Mathematical Statistics 1

Chapter 5. Consistency and Limiting Distributions

5.1. Convergence in Probability—Proofs of Theorems

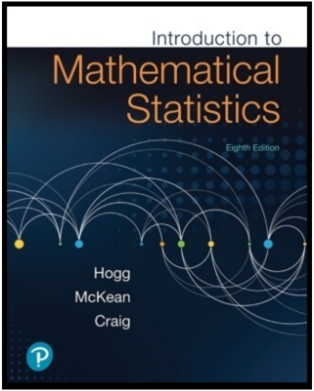

Table of contents

Theorem 5.1.1

Theorem 5.1.1. Weak Law of Large Numbers.
Let $\{X_n\}$ be a sequence of independent and identically distributed ("idd") random variables having common mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $\overline{X}_n = \left(\sum_{i=1}^n X_i\right)/n$ (this is the *sample mean*). Then $\overline{X}_n \xrightarrow{P} \mu$.

Proof. By Theorem 2.8.1, $E(\overline{X}_n) = \sum_{i=1}^n \mu/n = \mu$. By Corollary 2.8.2, $\text{Var}(\overline{X}_n) = \sum_{i=1}^n \sigma^2/n^2 = \sigma^2/n$. So by Chebychev's Inequality (Theorem 1.10.3; see Note 1.10.A), we have for every $\varepsilon > 0$

$$P(|\overline{X}_n - \mu| \geq \varepsilon) = P\left(|\overline{X}_n - \mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\text{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Theorem 5.1.1

Theorem 5.1.1. Weak Law of Large Numbers.
Let $\{X_n\}$ be a sequence of independent and identically distributed ("idd") random variables having common mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $\overline{X}_n = \left(\sum_{i=1}^n X_i\right)/n$ (this is the *sample mean*). Then $\overline{X}_n \xrightarrow{P} \mu$.

Proof. By Theorem 2.8.1, $E(\overline{X}_n) = \sum_{i=1}^n \mu/n = \mu$. By Corollary 2.8.2, $\text{Var}(\overline{X}_n) = \sum_{i=1}^n \sigma^2/n^2 = \sigma^2/n$. So by Chebychev's Inequality (Theorem 1.10.3; see Note 1.10.A), we have for every $\varepsilon > 0$

$$P(|\overline{X}_n - \mu| \geq \varepsilon) = P\left(|\overline{X}_n - \mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\text{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

For given $\varepsilon > 0$, $\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$ and so (by the Sandwich Theorem, say)

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) \to 0) \text{ and } \lim_{n\to\infty} P|\overline{X}_n - \mu| < \varepsilon) = 0.$$

Since $\varepsilon > 0$ is arbitrary, we have that $\overline{X}_n \xrightarrow{P} \mu$, as claimed. □

Theorem 5.1.1

Theorem 5.1.1. Weak Law of Large Numbers.
Let $\{X_n\}$ be a sequence of independent and identically distributed ("idd") random variables having common mean $\mu < \infty$ and variance $\sigma^2 < \infty$. Let $\overline{X}_n = \left(\sum_{i=1}^n X_i\right)/n$ (this is the *sample mean*). Then $\overline{X}_n \xrightarrow{P} \mu$.

Proof. By Theorem 2.8.1, $E(\overline{X}_n) = \sum_{i=1}^n \mu/n = \mu$. By Corollary 2.8.2, $\text{Var}(\overline{X}_n) = \sum_{i=1}^n \sigma^2/n^2 = \sigma^2/n$. So by Chebychev's Inequality (Theorem 1.10.3; see Note 1.10.A), we have for every $\varepsilon > 0$

$$P(|\overline{X}_n - \mu| \geq \varepsilon) = P\left(|\overline{X}_n - \mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\frac{\sigma}{\sqrt{n}}\right) \leq \frac{\text{Var}(\overline{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

For given $\varepsilon > 0$, $\lim_{n\to\infty} \dfrac{\sigma^2}{n\varepsilon^2} = 0$ and so (by the Sandwich Theorem, say)

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| \geq \varepsilon) \to 0) \text{ and } \lim_{n\to\infty} P|\overline{X}_n - \mu| < \varepsilon) = 0.$$

Since $\varepsilon > 0$ is arbitrary, we have that $\overline{X}_n \xrightarrow{P} \mu$, as claimed. □

Theorem 5.1.2

Theorem 5.1.2. Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then $X_n + Y_n \xrightarrow{P} X + Y$.

Proof. Let $\varepsilon > 0$ be given. Let $\mathcal{C}$ be the sample space on which the random variables are defined. Then for each $c \in \mathcal{C}$ we have by the Triangle Inequality on $\mathbb{R}$ that

$$|(X_n(c) + Y_n(c)) = (X(c) + Y(c)| \le |X_n(c) - X(c)| + |Y_n(c) - Y(c)|.$$

So

$$\{c \in \mathcal{C} \mid (X_n(c) + Y_n(c)) - (X(c) + Y(c))| \ge \varepsilon\}$$
$$\subseteq \{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \ge \varepsilon\}.$$

Theorem 5.1.2

Theorem 5.1.2. Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then $X_n + Y_n \xrightarrow{P} X + Y$.

Proof. Let $\varepsilon > 0$ be given. Let $\mathcal{C}$ be the sample space on which the random variables are defined. Then for each $c \in \mathcal{C}$ we have by the Triangle Inequality on $\mathbb{R}$ that

$$|(X_n(c) + Y_n(c)) = (X(c) + Y(c)| \leq |X_n(c) - X(c)| + |Y_n(c) - Y(c)|.$$

So

$$\{c \in \mathcal{C} \mid (X_n(c) + Y_n(c)) - (X(c) + Y(c))| \geq \varepsilon\}$$
$$\subseteq \{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\}.$$

Theorem 5.1.2 (continued 1)

Proof (continued). By Theorem 1.3.3, P is monotone so that

$$P(|(X_n + Y_n) - (X + Y)| \geq \varepsilon)$$

$$= P(\{c \in \mathcal{C} \mid (X_n(c) + Y_n(c)) - (X(c) + Y(c))| \geq \varepsilon\})$$

$$\leq P(\{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\})$$

$$= P(|X_n - X| + |Y_n - Y| \geq \varepsilon). \qquad (*)$$

Now for any $c \in \mathcal{C}$ such that $|X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon$, we must have either $X_n(c) - X(c)| \geq \varepsilon/2$ or $|Y_n - Y(c)| \geq \varepsilon/2$. That is,

$$\{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\}$$

$$\subseteq \{c \in \mathcal{C} \mid |X_n(c) - X(c)| \geq \varepsilon/2\} \cup \{c \in \mathcal{C} \mid |Y_n(c) = Y(c)| \geq \varepsilon/2\}.$$

Theorem 5.1.2 (continued 1)

Proof (continued). By Theorem 1.3.3, P is monotone so that

$$P(|(X_n + Y_n) - (X + Y)| \geq \varepsilon)$$

$$= P(\{c \in \mathcal{C} \mid (X_n(c) + Y_n(c)) - (X(c) + Y(c))| \geq \varepsilon\})$$

$$\leq P(\{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\})$$

$$= P(|X_n - X| + |Y_n - Y| \geq \varepsilon). \qquad (*)$$

Now for any $c \in \mathcal{C}$ such that $|X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon$, we must have either $X_n(c) - X(c)| \geq \varepsilon/2$ or $|Y_n - Y(c)| \geq \varepsilon/2$. That is,

$$\{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\}$$

$$\subseteq \{c \in \mathcal{C} \mid |X_n(c) - X(c)| \geq \varepsilon/2\} \cup \{c \in \mathcal{C} \mid |Y_n(c) = Y(c)| \geq \varepsilon/2\}.$$

Theorem 5.1.2 (continued 2)

Proof (continued). So by Theorem 1.3.3 (monotonicity or P) and Theorem 1.3.5 (which implies $P(A \cup B) \leq P(A) + P(B)$; this is called *subadditivity* in measure theory),

$$P(|X_n - X| + |Y_n - Y| \geq \varepsilon\})$$
$$= P(\{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\})$$
$$\leq P(\{c \in \mathcal{C} \mid |X_n(c) - X(c)| \geq \varepsilon/2\}) + P(\{c \in \mathcal{C} \mid |Y_n(c) - Y(c)| \geq \varepsilon/2\})$$
$$= P(|X_n - X| \geq \varepsilon/2) + P(|Y_n - Y| \geq \varepsilon/2).$$

Combining this with $(*)$ we have

$$P(|X_n + Y_n| - |X + Y|) \geq \varepsilon) \leq P(|X_n - X| \geq \varepsilon/2) + P(|Y_n - Y| \geq \varepsilon/2).$$

Since $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon/2) = 0 \text{ and } \lim_{n\to\infty} P(|Y_n - Y| \geq \varepsilon/2) = 0.$$

So $\lim_{n\to\infty} P(|X_n + Y_n| - |X + Y|) \geq \varepsilon) = 0$ and $X_n + Y_n \xrightarrow{P} X + Y$. □

Theorem 5.1.2 (continued 2)

Proof (continued). So by Theorem 1.3.3 (monotonicity or P) and Theorem 1.3.5 (which implies $P(A \cup B) \leq P(A) + P(B)$; this is called *subadditivity* in measure theory),

$$P(|X_n - X| + |Y_n - Y| \geq \varepsilon\})$$
$$= P(\{c \in \mathcal{C} \mid |X_n(c) - X(c)| + |Y_n(c) - Y(c)| \geq \varepsilon\})$$
$$\leq P(\{c \in \mathcal{C} \mid |X_n(c) - X(c)| \geq \varepsilon/2\}) + P(\{c \in \mathcal{C} \mid |Y_n(c) - Y(c)| \geq \varepsilon/2\})$$
$$= P(|X_n - X| \geq \varepsilon/2) + P(|Y_n - Y| \geq \varepsilon/2).$$

Combining this with $(*)$ we have

$$P(|X_n + Y_n| - |X + Y|) \geq \varepsilon) \leq P(|X_n - X| \geq \varepsilon/2) + P(|Y_n - Y| \geq \varepsilon/2).$$

Since $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ then

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon/2) = 0 \text{ and } \lim_{n\to\infty} P(|Y_n - Y| \geq \varepsilon/2) = 0.$$

So $\lim_{n\to\infty} P(|X_n + Y_n| - |X + Y|) \geq \varepsilon) = 0$ and $X_n + Y_n \xrightarrow{P} X + Y$. □

Theorem 5.1.3

Theorem 5.1.3. Suppose $X_n \xrightarrow{P} X$ and a is a constant. Then $aX_n \xrightarrow{P} aX$.

Proof. First, the result holds trivially if $a = 0$ so we can suppose without loss of generality that $a \neq 0$. We have

$$P(|aX_n - aX| \geq \varepsilon) = P(|a||X - X_n| \geq \varepsilon) = P(|X_n - X| \geq \varepsilon/|a|).$$

Since $X_n \xrightarrow{P} X$ the $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon/|a|) = 0$ so (by the Sandwich Theorem, say) $\lim_{n\to\infty} P(|aX_n - aX| \geq \varepsilon) = 0$ so that $aX_n \xrightarrow{P} aX$, as claimed. □

Theorem 5.1.3

Theorem 5.1.3. Suppose $X_n \xrightarrow{P} X$ and a is a constant. Then $aX_n \xrightarrow{P} aX$.

Proof. First, the result holds trivially if $a = 0$ so we can suppose without loss of generality that $a \neq 0$. We have

$$P(|aX_n - aX| \geq \varepsilon) = P(|a||X - X_n| \geq \varepsilon) = P(|X_n - X| \geq \varepsilon/|a|).$$

Since $X_n \xrightarrow{P} X$ the $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon/|a|) = 0$ so (by the Sandwich Theorem, say) $\lim_{n\to\infty} P(|aX_n - aX| \geq \varepsilon) = 0$ so that $aX_n \xrightarrow{P} aX$, as claimed. □

Theorem 5.1.4

Theorem 5.1.4. Suppose $X_n \xrightarrow{P} a$ and the real function g is continuous at a. Then $g(X_n) \xrightarrow{P} g(a)$.

Proof. Let $\varepsilon > 0$ be given. Then since g is continuous at a, by the definition of continuity there exists $\delta > 0$ such that if $|x - a| < \delta$ then $|g(x) - g(a)| < \varepsilon$. So for any x such that $|g(x) - g(a)| \geq \varepsilon$, we must have $|x - a| \geq \delta$. Let $\mathcal{C}$ be the sample space on which the random variables are defined. Then we have

$$\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\} \subseteq \{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}.$$

By Theorem 1.3.3, P is monotone so that

$$P(\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\}) \leq P(\{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}).$$

Theorem 5.1.4

Theorem 5.1.4. Suppose $X_n \xrightarrow{P} a$ and the real function g is continuous at a. Then $g(X_n) \xrightarrow{P} g(a)$.

Proof. Let $\varepsilon > 0$ be given. Then since g is continuous at a, by the definition of continuity there exists $\delta > 0$ such that if $|x - a| < \delta$ then $|g(x) - g(a)| < \varepsilon$. So for any x such that $|g(x) - g(a)| \geq \varepsilon$, we must have $|x - a| \geq \delta$. Let $\mathcal{C}$ be the sample space on which the random variables are defined. Then we have

$$\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\} \subseteq \{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}.$$

By Theorem 1.3.3, P is monotone so that

$$P(\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\}) \leq P(\{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}).$$

Since $X_n \xrightarrow{P} a$, then $\lim_{n\to\infty} P(\{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}) = 0$. So (by the Sandwich Theorem, say) $\lim_{n\to\infty} P(\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\}) = \lim_{n\to\infty} P(|g(X_n) - g(a)| \geq \varepsilon) = 0$ and $g(X_n) \xrightarrow{P} g(a)$, as claimed. □

Theorem 5.1.4

Theorem 5.1.4. Suppose $X_n \xrightarrow{P} a$ and the real function g is continuous at a. Then $g(X_n) \xrightarrow{P} g(a)$.

Proof. Let $\varepsilon > 0$ be given. Then since g is continuous at a, by the definition of continuity there exists $\delta > 0$ such that if $|x - a| < \delta$ then $|g(x) - g(a)| < \varepsilon$. So for any x such that $|g(x) - g(a)| \geq \varepsilon$, we must have $|x - a| \geq \delta$. Let $\mathcal{C}$ be the sample space on which the random variables are defined. Then we have

$$\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\} \subseteq \{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}.$$

By Theorem 1.3.3, P is monotone so that

$$P(\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\}) \leq P(\{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}).$$

Since $X_n \xrightarrow{P} a$, then $\lim_{n\to\infty} P(\{c \in \mathcal{C} \mid |X_n(c) - a| \geq \delta\}) = 0$. So (by the Sandwich Theorem, say) $\lim_{n\to\infty} P(\{c \in \mathcal{C} \mid |g(X_n(c)) - g(a)| \geq \varepsilon\}) = \lim_{n\to\infty} P(|g(X_n) - g(a)| \geq \varepsilon) = 0$ and $g(X_n) \xrightarrow{P} g(a)$, as claimed. □

Theorem 5.1.5

Theorem 5.1.5. Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then $X_n Y_n \xrightarrow{P} XY$.

Proof. First, $X_n Y_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2$.

We have by Theorem 5.1.2, Theorem 5.1.3, and Theorem 5.1.A that

$$X_n Y_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2 \xrightarrow{P} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X - Y)^2 = XY.$$

That is, $X_n Y_n \xrightarrow{P} XY$, as claimed. □

Theorem 5.1.5

Theorem 5.1.5. Suppose $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$. Then $X_n Y_n \xrightarrow{P} XY$.

Proof. First, $X_n Y_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2$.

We have by Theorem 5.1.2, Theorem 5.1.3, and Theorem 5.1.A that

$$X_n Y_n = \frac{1}{2}X_n^2 + \frac{1}{2}Y_n^2 - \frac{1}{2}(X_n - Y_n)^2 \xrightarrow{P} \frac{1}{2}X^2 + \frac{1}{2}Y^2 - \frac{1}{2}(X - Y)^2 = XY.$$

That is, $X_n Y_n \xrightarrow{P} XY$, as claimed. □

Theorem 5.1.B

Theorem 5.1.B. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution of X with finite mean μ and finite variance σ^2 where $E[X^4]$ is finite, then the sample variance

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

(where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$) is a consistent estimator of σ^2.

Proof. In Theorem 2.8.A we showed that S_n^2 is an unbiased estimator of σ^2 (that is, $E[S^2] = \sigma^2$). Here we need to show the convergence in probability. Since $E[X^4]$ is finite, then $\mathrm{Var}(S^2) < \infty$ so that the hypotheses of The Weak law of Large Numbers (Theorem 5.1.1) are satisfied. By Theorem 5.1.1, Theorem 5.1.2, Theorem 5.1.3, Theorem 5.1.A, and the fact that $\lim_{n\to\infty} n/(n-1) = 1$, we have...

Theorem 5.1.B

Theorem 5.1.B. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution of X with finite mean μ and finite variance σ^2 where $E[X^4]$ is finite, then the sample variance

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

(where $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$) is a consistent estimator of σ^2.

Proof. In Theorem 2.8.A we showed that S_n^2 is an unbiased estimator of σ^2 (that is, $E[S^2] = \sigma^2$). Here we need to show the convergence in probability. Since $E[X^4]$ is finite, then $\text{Var}(S^2) < \infty$ so that the hypotheses of The Weak law of Large Numbers (Theorem 5.1.1) are satisfied. By Theorem 5.1.1, Theorem 5.1.2, Theorem 5.1.3, Theorem 5.1.A, and the fact that $\lim_{n\to\infty} n/(n-1) = 1$, we have. . .

Theorem 5.1.B (continued)

Proof (continued).

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2\overline{X}_n}{n}\sum_{i=1}^n X_i + \frac{1}{n}(n\overline{X}_n)\right)$$

$$= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2\right) \xrightarrow{P} (1)(E[X^2] - \mu^2) = \sigma^2.$$

(We could use more details here on why $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} E[X^2]$; these details are to be given in Exercise 5.1.A.) That is, the sample variance S_n^2 is a consistent estimator of the variance σ^2, as claimed. □