THE SAMPLE MEAN & ITS PROPERTIES

Set-up: $X_1, X_2, \dots, X_n$ are iid (independent, identically distributed) r.v.'s with $EX_i = \mu$ & $\operatorname{var} X_i = \sigma^2$, $i = 1, \dots, n$

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean.

① $E\bar{X} = \mu$ ② $\operatorname{var} \bar{X} = \sigma^2/n$

CLT: ③ For large n, $\bar{X}$ has (approximately) normal distribution.

How to use ①-③: Suppose μ & σ^2 are known & the question is:

$$P(a \le \bar{X} \le b) = P\Big(\underbrace{\frac{a-\mu}{\sigma/\sqrt{n}}}_{A} \le \underbrace{\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}}_{Z''} \le \underbrace{\frac{b-\mu}{\sigma/\sqrt{n}}}_{B}\Big) \underset{③}{\approx} P(A \le Z \le B)$$

where Z is standard normal.

Special case: $X_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{else} \end{cases}$; then $\mu = p$ & $\sigma^2 = p(1-p)$ & continuity correction...

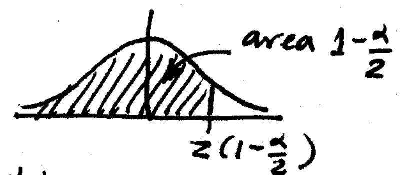
$\boxed{\text{STATISTICS}}$: INFERENCE FOR THE POPULATION FROM SAMPLE

If μ and σ^2 are unknown, estimate them by $\bar{X}$ & S^2 respectively where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance.

①: $\bar{X}$ is unbiased as an estimator of μ.

① & ②: $\bar{X}$ is consistent for μ, i.e. for any "tolerance" level ε, we have $P(\mu - \varepsilon \le \bar{X} \le \mu + \varepsilon) \to 1$ as $n \to \infty$.

③ is the Central Limit Theorem (CLT) which gives rise to $(1-\alpha)100\%$ confidence intervals for μ in the form: $\bar{X} \pm m$ where $m = z(1-\frac{\alpha}{2})\frac{\sigma}{\sqrt{n}}$ is the margin of error



[if σ^2 unknown, then $m = z(1-\frac{\alpha}{2})\, S/\sqrt{n}$]

and hypothesis testing of $H_0: \mu = \mu_0$ vs. $H_1: \mu \ne \mu_0$

→ Reject H_0 in favor of H_1 if μ_0 does not belong to the confidence interval.