Theoretical Statistics, STOR 655

Modes of Convergence, Laws of Large Numbers, Stochastic Order

Andrew Nobel

January 2022

Convergence of Random Vectors

Convergence of Random Vectors: Standard Notions

Setting: Sequence of random vectors $X_1, X_2, \ldots, X \in \mathbb{R}^d$ defined on the same probability space.

1. $X_n \to X$ with probability one (wp1, a.s., or a.e.) if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = \mathbb{P}\left(X_n \to X \text{ as } n \to \infty\right) = 1$$

2. $X_n \to X$ in quadratic mean (L_2) if

$$\mathbb{E}||X_n - X||^2 \to 0 \text{ as } n \to \infty$$

3. $X_n \to X$ in probability if for all $\delta > 0$

$$\mathbb{P}\left(||X_n - X|| > \delta\right) \to 0 \text{ as } n \to \infty$$

Relationships Between Modes of Convergence

Fact: Let $X_1, X_2, \ldots, X \in \mathbb{R}^d$ be random vectors

1. If $X_n \to X$ wp1 then $X_n \to X$ in probability
2. If $X_n \to X$ in L_2 then $X_n \to X$ in probability

In general, other implications do not hold

- convergence in probability does not imply convergence in L_2 or wp1
- convergence in L_2 does not imply convergence wp1

Operations Extending Convergence

1. *Gluing.* If $X_n \to X$ in probability and $Y_n \to Y$ in probability then

$$\begin{bmatrix} X_n \\ Y_n \end{bmatrix} \to \begin{bmatrix} X \\ Y \end{bmatrix} \text{ in probability}$$

2. *Continuous transformation.* If $X_n \to X$ in probability and $g : \mathbb{R}^d \to \mathbb{R}^k$ is continuous, then $g(X_n) \to g(X)$ in probability.

Note:

- Continuous transformation extension also holds if X_n, X take values in an open subset $\mathcal{X}$ of $\mathbb{R}^d$ and $g : \mathcal{X} \to \mathbb{R}^k$ is continuous
- Same properties hold for convergence with probability one

Scheffé's Theorems

Thm: Let $X_1, X_2, \ldots \in \mathbb{R}$ be non-negative r.v. such that as n tends to infinity

- $X_n \to X$ wp1 and
- $\mathbb{E}X_n \to \mathbb{E}X < \infty$

Then $\mathbb{E}|X_n - X| \to 0$

Thm: If $g_1, g_2, \ldots, g$ are probability densities on $\mathbb{R}^d$ such that $g_n(x) \to g(x)$ for Lebesgue almost every x then $\int |g_n(x) - g(x)|\, dx \to 0$.

Law of Large Numbers

Law of Large Numbers

Idea: An average of iid random vectors converges to their common expectation

Thm: Let $X_1, X_2, \ldots, X \in \mathbb{R}^d$ be iid. For $n \geq 1$ define $\overline{X}_n := n^{-1} \sum_{i=1}^n X_i$

1. If $\mathbb{E}||X|| < \infty$ then $\overline{X}_n \to \mathbb{E}X$ wp1

2. If $\mathbb{E}||X|| < \infty$ then $\overline{X}_n \to \mathbb{E}X$ in probability

3. If $\mathbb{E}||X||^2 < \infty$ then $\overline{X}_n \to \mathbb{E}X$ in L_2

Application: Consistency of Sample Variance Matrix

Example: Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be iid with $\mathbb{E}(X_i) = \mu$ and $\mathrm{Var}(X_i) = \Sigma$. Wish to estimate Σ from $X_1, \ldots, X_n$

Definition: The sample variance matrix of $X_1, \ldots, X_n \in \mathbb{R}^d$ is

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

Fact

1. $S_n = n^{-1} \sum_{i=1}^n X_i X_i^t - (\overline{X}_n)(\overline{X}_n)^t$
2. $S_n \to \Sigma$ wp1 as $n \to \infty$

Empirical CDF and the Glivenko-Cantelli Theorem

The Empirical CDF

Setting: Observations $X_1, X_2, \ldots, X \in \mathbb{R}$ iid with CDF $F(x) = \mathbb{P}(X \leq x)$

Definition: The *empirical CDF* of $X_1, \ldots, X_n$ is given by

$$\hat{F}_n(t) \;=\; \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(X_i \leq t) \quad t \in \mathbb{R}$$

Note: For fixed t, $\hat{F}_n(t)$ is the fraction of points $X_1, \ldots, X_n$ that are less than or equal to t. In particular, $n\hat{F}_n(t) \stackrel{\mathsf{d}}{=} \text{Bin}(n, F(t))$ and

$$\mathbb{E}\hat{F}_n(t) = F(t) \;\text{ and }\; \mathrm{Var}(\hat{F}_n(t)) = F(t)(1 - F(t))/n$$

Glivenko-Cantelli Theorem

Question: The LLN ensures that $\hat{F}_n(t) \to F(t)$ wp1 for each fixed $t \in \mathbb{R}$. Is this convergence *uniform* over all t?

Theorem: If $X_1, X_2, ... \in \mathbb{R}$ are iid with $X_i \sim F$ then as $n \to \infty$

$$\sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)| \ \to \ 0 \ \text{ wp1}$$

Note: As $\hat{F}_n, F$ are right continuous, the sup is unchanged if we replace $\mathbb{R}$ by the rationals. As the rationals are countable, the sup is measurable.

Proof idea: First establish the result for continuous F by approximation, then discrete F by a direct argument, then combine these to get the general result.

Application of Glivenko-Cantelli

Moving-Target: In many problems one needs to analyze the behavior of objects where two or more random quantities interact. Uniform laws of large numbers like the G-C theorem can be useful for such analyses.

Example: Let $X_1, X_2, \ldots \in \mathbb{R}$ iid with $X_i \sim F$ continuous. For $n \geq 1$ let

- $\hat{\theta}_n = \hat{\theta}_n(X_1^n) \in \mathbb{R}$ point estimate of parameter θ_0 of interest
- $\hat{F}_n =$ empirical CDF of $X_1, \ldots, X_n$

Question: If $\hat{\theta}_n \to \theta_0$, does $\hat{F}_n(\hat{\theta}_n) \to F(\theta_0)$?

Estimating of Percentiles

Recall: The *percentile function* of a CDF F is defined for $p \in (0, 1)$ by

$$F^{-1}(p) := \inf\{u : F(u) \geq p\}$$

Fact (Basic properties of percentile function)

1. F^{-1} is the usual inverse if F is continuous and strictly increasing
2. F^{-1} is non-decreasing
3. $F^{-1}(p) \leq u$ if and only if $F(u) \geq p$
4. If $U \sim \text{Uniform}(0, 1)$ then $X = F^{-1}(U)$ has CDF F
5. If $X \sim F$ and F is continuous then $F(X) \sim \text{Uniform}(0, 1)$

Consistent Estimation of Percentiles

Percentile Inference problem

- Observe $X_1, X_2, \ldots, X_n \in \mathbb{R}$ iid with unknown CDF F
- Goal: Estimate percentile $F^{-1}(p)$ for fixed $p \in (0,1)$ of interest
- Candidate estimate: percentile $\hat{F}_n^{-1}(p)$ of empirical CDF

Fact: If F is continuous and increasing in a neighborhood of $F^{-1}(p)$ then $\hat{F}_n^{-1}(p) \to F^{-1}(p)$ wp1 as $n \to \infty$.

Standard and Stochastic Order Relations

Standard Order Relations

Given: Numerical sequences $\{a_n : n \geq 1\}$ and $\{b_n : n \geq 1\}$

Definition

- $a_n = O(b_n)$ if there exists $C < \infty$ such that $|a_n| \leq C|b_n|$ for all $n \geq 1$
- $a_n = o(b_n)$ if $|a_n|/|b_n| \to 0$ as $n \to \infty$
- $a_n = \Omega(b_n)$ if there exists $C > 0$ such that $|a_n| \geq C\,|b_n|$ for all $n \geq 1$
- $a_n = \Theta(b_n)$ if $a_n = O(b_n)$ and $a_n = \Omega(b_n)$

Absolute Order Relations

Given: Numerical sequence $\{a_n : n \geq 1\}$

Absolute Order: Special case where $b_n \equiv 1$

- $a_n = O(1)$ if there exists $C < \infty$ such that $|a_n| \leq C$ for all $n \geq 1$
- $a_n = o(1)$ if $|a_n| \to 0$ as $n \to \infty$
- $a_n = \Omega(1)$ if there exists $C > 0$ such that $|a_n| \geq C$ for all $n \geq 1$
- $a_n = \Theta(1)$ if there exist $C > 0$ such that $C^{-1} \leq |a_n| \leq C$ for all $n \geq 1$

Standard Order Relations: Formal Interpretation

Definition: $O(b_n)$ is the family of all sequences $\{a_n\}$ for which there exists constant $C < \infty$, depending on $\{a_n\}$, such that $|a_n| \leq C|b_n|$ for all $n \geq 1$

Families $o(b_n)$, $\Omega(b_n)$, and $\Theta(b_n)$ are defined in a similar fashion

Interp'n. Equations involving order notation are regarded as inclusions: family of sequences on left of equality is contained in family on right

- $1^2 + 2^2 + \cdots + n^2 = O(n^3)$
- $3n + O(n^2) = O(n^2)$
- $O(n) + o(n^2) = o(n^2)$
- $O(n^2) = O(n^3)$
- $O(a_n)O(b_n) = a_n O(b_n)$

Stochastic Order Relations

Definition: Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be random vectors defined on the same probability space

1. $X_n = O_p(1)$ if for every $\epsilon > 0$ there exists $M = M(\epsilon) < \infty$ such that

$$\mathbb{P}(||X_n|| \geq M) \ \leq \ \epsilon \ \text{ for every } n \geq 1$$

   In this case $\{X_n\}$ is said to be *stochastically bounded* or *tight*

2. $X_n = o_p(1)$ if $||X_n|| \to 0$ in probability

Stochastic Order Relations: Examples

1. If X is any random vector, the constant sequence $X_n = X$ is tight
2. If $X_n \to X$ in probability, then $\{X_n\}$ is tight
3. The sequences $X_n \sim \mathcal{N}(0, n)$ and $Y_n \sim \mathcal{N}(n, 1)$ are not tight

Idea: A sequence is tight if the probability mass of the vectors X_n does not “escape” to plus or minus infinity

Stochastic Order Relations

Formal interpretation of $O_p(1)$ and $o_p(1)$ is similar to standard order symbols

- $O_p(1)$ is the family of all stochastically bounded sequences of random vectors

- $o_p(1)$ is the family of all sequences of random vectors whose norms converge to zero in probability,

Note: Random vectors may be defined on different probability spaces

Equations involving stochastic order notation understood as inclusions: the set of sequences on the right of the equality is contained in the set on the left

Stochastic Order Relations: Basic Properties

Fact: Under the set-inclusion interpretation, the following relations hold

1. $o_p(1) = O_p(1)$
2. $O_p(1) + O_p(1) = O_p(1)$
3. $O_p(1) + o_p(1) = O_p(1)$
4. $o_p(1) + o_p(1) = o_p(1)$
5. $O_p(1) O_p(1) = O_p(1)$
6. $O_p(1) o_p(1) = o_p(1)$

Relative Stochastic Order Relations

Definition: Let $X_1, X_2, \ldots \in \mathbb{R}^d$ and $Y_1, Y_2, \ldots \in \mathbb{R}^d$ be random vectors defined on the same probability space

1. $X_n = O_P(Y_n)$ if for every $\epsilon > 0$ there exists $M < \infty$ such that $\mathbb{P}(||X_n|| \geq M||Y_n||) \leq \epsilon$ when n is sufficiently large

2. $X_n = o_P(Y_n)$ if the ratio $||X_n||/||Y_n||$ converges to zero in probability

In principal, X_i and Y_i could be of different dimensions