STAT 135 Lecture 2

Henry Liev

25 June 2025

Example 0.1

Sample Proportion/Fraction

Biased coin. Lands heads with prob p. Observe X heads in n flips.
Sample proportion of heads: $\hat{p} = \frac{X}{n}$. $E(\hat{p}) = \frac{1}{n}E(X) \to X \sim \text{Binomial}(n, p)$.
$E(X) = np \to E(\hat{p}) = p$.
$Var(\hat{p}) = \frac{1}{n^2}Var(X) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$./ $Var(\bar{X}) = \frac{\sigma^2}{n}$.
$SE(\hat{p}) = \sqrt{\frac{p(1-p)}{n}}$.
$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
$E(\bar{X}) = \mu$ (unbiased)
$SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.
$Var(\bar{X}) = \frac{\sigma^2}{n}$

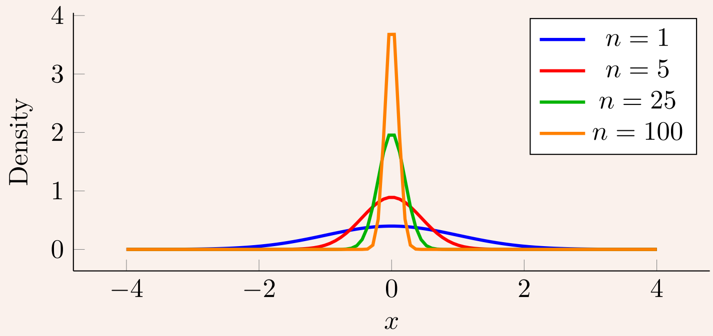
Definition 0.2

Standard Error

The standard deviation of an estimator. Gives us a measure that is in the same units as the mean.

Example 0.3

Population with mean μ, variance σ^2.
Estimating it from iid samples, $x_1, x_2, \ldots, x_n$.
$E(X) = \mu$
$Var(X) = \sigma^2$
What happens as $n \to \infty$?
$E(\bar{X}_n) = \mu$
$Var(\bar{X}_n) = \frac{\sigma^2}{n} \to 0.$
$SE(\bar{X}_n) = \frac{\sigma}{\sqrt{n}} \to 0.$
$\bar{X}_n$ "tends to" μ as $n \to \infty$?
$\bar{X}$ converges in probability to μ.
$\forall \epsilon > 0, P(|\bar{X}_n - \mu| > \epsilon) \to 0$ as $n \to \infty$.



Definition 0.4

Consistency
Estimator converges to the right thing. 'LOW BAR'

Example 0.5

$Var\left(\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}\right) = 1$

Knows the variance is constant at 1, but as $n \to \infty$, $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0, 1)$ by CLT
Convergence in distribution (cdf $F_n(x) \to \Phi(x)$ pointwise)
$\bar{X}$ is 'approximately normal' for large n.
Recap:

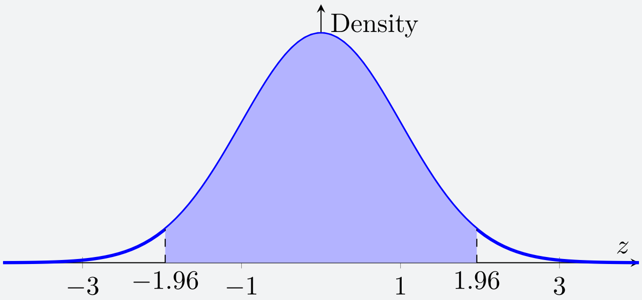
- $E(\bar{X}_n) = \mu$ (unbiased)
- $\bar{X}_n$ is consistent (converges in probability to μ)

Definition 0.6

Confidence Interval

$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim N(0,1)$ for large n by CLT.

$\frac{\sqrt{n}(\mu - \bar{X}_n)}{\sigma} \sim N(0,1)$ for large n by CLT.



The shaded region between $z = -1.96$ and $z = 1.96$ contains approximately 95% of the area under the standard normal curve.

$P(-1.96 \leq \frac{\sqrt{n}(\mu - \bar{X}_n)}{\sigma} \leq 1.96) = 0.95$

$P(\bar{X}_n - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}) \approx 0.95$

'Reasonable Range' for μ?

$P(\bar{X}_n - \frac{1.96\hat{\sigma}}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{1.96\hat{\sigma}}{\sqrt{n}}) \approx 0.95$

Example 0.7

Sample proportion

Confidence interval for p?

$P\left(\hat{p} - 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 0.95$

We don't know p

1. We can plug in $\hat{p}$ for p
2. If we want to be conservative we can plug in $p = 0.5$ since $p(1-p)$ is maximized at $p = 0.5$.

Example 0.8

Confidence Interval Example

Sample survey of 1600 people, sampled with replacement about future elections, JD Vance or AOC. 53% of respondents would vote for JD Vance.

Confidence interval: $0.53 \pm 1.96\sqrt{\frac{0.53(1-0.53)}{1600}} = (0.505, 0.554)$

We cannot say $P(0.505 \leq p \leq 0.554) \approx 0.95$ because the population proportion is not random.

Example 0.9

Estimating the Population
$X_1, X_2, \ldots, X_n$ drawn from a population and iid with mean μ, variance σ^2.
Want to estimate σ^2. Finite population, $x_1, x_2, \ldots, x_N$ $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$
Plug in estimate:
$Y = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \leftarrow$ Little too small on average since x_i is used to create $\bar{X}$
$T = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$
$E(T) = \frac{1}{n}\sum_{i=1}^{n}E(x_i - \mu)^2 = \frac{1}{n}n\sigma^2 = \sigma^2$
$\frac{1}{n}\sum_{i=1}^{n}(x_i - c)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x} + \bar{x} - c)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - c) + (\bar{x} - c)^2 =$
$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 + (\bar{x} - c)^2$

Takeaways

- CLT gives us useful infor about sampling distribution of $\bar{X}_n$ for large n.
  - Get approximate intervals