Convergence Concepts

Definition 5.8
Suppose that $X_1, X_2, \ldots$ is a sequence of random variables. We say that this sequence converges in distribution *to a random variable X if*

$$\lim_{n \to \infty} P(X_n \leqslant x) \ = \ P(X \leqslant x)$$

at all points x at which $F_X(x) = P(X \leqslant x)$ is continuous.

- This is quite a weak form of convergence since it only says that the distribution functions converge.

Definition 5.9

A sequence of random variables, $X_1, X_2, \ldots$ *converges in probability to a random variable* X *if, for every* $\epsilon > 0$

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

We will generally write this as

$$X_n \xrightarrow{p} X$$

Theorem 5.19 (Chebychev's Inequality)

Let X *be a random variable and let* $g(x)$ *be any non-negative function then*

$$P(g(X) \geqslant r) \leqslant \frac{\mathsf{E}[g(X)]}{r}$$

Theorem 5.20 (Weak Law of Large Numbers)
Let $X_1, X_2, \ldots$ *be* iid *random variables with* $\mathrm{E}[X_i] = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2 < \infty$. *Define the sample mean* $\overline{X}_n = n^{-1} \sum_1^n X_i$. *Then the sequence of random variables* $\overline{X}_1, \overline{X}_2, \ldots$ *converges in probability to the constant* μ. *That is for every* $\epsilon > 0$

$$\lim_{n \to \infty} P(|\overline{X}_n - \mu| < \epsilon) = 1$$

- Using characteristic functions we can prove that the requirement for finite variance is not needed.

Definition 5.10
Suppose that $T_n = T(X_1, \ldots, X_n)$ *is an estimator of a parameter* θ. *Then* T_n *is said to be a consistent estimator if* T_n *converges in probability to* θ.

Theorem 5.21

Suppose that $X_1, X_2, \ldots$ *converges in probability to a random variable* X *and that* h *is a continuous function. Then*

$$h(X_n) \xrightarrow{p} h(X)$$

Theorem 5.22

If a sequence of random variables $X_1, X_2, \ldots$ *converges in probability to a random variable* X *then the sequence also converges in distribution to* X*.*

Theorem 5.23

A sequence of random variables $X_1, X_2, \ldots$ *converges in probability to a constant* μ *if, and only if, the sequence converges in distribution to* μ*.*

Definition 5.11

A sequence of random variables $X_1, X_2, \ldots$ *converges almost surely to a random variable* X *if, for every* $\epsilon > 0$*,*

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

Theorem 5.24

If a sequence of random variables $X_1, X_2, \ldots$ *converges almost surely to a random variable* X*, it converges in probability to* X *also.*

Theorem 5.25 (Strong Law of Large Numbers)

Let $X_1, X_2, \ldots$ *be* iid *random variables with* $\mathrm{E}[X_i] = \mu$ *and* $\mathrm{Var}(X_i) = \sigma^2 < \infty$*. Define the sample mean* $\overline{X}_n = n^{-1} \sum_1^n X_i$*. Then the sequence of random variables* $\overline{X}_1, \overline{X}_2, \ldots$ *converges almost surely to the constant* μ*. That is for every* $\epsilon > 0$

$$P(\lim_{n\to\infty} |\overline{X}_n - \mu| < \epsilon) = 1$$

Theorem 5.26 (Central Limit Theorem)
Let $X_1, X_2, \ldots$ *be a sequence of* iid *random variables with* $E[X_i] = \mu$ *and* $\text{Var}(X_i) = \sigma^2 < \infty$. *Define the sample mean* $\overline{X}_n = n^{-1} \sum_1^n X_i$ *and let* $F_n(x)$ *denote the cdf of the random variable*

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$$

Then for any $x \in \mathbb{R}$,

$$\lim_{n \to \infty} F_n(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \, dz$$

Theorem 5.27 (Slutsky's Theorem)

If $X_n \xrightarrow{d} X$ *and* $Y_n \xrightarrow{p} a$ *where* a *is a constant then*

(i) $X_n Y_n \xrightarrow{d} aX$

(ii) $X_n + Y_n \xrightarrow{d} X + a$

Theorem 5.28 (Delta Method)

Let $Y_1, Y_2, \ldots$ *be a sequence of random variables such that* $\sqrt{n}(Y_n - \theta)$ *converges in distribution to a normal*$(0, \sigma^2)$ *random variable. Suppose that* g *is a function such that* $g'(\theta)$ *exists and is not 0, then*

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{d} \textit{normal}(0, [g'(\theta)]^2\sigma^2)$$

Theorem 5.29 (Second Order Delta Method)

Let $Y_1, Y_2, \ldots$ *be a sequence of random variables such that* $\sqrt{n}(Y_n - \theta)$ *converges in distribution to a normal*$(0, \sigma^2)$ *random variable. Suppose that* g *is a function such that* $g'(\theta) = 0$ *and* $g''(\theta)$ *exists and is not 0, then*

$$n[g(Y_n) - g(\theta)] \xrightarrow{d} \sigma^2 \frac{g''(\theta)}{2} \chi^2_1$$

6. Principles of Data Reduction

- In statistics we use a sample $X_1, \ldots, X_n$ to make inference about a parameter θ through the use of a statistic $T(\boldsymbol{X})$.

- The aim of data reduction is to keep all of the relevant information in the sample through a smaller number of statistics.

- Any statistic $T(\boldsymbol{X})$ defines a partition of the sample space into sets

$$A_t = \{\boldsymbol{x} : T(\boldsymbol{X} = t)\}$$

- Within such a partition we treat two samples, $\boldsymbol{x}$ and $\boldsymbol{y}$ as equal if $T(\boldsymbol{x}) = T(\boldsymbol{y})$.

The Sufficiency Principle

- The sufficiency principle relies on the concept of a sufficient statistic.

- Such statistics are functions of the data which contain all of the information about the parameter of interest.

Definition 6.1
A statistic $T(\boldsymbol{x})$ is a sufficient statistic for a parameter θ if the conditional distribution of the sample $\boldsymbol{X}$ given the value of $T(\boldsymbol{X})$ does not depend on θ.

The Sufficiency Principle

If $T(\boldsymbol{X})$ is a sufficient statistic for θ then any inference about θ should depend on the sample $\boldsymbol{X}$ only through the value of $T(\boldsymbol{X})$. If two sample points $\boldsymbol{x}$ and $\boldsymbol{y}$ have $T(\boldsymbol{x}) = T(\boldsymbol{y})$ then inference about θ should be the same whether $\boldsymbol{X} = \boldsymbol{x}$ or $\boldsymbol{X} = \boldsymbol{y}$.

Theorem 6.1

Let $\boldsymbol{X}$ be a random vector with pdf or pmf $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$ and let $T(\boldsymbol{X})$ be a statistic with pdf or pmf $f_T(t \mid \theta)$. Then $T(\boldsymbol{X})$ is a sufficient statistic if, for every $\boldsymbol{x}$ with $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta > 0)$, the ratio

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)}{f_T(T(\boldsymbol{x}) \mid \theta)}$$

is constant as a function of θ.

Theorem 6.2 (Factorization Criterion)

Let $f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$ be the joint pdf (pmf) of a sample $\boldsymbol{X}$. A statistic $T(\boldsymbol{X})$ is a sufficient statistic for θ if, and only if, there exist two non-negative functions $g(t, \theta)$ and $h(\boldsymbol{x})$ such that $h(\boldsymbol{x})$ is free of θ and for all sample points $\boldsymbol{x}$

$$f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) = g(T(\boldsymbol{x}), \theta) h(\boldsymbol{x}).$$

Theorem 6.3

If $X_1, \ldots, X_n$ is a random sample from a distribution having an exponential family form

$$f(x; \boldsymbol{\theta}) = h(x) c(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^{k} w_j(\boldsymbol{\theta}) t_j(x) \right\}$$

Then a sufficient statistic for $\boldsymbol{\theta}$ is the vector

$$T(X_1, \ldots, X_n) = \left\{ \sum_i t_1(x_i), \ldots, \sum_i t_k(x_i) \right\}$$

- In general, there are many sufficient statistics available for any model.

Theorem 6.4

Any one-to-one function of a sufficient statistic is also a sufficient statistic.

Definition 6.2

A sufficient statistic $T(\boldsymbol{X})$ is minimal sufficient if, and only if, for every other sufficient statistic $T'(\boldsymbol{X})$, $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$.

By saying that $T(\boldsymbol{X})$ is a function of $T'(\boldsymbol{X})$ we simply mean that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are sample points such that $T'(\boldsymbol{x}) = T'(\boldsymbol{y})$ then $T(\boldsymbol{x}) = T(\boldsymbol{y})$.

Theorem 6.5

Let $f(\boldsymbol{x} \mid \theta)$ be the pdf of a sample $\boldsymbol{X}$ and suppose that there exists a function $T(\boldsymbol{x})$ such that, for any two points $\boldsymbol{x}$ and $\boldsymbol{y}$, the ratio $f(\boldsymbol{x} \mid \theta)/f(\boldsymbol{y} \mid \theta)$ is constant in θ if, and only if, $T(\boldsymbol{x}) = T(\boldsymbol{y})$. Then $T(\boldsymbol{X})$ is a minimal sufficient statistic for θ.

Definition 6.3

Suppose that $\boldsymbol{X}$ is a random sample from a distribution depending on a parameter θ. A statistic $S(\boldsymbol{X})$ is called an ancillary statistic if the sampling distribution of $S(\boldsymbol{X})$ does not depend on θ.

Definition 6.4

Suppose $f(t \mid \theta)$ is the family of distributions for a statistic $T(\boldsymbol{X})$ indexed by the parameter θ. This family of distributions is called **complete** *if*

$$\mathsf{E}_\theta(g(T)) = 0 \implies \mathsf{P}_\theta(g(T) = 0) = 1 \text{ for every } \theta.$$

- A statistic $T(\boldsymbol{X})$ from a complete family of pdfs or pmfs is usually called a complete statistic.

- Unfortunately, it is often quite difficult to prove completeness.

Theorem 6.6

Suppose that $X_1, \ldots, X_n$ *is a random sample from an exponential family with pdf or pmf given by*

$$f(x \mid \boldsymbol{\theta}) \;=\; h(x)c(\boldsymbol{\theta}) \exp\left(\sum_{i=1}^{k} w_i(\boldsymbol{\theta}) t_i(x)\right)$$

and further suppose that the set $\{w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta})\}$ *contains an open subset in* $\mathbb{R}^k$.

Then the statistic

$$T(X_1, \ldots, X_n) = \left\{\sum_i t_1(x_i), \ldots, \sum_i t_k(x_i)\right\}$$

is complete and minimal sufficient.

Theorem 6.7 (Bahadur's Theorem)
If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.

Theorem 6.8 (Basu's Theorem)
If $T(\boldsymbol{X})$ *is a complete and minimal sufficient statistic then* $T(\boldsymbol{X})$ *is independent of any ancillary statistic.*