1 The Glivenko-Cantelli Theorem

Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution function $F$ on $\mathbb{R}$. The empirical distribution function is the function of $x$ defined by

$$\hat{F}_n(x) = \frac{1}{n} \sum_{1 \leq i \leq n} I\{X_i \leq x\}.$$ 

For a given $x \in \mathbb{R}$, we can apply the strong law of large numbers to the sequence $I\{X_i \leq x\}, i = 1, \ldots, n$ to assert that

$$\hat{F}_n(x) \rightarrow F(x) \quad \text{a.s.}$$

(a.s in order to apply the strong law of large numbers we only need to show that $E[|I\{X_i \leq x\}|] < \infty$, which in this case is trivial because $|I\{X_i \leq x\}| \leq 1$). In this sense, $\hat{F}_n(x)$ is a reasonable estimate of $F(x)$ for a given $x \in \mathbb{R}$.

But is $\hat{F}_n(x)$ a reasonable estimate of the $F(x)$ when both are viewed as functions of $x$?

The Glivenko-Cantelli Theorem provides an answer to this question. It asserts the following:

**Theorem 1.1** Let $X_i, i = 1, \ldots, n$ be an i.i.d. sequence of random variables with distribution function $F$ on $\mathbb{R}$. Then,

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \rightarrow 0 \quad \text{a.s.} \quad (1)$$

This result is perhaps the oldest and most well known result in the very large field of empirical process theory, which is at the center of much of modern econometrics. The statistic (1) is an example of a Kolmogorov-Šnirnov statistic.

We will break the proof up into several steps.

**Lemma 1.1** Let $F$ be a (nonrandom) distribution function on $\mathbb{R}$. For each $\epsilon > 0$ there exists a finite partition of the real line of the form $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \leq j \leq k - 1$

$$F(t_{j+1}^-) - F(t_j) \leq \epsilon.$$
Proof: Let $\epsilon > 0$ be given. Let $t_0 = -\infty$ and for $j \geq 0$ define

$$t_{j+1} = \sup \{ z : F(z) \leq F(t_j) + \epsilon \} .$$

Note that $F(t_{j+1}) \geq F(t_j) + \epsilon$. To see this, suppose that $F(t_{j+1}) < F(t_j) + \epsilon$. Then, by right continuity of $F$ there would exist $\delta > 0$ so that $F(t_{j+1} + \delta) < F(t_j) + \epsilon$, which would contradict the definition of $t_{j+1}$. Thus, between $t_j$ and $t_{j+1}$, $F$ jumps by at least $\epsilon$. Since this can happen at most a finite number of times, the partition is of the desired form, that is $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ with $k < \infty$. Moreover, $F(t_{j+1}^-) \leq F(t_j) + \epsilon$. To see this, note that by definition of $t_{j+1}$ we have $F(t_{j+1}^-) \leq F(t_j) + \epsilon$ for all $\delta > 0$. The desired result thus follows from the definition of $F(t_{j+1}^-)$.

Lemma 1.2 Suppose $F_n$ and $F$ are (nonrandom) distribution functions on $\mathbb{R}$ such that $F_n(x) \to F(x)$ and $F_n(x^-) \to F(x^-)$ for all $x \in \mathbb{R}$. Then

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0 .$$

Proof: Let $\epsilon > 0$ be given. We must show that there exists $N = N(\epsilon)$ such that for $n > N$ and any $x \in \mathbb{R}$

$$|F_n(x) - F(x)| < \epsilon .$$

Let $\epsilon > 0$ be given and consider a partition of the real line into finitely many pieces of the form $-\infty = t_0 < t_1 < \cdots < t_k = \infty$ such that for $0 \leq j \leq k - 1$

$$F(t_{j+1}^-) - F(t_j) \leq \frac{\epsilon}{2} .$$

The existence of such a partition is ensured by the previous lemma.

For any $x \in \mathbb{R}$, there exists $j$ such that $t_j \leq x < t_{j+1}$. For such $j$,

$$F_n(t_j) \leq F_n(x) \leq F_n(t_{j+1})$$

and

$$F(t_j) \leq F(x) \leq F(t_{j+1}) ,$$

which implies that

$$F_n(t_j) - F(t_{j+1}^-) \leq F_n(x) - F(x) \leq F_n(t_{j+1}^-) - F(t_j) .$$
Furthermore,
\[ F_n(t_j) - F(t_j) + F(t_j) - F(t_{j+1}^-) \leq F_n(x) - F(x) \]
\[ F_n(t_{j+1}^-) - F(t_{j+1}^-) + F(t_{j+1}^-) - F(t_j) \geq F_n(x) - F(x) . \]

By construction of the partition, we have that
\[ F_n(t_j) - F(t_j) - \frac{\epsilon}{2} \leq F_n(x) - F(x) \]
\[ F_n(t_{j+1}^-) - F(t_{j+1}^-) + \frac{\epsilon}{2} \geq F_n(x) - F(x) . \]

For each \( j \), let \( N_j = N_j(\epsilon) \) be such that for \( n > N_j \)
\[ F_n(t_j) - F(t_j) > -\frac{\epsilon}{2} \]
and let \( M_j = M_j(\epsilon) \) be such that for \( n > M_j \)
\[ F_n(t_{j+1}^-) - F(t_{j+1}^-) < \frac{\epsilon}{2} . \]

Let \( N = \max_{1 \leq j \leq k} \max\{N_j, M_j\} \). For \( n > N \) and any \( x \in \mathbb{R} \), we have that
\[ |F_n(x) - F(x)| < \epsilon . \]

The desired result follows. ■

**Lemma 1.3** Suppose \( F_n \) and \( F \) are (nonrandom) distribution functions on \( \mathbb{R} \) such that \( F_n(x) \to F(x) \) for all \( x \in \mathbb{Q} \). Suppose further that \( F_n(x) - F_n(x^-) \to F(x) - F(x^-) \) for all jump points of \( F \). Then, for all \( x \in \mathbb{R} \)
\( F_n(x) \to F(x) \) and \( F_n(x^-) \to F(x^-) \).

**Proof:** Let \( x \in \mathbb{R} \). We first show that \( F_n(x) \to F(x) \). Let \( s, t \in \mathbb{Q} \) such that \( s < x < t \). First suppose \( x \) is a continuity point of \( F \). Since \( F_n(s) \leq F_n(x) \leq F_n(t) \) and \( s, t \in \mathbb{Q} \), it follows that
\[ F(s) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(t) . \]

Since \( x \) is a continuity point of \( F \),
\[ \lim_{s \to x^-} F(s) = \lim_{t \to x^+} F(t) = F(x) , \]

\[ \]
from which the desired result follows. Now suppose \( x \) is a jump point of \( F \).

Note that
\[
F_n(s) + F_n(x) - F_n(x^-) \leq F_n(x) \leq F_n(t) .
\]

Since \( s, t \in \mathbb{Q} \) and \( x \) is a jump point of \( F \),
\[
F(s) + F(x) - F(x^-) \leq \liminf_{n \to \infty} F_n(x) \leq \limsup_{n \to \infty} F_n(x) \leq F(t) .
\]

Since
\[
\lim_{s \to x^-} F(s) = F(x^-) \quad \lim_{t \to x^+} F(t) = F(x) ,
\]
the desired result follows.

We now show that \( F_n(x^-) \to F(x^-) \). First suppose \( x \) is a continuity point of \( F \). Since \( F_n(x^-) \leq F_n(x) \),
\[
\limsup_{n \to x^-} F_n(x^-) \leq \limsup_{n \to x^-} F_n(x) = F(x) = F(x^-) .
\]

For any \( s \in \mathbb{Q} \) such that \( s < x \), we have \( F_n(s) \leq F_n(x^-) \), which implies that
\[
F(s) \leq \liminf_{n \to \infty} F_n(x^-) .
\]

Since
\[
\lim_{s \to x^-} F(s) = F(x^-) ,
\]
the desired result follows. Now suppose \( x \) is a jump point of \( F \). By assumption, \( F_n(x) - F_n(x^-) \to F(x) - F(x^-) \), and, by the above argument, \( F_n(x) \to F(x) \). The desired result follows. ■

**Proof of Theorem 1.1:** If we can show that there exists a set \( N \) such that \( \Pr\{N\} = 0 \) and for all \( \omega \notin N \) (i) \( \hat{F}_n(x, \omega) \to F(x) \) for all \( x \in \mathbb{Q} \) and (ii) \( \hat{F}_n(x, \omega) - F_n(x, \omega^-) \to F(x) - F(x^-) \) for all jump points of \( F \), then the result will follow from an application of Lemmas 1.2 and 1.3.

For each \( x \in \mathbb{Q} \), let \( N_x \) be a set such that \( \Pr\{N_x\} = 0 \) and for all \( \omega \notin N_x \), \( \hat{F}_n(x, \omega) \to F(x) \). Let \( N_1 = \bigcup_{x \in \mathbb{Q}} N_x \). Then, for all \( \omega \notin N_1 \), \( \hat{F}_n(x, \omega) \to F(x) \) by construction. Moreover, since \( \mathbb{Q} \) is countable, \( \Pr\{N_1\} = 0 \).
For integer \( i \geq 1 \), let \( J_i \) denote the set of jump points of \( F \) of size at least \( 1/i \). Note that for each \( i \), \( J_i \) is finite. Next note that the set of all jump points of \( F \) can be written as \( J = \bigcup_{1 \leq i < \infty} J_i \). For each \( x \in J \), let \( M_x \) denote a set such that \( \Pr\{M_x\} = 0 \) and for all \( \omega \not\in M_x \), \( \hat{F}_n(x, \omega) - F_n(x^-, \omega) \to F(x) - F(x^-) \). Let \( N_2 = \bigcup_{x \in J} M_x \). Since \( J \) is countable, \( \Pr\{N_2\} = 0 \).

To complete the proof, let \( N = N_1 \cup N_2 \). By construction, for \( \omega \not\in N \), (i) and (ii) hold. Moreover, \( \Pr\{N\} = 0 \). The desired result follows. ■

## 2 The Sample Median

We now give a brief application of the Glivenko-Cantelli Theorem. Let \( X_i, i = 1, \ldots, n \) be an i.i.d. sequence of random variables with distribution \( F \). Suppose one is interested in the median of \( F \). Concretely, we will define

\[
\text{Med}(F) = \inf\{x : F(x) \geq \frac{1}{2}\}.
\]

A natural estimator of \( \text{Med}(F) \) is the sample analog, \( \text{Med}(\hat{F}_n) \). Under what conditions is \( \text{Med}(\hat{F}_n) \) a reasonable estimate of \( \text{Med}(F) \)?

Let \( m = \text{Med}(F) \) and suppose that \( F \) is well behaved at \( m \) in the sense that \( F(t) > \frac{1}{2} \) whenever \( t > m \). Under this condition, we can show using the Glivenko-Cantelli Theorem that \( \text{Med}(\hat{F}_n) \to \text{Med}(F) \) a.s. We will now prove this result.

Suppose \( F_n \) is a (nonrandom) sequence of distribution functions such that

\[
\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \to 0.
\]

Let \( \epsilon > 0 \) be given. We wish to show that there exists \( N = N(\epsilon) \) such that for all \( n > N \)

\[
|\text{Med}(F_n) - \text{Med}(F)| < \epsilon.
\]

Choose \( \delta > 0 \) so that

\[
\delta < \frac{1}{2} - F(m - \epsilon)
\]

and

\[
\delta < F(m + \epsilon) - \frac{1}{2},
\]

which implies that

\[
\text{Med}(F_n) < m + \epsilon.
\]
which in turn implies that

\[ F(m - \epsilon) \ < \ \frac{1}{2} - \delta \]
\[ F(m + \epsilon) \ > \ \frac{1}{2} + \delta . \]

(It might help to draw a picture to see why we should pick \( \delta \) in this way.)

Next choose \( N \) so that for all \( n > N \),

\[ \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| < \delta . \]

Let \( m_n = \text{Med}(F_n) \). For such \( n \), \( m_n > m - \epsilon \), for if \( m_n \leq m - \epsilon \), then

\[ F(m - \epsilon) > F_n(m - \epsilon) - \delta \geq \frac{1}{2} - \delta , \]

which contradicts the choice of \( \delta \). We also have that \( m_n < m + \epsilon \), for if \( m_n \geq m + \epsilon \), then

\[ F(m + \epsilon) < F_n(m + \epsilon) + \delta \leq \frac{1}{2} + \delta , \]

which again contradicts the choice of \( \delta \). Thus, for \( n > N \), \( |m_n - m| < \epsilon \), as desired.

By the Glivenko-Cantelli Theorem, it follows immediately that \( \text{Med}(\hat{F}_n) \rightarrow \text{Med}(F) \) a.s.