Limit Theorems[1]
Sections 4.2 and 4.4
STA 256: Fall 2019

[1] This slide show is an open-source document. See last slide for copyright information.

Overview

Infinite Sequence of random variables

$T_1, T_2, \ldots$

- We are interested in what happens to T_n as $n \to \infty$.
- Why even think about this?
- For fun.
- And because T_n could be a sequence of *statistics*, numbers computed from sample data.
- For example, $T_n = \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$.
- n is the sample size.
- $n \to \infty$ is an approximation of what happens for large samples.
- Good things should happen when estimates are based on more information.

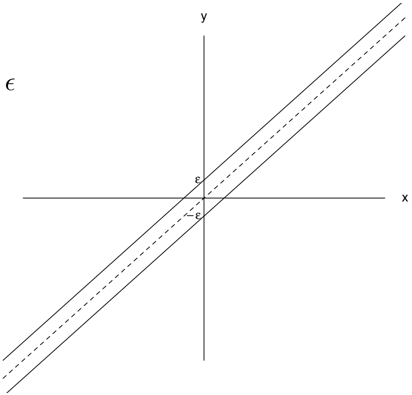
Convergence

- Convergence of T_n as $n \to \infty$ is not an ordinary limit, because probability is involved.
- There are several different types of convergence.
- We will work with *convergence in probability* and *convergence in distribution*.

Convergence in Probability to a random variable

Definition: The sequence of random variables $X_1, X_2, \ldots$ is said to converge in probability to the random variable Y if for all $\epsilon > 0$, $\lim_{n\to\infty} P\{|X_n - Y| \geq \epsilon\} = 0$, and we write $X_n \xrightarrow{p} Y$.

$$\begin{aligned} |X_n - Y| < \epsilon \quad &\Leftrightarrow \quad -\epsilon < X_n - Y < \epsilon \\ &\Leftrightarrow \quad Y - \epsilon < X_n < Y + \epsilon \end{aligned}$$



Convergence in Probability to a constant

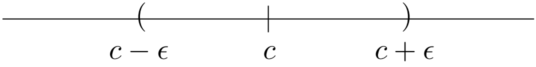
More immediate applications in statistics: We will focus on this.

Definition: The sequence of random variables $T_1, T_2, \ldots$ is said to converge in probability to the constant c if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\{|T_n - c| \geq \epsilon\} = 0$$

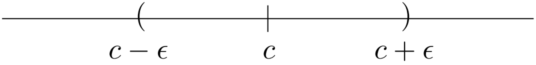
and we write $T_n \xrightarrow{p} c$.

$$\begin{aligned} |T_n - c| < \epsilon \quad & \Leftrightarrow \quad -\epsilon < T_n - c < \epsilon \\ & \Leftrightarrow \quad c - \epsilon < T_n < c + \epsilon \end{aligned}$$



Example: $T_n \sim U(-\frac{1}{n}, \frac{1}{n})$

Convergence in probability means $\lim_{n\to\infty} P\{|T_n - c| \geq \epsilon\} = 0$



- T_1 is uniform on $(-1, 1)$. Height of the density is $\frac{1}{2}$.
- T_2 is uniform on $(-\frac{1}{2}, \frac{1}{2})$. Height of the density is 1.
- T_3 is uniform on $(-\frac{1}{3}, \frac{1}{3})$. Height of the density is $\frac{3}{2}$.
- Eventually, $\frac{1}{n} < \epsilon$ and $P\{|T_n - 0| \geq \epsilon\} = 0$, forever.
- Eventually means for all $n > \frac{1}{\epsilon}$.

Example: $X_1, \ldots, X_n$ are independent $U(0, \theta)$

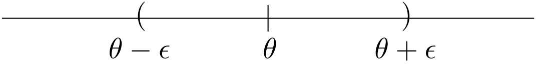
Convergence in probability means $\lim_{n\to\infty} P\{|T_n - c| \geq \epsilon\} = 0$

For $0 < x < \theta$,

$F_{X_i}(x) = \int_0^x \frac{1}{\theta}\, dt = \frac{x}{\theta}$.

$Y_n = \max_i(X_i)$.

$F_{Y_n}(y) = \left(\frac{y}{\theta}\right)^n$



$$
\begin{aligned}
P\{|Y_n - \theta| \geq \epsilon\} &= F_{Y_n}(\theta - \epsilon) \\
&= \left(\frac{\theta - \epsilon}{\theta}\right)^n \\
&\to 0 \quad \text{because } \frac{\theta - \epsilon}{\theta} < 1.
\end{aligned}
$$

So the observed maximum data value goes in probability to θ, the theoretical maximum data value.

Markov's inequality: Theorem 3.6.1

A stepping stone

Let Y be a random variable with $P(Y \geq 0) = 1$. Then for any $a > 0$, $E(Y) \geq a\,P(Y \geq a)$.

Proof (for continuous random variables):

$$
\begin{aligned}
E(Y) &= \int_0^\infty y f(y)\,dy \\
&= \int_0^a y f(y)\,dy + \int_a^\infty y f(y)\,dy \\
&\geq \int_a^\infty y f(y)\,dy \\
&\geq \int_a^\infty a f(y)\,dy \\
&= a \int_a^\infty f(y)\,dy \\
&= a\,P(Y \geq a) \qquad \blacksquare
\end{aligned}
$$

The Variance Rule

Not in the text, I believe

Let $T_1, T_2, \ldots$ be a sequence of random variables, and let c be a constant. If

- $\lim_{n \to \infty} E(X_n) = c$ and
- $\lim_{n \to \infty} Var(X_n) = 0$

Then $T_n \xrightarrow{p} c$.

Proof of the Variance Rule

Using Markov's inequality: $E(Y) \geq a\, P(Y \geq a)$

Seek to show $\forall \epsilon > 0$, $\lim_{n \to \infty} P\{|T_n - c| \geq \epsilon\} = 0$. Denote $E(T_n)$ by μ_n.

In Markov's inequality, let $Y = (T_n - c)^2$, and $a = \epsilon^2$.

$$
\begin{aligned}
E[(T_n - c)^2] &\geq \epsilon^2 P\{(T_n - c)^2 \geq \epsilon^2\} \\
&= \epsilon^2 P\{|T_n - c| \geq \epsilon\}, \text{ so} \\
0 &\leq P\{|T_n - c| \geq \epsilon\} \leq \frac{1}{\epsilon^2} E[(T_n - c)^2] \\
&= \frac{1}{\epsilon^2} E[(T_n - \mu_n + \mu_n - c)^2] \\
&= \frac{1}{\epsilon^2} E[(T_n - \mu_n)^2 + 2(T_n - \mu_n)(\mu_n - c) + (\mu_n - c)^2] \\
&= \frac{1}{\epsilon^2} \left(E(T_n - \mu_n)^2 + 2(\mu_n - c) E(T_n - \mu_n) + E(\mu_n - c)^2 \right) \\
&= \frac{1}{\epsilon^2} \left(E(T_n - \mu_n)^2 + 2(\mu_n - c)(E(T_n) - \mu_n) + (\mu_n - c)^2 \right) \\
&= \frac{1}{\epsilon^2} \left(E(T_n - \mu_n)^2 + 0 + (\mu_n - c)^2 \right)
\end{aligned}
$$

Continuing the proof

Have

$$\begin{aligned}
0 &\leq P\{|T_n - c| \geq \epsilon\} \\
&\leq \frac{1}{\epsilon^2}\left(E(T_n - \mu_n)^2 + (\mu_n - c)^2\right) \\
&= \frac{1}{\epsilon^2}\left(Var(T_n) + (\mu_n - c)^2\right), \text{ so that} \\
0 &\leq \lim_{n\to\infty} P\{|T_n - c| \geq \epsilon\} \\
&\leq \lim_{n\to\infty} \frac{1}{\epsilon^2}\left(Var(T_n) + (\mu_n - c)^2\right) \\
&= \frac{1}{\epsilon^2}\left(\lim_{n\to\infty} Var(T_n) + \lim_{n\to\infty} (\mu_n - c)^2\right) \\
&= \frac{1}{\epsilon^2}\left(\lim_{n\to\infty} Var(T_n) + \left(\lim_{n\to\infty} \mu_n - \lim_{n\to\infty} c\right)^2\right) \\
&= \frac{1}{\epsilon^2}\left(0 + (c - c)^2\right) = 0
\end{aligned}$$

Squeeze. ■

The Law of Large Numbers

That is, the "Weak" Law of Large Numbers

Theorem: Let $X_1, \ldots, X_n$ be independent random variables with expected value μ and variance σ^2. Then the sample mean

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

Proof: $E(\overline{X}_n) = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$.
As $n \to \infty$, $E(\overline{X}_n) \to \mu$ and $Var(\overline{X}_n) \to 0$.
So by the Variance Rule, $\overline{X}_n \xrightarrow{p} \mu$. ■

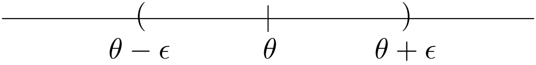
The implications are huge.

Probability is long-run relative frequency

Sometimes offered as a *definition* of probability!

This follows from the Law of Large Numbers.
Repeat some process over and over a lot of times, and count how many times the event A occurs. Independently for $i = 1, \ldots, n$,

- Let $X_i(s) = 1$ if $s \in A$, and $X_i(s) = 0$ if $s \notin A$.
- So X_i is an *indicator* for the event A.
- X_i is Bernoulli, with $P(X_i = 1) = \theta = P(A)$.
- $E(X_i) = \sum_{x=0}^{1} x\, p(x) = 0 \cdot (1 - \theta) + 1 \cdot \theta = \theta$.
- $\overline{X}_n$ is the proportion of times the event occurs in n independent trials.
- The proportion of successes converges in probability to $P(A)$.



More comments

- Law of Large Numbers is the basis of using *simulation* to estimate probabilities.
- Have things like $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \stackrel{p}{\rightarrow} E(X^2)$
- In fact, $\frac{1}{n}\sum_{i=1}^{n} g(X_i) \stackrel{p}{\rightarrow} E[g(X)]$
- Convergence in probability also applies to *vectors* of random variables, like $(X_n, Y_n) \stackrel{p}{\rightarrow} (c_1, c_2)$.

Theorem

Continuous Mapping Theorem for convergence in probability

Let $g(x)$ be a function that is continuous at $x = c$. If $T_n \xrightarrow{p} c$, then $g(T_n) \xrightarrow{p} g(c)$.

Examples:

- A Geometric distribution has expected value $\frac{1-\theta}{\theta}$. $g(\overline{X}_n) = 1/(1 + \overline{X}_n)$ converges in probability to

$$\begin{aligned} \frac{1}{1 + E(X_i)} &= \frac{1}{1 + \frac{1-\theta}{\theta}} \\ &= \theta \end{aligned}$$

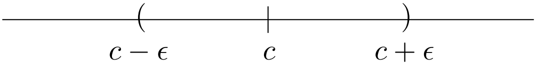
- A Uniform$(0, \theta)$ distribution has expected value $\theta/2$. So $2\overline{X}_n \xrightarrow{p} 2E(X_i) = 2\frac{\theta}{2} = \theta$

Background

For the proof of the continuous mapping theorem

- $T_n \xrightarrow{p} c$ means that for all $\epsilon > 0$,

$$\begin{aligned} & \lim_{n \to \infty} P\{|T_n - c| \geq \epsilon\} = 0 \\ \Leftrightarrow \quad & \lim_{n \to \infty} P\{|T_n - c| < \epsilon\} = 1 \end{aligned}$$



- $g(x)$ continuous at c means that for all $\epsilon > 0$, there exists $\delta > 0$ such that if $|x - c| < \delta$, then $|g(x) - g(c)| < \epsilon$.

Proof of the Continuous Mapping Theorem

For convergence in probability

Have $T_n \stackrel{p}{\rightarrow} c$ and $g(x)$ continuous at c. Seek to show that for all $\epsilon > 0$, $\lim_{n\rightarrow\infty} P\{|g(T_n) - g(c)| < \epsilon\} = 1$. Let $\epsilon > 0$ be given.
$g(x)$ continuous at c means there exists $\delta > 0$ such that for $s \in S$, if $|X_n(s) - c| < \delta$, then $|g(X_n(s)) - g(c)| < \epsilon$. That is,

If $s_0 \in \{s : |X_n(s) - c| < \delta\}$, then $s_0 \in \{s : |g(X_n(s)) - g(c)| < \epsilon\}$.
This is the definition of containment:

$$
\begin{aligned}
& \{s : |X_n(s) - c| < \delta\} \subseteq \{s : |g(X_n(s)) - g(c)| < \epsilon\} \\
\Rightarrow \quad & P(|X_n - c| < \delta) \leq P(|g(X_n) - g(c)| < \epsilon) \leq 1 \\
\Rightarrow \quad & \underbrace{\lim_{n\rightarrow\infty} P(|X_n - c| < \delta)}_{=\,1} \leq \lim_{n\rightarrow\infty} P(|g(X_n) - g(c)| < \epsilon) \leq 1
\end{aligned}
$$

Squeeze ■

Convergence in distribution

Another mode of convergence

Definition: Let the random variables $X_1, X_2 \ldots$ have cumulative distribution functions $F_{X_1}(x), F_{X_2}(x)\ldots$, and let the random variable X have cumulative distribution function $F_X(x)$. The (sequence of) random variable(s) X_n is said to *converge in distribution* to X if

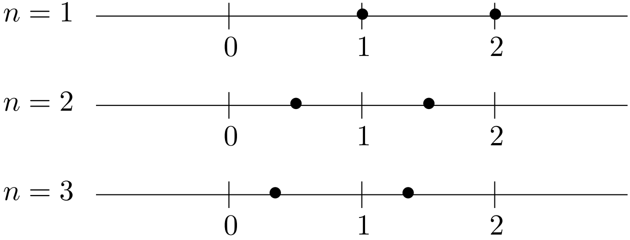
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

at every point where $F_X(x)$ is continuous, and we write $X_n \xrightarrow{d} X$.

Example: Convergence to a Bernoulli with $p = \frac{1}{2}$

$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all continuity points of $F_X(x)$

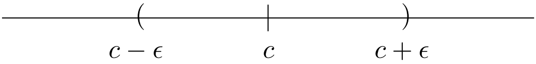
$$p_{X_n}(x) = \begin{cases} 1/2 & \text{for } x = \frac{1}{n} \\ 1/2 & \text{for } x = 1 + \frac{1}{n} \\ 0 & \text{Otherwise} \end{cases}$$



- For $x < 0$, $\lim_{n\to\infty} F_{X_n}(x) = 0$
- For $0 < x < 1$, $\lim_{n\to\infty} F_{X_n}(x) = \frac{1}{2}$
- For $x > 1$, $\lim_{n\to\infty} F_{X_n}(x) = 1$
- What happens at $x = 0$ and $x = 1$ does not matter.

Convergence to a constant

Consider a "degenerate" random variable X with $P(X = c) = 1$.



Suppose X_n converges in probability to c.

- Then for any $x > c$, $F_{X_n}(x) \rightarrow 1$ for ϵ small enough.
- And for any $x < c$, $F_{X_n}(x) \rightarrow 0$ for ϵ small enough.
- So X_n converges in distribution to c.

Suppose X_n converges in distribution to c, so that $F_{X_n}(x) \rightarrow 1$ for all $x > c$ and $F_{X_n}(x) \rightarrow 0$ for all $x < c$. Let $\epsilon > 0$ be given.

$$\begin{aligned} P\{|X_n - c| < \epsilon\} &= P\{c - \epsilon < X_n < c + \epsilon\} \\ &= F_{X_n}(c+\epsilon) - F_{X_n}(c-\epsilon) \text{ so} \\ \lim_{n\rightarrow\infty} P\{|X_n - c| < \epsilon\} &= \lim_{n\rightarrow\infty} F_{X_n}(c+\epsilon) - \lim_{n\rightarrow\infty} F_{X_n}(c-\epsilon) \\ &= 1 - 0 = 1 \end{aligned}$$

And X_n converges in probability to c.

Comment

- Convergence in probability might seem redundant, because it's just convergence in distribution to a constant.
- But that's only true when the convergence is to a constant.
- Convergence in probability to a non-degenerate random variable implies convergence in distribution.
- But convergence in distribution does not imply convergence in probability when the convergence is to a non-degenerate variable.

Big Theorem about convergence in distribution

Theorem 4.4.2 in the text

Let the random variables $X_1, X_2 \dots$ have cumulative distribution functions $F_{X_1}(x), F_{X_2}(x) \dots$ and moment-generating functions $M_{X_1}(t), M_{X_2}(t) \dots$.
Let the random variable X have cumulative distribution function $F_X(x)$ and moment-generating function $M_X(t)$.
If

$$\lim_{n \to \infty} M_{X_n}(t) = M_X(t)$$

for all t in an open interval containing $t = 0$, then X_n converges in distribution to X.

The idea is that convergence of moment-generating functions implies convergence of distribution functions. This makes sense because moment-generating functions and distribution functions are one-to-one.

Example: Poisson approximation to the binomial

We did this before with probability mass functions and it was a challenge.

Let X_n be a binomial (n, p_n) random variable with $p_n = \frac{\lambda}{n}$, so that $n \to \infty$ and $p \to 0$ in such a way that the value of $n\, p_n = \lambda$ remains fixed. Find the limiting distribution of X_n.

Recalling that the MGF of a Poisson is $e^{\lambda(e^t-1)}$ and $\left(1 + \frac{x}{n}\right)^n \to e^x$,

$$\begin{aligned}
M_{X_n}(t) &= (\theta e^t + 1 - \theta)^n \\
&= \left(\frac{\lambda}{n} e^t + 1 - \frac{\lambda}{n}\right)^n \\
&= \left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n \\
&\to e^{\lambda(e^t-1)}
\end{aligned}$$

MGF of Poisson(λ).

The Central Limit Theorem

Proved using limiting moment-generating functions

Let $X_1, \ldots, X_n$ be independent random variables from a distribution with expected value μ and variance σ^2. Then

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} Z \sim N(0,1)$$

In practice, Z_n is often treated as standard normal for $n > 25$, although the n required for an accurate approximation really depends on the distribution.

Sometimes we say the distribution of the sample mean is approximately normal, or "asymptotically" normal.

- This is justified by the Central Limit Theorem.
- But it does *not* mean that $\overline{X}_n$ converges in distribution to a normal random variable.
- The Law of Large Numbers says that $\overline{X}_n$ converges in probability to a constant, μ.
- So $\overline{X}_n$ converges to μ in distribution as well.
- That is, $\overline{X}_n$ converges in distribution to a degenerate random variable with all its probability at μ.

Why would we say that for large n, the sample mean is approximately $N(\mu, \frac{\sigma^2}{n})$?

Have $Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$ converging to $Z \sim N(0,1)$.

$$
\begin{aligned}
Pr\{\overline{X}_n \leq x\} &= Pr\left\{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\
&= Pr\left\{Z_n \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \approx \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right)
\end{aligned}
$$

Suppose Y is *exactly* $N(\mu, \frac{\sigma^2}{n})$:

$$
\begin{aligned}
Pr\{Y \leq x\} &= Pr\left\{\frac{\sqrt{n}(Y - \mu)}{\sigma} \leq \frac{x-\mu}{\sigma/\sqrt{n}}\right\} \\
&= Pr\left\{Z_n \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} = \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right)
\end{aligned}
$$

Copyright Information