Math 505 Notes

Chapter 4

Jesse Crawford

Department of Mathematics
Tarleton State University

Fall 2009

Outline

Definition

- Let $\Omega \subseteq \mathbb{R}^p$
- For every $\theta \in \Omega$, suppose P_θ is a probability measure
- $\{P_\theta \mid \theta \in \Omega\}$ is a *statistical model*

- θ is called the *parameter*, and it is considered to be unknown
- Ω is called the *parameter space*

- Suppose $X_1, \ldots, X_n$ are i.i.d. random variables such that $X_i \sim P_\theta$
- Then $X_1, \ldots, X_n$ is a *random sample* from the above model.

- A function $T = T(X_1, \ldots, X_n)$ of the sample is called a *statistic*.
- If T is intended to estimate the unknown parameter θ, T is called an *estimator* of θ, often denoted $\hat{\theta}$.

- If $E(\hat{\theta}) = \theta$, for every $\theta \in \Omega$, then $\hat{\theta}$ is called *unbiased*.

Outline

Theorem

- $$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

- $$cov\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j cov(X_i, Y_j).$$

- *If the random variables are independent,*

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i).$$

Example

The statistical model

$$\{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$$

represents a normal distribution with unknown mean μ and variance σ^2.

Given a sample $X_1, \ldots, X_n$, estimators for μ and σ^2 are

$$\hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

These estimators are unbiased.

Outline

Definition

- Let $\{X_n\}$ be a sequence of random variables
- Let X be a random variable
- All defined on the same sample space
- Then X_n *converges in probability* to X if for every $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0,$$

- or equivalently

$$\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1.$$

- $X_n \xrightarrow{P} X$

Definition

- Let $\{P_\theta \mid \theta \in \Omega\}$ be a statistical model.
- Let $\hat{\theta}$ be an estimator for θ.
- If $\hat{\theta} \xrightarrow{P} \theta$, for every $\theta \in \Omega$, then $\hat{\theta}$ is called a *consistent* estimator.

Theorem (Weak Law of Large Numbers)

Let $\{X_n\}$ be a sequence of i.i.d. random variables with mean μ and $\sigma^2 < \infty$. Then

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.$$

In particular, $\bar{X}_n$ is a consistent estimator of μ.

Theorem

- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$.
- If $X_n \xrightarrow{P} X$ and $a \in \mathbb{R}$, then If $aX_n \xrightarrow{P} aX$.
- If $X_n \xrightarrow{P} a$, and g is continuous at a, then $g(X_n) \xrightarrow{P} g(a)$.
- If $X_n \xrightarrow{P} X$, and g is continuous, then $g(X_n) \xrightarrow{P} g(X)$.
- If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n Y_n \xrightarrow{P} XY$.

Outline

Definition

- Let X_n be a random variable with c.d.f. F_n, for $n = 1, 2, \ldots$.
- Let X be a random variable with c.d.f. F.
- If $\lim_{n\to\infty} F_n(x) = F(x)$ at every $x \in \mathbb{R}$ where F is continuous, then we say X_n *converges in distribution* (converges weakly) to X.
- Denoted $X_n \xrightarrow{D} X$

Example

Suppose X_n, $n = 1, 2, \ldots$, and X are continuous random variables, and $X_n \xrightarrow{D} X$. Then

$$P(a < X_n < b) \to P(a < X < b),$$

for all $a, b \in \mathbb{R}$.

Theorem

- *If* $X_n \xrightarrow{P} X$, *then* $X_n \xrightarrow{D} X$.
- *If a is a constant, then* $X_n \xrightarrow{P} a$ *if and only if* $X_n \xrightarrow{D} a$.

Theorem

- *If* $X_n \xrightarrow{D} X$, *and* $Y_n \xrightarrow{D} 0$, *then* $X_n + Y_n \xrightarrow{D} X$.
- *If* $X_n \xrightarrow{D} X$, *and g is a continuous function on the support of* X, *then* $g(X_n) \xrightarrow{D} g(X)$.
- *If* $X_n \xrightarrow{D} X$, $A_n \xrightarrow{D} a$, *and* $B_n \xrightarrow{D} b$, *where* $a, b \in \mathbb{R}$, *then*

$$A_n + B_n X_n \xrightarrow{D} a + bX.$$

Theorem

- *Suppose the m.g.f. of X_n, $M_n(t)$, is defined for $-h < t < h$, for all $n \in \mathbb{N}$, and*
- *the m.g.f. of X, $M(t)$, is defined for $-h < t < h$.*
- *If $M_n(t) \to M(t)$ for all $t \in (-h, h)$, then $X_n \xrightarrow{D} X$.*

Theorem (The Central Limit Theorem)

Suppose $\{X_i\}$ is an i.i.d. sequence of random variables with mean μ and finite variance $\sigma^2 > 0$. Then

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1).$$

- $\bar{X}_n \approx N(\mu, \sigma^2/n)$
- $\sum_{i=1}^{n} X_i \approx N(n\mu, \sigma^2 n)$

Theorem

If the moment-generating function of X, $M(t)$, exists on the interval $-h < t < h$, then

- *X has finite moments of all orders, i.e.,*

$$E(|X|^k) < \infty, \text{ for every } k = 1, 2, \ldots.$$

- *$M(t)$ has the power series representation*

$$M(t) = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k.$$

- *$M(t)$ is C^∞, and*

$$M^{(k)}(0) = E(X^k).$$

Theorem (Taylor's Theorem with Lagrange's Remainder)

- *Let f be a k times differentiable function on the interval I.*
- *Let $a \in I$.*
- *Then, for any $x \in I$, there exists c_x between a and x such that*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(k-1)}(a)}{(k-1)!}(x-a)^{k-1} + \frac{f^{(k)}(c_x)}{k!}(x-a)^k.$$

Outline

Definition

- Consider a statistical model with unknown parameter $\theta \in \mathbb{R}$.
- Let $X_1, \ldots, X_n$ be a sample from the model.
- Suppose (a, b) is a random interval based on this sample such that
$$P(a < \theta < b) = 1 - \alpha, \text{ where } \alpha \in [0, 1].$$
- Then (a, b) is a *confidence interval for* θ with *confidence level* or *confidence coefficient* $1 - \alpha$.

Example

- Consider a statistical model with finite positive variance σ^2 and mean μ.
- Let $X_1, \ldots, X_n$ be a large sample ($n \geq 30$).
- Let $\bar{X}_n$ and S_n be the sample mean and sample standard deviation, respectively.
- An *approximate* $1 - \alpha$ confidence interval for μ is

$$\left(\bar{X}_n - z_{\alpha/2} \frac{S_n}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2} \frac{S_n}{\sqrt{n}} \right),$$

- where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile of a standard normal distribution, i.e.,

$$P(Z \leq z_{\alpha/2}) = 1 - \alpha/2.$$

Example

- Suppose the population is normally distributed, i.e., the model is

$$\{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}.$$

- Then,

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

has a t-distribution with $n-1$ degrees of freedom (see Section 3.6).

- An *exact* $1-\alpha$ confidence interval for μ is

$$\left(\bar{X}_n - t_{\alpha/2,n-1}\frac{S_n}{\sqrt{n}}, \bar{X}_n + t_{\alpha/2,n-1}\frac{S_n}{\sqrt{n}}\right),$$

- where $t_{\alpha/2,n-1}$ is the $1-\alpha/2$ quantile of a t-distribution with $n-1$ degrees of freedom.
- This confidence interval is valid for any sample size $n \geq 2$.

Definition

Suppose Z and V are independent random variables such that $Z \sim N(0,1)$ and $V \sim \chi^2(n-1)$. Then the random variable

$$T = \frac{Z}{\sqrt{V/(n-1)}}$$

has a t-distribution with $n-1$ degrees of freedom.

Theorem (Student's Theorem)

Let $X_1, \ldots, X_n$ be a random sample from the distribution $N(\mu, \sigma^2)$, and let $\bar{X}$ and S^2 be the sample mean and variance.

- $\bar{X} \sim N(\mu, \sigma^2/n)$ *and* $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$
- $\bar{X}$ *and* S^2 *are independent*
- $$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

  has a t-distribution with $n-1$ degrees of freedom.

Example

- Consider two normally distributed populations with means μ_1 and μ_2 and the same variance σ^2.
- Let $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$ be independent samples from these populations, and let $n = n_1 + n_2$.
- Define the *pooled sample variance* to be

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n - 2}.$$

- Then a $1 - \alpha$ confidence interval for $\mu_1 - \mu_2$ is

$$\bar{X} - \bar{Y} \pm t_{\alpha/2, n-2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

Example

- Let $Y \sim b(n, p)$, where $n \in \mathbb{N}$ is known and $p \in (0, 1)$ is unknown.
- This is the setting of estimating an unknown population proportion p based on a sample of size n.
- For large sample sizes, an approximate $1 - \alpha$ confidence interval for p is

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},$$

- where $\hat{p} = Y/n$ is the proportion of successes in the sample.
- (The approximation is conventionally considered valid when $n\hat{p} \geq 5$ and $n(1-\hat{p}) \geq 5$, although some authors replace 5 with a more conservative value, such as 15.)
- Confidence interval for the difference of two proportions based on two independent samples:

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$