Topics Covered

4.1 - 4.9 HTs + CIs

- Measurement Error ✓ (bias + variance)
- Distributions
  - Bernoulli ✓
  - Poisson ✓
  - Normal ✓
  - Binomial ✓
  - Exponential ✓
- Central Limit Theorem ✓
- Point Estimation ✓
- Hypothesis tests for pop. mean
- Confidence Intervals for pop mean

6 PM, 1800 Chem

Notes

- Error = Bias + Variance

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)\,P(A)}{P(B)} = \frac{P(B \mid A)\,P(A)}{P(B \mid A)P(A) + P(B \mid A^c)P(A^c)}$$

Bernoulli:

Each trial is success ($Pr(S) = p$) or failure ($Pr = 1 - p = q$).

Bernoulli R.V. is 1 for success, 0 for failure

$E[X] = p$; $Var[X] = p(1-p)$

Binomial:

\# of successes in n Bernoulli trials ($X \sim Bin(n, p)$)

5% rule for sampling w/ binomial modeling

$$P(X = x) = p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

remember to do bounds shit for full credit

$E[X] = np$

$Var[X] = np(1-p)$

Sample proportion $x/n = \hat{p}$ is unbiased estimator; $E[\hat{p}] = p$; $SE(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

Poisson

Count # of successes in some continuous interval (like time)

$$P(X_t = x) = p(x) = \frac{e^{-\lambda_t} (\lambda_t)^x}{x!} \quad \text{for } x = 0, 1, \dots$$

(λ_t : mean # of events in t units of time/space)

$E[X_t] = \mu_{X_t} = \lambda_t$

$Var(X_t) = \sigma_{X_t}^2 = \lambda_t$

$\lambda_1 = \frac{\lambda_t}{t} = \frac{\lambda_{nt}}{nt}$ (rescale λ for time interval)

Exponential

Model waiting time before event occurs

PDF: $f(t) = \begin{cases} \lambda_1 e^{-\lambda_1 t} & t > 0 \\ 0 & \text{otherwise} \end{cases}$; λ_t is some parameter ($\approx$ probability of event)

$E[T] = \frac{1}{\lambda_1} = SD[T]$

$Var[T] = \frac{1}{\lambda_1^2}$

CDF: $F_T(t) = P(T \le t) = \begin{cases} 1 - e^{-\lambda_1 t} & t \ge 0 \\ 0 & t \le 0 \end{cases}$

Note: Same λ as Poisson.

If events follow Poisson w/ $\lambda_t = \lambda_1 t$,

then wait time is $T \sim Exp(\lambda_1)$

Note: "Memoryless Property": $P(T > t + s \mid T > s) = P(T > t)$

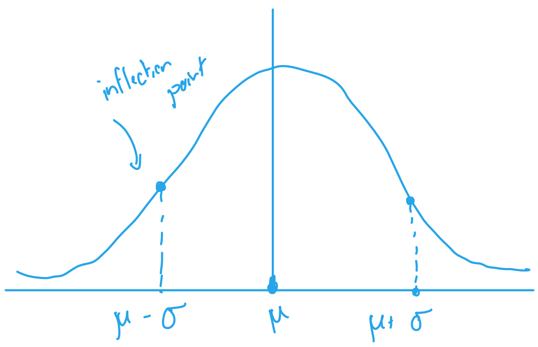
Waiting for additional t mins after waiting s mins is same probability as waiting t mins from start

Normal:

Model for many physical measurements

model errors in measurements

model sums + averages of R.V.s (CLT)



$X \sim N(\mu, \sigma^2)$

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $-\infty < x < \infty$

$z = \frac{x - \mu}{\sigma}$ (z-score) — $z \sim N(\mu = 0, \sigma^2 = 1)$

68 - 95 - 99.7

$\mu \pm \sigma$, $\mu \pm 2\sigma$, $\mu \pm 3\sigma$

If $X_1, X_2, \dots, X_n$ is R.S. from $N(\mu, \sigma^2)$:

estimate $\mu \approx \bar{x}$; $\sigma^2 \approx S^2$

uncertainty: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{S}{\sqrt{n}}$

Linear Combinations of Normal RVs = Normal RVs

$\mu = c_1\mu_1 + c_2\mu_2 + \dots + c_n\mu_n$

$\sigma^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + c_3^2\sigma_3^2 + \dots$

Distribution of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be R.S. from pop w/ μ, σ^2

$\bar{X}$ has mean $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$

If population dist. normal, $\bar{X}$ dist. normal

Probability Plots:

- Dots look close to line if ~ normal
- Weird, systematically based curve if not.

Central Limit Theorem

Sums and AVERAGES of independent normal vars are normally distributed

$T = X_1 + X_2 + \ldots + X_n \sim N(n\mu, n\sigma^2)$

$\bar{X} = \frac{1}{n} T \sim N(\mu, \sigma^2/n)$

"sufficiently large" n — $n \geq 30$, may be higher for skewed

If population dist normal, then approximation is exact

Point Estimation

- Sample Statistic to estimate Population Parameter
- Accuracy of estimator = BIAS
- Precision of estimator = STDEV

Bias $\hat{\theta} = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$

Uncertainty $= SD(\hat{\theta}) = \sigma_{\hat{\theta}}$

$MSE = E[\hat{\theta} - \theta]^2 + \sigma_{\hat{\theta}}^2$

Method of Maximum Likelihood:

$$L(\lambda) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot \ldots P(X_n = x_n)$$

$$= \prod_{i=1}^{n} f(x_i, \theta)$$

Take ln to make into sum

Find maximizing λ by taking derivative w.r.t λ = 0

MLEs invariant to functions

If $\hat{\theta}_1$, $\hat{\theta}_2$ MLEs of θ_1, θ_2, then $h(\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_m) = h(\theta_1, \theta_2, \theta_{\ldots} \theta_m)$

Hypothesis Tests

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$ ✱✱✱

P : $P(Z < z)$ or $P(Z > z)$ or $2P(Z \geq |z|)$

"at least as unusual"

Template :

We are interested in μ, the population mean of [. . . . . —].

H_0 : μ = . . .

H_1 : $\mu \lessgtr$ . . .

We would like to have a random sample from a normally-distributed population.

[Check R.S.]

- Stated
- If not, is independence of observations reasonable?
- 5% rule
- can always be like "we hope"

[Check UPDN]

- Stated
- Look @ Hist
- If no data, sample size ≥ 30? (CLT)
- t-procedures robust to violations of normality

Perform test procedure (test statistic + p-value)

Make conclusion

CONFIDENCE INTERVALS

$$\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

invNorm(0.025) __ ___

T - Procedures:

$n > 40$: data safe if not terribly skewed

$15 < n < 40$: unimodal, relatively symm data

$n < 15$: if data come from normally dist. populn

$$\bar{X} \pm t_{n-1,\, \alpha/2} \frac{s}{\sqrt{n}}$$