Estimating a mean

Probably the most widely used formula for a confidence interval is

$$\bar{x} \pm 2\sqrt{s^2/n}$$

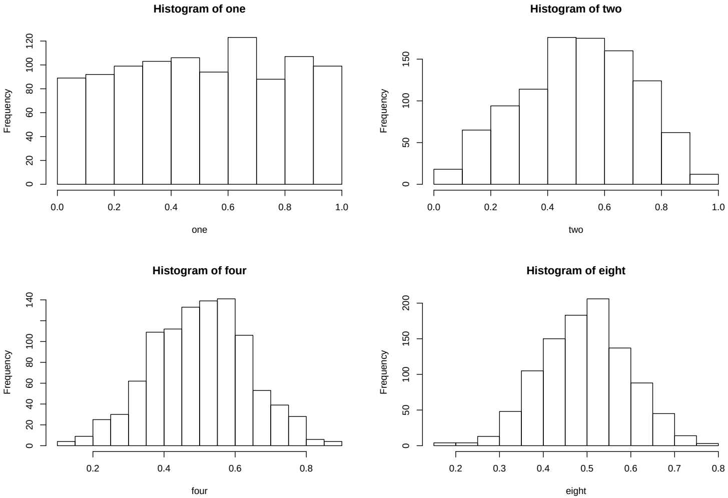
- $\bar{x}$ is the sample mean
- s^2 is the sample variance
- n is the sample size
- 2 is PJD's approximation to 1.96

Strictly, you should use a value c_n which depends on n, but is approximately 2 for reasonably large n, for example:

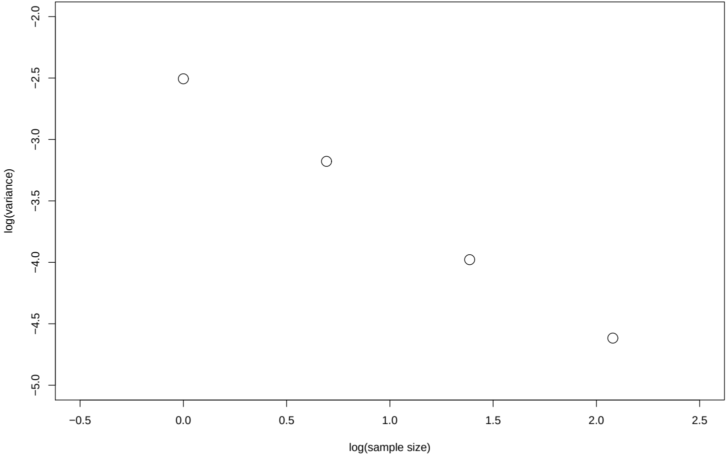
n	c_n
5	2.78
10	2.26
20	2.09
50	2.01
∞	1.96

Where does the formula for the confidence interval come from?

This diagram shows how the **distribution** of the sample mean changes with the sample size:



And this diagram shows how the variance of the sample mean changes with the sample size:



Conclusions

- sample means are approximately Normally distributed (symmetric, bell-shaped histogram)
- larger samples lead to more precise estimates
- the variance of an estimate is inversely proportional to the sample size, n
- the standard error of an estimate is therefore inversely proportional to $\sqrt{n}$
- Murphy's law of diminishing returns – doubling the sample size does not double the precision of your estimate