AMS 572 Data Analysis I
Basic Concepts of Inference

Pei-Fen Kuan

Applied Math and Stats, Stony Brook University

Sampling

- A function of statistics is the provision of techniques for making inductive inferences, i.e, generalizations beyond the actual data in hand
- The goal of inductive inference is to find out something about a target population by examining a sample of it. (Read Chapter 3 of sampling designs).
- Inductive inference is accomplished by constructing a model that describes the origin of the data and a model for data collection (sampling)

Definition: The random variables $X_1, \ldots, X_n$ are called a *random sample* of size n from the population f if $X_1, \ldots, X_n$ are independent and identically distributed (iid) with marginal pdf or pmf f.

Note: The definition implies that the joint pdf or pmf of $X_1, \ldots, X_n$ is

Statistics and sample moments

- A *statistic* is a function of observable random variables. The probability distribution of a statistic is called its sampling distribution
- Sample mean:
- Sample variance:

Sampling from normal distribution

Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Then $\bar{X}_n$ has $N(\mu, \sigma^2/n)$ distribution.

Statistical Inference

- Statistical inference: Using statistics and probability theory to draw conclusions about parameters
- Two modes of inference:
  - Estimation: attempt to estimate value of parameter(s) and quantify uncertainty about these estimate(s)
  - Hypothesis testing: posit certain values for parameters and test whether the observed data are consistent with the hypothesis

- *Estimand*: parameter of interest we are trying to estimate; a constant; Eg
- *Estimator*: the statistic used to estimate the estimand; a random variable; Eg
- *Estimate*: a realization of an estimator from an observed data set; Eg

Point estimation:
Methods of Maximum Likelihood and Moment Estimator

Definition: Let $X_1, \ldots, X_n$ be a sample with joint p.d.f or p.m.f $f(\mathbf{x}|\theta)$. Given that $\mathbf{X} = \mathbf{x}$ is observed, the function of θ defined by

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$$

is called the likelihood function.

Point estimation: Method of Maximum Likelihood

Definition: For a given observed sample $\mathbf{x}$, let $\hat{\theta(\mathbf{x})}$ be a value in the parameter space at which $L(\theta|\mathbf{x})$ attains its maximum. $\hat{\theta(\mathbf{x})}$ is called the maximum likelihood estimator (MLE) of θ.

Example: Suppose $X_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Derive the MLE for μ and σ^2.

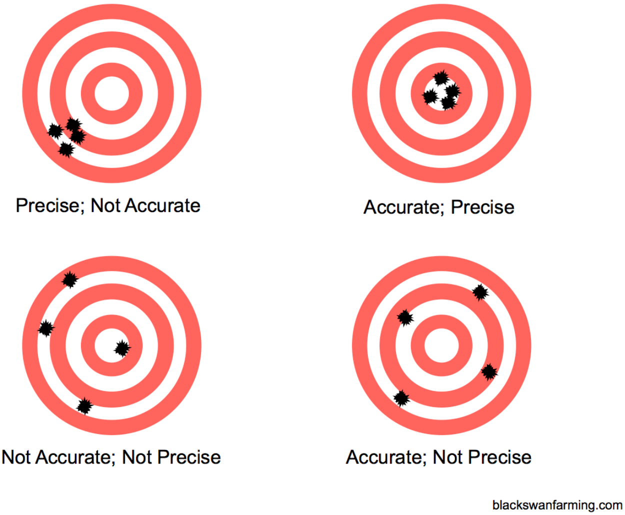
Point estimation: Method of Moments

Method of moment method is one of the oldest method, and is simple to use. This method almost always yields some sort of estimate (MME).
Work by equating the sample moments to the population moments:

Point estimation: Method of Moments

Example: Suppose $X_i \sim N(\mu, \sigma^2)$ for $i = 1, \ldots, n$. Derive the MME for μ and σ^2.

Methods of Evaluating Estimators



Methods of Evaluating Estimators

Definition: The bias of an estimator $\hat{\theta}$ of θ is

$$\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$

(measures accuracy)

Example: Are the MLEs of μ and σ for normal distribution unbiased? If not, find an unbiased estimator for these parameters.

Methods of Evaluating Estimators

Definition: The mean squared error (MSE) of an estimator $\hat{\theta}$ of θ is

$$\text{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

(measures precision and accuracy)

In addition, $\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2$

Methods of Evaluating Estimators

Example: Compare the MSE of these estimators for σ^2,
$S_1^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n-1)$ and $S_2^2 = \sum_{i=1}^n (X_i - \bar{X})^2/n$.

- On average S_2^2 will be closer to σ^2 if MSE is used as criterion
- On average S_2^2 underestimates σ^2

Confidence Interval

Definition: Let $X_1, \dots, X_n$ be random variables with joint p.d.f or p.m.f $f(\mathbf{x}|\theta)$. The random interval $(T_1(\mathbf{X}), T_2(\mathbf{X}))$ is called a $100(1-\alpha)\%$ confidence interval for θ if

$$P(T_1(\mathbf{X}) < \theta < T_2(\mathbf{X})) = 1 - \alpha$$

where $0 < \alpha < 1$.

- $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ are called the lower and upper confidence limits, respectively.
- $1 - \alpha$ is called the confidence coefficient.

width of the C.I is $T_2(x) - T_1(x)$

Confidence Interval Interpretation

- If we draw 100 different random samples, on average $100(1-\alpha)\%$ of them will contain θ
- How can one decrease the width of confidence interval?
  - Increase sample size "n"
  - Increase "2"

width $2 z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

We will study the pivotal quantity method for deriving confidence interval in Chapter 7 lecture notes.

Hypothesis testing

1. Set up a hypothesis
2. Collect data
3. Infer from the data whether hypothesis is plausible

- Examples:
  Is BP the same in diabetics and non-diabetics?
  Will folic acid supplementation reduce the risk of stroke?

Hypothesis testing

- Null hypothesis H_0: hypothesis to be tested
- Example of null hypothesis for folic acid study:
  The incidence of stroke will be the same in those taking folic acid supplements and those not taking folic acid supplements

Null and Alternative

- In a test of a hypothesis, we are testing whether some population parameter has a particular value
- For example,
$$H_0 : \theta = \theta_0$$
where θ_0 is a known constant
- The alternative hypothesis is complement of null hypothesis
$$H_a : \theta \neq \theta_0$$

Test statistic

- Once the data are collected, we will compute a *test statistic* related to θ, say $S(\hat{\theta})$
- $S(\hat{\theta})$ is a random variable, since it is computed from a sample
- $S(\hat{\theta})$ will have a particular probability distribution under the assumption H_0, say $F_0[S(\hat{\theta})]$

Test statistic

- Under F_0, we compute the probability that we would observe $S(\hat{\theta})$ or a value more extreme than $S(\hat{\theta})$ if the null H_0 was true
- If this probability is large, the data are consistent with H_0
- If this probability is small, H_0 is probably not true

Interpretation

- Usually if the probability is small, we conclude H_0 is not true; i.e., we "reject" H_0
- If the probability is large, we have not proven H_0. We say that "we failed to reject H_0"
- We can never prove H_0 is true!

Significance Level

- How do we decide if the probability is too small?
- Prior to seeing the data, we select a value α such that: if the computed probability is less than or equal to α, we reject H_0
- α is known as *significance level*

Critical Region and Value

- We have a statistic $S(\hat{\theta})$ with distribution F_0 under the null hypothesis
- We specify α and under F_0 determine a *critical region* or *rejection region* C_α such that

$$P(S(\hat{\theta}) \in C_\alpha \mid H_0) = \alpha$$

in

$$\leq \alpha$$

- Values at the boundaries of C_α are call *critical values*

Critical Region and Value

If $S(\hat{\theta}) \in C_\alpha$, then we reject H_0.

If $S(\hat{\theta}) \notin C_\alpha$, then we do not reject H_0

- From the data we compute $S(\hat{\theta})$

Type I error, Type II error and Power

		Truth	
		H_0	H_a
Decision	Do not reject	✓	Type II error
	Reject	Type I error	Power

$\alpha = P(\text{Type I error}) = P(\text{reject } H_0 \mid H_0)$

$\beta = P(\text{Type II error}) = P(\text{do not reject } H_0 \mid H_a)$

$\text{Power} = 1-\beta = P(\text{reject } H_0 \mid H_a)$

Tests of Hypotheses: Seven Steps

1. Design study
2. Establish null hypothesis
3. Determine test statistic to be employed
4. Choose significance level α and establish C_α
5. Carry out study and collect data
6. Compute statistic from data
7. If statistic is in C_α, reject H_0

We will study step by step of setting up a hypothesis test in Chapter 7 slides.