Limit Theorems[1]

STA 256: Fall 2018

[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.
- In this class, we will work with *convergence in probability* and *convergence in distribution*.

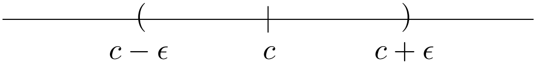
Convergence in Probability

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

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

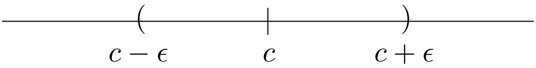
Observe

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



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 $U(0, \theta)$

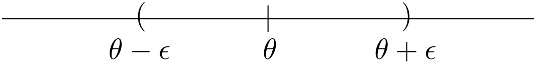
Convergence in probability means $\lim_{n \to \infty} P\{|T_n - c| \geq \epsilon\} = 0$

For $0 < x < \theta$,

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

$Y_n = \max_i(X_i)$.

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



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

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

The Law of Large Numbers

Theorem: Let $X_1, \ldots, X_n$ be independent random variables with expected value μ and variance σ^2. Then $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ converges in probability to μ.

- This is not surprising, because $E(\overline{X}_n) = \mu$ and
- $Var(\overline{X}_n) = \frac{\sigma^2}{n}$

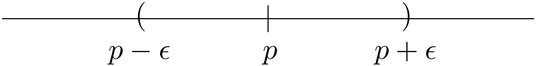
$$\begin{aligned} Var\left(\frac{1}{n}\sum_{i=1}^n X_i\right) &= \frac{1}{n^2} Var\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n} \downarrow 0. \end{aligned}$$

- And the implications are huge.

Probability is long-run relative frequency

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



Proof of the Law of Large Numbers

Using $E(\overline{X}_n) = \mu$ and $Var(\overline{X}_n) = \frac{\sigma^2}{n}$

- Chebyshev's inequality says $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$
- Here, X is replaced by $\overline{X}_n$ and σ is replaced by $\frac{\sigma}{\sqrt{n}}$.
- So Chebyshev's inequality becomes $P(|\overline{X}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}) \leq \frac{1}{k^2}$.
- $k > 0$ is arbitrary, so set $\frac{k\sigma}{\sqrt{n}} = \epsilon$.
- Then $k = \frac{\epsilon\sqrt{n}}{\sigma}$ and $\frac{1}{k^2} = \frac{\sigma^2}{\epsilon^2 n}$.
- Thus,

$$0 \leq P\{|\overline{X}_n - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2 n} \downarrow 0$$

Squeeze. ∎

Theorem

Proof omitted in 2018

Let $g(x)$ be a function that is continuous at $x = c$. If T_n converges in probability to c, then $g(T_n)$ converges in probability to $g(c)$.

Examples:

- A Geometric distribution has expected value $1/p$. $1/\overline{X}_n$ converges in probability to $1/E(X_i) = p$.
- A Uniform$(0, \theta)$ distribution has expected value $\theta/2$. $2\overline{X}_n$ converges in probability to $2E(X_i) = 2\frac{\theta}{2} = \theta$.

Convergence in distribution

Another mode of convergence

Definition: Let the random variables $X_1, X_2 \ldots$ have cumulative distribution functions $F_1(x), F_2(x) \ldots$, and let the random variable X have cumulative distribution function $F(x)$. The (sequence of) random variable X_n is said to *converge in distribution* to X if

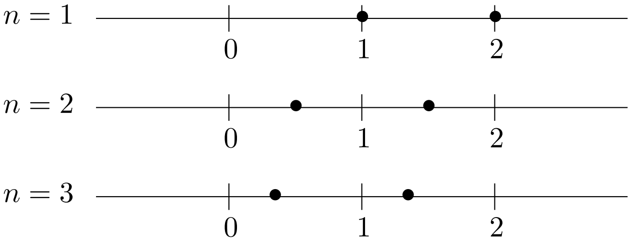
$$\lim_{n \to \infty} F_n(x) = F(x)$$

at every point where $F(x)$ is continuous.

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

$\lim_{n\to\infty} F_n(x) = F(x)$ at all continuity points of $F(x)$

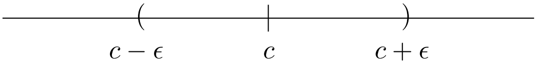
$$p_n(x) = \begin{cases} 1/2 & \text{for } x = \frac{1}{n} \\ 1/2 & \text{for } x = 1 + \frac{1}{n} \\ 0 & \text{Otherwise} \end{cases}$$



- For $x < 0$, $\lim_{n\to\infty} F_n(x) = 0$
- For $0 < x < 1$, $\lim_{n\to\infty} F_n(x) = \frac{1}{2}$
- For $x > 1$, $\lim_{n\to\infty} F_n(x) = 1$
- 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 $x > c$, $F_n(x) \to 1$ for ϵ small enough.
- And for any $x < c$, $F_n(x) \to 0$ for ϵ small enough.
- So X_n converges in distribution to c.

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

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

And X_n converges in distribution 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

Book calls it the "Continuity Theorem"

Let the random variables $X_1, X_2 \ldots$ have cumulative distribution functions $F_1(x), F_2(x) \ldots$ and moment-generating functions $M_1(t), M_2(t) \ldots$. Let the random variable X have cumulative distribution function $F(x)$ and moment-generating function $M(t)$. If

$$\lim_{n \to \infty} M_n(t) = M(t)$$

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.

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 $\left(1+\frac{x}{n}\right)^n \to e^x$,

$$
\begin{aligned}
M_n(t) &= (pe^t + 1 - p)^n \\
&= \left(\frac{\lambda}{n}e^t + 1 - \frac{\lambda}{n}\right)^n \\
&= \left(1 + \frac{\lambda(e^t - 1}{n}\right)^n \\
&\to e^{\lambda(e^t-1)}
\end{aligned}
$$

MGF of Poisson(λ).

The Central Limit Theorem

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

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

converges in distribution to $Z \sim \text{Normal}(0,1)$.

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