Suppose we are doing a binomial experiment with estimator

$$\tilde{\theta} = \frac{Y}{n}$$

And we want to calculate the probability

$$P(|\tilde{\theta} - \theta| \leq d) \quad \text{for some } d$$

↳ true mean (can be unknown)

Here, we are essentially asking: what is the probability that the difference between the estimator's output and the true mean (which is unknown) is within some threshold d?

- Suppose we collect 100 samples, all with size n. If p is the probability value calculated above, then in 100p samples, the point estimate provided by the estimator $\tilde{\theta}$ will be within d units of the true mean.

In most of these problems, this is done by approximating the binomial distribution using the normal/ Gaussian distribution.

Recall that in the Gaussian distribution, variables Y are standardized to z-scores using the formula

$$z = \frac{Y - \mu}{\sigma} \sim G(0, 1)$$

Here, $Y \to \tilde{\theta}$; $\mu \to \theta$ (sample mean)

To complete the standardization, we need to compute the standard deviation (denominator).

For binomial distributions:

$$sd(X) = \sqrt{Var(Y)} = \sqrt{n\theta(1-\theta)}$$

However, we are calculating the standard deviation of the *estimator*, which is equal to Y/n. Then:

$$\begin{aligned} sd(\tilde{\theta}) = sd(Y/n) = Var(Y/n)^{1/2} &= (n^{-2} Var(Y))^{1/2} \\ &= (n^{-2} n\theta(1-\theta))^{1/2} \\ &= \sqrt{\frac{\theta(1-\theta)}{n}} = \sigma \end{aligned}$$

$$\text{Then: } P(|\tilde{\theta} - \theta| \leq d) = P\left(-\frac{d}{\sigma} \leq \frac{\tilde{\theta} - \theta}{\sigma} \leq \frac{d}{\sigma}\right)$$

As such, an *asymptotic pivotal quantile* for binomial distributions is

$$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$$

Confidence Intervals

General formula: $\pm z \cdot sd(\tilde{\theta})$

Ex. For binomial, a 95% CI is $\pm 1.96\sqrt{\frac{\theta(1-\theta)}{n}}$

Proof: An asymptotic pivotal quantity for $\tilde{\theta}$ is

$$\frac{\tilde{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}$$

which follows $G(0,1)$ as $n \to \infty$.
As such, we can set

$$P\left(-d \leq \frac{\tilde{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \leq d\right) \approx 0.95 \quad \text{for some } d$$

Then, we can set $d = 1.96$, since $P(-1.96 \leq z \leq 1.96) \approx 0.95$ for any z on the Gaussian dist.

$$\Rightarrow P\left(-1.96 \leq \frac{\tilde{\theta}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \leq 1.96\right) \approx 0.95$$

$$\Rightarrow P\left(-1.96\sqrt{\frac{\theta(1-\theta)}{n}} \leq \tilde{\theta}-\theta \leq 1.96\sqrt{\frac{\theta(1-\theta)}{n}}\right) \approx 0.95$$

$$\Rightarrow P\left(-1.96\sqrt{\frac{\theta(1-\theta)}{n}} - \tilde{\theta} \leq \theta \leq 1.96\sqrt{\frac{\theta(1-\theta)}{n}} - \tilde{\theta}\right) \approx 0.95$$

$$\Rightarrow P\left(\tilde{\theta}-1.96\sqrt{\frac{\theta(1-\theta)}{n}} \leq \theta \leq \tilde{\theta}+1.96\sqrt{\frac{\theta(1-\theta)}{n}}\right) \approx 0.95$$

$\Rightarrow P\left(\theta \in \left[\tilde{\theta} \pm 1.96\sqrt{\frac{\theta(1-\theta)}{n}}\right]\right) \approx 0.95$ ☆ If on test, end proof here

As such, the probability that the true mean is within 1.96 • sqrt(...) of the estimator is 0.95.

Thus, that is a 95% confidence interval.

PCs:

If σ is known: $\frac{\tilde{\theta} - \theta}{\sigma / \sqrt{n}} \sim G(0, 1)$ σ is population sd

If σ is not known: $\frac{\tilde{\theta} - \theta}{\tilde{\sigma}} \sim G(0, 1)$

Then get estimator for σ