Nature of sampling distributions

- The sampling distribution depends on what X is. For sum, mean, proportion, median the sampling distribution is normal for sufficiently large N.
- For variance, it is a χ^2 distribution.
- More generally, CLT says that a sum of many independent and identically distributed (iid) random variables (or alternatively, random variables with specific types of dependence) will tend to be distributed according to one of a small set of attractor distributions (normal is the most common one).

Population size and sampling without replacement

- For a finite population of size N_p where we sample with sample size $N < N_p$ we have a different result:

$$\mu_{sd} = \mu, \quad \sigma_{sd} = \frac{\sigma}{\sqrt{N}} \sqrt{\frac{N_p - N}{N_p - 1}}$$

- If population is infinite (practically $N < 0.05 N_p$) or sampling is with replacement then we have the earlier values: $\mu_{sd} = \mu$ and $\sigma_{sd} = \frac{\sigma}{\sqrt{N}}$.
- For large $N \geq 30$ the sampling distribution is approximately normal with mean μ_{sd} and std. dev. σ_{sd} when population mean and std. dev. is finite and population is greater than twice sample size.

Sampling distribution rules of thumb

Approx. normal sampling distribution will have $\mu_{sd} = \mu$, $\sigma^2_{sd} = \frac{\sigma^2}{N}$.

Population	**Sampling type**	**Sampling distr.**	**N (sample size)**
Finite	with replacement	Approx. normal	$N \geq 30$
Finite	w/o replacement	Approx. normal	$N \geq 30,\ N \leq .05N_p$
Finite	w/o replacement	Approx. normal $\sigma^2_{sd} = \frac{\sigma^2}{N}(\frac{N_p - N}{N_p - 1})$	$N \geq 30,\ N > .05N_p$
Infinite	-	Approx. normal	$N \geq 30$

Application of sampling distributions

- Behaviour of sample mean - for example probability for bounds.
- Test claims. Our major use case.

Example 1: CLT appln

- Let $p(x) = \frac{3}{2}x^2$, sample size $n = 15$. For $-1 < x < 1$ what is the probability the sample mean falls between $-\frac{2}{5}$ and $\frac{1}{5}$?
- Using definition (needs integration) $\mu = 0$, $\sigma^2 = \frac{3}{5}$.

Example 1: CLT appln

- Let $p(x) = \frac{3}{2}x^2$, sample size $n = 15$. For $-1 < x < 1$ what is the probability the sample mean falls between $-\frac{2}{5}$ and $\frac{1}{5}$?
- Using definition (needs integration) $\mu = 0$, $\sigma^2 = \frac{3}{5}$.
- From CLT $\hat{\mu} = \mu = 0$ $\hat{\sigma^2} = \frac{\frac{3}{5}}{15} = \frac{1}{25}$.
- We need $P(-\frac{2}{5} < \bar{X} < \frac{1}{5}) = P(-2 < Z < 1)$ that is $0.8413 - 0.0228 = 0.8186$

Example 2: CLT appln

- Waiting time for ith person in a food queue, X_i. Assume claim is average waiting time is 2 mins.
- Distribution is a negative exponential distribution (point Poisson process).
- You take a sample of 36 persons and find average waiting time is 3.2mins.
- Is the earlier claim reasonable? What is the probability that you will obtain a sample mean ≥ 3.2.
- Using CLT $\hat{\mu} = 2$ and $\hat{\sigma}^2 = \frac{4}{36} = \frac{1}{9}$.
- $P(\hat{X} > 3.2) = P(Z > 3.6) = 0.0002$. So, highly unlikely.