Weak Convergence in Metric Spaces

Shiyuan Fan | 25. May 2010

- In measure theory, there are various notions of the convergence of measures. Broadly speaking, there are two kinds of convergence, strong convergence and weak convergence.

- **strong convergence**: If the collection of all measures on a measurable space can be given some kind of metric, then convergence in this metric is usually referred to as strong convergence.

- **weak convergence**:

S metric space $\wp$ class of Borel sets in S P probability measure on $\wp$

If $\int_S f\,dP_n \to \int_S f\,dP$ for every bounded, continuous real function f on S, we write $P_n \Rightarrow P$

Measures in Metric Spaces

Properties of Weak Convergence

Some Special Cases

Convergence in Distribution

Weak Convergence and Mappings

Applications

- The Central Limit Theorem
- DeMoivre-Laplace Theorem
- Proposition of Slutsky

Measures and Integrals

- Every probability measure on $(S, \mathcal{S})$ is regular; that is,
  if $A \in \mathcal{S}$ and $\varepsilon > 0$, then there exist a closed set F and an open set G such that $F \subset A \subset G$ and $P(G - F) < \varepsilon$.

- Probability measures P and Q on $(S, \mathcal{S})$ coincide if
  $$\int f\,dP = \int f\,dQ \qquad \text{for each } f \text{ in } C(S).$$

- If F is closed and ε positive, there is a function f in $C(S)$ s.t.
  $f(x) = 1$ if $x \in F$, $f(x) = 0$ if $\rho(x, F) \geq \varepsilon$, and $0 \leq f(x) \leq 1$ for all x.
  The function f may be taken to be uniformly continuous.

Measures in Metric Spaces

Properties of Weak Convergence

Some Special Cases

Convergence in Distribution

Weak Convergence and Mappings

Applications

- The Central Limit Theorem
- DeMoivre-Laplace Theorem
- Proposition of Slutsky

Portmanteau Theorem

Let P_n, P be probability measures on $(S, \mathcal{S})$. These five conditions are equivalent:

(i) $P_n \Rightarrow P$.

(ii) $\lim_n \int f dP_n = \int f dP$ for all bounded, uniformly continuous real f.

(iii) $\limsup_n P_n(F) \leq P(F)$ for all closed F.

(iv) $\liminf_n P_n(G) \geq P(G)$ for all open G.

(v) $\lim_n P_n(A) = P(A)$ for all P-continuity sets A.

Other Criteria

- Let $\mathcal{U}$ be a class of sets s.t.

  (i) $\mathcal{U}$ is closed under the formation of finite intersections;

  (ii) each open set in S is a finite or countable union of elements of $\mathcal{U}$

  If $P_n(A) \to P(A)$ for every A in $\mathcal{U}$, then $P_n \Rightarrow P$.

- Corollary:

  Let $\mathcal{U}$ be a class of sets s.t.

  (i) $\mathcal{U}$ is closed under the formation of finite intersections;

  (ii) for every x in S and every positive ε there is an A in $\mathcal{U}$ with

  $$x \in \text{int}(A) \subset A \subset S(x, \varepsilon).$$

  If S is separable and if $P_n(A) \to P(A)$ for every A in $\mathcal{U}$, then $P_n \Rightarrow P$.

Other Criteria

- Corollary:

  Suppose that, for each finite intersection A of open spheres, we have $P_n(A) \to P(A)$, provided A is a P-continuity set. If S is separable, then $P_n \Rightarrow P.$

- Another condition for weak convergence:

  We have $P_n \Rightarrow P$ if and only if each subsequence $\{P_{n'}\}$ contains a further subsequence $\{P_{n''}\}$ s.t. $P_{n''} \Rightarrow P.$

Measures in Metric Spaces

Properties of Weak Convergence

Some Special Cases

Convergence in Distribution

Weak Convergence and Mappings

Applications

- The Central Limit Theorem
- DeMoivre-Laplace Theorem
- Proposition of Slutsky

Euclidean Space

- We can relate weak convergence $P_n \Rightarrow P$ to the usual notion of convergence for the corresponding distribution functions F_n , F.
- R^K k-dimensional Euclidean space

  $\rho(x, y)$ ordinary metric which equals $|x-y| = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$

  $\mathcal{R}^K$ the class of Borel sets

- The general probability measure P on $(R^K , \mathcal{R}^K)$ has a distribution function F:

$$F(x) = P\{y : y \le x\}, x \in R^K .$$

- For distribution functions F_n and F, define $F_n \Rightarrow F$ to mean that

  $F_n(x) \to F(x)$ at continuity points x of F.

- We can prove that if $P_n \Rightarrow P$, then $F_n \Rightarrow F$

The Circle

- S is the unit circle in the complex plane.
- $P_n \Rightarrow P$ if and only if $P_n(A) \to P(A)$ for every arc A whose endpoints have *P*-measure 0.

- Measures in Metric Spaces
- Properties of Weak Convergence
- Some Special Cases
- **Convergence in Distribution**
- Weak Convergence and Mappings
- Applications
  - The Central Limit Theorem
  - DeMoivre-Laplace Theorem
  - Proposition of Slutsky

Random Elements

- X is a mapping from a probability space $(\Omega, \mathcal{B}, \mathsf{P})$ into a metric space S.

  If X is measurable, we call it a random element.

- The distribution of X is the probability measure P on $(S, \mathcal{S})$:

$$P(A) = \mathsf{P}(X^{-1}A) = \mathsf{P}(\omega : X(\omega) \in A) = \mathsf{P}(X \in A)$$

- Note that P is a probability measure on a space of an arbitrary nature, whereas P is always defined on a metric space. For many questions, the distribution P contains all relevant information about the random element X.

- If h is a measurable function on S, then by change-of-variable formula

$$\int h(X)d\mathsf{P} = \int h dP$$

  In the usual expected-value notation,

$$\int h dP = \mathrm{E}[h(X)]$$

Convergence in Distribution

- A sequence $\{X_n\}$ of random elements converges in distribution to the random element X: $X_n \xrightarrow{\mathcal{D}} X$, if the distributions P_n of the X_n converge weakly to the distribution P of X: $P_n \Rightarrow P$.
- The underlying probability spaces (the domains) may be all distinct.
- Each theorem about weak convergence can be similarly recast.

Convergence in Distribution

(i) $P_n \Rightarrow P$.

(ii) $\lim_n \int f dP_n = \int f dP$ for all bounded, uniformly continuous real f.

(iii) $\limsup_n P_n(F) \leq P(F)$ for all closed F.

(iv) $\liminf_n P_n(G) \geq P(G)$ for all open G.

(v) $\lim_n P_n(A) = P(A)$ for all P-continuity sets A.

(i) $X_n \xrightarrow{\mathcal{D}} X$

(ii) $\lim_n \mathrm{E}[f(X_n)] = \mathrm{E}[f(X)]$ for all bounded, uniformly continuous real f.

(iii) $\limsup_n \mathsf{P}(X_n \in F) \leq \mathsf{P}(X \in F)$ for all closed F.

(iv) $\liminf_n \mathsf{P}(G_n \in F) \geq \mathsf{P}(X \in G)$ for all open G.

(v) $\lim_n \mathsf{P}(X_n \in A) = P(X \in A)$ for all X-continuity sets A.

Convergence in Distribution

- Hybrid terminology:

  We say the X_n converge in distribution to P, and write

$$X_n \xrightarrow{\mathcal{D}} P,$$

  in case $P_n \Rightarrow P$.

- It is great convenience to be able to pass from one to another of three equivalent concepts. This is largely a matter of expedient phraseology.

- Example:

$$X_n \xrightarrow{\mathcal{D}} N(\mu,\sigma^2).$$

Convergence in Probability

- If, for an element a of S,

$$\mathsf{P}\{\rho(X_n, a) \geq \varepsilon\} \to 0$$

for each positive ε, we say X_n converges in probability to a and write

$$X_n \xrightarrow{\mathcal{P}} a.$$

- If a is conceived as a constant-valued random element, then

$X_n \xrightarrow{\mathcal{P}} a$ if and only if $X_n \xrightarrow{\mathcal{D}} a$.

- Alternatively, $X_n \xrightarrow{\mathcal{P}} a$ if and only if the distribution of X_n converges weakly to the probability measure corresponding to a mass of 1 at the point a.

Convergence in Probability

- If, for an element *a* of *S*,

$$\mathsf{P}\{\rho(X_n, a) \geq \varepsilon\} \rightarrow 0$$

for each positive ε, we say X_n converges in probability to *a* and write

$$X_n \xrightarrow{\mathcal{P}} a.$$

- If *a* is conceived as a constant-valued random element, then

$$X_n \xrightarrow{\mathcal{P}} a \text{ if and only if } X_n \xrightarrow{\mathcal{D}} a \quad .$$

- Alternatively, $X_n \xrightarrow{\mathcal{P}} a$ if and only if the distribution of X_n

converges weakly to the probability measure corresponding to a mass of 1 at the point a.

Measures in Metric Spaces

Properties of Weak Convergence

Some Special Cases

Convergence in Distribution

Weak Convergence and Mappings

Applications

- The Central Limit Theorem
- DeMoivre-Laplace Theorem
- Proposition of Slutsky

Continuous Mappings

- h is a measurable mapping $: S \to S'$, then each probability measure P on $(S,\mathcal{S})$ induces on $(S',\mathcal{S}')$ a unique probability measure Ph^{-1}

- $Ph^{-1}(A) = P(h^{-1}A), A \in \mathcal{S}'$

- If h is a continuous mapping, $P_n \Rightarrow P$ implies $P_n h^{-1} \Rightarrow Ph^{-1}$.

Main Theorem

- We can weaken the continuity assumption of *h*.
- Assume only *h* is measurable and let D_h be the set of discontinuities of *h*. Then we can prove $D_h \in \wp$.
- If $P_n \Rightarrow P$ and $P(D_h) = 0$,then $P_n h^{-1} \Rightarrow Ph^{-1}$.

Measures in Metric Spaces

Properties of Weak Convergence

Some Special Cases

Convergence in Distribution

Weak Convergence and Mappings

Applications

- **The Central Limit Theorem**
- DeMoivre-Laplace Theorem
- Proposition of Slutsky

- $\xi_{n1},\dots,\xi_{nk_n}$ are independent random variables with mean 0 and finite variance σ_{nk}^2. The probability space on which the variables are defined may vary with n.
- $S_n = \xi_{n1} + \dots + \xi_{nk_n}$ and suppose its variance $s_n^2 = \sigma_{n1}^2 + \dots + \sigma_{nk_n}^2$ is positive.
- N is a random variable normally distributed with mean 0 and variance 1.
- Lindeberg's theorem:

If

$$\frac{1}{s_n^2}\sum_{k=1}^{k_n}\int_{\{|\xi_{nk}|\geq \varepsilon s_n\}} \xi_{nk}{}^2 dP \rightarrow 0 (n \rightarrow \infty)$$

for each positive ε, then

$$S_n / s_n \xrightarrow{\mathcal{D}} N.$$

- From the last theorem we can deduce the lindeberg-Levy theorem:

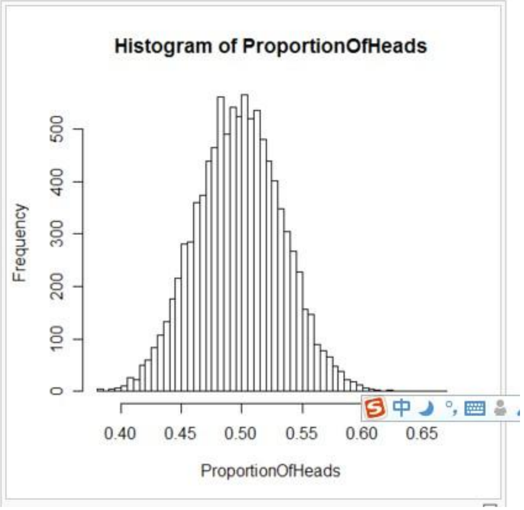
If $\xi_1, \xi_2, \cdots$ are independent and identical distributed with mean 0 and finite variance $\sigma^2 > 0$, then

$$\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}\xi_k \xrightarrow{\mathcal{D}} N.$$

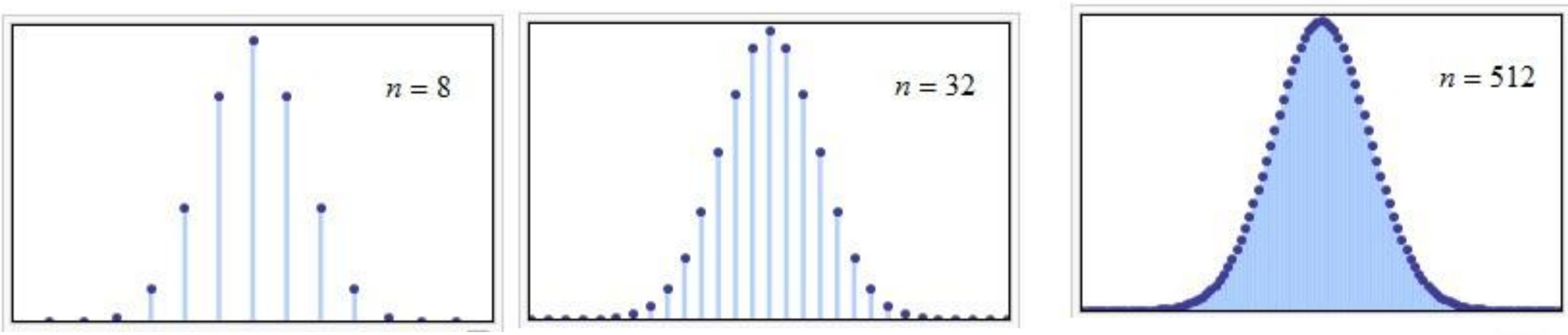
- To prove this result, take $k_n = n, \xi_{ni} = \xi_i$; the sum in former equation is at most ξ_1^2/σ^2 integrated over $\{|\xi_1| \geq \varepsilon\sigma\sqrt{n}\}$.

- A **central limit theorem** is any of a set of weak-convergence theories. They all express the fact that a sum of many independent random variables will tend to be distributed according to one of a small set of "attractor" (i.e. stable) distributions.

  - the central limit theorem states conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed



- The **de Moivre–Laplace theorem** is a normal approximation to the binomial distribution. It is a special case of the central limit theorem. It states that the binomial distribution of the number of "successes" in *n* independent trials with probability *p* of success on each trial is approximately a normal distribution with mean *np* and standard deviation , if *n* is very large and some conditions are satisfied.



Measures in Metric Spaces

Properties of Weak Convergence

Some Special Cases

Convergence in Distribution

Weak Convergence and Mappings

Applications

- The Central Limit Theorem
- DeMoivre-Laplace Theorem
- **Proposition of Slutsky**

Given $X_n \xrightarrow{\mathcal{D}} X, Y_n \xrightarrow{\mathcal{P}} c$, where c is a constant, then

$$X_n + Y_n \xrightarrow{\mathcal{D}} X + c;$$

$$X_n Y_n \xrightarrow{\mathcal{D}} cX;$$

$$Y_n^{-1} X_n \xrightarrow{\mathcal{D}} c^{-1} X.$$

The theorem remains valid if we replace all convergences in distribution with convergences in probability

Thank You!