ESTIMATION

Point estimation

Goal of statistical inference

events

- The goal of statistical inference is to use information from observed <u>sample data</u> to make statements about some <u>population</u> of interest.
- We will choose some probability distribution to be a <u>model</u> for the data, and then use the data to estimate the unknown <u>parameters</u> of the distribution.

Modeling examples

What are the models and parameters in each of these situations?

1. To estimate the effectiveness of a drug we might sample n patients, administer the drug to each one, and count the number of patients who respond to the treatment.

2. To assess the risk of suicide, we might record the weekly number of attempts among some population.

3. To study the blood pressure of low birth weight infants, we might sample some number of such infants and measure the systolic blood pressure of each.

Two kinds of estimates

7.14

$\bar{x}$ $\hat{p}$

- <u>Definition 1</u>: A <u>POINT ESTIMATE</u> is a single number. It may be thought to represent the "best guess" of the unknown parameter.

- <u>Definition 2</u>: An <u>INTERVAL ESTIMATE</u> consists of a range of numbers, thought to be the most "plausible" values of the unknown parameter.

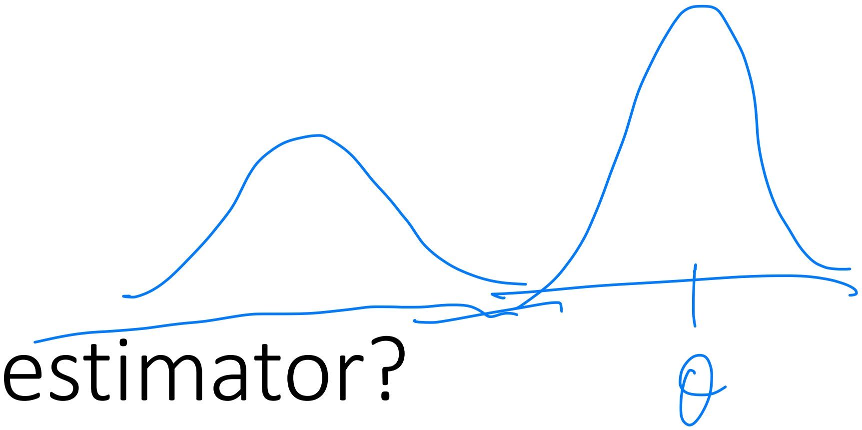
Estimator vs. estimate

$\bar{Y}$, s^2

- An <u>estimator</u> is a random variable, a formula, a function of the data. We will talk about distributions of estimators, expected values, variances, etc. We will focus on the *process* of creating estimates.
- An <u>estimate</u> is a number, the result of applying an estimator to a specific set of observed data.

7.91

0.4306



What makes for a good estimator?

- A good estimator is one that will result in estimates that are close to the target parameter with high probability.
- Thinking of an estimator as being a random variable, we want its distribution to be close to the target parameter.
- An estimator is unbiased if it doesn't tend to either overestimate or underestimate the target parameter.
- All other things being equal, we would prefer an estimator with good precision.
- We will next formalize these concepts.

Bias

$\mu, \sigma^2, \lambda, p$

- Let's use θ to denote our target parameter.
- Let's use $\hat{\theta}$ to denote an estimator of θ.
- Since $\hat{\theta}$ is a random variable, we can speak of $E[\hat{\theta}]$ and $Var(\hat{\theta})$.
- Definition 3: The BIAS of an estimator $\hat{\theta}$ for θ is $E[\hat{\theta}] - \theta$.
- Definition 4: An estimator $\hat{\theta}$ is UNBIASED for θ if $E[\hat{\theta}] = \theta$.
- It's good for an estimator to be unbiased, but that's not the only consideration!

Mean squared error

- In addition to having small bias, we would also like for our estimator to have small variance.
- <u>Definition 5</u>: The <u>MEAN SQUARED ERROR</u> of an estimator $\hat{\theta}$ for θ is

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right]$$

- The MSE measures a combination of the bias and the variance. of the estimator

$$\boxed{MSE(\hat{\theta})} = E\left[(\hat{\theta} - \theta)^2\right] = E\left[\left(\underbrace{\hat{\theta} - E[\hat{\theta}]} + \underbrace{E[\hat{\theta}] - \theta}_{\text{constant}}\right)^2\right]$$

$$= E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2 + \left(E[\hat{\theta}] - \theta\right)^2 + 2\left(\hat{\theta} - E[\hat{\theta}]\right)\left(E[\hat{\theta}] - \theta\right)\right]$$

$$= E\left[\left(\hat{\theta} - E[\hat{\theta}]\right)^2\right] + \left(E[\hat{\theta}] - \theta\right)^2 + 2\left(E[\hat{\theta}] - \theta\right)E\left[\hat{\theta} - E[\hat{\theta}]\right]$$

$$= \boxed{Var(\hat{\theta}) + bias^2(\hat{\theta})} + \underbrace{2\, bias(\hat{\theta}) \left(E[\hat{\theta}] - E[\hat{\theta}]\right)}_{=0} = 0$$

Mean squared error: Example

- Example 1: To estimate p, the "true probability of 'success'" in a binomial experiment, we repeat the trial n times and observe $Y \sim Bi(n, p)$, the number of "successes".
  - A natural estimator is $\hat{p} = \frac{Y}{n}$.
  - An alternative estimator is $\hat{p}^* = \frac{Y+1}{n+1}$??
  - How to decide which one is "better"?

$E[Y] = np$

$Var(Y) = np(1-p)$

$E[\hat{p}] = E\left[\frac{Y}{n}\right] = \frac{1}{n} E[Y] = \frac{1}{\cancel{n}} \cdot \cancel{n}p = p$

UNBIASED

$bias(\hat{p}) = 0$

$Var(\hat{p}) = Var\left(\frac{Y}{n}\right) = \frac{1}{n^2} Var(Y)$

$= \frac{1}{n^{\cancel{2}}} \cancel{n}p(1-p) = \frac{p(1-p)}{n}$

$MSE(\hat{p}) = Var(\hat{p}) + bias^2(\hat{p}) = \frac{p(1-p)}{n} + 0 = \frac{p(1-p)}{n}$

$$E[\hat{p}^*] = E\left[\frac{Y+1}{n+1}\right] = E\left[\frac{Y}{n+1} + \frac{1}{n+1}\right] = \frac{1}{n+1}E[Y] + \frac{1}{n+1}$$

$$= \frac{np}{n+1} + \frac{1}{n+1} = \frac{np+1}{n+1} \overset{??}{=} p \quad \underline{\underline{\text{NO}}} \qquad \underline{\underline{\text{BIASED}}}\ !!$$

$$\text{bias}(\hat{p}^*) = \frac{np+1}{n+1} - p = \frac{\cancel{np}+1-\cancel{np}-p}{n+1} = \frac{1-p}{n+1}$$

BIAS is small $\underline{\underline{\text{if}}}$ $n+1$ large

or if p is close to 1.

$$\text{Var}(\hat{p}^*) = \text{Var}\left(\frac{Y}{n+1} + \frac{1}{n+1}\right) = \text{Var}\left(\frac{Y}{n+1}\right)$$

$$= \frac{1}{(n+1)^2}\text{Var}(Y) = \frac{np(1-p)}{(n+1)^2}$$

$$\text{MSE}(\hat{p}^*) = \frac{np(1-p)}{(n+1)^2} + \frac{(1-p)^2}{(n+1)^2} = \frac{(1-p)(np+1-p)}{(n+1)^2} = \frac{(1-p)(1+(n-1)p)}{(n+1)^2}$$

Which estimator has smaller MSE?

May depend on n, p

$p = \frac{1}{2} \quad n = 10$

$$\text{MSE}(\hat{p}) = \frac{\frac{1}{2}\cdot\frac{1}{2}}{10} = \frac{1}{40} = 0.025$$

$$\text{MSE}(\hat{p}^*) = \frac{\frac{1}{2}(1+9(\frac{1}{2}))}{11^2} = 0.023$$

Some common unbiased point estimators

model

population mean

- Situation 1: $Y_1, Y_2, \dots, Y_n \sim iid\ N(\mu, \sigma^2)$. How to estimate μ?

$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i$

$Var(\bar{Y}) = \frac{\sigma^2}{n}$

$E[\bar{Y}] = E\left[\frac{1}{n} \sum Y_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[Y_i]$

$= \frac{1}{n}(\mu + \mu + \dots + \mu) = \frac{n\mu}{n} = \mu$

- Situation 2: $X_1, X_2, \dots, X_n \sim iid\ Bernoulli(p)$ and $Y = \sum_{i=1}^{n} X_i$. How to estimate p?

Binomial(1, p)

$Y \sim binomial(n, p)$

don't 0 or 1 respond to trt.

$E[Y] = np$

$\hat{p} = \frac{Y}{n}$

$E[\hat{p}] = \frac{np}{n} = p$

unbiased for p

Another common unbiased point estimator

- Situation 3: $Y_{1,1} Y_{1,2}, \dots, Y_{1,n_1} \sim iid\ N(\mu_1, \sigma_1^2)$, $Y_{2,1} Y_{2,2}, \dots, Y_{2,n_2} \sim iid\ N(\mu_2, \sigma_2^2)$, mutually independent. How to estimate $\mu_1 - \mu_2$?

$$\bar{Y}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_{1i}$$

$$\bar{Y}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_{2i}$$

$$E[\bar{Y}_1] = \mu_1$$

$$E[\bar{Y}_2] = \mu_2$$

$$E[\bar{Y}_1 - \bar{Y}_2] = \mu_1 - \mu_2$$

UNBIASED!

$$\widehat{\mu_1 - \mu_2} = \hat{\mu}_1 - \hat{\mu}_2 = \bar{Y}_1 - \bar{Y}_2$$

$$E[\widehat{\mu_1 - \mu_2}] = E[\bar{Y}_1 - \bar{Y}_2] = E[\bar{Y}_1] - E[\bar{Y}_2] = \mu_1 - \mu_2$$

And another common unbiased point estimator

placebo new drug

- Situation 4: $X_{1,1}X_{1,2}, \ldots, X_{1,n_1} \sim iid\ Bernoulli(p_1)$, $X_{2,1}, X_{2,2}, \ldots, X_{2,n_2} \sim iid\ Bernoulli(p_2)$, mutually independent. How to estimate $p_1 - p_2$?

$Y_1 = \sum X_{1i}$ $\quad Y_2 = \sum X_{2i}$

$Y_1 \sim Binomial(n_1, p_1)$

$Y_2 \sim Binomial(n_2, p_2)$

$\hat{p}_1 = \frac{Y_1}{n_1}$ unbiased for p_1

$\hat{p}_2 = \frac{Y_2}{n_2}$ " " p_2

$E[\hat{p}_1 - \hat{p}_2] = E(\hat{p}_1) - E(\hat{p}_2) = p_1 - p_2$ UNBIASED!

$Var(\hat{p}_1 - \hat{p}_2) =$

Yet another common unbiased point estimator

properties of χ^2 dist.

- Back to Situation 1: $Y_1, Y_2, \dots, Y_n \sim iid\ N(\mu, \sigma^2)$, but how to estimate σ^2?

Claim: The sample variance $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ is an unbiased estimator of σ^2.

We saw before $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ So $E\left[\frac{(n-1)s^2}{\sigma^2}\right] = n-1$

$\frac{\cancel{n-1}}{\sigma^2} E[s^2] = \cancel{n-1}$

$E[s^2] = \sigma^2$

UNBIASED!

What about $\hat{\sigma}^{2*} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$?

$\hat{\sigma}^{*2} = \frac{n-1}{n} s^2$ $E[\hat{\sigma}^{*2}] = \frac{n-1}{n} E[s^2]$ $= \frac{n-1}{n}\sigma^2$ BIASED

And what about $\hat{\sigma} = s$ as an estimator for σ?

Yeah, sure, but $E[s] \neq \sigma$ in general

BIASED!! there is no estimator of σ that is unbiased for all σ!!

Point estimates: What are they good for?

- A point estimate provides a "best guess" as to the unknown parameter value.
- Joe gets a sample and his estimate of μ is 41.9.
- Dustin gets another sample and his estimate of μ (the same one) is 48.5.
- Faye gets her own sample, and her estimate of μ (again, the same one) is 62.5.
- Who is right? Who is wrong? Whom to believe?

A point estimate: It is what it is.

- Joe sampled 50 patients; Dustin sampled 65. Faye only sampled 15.
- Now whom do we believe (most)?