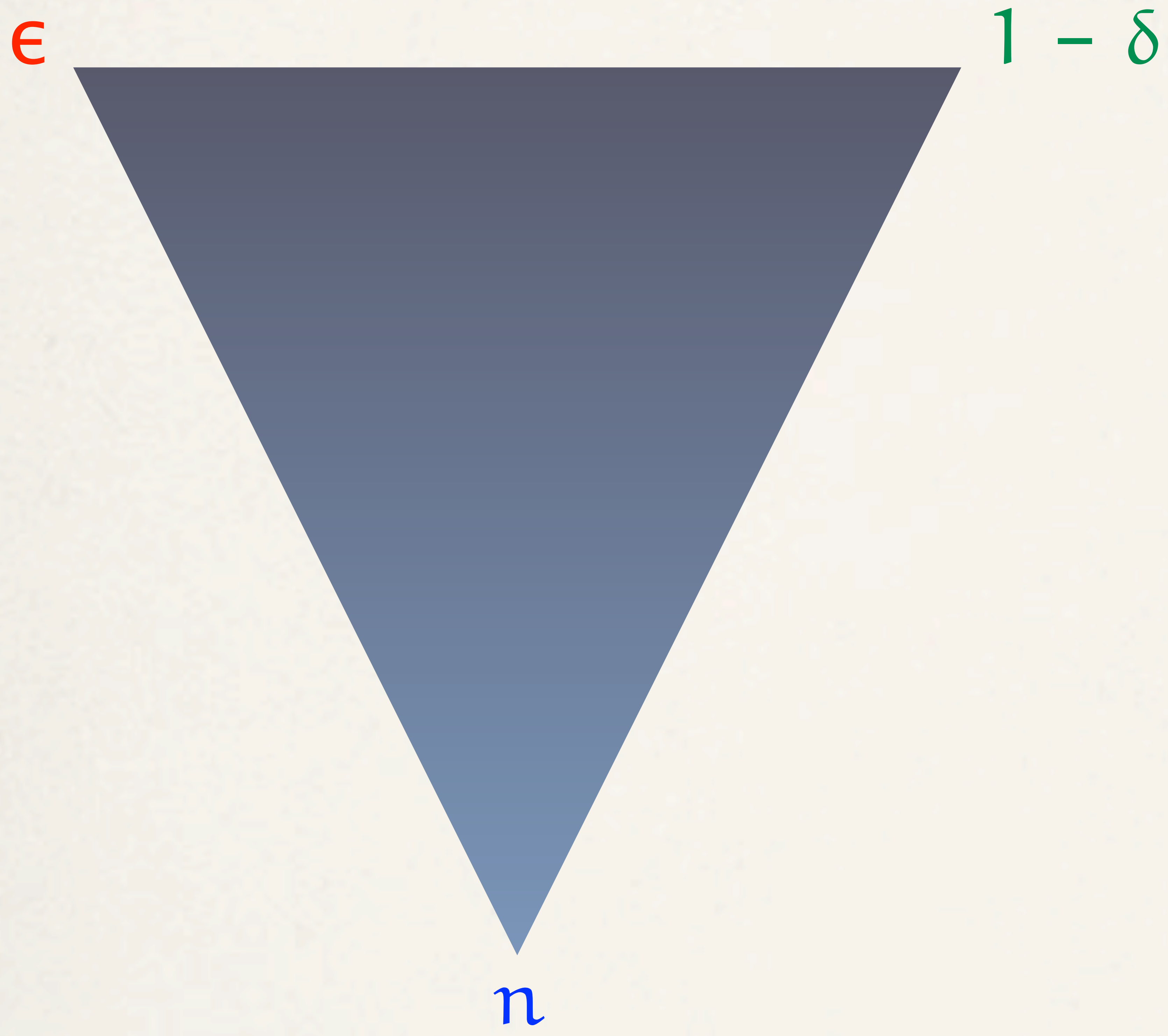


How are the error ϵ, the confidence $1 - \delta$, and the sample size n related?



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

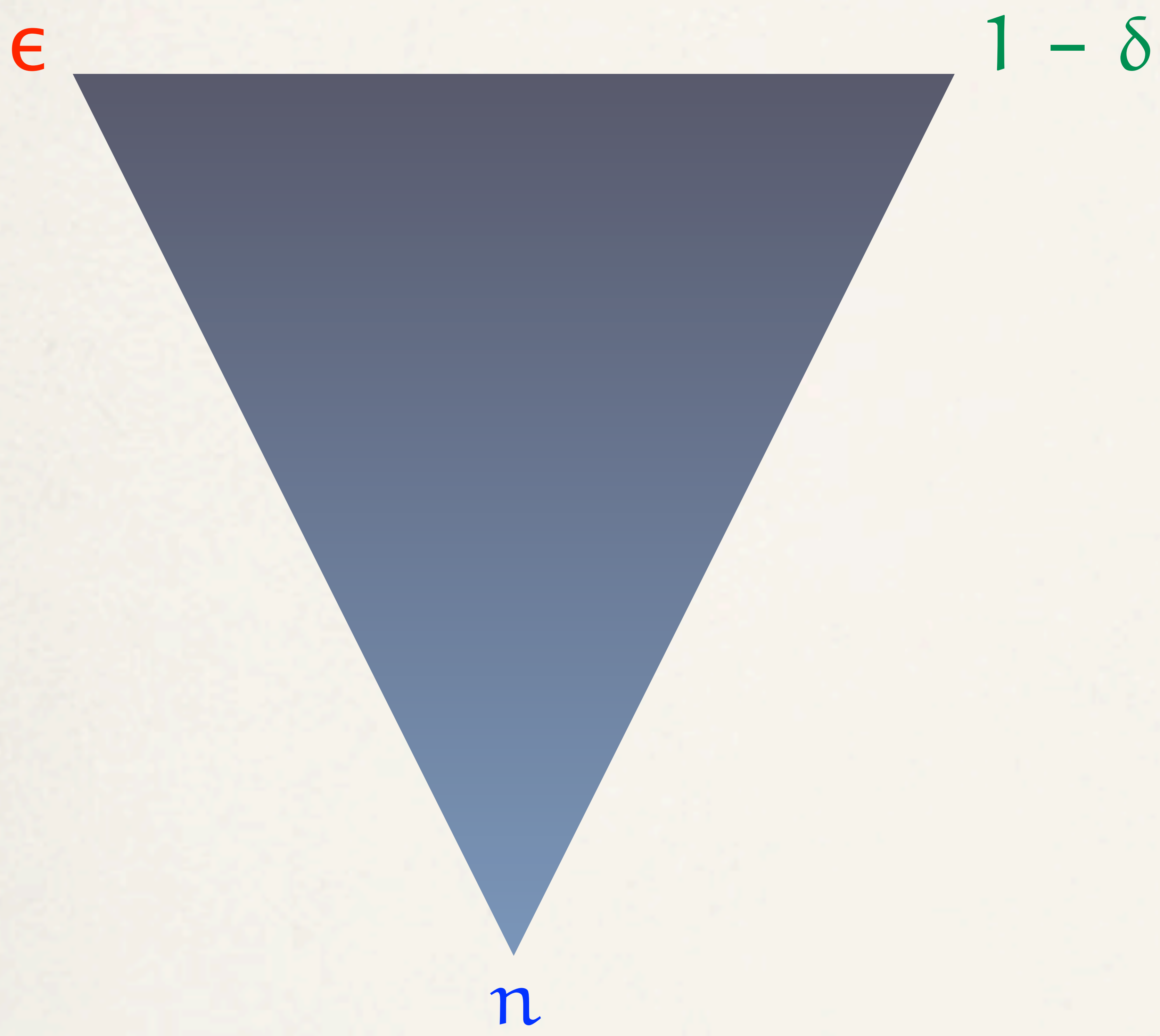
How are the error ϵ, the confidence $1 - \delta$, and the sample size n related?



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

Chebyshev!

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

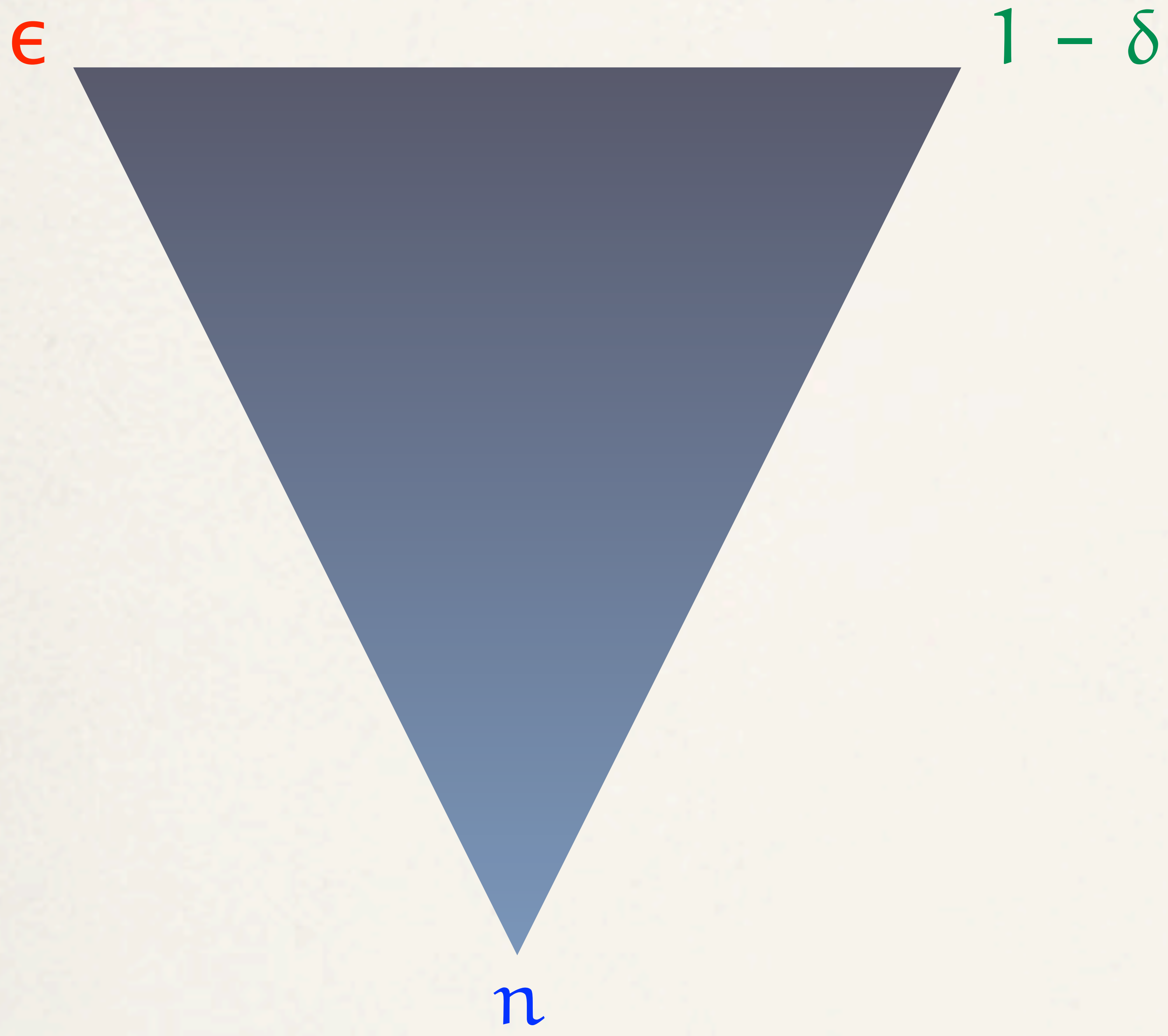


$$\mathbf{P}\left\{\left|\frac{S_n}{n} - p\right| > \epsilon\right\} \le \frac{1}{4n\epsilon^2} \le \delta$$

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

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

Chebyshev!

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

Slogan

A relatively small, honest, random sample *whose size does not depend upon the size of the underlying population or its composition* gives a good estimate of the underlying population proportions (sentiments).