STAT 703/J703
February 8th, 2005
-*Lecture 9*-

Instructor: Brian Habing
Department of Statistics
LeConte 203
Telephone: 803-777-3578
E-mail: habing@stat.sc.edu

Today

- Homework 3 Solutions
- Basics of Hypothesis Testing (Chapter 9)

Ch.8#5 Suppose that X follows a geometric distribution,

$$P[X = k] = p(1-p)^{k-1}$$

And assume an i.i.d. sample of size *n*.

b) Find the mle of *p*.

c) Find the asymptotic variance of the MLE.

<u>Ch.8#6</u> Consider the data

# Hops	Freq	# Hops	Freq
1	48	7	4
2	31	8	2
3	20	9	1
4	9	10	1
5	6	11	2
6	5	12	1

a) Fit a geometric distribution

b) Find an approximate 95% confidence interval for p.

Ch8#44c Let $X_1,\ldots X_n$ be i.i.d random variables with density function

$$f(x\mid\theta)=(\theta+1)x^{\theta}, 0\le x\le 1$$

Find the asymptotic variance of the MLE.

Statistical Inference – Confidence Intervals and Tests of Hypotheses

Test of hypothesis – general method to distinguish between 2 (or more) probability distributions (or models), based on a sample $X_1, \ldots, X_n$ assumed to come from one of them.

In particular, based on $X_1, \ldots, X_n$, decide whether $f_1(x)$ or $f_2(x)$ is the pdf (or population) from which the sample came.

More specifically, suppose we think the sample is from a normal population with mean either μ=5 or μ=10 with variance 4. (Or, more generally, $\mu=\mu_1$ vs. $\mu=\mu_2$).

To do this, we follow the "Neyman-Pearson Paradigm" and discuss "significance tests".

Neyman-Pearson Approach:
Group the possible hypothesized distributions into two categories, the null hypothesis, H_0, and the alternative hypothesis, H_A (research hypothesis).

E.g. Observe $X_1, \ldots, X_n$.
Either H_0: $N(\mu_1, \sigma^2)$ or H_A: $N(\mu_2, \sigma^2)$
or H_0: $\mu=\mu_1$ vs. H_A: $\mu=\mu_2$,
μ is the mean of $N(\mu, \sigma^2)$.

Here, if σ^2 is known, each of these hypotheses completely specifies the distribution or population. So, H_0 and H_A are called simple hypotheses.

If H_0: $\mu=0$ vs. H_A: $\mu > 0$ (μ in $N(\mu, \sigma^2)$, with σ^2 known).

Then H_0 is simple and H_A is a composite hypothesis, i.e. several normal distributions would satisfy it.

H_A is also referred to as a one-sided hypothesis.

If H_0: $\mu=0$ vs. H_A: $\mu\neq0$ (μ in $N(\mu, \sigma^2)$, σ^2 known), then H_A is a two-sided (composite) hypothesis.

Next, we set up the framework for "testing" H_0 and H_A based on the sample.

9.2 Neyman-Pearson Paradigm

Let $\underline{X} = (X_1, \ldots, X_n)$ denote a sample from population $f(x|\theta)$. Decide on H_0 vs. H_A based on the sample.

A decision on whether or not to reject H_0 in favor of H_A is made on the basis of a statistic

$$T=T(\underline{X})=T(X_1, \ldots, X_n).$$

The set of values of T for which H_0 is accepted is called the acceptance region and the set of values of T for which H_0 is rejected is the rejection region of the test.

Two kinds of error may occur:

1. H_0 is rejected when it is true: Type I error.

P(type I error) = α
= P(T $\in$ rejection region | H_0 true).

If H_0 is simple, α is called the significance level of the test.

STAT 703/J703 B.Habing Univ. of SC 16

2. H_0 is accepted when it is false: Type II error.

P(type II error) = β
= P(T in acceptance region | H_0 false)

If H_A is composite, β depends on which member of H_A is the true pdf.

STAT 703/J703 B.Habing Univ. of SC 17

Power of the test =P(H_0 is rejected when false)
= 1 - P(H_0 is accepted | H_0 false)
= 1 - β.

Ideally, we would want $\alpha=\beta=0$, but this not possible since the decision is based on data.

STAT 703/J703 B.Habing Univ. of SC 18