# MATH324 Crib Sheet

Sam Zhang

# 1 Properties of Estimators and Statistics

## 1.1 Likelihood

The likelihood function is defined as

$$\begin{aligned} L(\theta) &= f(y_1, \ldots, y_n|\theta) \\ &= \prod_{i=1}^{n} f(y_i|\theta) \end{aligned} \tag{1}$$

Given that $y_i$ are independent and identically distributed (i.i.d.).

## 1.2 Biasedness

An estimator $\hat{\theta}$ is biased if $\mathbb{E}(\hat{\theta}) \neq \theta$.
**Example** $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is biased if $X_i$ are i.i.d. with mean $\mu$ and variance $\sigma^2$.

$$\mathbb{E}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} n\mu = \mu \tag{2}$$

The bias is determined by the equation

$$\mathbb{B}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta \tag{3}$$

We can generally find an unbiased estimator from a biased estimator by eliminating the constants surrounding the estimator, such that $\mathbb{E}(\hat{\theta}) \to \theta$.

## 1.3 Consistency

An estimator $\hat{\theta}_n$ is consistent if $\hat{\theta}$ converges in probability to $\theta$ as $n \to \infty$.

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0 \tag{4}$$

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1 \tag{5}$$

This is equivalent to the following:

$$\lim_{n\to\infty} \mathbb{V}(\hat{\theta_n}) = 0 \tag{6}$$

## 1.4 Asymptotic Normality

An estimator $\hat{\theta}$ is asymptotically normal if $\hat{\theta}$ converges in distribution to a normal distribution as $n \to \infty$.
**Example** $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is asymptotically normal if $X_i$ are independent and identically distributed (i.i.d.) with mean $\mu$ and variance $\sigma^2$. For large samples:

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \tag{7}$$

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0, 1) \tag{8}$$

### 1.5 Sufficiency

Given a random sample $Y_1 \dots Y_n$ with the parameter $\theta$, a statistic $T$ is sufficient for $\theta$ if $T$ contains all the information about $\theta$. This implies that $\theta$ can be uniquely determined from an estimator based on $T$ without any loss of information.
This is true iff the distribution of $Y$ given $T$ is does not depend on $\theta$.

#### 1.5.1 Fisher-Neyman Theorem

Let $U$ be a statistic of the random $Y_1 \dots Y_n$. $U$ is sufficient for $\theta$ iff $L(\theta)$ can be writte as

$$L(\theta) = g(u, \theta) \cdot h(y_1, y_2, \dots, y_n | u) \tag{9}$$

where $g(u, \theta)$ is a function of $u$ and $\theta$ and $h(y_1, y_2, \dots, y_n)$ is not a function of $\theta$

### 1.6 Efficiency

An estimator $\hat{\theta}$ is efficient if $\hat{\theta}$ has the smallest variance among all unbiased estimators of $\theta$. the efficiency of two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ is given by

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\mathbb{V}(\hat{\theta}_1)}{\mathbb{V}(\hat{\theta}_2)} \tag{10}$$

#### 1.6.1 The Rao Blackwell Theorem

Let $\hat{\theta}$ be an unbiased estimator of $\theta$ such that $\mathbb{V}(\hat{\theta}) < \infty$. If $U$ is a sufficient statistic for $\theta$, define $\hat{\theta}^* = \mathbb{E}(\hat{\theta}|U)$. Then $\forall\theta$:

$$\mathbb{E}(\hat{\theta}^*) = \theta \quad \text{and} \quad \mathbb{V}(\hat{\theta}^*) \leq \mathbb{V}(\hat{\theta})$$

**Remark** The result of the Rao Blackwell Theorem is the *minimum-variance unbiased estimator* of $\theta$. (MVUE)

## 2 Hypothesis Testing

### 2.1 Terminologies

- **Null Hypothesis** $\rightarrow H_0 : \theta = \theta_0$
- **Alternative Hypothesis**$\rightarrow H_a : \theta \neq \theta_0$
- **Type I Error**$\rightarrow \alpha = P(\text{Reject } H_0 \text{ when } H_0 \text{ is true})$ i.e. $P(T \in RR|H_0)$
- **Type II Error**$\rightarrow \beta = P(\text{Fail to reject } H_0 \text{ when } H_1 \text{ is true})$ i.e. $P(T \notin RR|H_1)$

### 2.2 Rejection Regions

A rejection region is a set of values of the test statistic $T$ such that if $T$ falls in the rejection region, we reject the null hypothesis.
**Example** Let $X_i$ be i.i.d. with mean $\mu$ and variance $\sigma^2$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$. We want to test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu \neq \mu_0$.
We can use the following rejection region:

$$R = \left\{ T \in \mathbb{R} : |T - \mu_0| > c\sqrt{\frac{\sigma^2}{n}} \right\} \tag{11}$$

where $c$ is a constant.
**Remark** This is in fact a two-sided T-test for the population mean.

## 2.3 The T-test

### 2.3.1 Large-Sample Hypothesis Testing

Large sample hypothesis testing is based on the central limit theorem.
Given an estimator $\hat{\theta}$ that is asymptotically normal in regards to $\theta$, we know the following:

$$Z = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{\mathbb{V}(\hat{\theta}_n)}{n}}} \sim N(0,1) \tag{12}$$

We can make a comparison with the standard normal distribution's rejection region in regards to a chosen $\alpha$, e.g. $Z_\alpha = Z_{0.05}$, and see if $Z$ falls in the rejection region $Z_{0.05}$.
Alternatively, a clearer way is to use the *p-value*, which is the probability of observing a value of $Z$ as extreme as the one observed, given that $H_0$ is true.
We can obtain the p-value by using the standard normal distribution's CDF, but this is generally simplified into a table or a software.
We reject $H_0$ if $p < \alpha$. Otherwise, we fail to reject $H_0$.

### 2.3.2 Small-Sample Hypothesis Test

The small-sample hypothesis test is based on the t-distribution, a distribution similiar to the standard normal distribution, but with heavier tails.
The t-distribution is defined as follows:

$$T = \frac{\hat{\theta}_n - \theta_0}{\sqrt{\frac{\mathbb{V}(\hat{\theta}_n)}{n}}} \sim t(n-1) \tag{13}$$

The t-distribution is a similiar distribution to $Z$, with different parameters. The parameter $n-1$ is the degrees of freedom.
**Remark** The t-distribution is used in the same way as the standard normal distribution, except that the rejection region is defined by the t-distribution instead of the standard normal distribution.

### 2.3.3 F-test for Variance

The F-test is used to test the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against the alternative hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$.
The test statistic is defined as follows:

$$F = \frac{S_1}{S_2} \tag{14}$$

And we can make a conclusion using the F-table like the t-table.

## 2.4 The Likelihood Ratio Test

### 2.4.1 The Likelihood Ratio Test for simple hypothesis

**The Neyman-Pearson Lemma** The Neyman-Pearson lemma is a special case of the Likelihood Ratio Test. It is applicable when we wish to test the simple null hypothesis $H_0 : \theta = \theta_0$ against the alternative hypothesis $H_1 : \theta \neq \theta_0$.
The test statistic $k$ is the rejection region, it is defined as follow:

$$\frac{L(\theta_0)}{L(\hat{\theta})} < k \tag{15}$$

We can compute $k$ given an $\alpha$.

### 2.4.2 The Likelihood Ratio Test for composite hypothesis

Define:
$\Theta$ = the vector of all $k$ parameters $(\theta_1, \theta_2, \ldots, \theta_k)$
$\Omega_0$ = The set of possible values that $\Theta$ may lie in given $H_0$
$\Omega_a$ = The set of possible values that $\Theta$ may lie in given $H_a$
$\Omega = \Omega_0 \cup \Omega_a$
The test statistic $\lambda$ is as follows:

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} = \frac{\max_{\Theta \in \Omega_0} L(\Theta)}{\max_{\Theta \in \Omega} L(\Theta)} \tag{16}$$

And the rejection region is determined by $\lambda \leq k$
**Thereom** Let $Y_1, \ldots, Y_n$ have joint likelihood function $L(\Theta)$. Let $r_0$ denote the number of free parameters given $H_0$ and $r$ denote the number of free parameters given $\Theta \in \Omega$. Then for large $n$:

$$-2 \ln \lambda \sim \chi^2(r_0 - r) \tag{17}$$

# 3 Linear Regression

## 3.1 Parameters of a Linear Model

A single parameter linear model can be defined as follow:

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n + \epsilon \tag{18}$$

Define:
$\beta_0$ = the intercept
$\beta_1 \ldots \beta_n$ = the slope (s)
$\epsilon$ = the residual error
**Remark** The residual error is the difference between the observed value and the predicted value.

Define the sum of squares error as

$$SSE = \sum_{i=1}^{n} (y - \hat{y}_i)^2. \tag{19}$$

Where $\hat{y}_i$ is the predicted value of $y_i$.
We estimate the parameters $\beta_0, \beta_1, \ldots, \beta_n$ by minimizing the sum of squares of the residuals.

## 3.2 The Least Squares Estimator

The least-squares estimator is defined as follow:

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \tag{20}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{21}$$

Where $S_x y = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$ and $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

## 3.3 The correlation coefficient

The correlation coefficient $r$ is defined as follow:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \hat{\beta}_1\sqrt{\frac{S_{xx}}{S_{yy}}} \tag{22}$$

## 3.4 Inferences for Single Features

### 3.4.1 Testing for Large Samples

Given the null hypothesis $H_0 : \beta_1 = 0$ and the alternative hypothesis $H_a : \beta_1 \neq 0$, we can test the effectiveness of an individual slope or intercept with:

$$Z = \frac{\hat{\beta}_i - \beta_{i0}}{\sigma\sqrt{c_{ii}}} \tag{23}$$

Where

$$c_{00} = \frac{\sum x_i^2}{nS_{xx}} \text{ and } c_{ii} = \frac{1}{nS_{xx}} \tag{24}$$

### 3.4.2 Testing for Small Samples

Testing for small samples is a similiar process

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{S\sqrt{c_{ii}}} \tag{25}$$

Where we approximate $\sigma$ with $S = \sqrt{\frac{SSE}{(n-2)}}$, returning a T-distribution with (n-2) df.

## 3.5 Inferencces about The Model

Define
$\theta = a_0\beta_0 + a_1\beta_1$
$\hat{\theta} = a_0\beta_0 + a_1\beta_1$ And we want to test for

$$H_0 : \theta = \theta_0 \text{ vs } H_a : \theta \neq \theta_0$$

### 3.5.1 Testing for Large Samples

$$Z = \frac{\hat{\theta} - \theta_0}{\sigma\sqrt{\frac{a_0^2\frac{\sum x_i^2}{n} + a_1^2 - 2a_0a_1\bar{x}}{S_{xx}}}} \tag{26}$$

### 3.5.2 Testing for Small Samples

$$T = \frac{\hat{\theta} - \theta_0}{\sigma\sqrt{\frac{a_0^2\frac{\sum x_i^2}{n} + a_1^2 - 2a_0a_1\bar{x}}{S_{xx}}}} \tag{27}$$

where $df = n - 2$

### 3.6 Predicting Values

We can establish a confidence interval to predict the value of $Y$ for a given $x = x*$.

$$CI = \hat{\beta}_0 + \hat{\beta}_1 x^* \pm t_{\alpha/2,n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{S_{xx}}} \tag{28}$$

# Formulas, Tables, and Other Tools

## Theorem: Convergence in Probability

Suppose that $\hat{X}_n \to X$ in probability and $\hat{Y}_n \to Y$ in probability. Then:

- $\hat{X}_n + \hat{Y}_n \to X + Y$ in probability
- $\hat{X}_n \cdot \hat{Y}_n \to X \cdot Y$ in probability
- $Y \neq 0 \implies \frac{\hat{X}_n}{\hat{Y}_n} \to \frac{X}{Y}$ in probability
- $g(\cdot)$ is a continuous function at $X \implies g(\hat{X}_n) \to g(X)$ in probability

Suppose that $U_n$ converges to a standard normal as $n \to \infty$ and $W_n$ converges to 1. Then:

$$\frac{U_n}{W_n} \to N(0,1) \tag{29}$$

## Common T and Z hypothesis tests

| Test Parameter | Sample Size | Point Estimator | Standard Error |
|---|---|---|---|
| $\mu$ | $n$ | $\bar{X}$ | $\frac{\sigma}{\sqrt{n}}$ |
| $p$ | $n$ | $\hat{p} = \frac{X}{p}$ | $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ |
| $\mu_1 - \mu_2$ | $n_1 + n_2$ | $\bar{X}_1 - \bar{X}_2$ | $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ |
| $p_1 - p_2$ | $n_1 + n_2$ | $\hat{p}_1 - \hat{p}_2$ | $\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ |

$S \approx \sigma$, but given a small sample size ($n \leq 30$), add the extra parameter $df = n - 1$ to the t-distribution.

## Chi-Square distribution and Variance

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)} \tag{30}$$

## F-distribution in regards to chi-squares

$$F = \frac{W_1/df_1}{W_2/df_2} \sim F(df_1, df_2) \tag{31}$$

Where $W_1$ and $W_2$ are chi-squared random variables with $df_1$ and $df_2$.

## R-Scripts

All R-scripts below are available at https://github.com/SamZhang02/math324/tree/main/src/r_tools.

- Single/Multiple Linear Regression
- Hypothesis Testing