STA 257S
Some Statistical Applications

If $X_1, X_2, \ldots, X_n$ are i.i.d. we call $X_1, X_2, \ldots, X_n$ a random sample of the random variable X. Any function of the elements of a random sample, which does not depend on any unknown parameters, is called a statistic.

Estimators: An estimator $\hat{\theta}_n$ of an unknown quantity θ is a statistic which attempts to estimate θ. The estimator is unbiased if $E(\hat{\theta}_n) = \theta$. (Note that unbiased estimators are not always desirable. They may not exist or may be inappropriate for the problem. They also may be in conflict with other desirable properties of estimators such as minimum error.) An estimator is consistent if for any $\epsilon > 0$, $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0$. The mean square error of an estimate $\hat{\theta}_n$ is $MSE(\hat{\theta}_n) = E[(\hat{\theta}_n - \theta)^2]$.

If $X_1, X_2, \ldots, X_n$ are i.i.d. with $E(X_i) = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$, then we can estimate μ by $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$, the average of the observed values. This estimator is unbiased and consistent. Furthermore, we can estimate σ^2 by $\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$ if μ is known, or by $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ if μ is unknown. Both of these estimators are also unbiased and consistent. For $X_1, X_2, \ldots, X_n$ i.i.d. normal random variables, it can be shown that $\overline{X}$ and s^2 are independent.

Distributions Derived from the Normal Distribution:

- If Z is a standard normal random variable, the distribution of $U = Z^2$ is the chi-square distribution with 1 degree of freedom.
- If $U_1, U_2, \ldots, U_n$ are independent chi-square random variables with 1 degree of freedom, the distribution of $V = U_1 + U_2 + \ldots + U_n$ is the chi-square distribution with n degrees of freedom and is denoted by χ^2_n.
- If $X_1, X_2, \ldots, X_n$ are a random sample from a $N(\mu, \sigma^2)$ distribution and $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, then $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$.
- If $Z \sim N(0,1)$ and $U \sim \chi^2_n$ and Z and U are independent, then the distribution of $Z/\sqrt{U/n}$ is the t distribution with n degrees of freedom.
- If $X_1, X_2, \ldots, X_n$ are a random sample from a $N(\mu, \sigma^2)$ distribution and $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, then $(\overline{X} - \mu)/\sqrt{\frac{s^2}{n}} \sim t_{n-1}$.
- Let U and V be independent chi-square random variables with m and n degrees of freedom, respectively. The distribution of

  $$W = \frac{U/m}{V/n}$$

  is the F distribution with m and n degrees of freedom and is denoted by $F_{m,n}$.

Problems

1. Let $X_1, X_2, \ldots, X_n$ be i.i.d. with mean μ. What conditions must hold on the real numbers $a_1, a_2, \ldots, a_n$ such that $\sum_{i=1}^n a_i X_i$ is an unbiased estimator for μ? [Ans: $\sum_{i=1}^n a_i = 1$]

2. Let $X_1, X_2, \ldots, X_n$ be independent $N(\mu, \sigma^2)$. Suppose we estimate μ by X_n (*i.e.* the last measurement) and estimate σ^2 by $(X_n - \overline{X})^2$. Are these estimators unbiased? Are they consistent? [Ans: $\hat{\mu}$ unbiased; $\hat{\sigma^2}$ biased ($E(\hat{\sigma^2}) = \frac{n-1}{n}\sigma^2$); both not consistent]

3. Let $X_1, X_2, \ldots, X_n$ be a random sample of a Poisson random variable with parameter μ and define
$$M = d[X_1(X_1 - X_2) + X_2(X_2 - X_3) + \ldots + X_{n-1}(X_{n-1} - X_n)]$$
Find the constant d that makes M an unbiased estimator for μ. [Ans: $d = \frac{1}{n-1}$]

4. Suppose $X_1, X_2, \ldots, X_n$ is a random sample of a $N(\mu, \sigma^2)$ random variable. Determine c such that
$$c[(X_1 - X_2)^2 + (X_3 - X_4)^2 + \ldots + (X_{n-1} - X_n)^2]$$
is an unbiased estimator of σ^2. [Ans: $c = \frac{1}{n}$]

5. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from a $N(\mu, \sigma^2)$ distribution. Is $s = \sqrt{s^2}$ an unbiased estimate of σ? If not, evaluate $E(s)$. [Ans: no; $E(s) = \sqrt{\frac{2}{n-1}}\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}\sigma$]

6. If Y has a binomial distribution with parameters n and p, then $\hat{p}_1 = \frac{Y}{n}$ is an unbiased estimator of p. Another estimator of p is $\hat{p}_2 = \frac{Y+1}{n+2}$.

   (a) Find $E(\hat{p}_2)$.

   (b) Derive $MSE(\hat{p}_1)$ and $MSE(\hat{p}_2)$. For what values of p is $MSE(\hat{p}_2) < MSE(\hat{p}_1)$.

   [Ans: (a) $\frac{np+1}{n+2}$ (b) $\frac{p(1-p)}{n}$; $\frac{[np(1-p)+(1-2p)^2]}{(n+2)^2}$; p near $\frac{1}{2}$]

7. Show that if X and Y are independent exponential random variables with $\lambda = \frac{1}{2}$, then $\frac{X}{Y}$ follows an F distribution. Identify the degrees of freedom. [Ans: $F_{2,2}$]

8. Suppose Y, $X_1, X_2, \ldots, X_n$ are independent $N(\mu, \sigma^2)$. Set $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$, as usual. For each of the following statements, find values of a, b, c, and d to make them true.

   (a) $a\overline{X} + b \sim N(0,1)$

   (b) $a\sum_{i=1}^n (X_i + b)^2 \sim \chi^2_c$

   (c) $(aY + b)/\sqrt{s^2} \sim t_c$

   (d) $(aY + b)^2/s^2 \sim F_{c,d}$

   (e) $a\left((X_1 - X_2)^2 + (X_3 - X_4)^2 + \ldots + (X_{n-1} - X_n)^2\right) \sim \chi^2_b$

   [Ans: (a) $a = \frac{\sqrt{n}}{\sigma}$, $b = -a\mu$ (b) $a = \frac{1}{\sigma^2}$, $b = -\mu$, $c = n$ (c) $a = 1$, $b = -\mu$, $c = n-1$ (d) $a = 1$, $b = -\mu$, $c = 1$, $d = n-1$ (e) $a = \frac{1}{2\sigma^2}$, $b = \frac{n}{2}$]