Convergence theorems

EM13-Probability and statistics: Courses 07-08

September 2014

Manuel SAMUELIDES[1] Zhigang SU[2]

[1]Professor
Institut Supereur de l'Aeronautique et de l'Espace

[2]Professor
Sino-European Institute of Aviation Engineering
Civil Aviation University of China

Homework–20140923

1. Large deviation Cramer’ s theorem and exponential convergence

(1) It is easy to check from a power series development that $\cos\alpha \leqslant \exp\frac{\alpha^2}{2}$. Use this inequality to show that if X is a uniform binary variable, i.e. $Pr(X=1)=Pr(X=-1)$ $=0.5$, then

$$\mathbf{E}\{\exp(\alpha X)\} \leqslant \exp\frac{\alpha^2}{2} \qquad \forall\alpha > 0$$

(2) Let $(X_1,\cdots,X_n)$ be a size n iid sample of the uniform binary law. We set $S_n = X_1+\cdots+X_n$ and let $\overline{X}_n = \frac{S_n}{n}$ be the empirical mean of the iid sample. Show that

$$\begin{aligned} Pr(S_n>\lambda) &\leqslant \exp(-\alpha\lambda)\mathbf{E}\{\exp(\alpha S_n)\} \leqslant \exp(\frac{n\alpha^2}{2}-\alpha\lambda) \\ &\qquad\qquad \forall\alpha>0,\forall\lambda>0 \\ Pr(S_n>\lambda) &\leqslant \exp(\frac{\lambda^2}{2n}) \qquad\qquad \forall\lambda \end{aligned}$$

(3) Use the previous question to show that

$$Pr(\overline{X}_n > \lambda) \leqslant \exp(\frac{n\lambda^2}{2})$$

2. Biased estimator of variance

(1) Compute the estimator of expectation $\hat{m}$ and variance $\hat{v}$ of an iid sample $(X_1, \cdots, X_n)$ by the method of moments and show

$$\begin{cases} \hat{m} = \overline{X} \\ \hat{v} = \overline{X^2} - \overline{X}^2 \end{cases}$$

where $\overline{X} = \frac{1}{n}\sum_k X_k$ and $\overline{X^2} = \frac{1}{n}\sum_k X_k^2$.

(2) Show that the estimator $\hat{v}$ is biased and compute its bias.

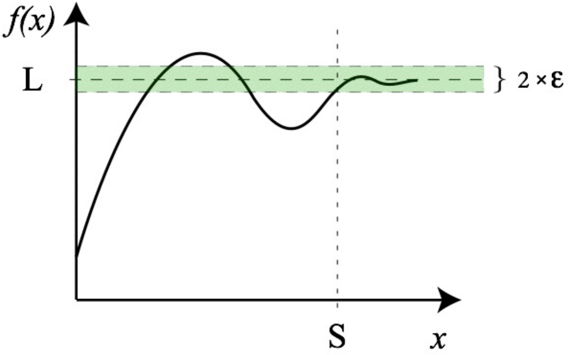
(3) Propose an unbiased estimator.

(4) Show that $\hat{v}$ is equal to $\frac{1}{n}\sum_{k=1}^{n}[x_k - \hat{m}]^2$.

常序列的收敛

Definition

对于一个普通序列$\{x_n\}$，x是常数，若对任意正数$\varepsilon > 0$恒存在自然数N，使$n > N$时，有$|x_n - x| < \varepsilon$，称n趋于无穷时，x_n以x为极限，或x_n收敛于x。



Sure convergence/处处收敛

Definition

To say that the sequence of random variables (X_n) defined over the same probability space (i.e., a random process) converges **surely**(确定) or **everywhere**(处处) or **pointwise**(逐点) towards X means

$$\lim_{n\to\infty} X_n(e) = X(e) \qquad e \in \Omega$$

where Ω is the sample space of the underlying probability space over which the random variables are defined.

Almost sure convergence/几乎处处收敛/以概率1收敛

Manuel SAMUELIDES, Zhigang SU

中国民航大学

Definition

To say that the sequence X_n converges almost surely or almost everywhere or with probability 1 or strongly towards X means that

$$Pr\{\lim_{n\to\infty} X_n(e) = X(e)\} = 1$$

This means that the values of X_n approach the value of X, in the sense (see almost surely) that events for which X_n does not converge to X have probability 0.

Remark

Almost sure convergence is often denoted by adding the letters a.s. over an arrow indicating convergence:

$$X_n \xrightarrow{a.s.} X$$

Convergence in mean/依均值收敛

Definition

Given a real number $r \geqslant 1$, we say that the sequence X_n converges in the r-th mean (or in the L^r-norm) towards the random variable X, if the r-th absolute moments $E(|Xn|^r)$ and $E(|X|^r)$ of X_n and X exist, and

$$\lim_{n\to\infty} \mathrm{E}\left(|X_n - X|^r\right) = 0$$

where the operator E denotes the expected value.

Convergence in r-th mean tells us that the expectation of the r-th power of the difference between Xn and X converges to zero.

Remark

Convergence in mean is often denoted by adding the letters L^r over an arrow indicating convergence:

$$X_n \xrightarrow{L^r} X$$

Manuel SAMUELIDES, Zhigang SU

中国民航大学

The most important cases of convergence in r-th mean are:

- When X_n converges in r-th mean to X for $r = 1$, we say that X_n **converges in mean**(依均值收敛) to X.
- When X_n converges in r-th mean to X for $r = 2$, we say that X_n **converges in mean square**(均方收敛) to X.

Convergence in probability

Definition

A sequence $\{X_n\}$ of random variables converges in probability towards the random variable X if for all $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\left(|X_n - X| \geqslant \varepsilon\right) = 0$$

The basic idea behind this type of convergence is that the probability of an “unusual” outcome becomes smaller and smaller as the sequence progresses.

Remark

Convergence in probability is often denoted by adding the letters p over an arrow indicating convergence:

$$X_n \xrightarrow{p} X$$

Convergence in distribution

Definition

A sequence $X_1, X_2, \cdots$ of random variables is said to converge in distribution, or converge weakly, or converge in law to a random variable X if

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

for every number $x \in R$ at which F is continuous. Here F_n and F are the cumulative distribution functions of random variables X_n and X, respectively.

Remark

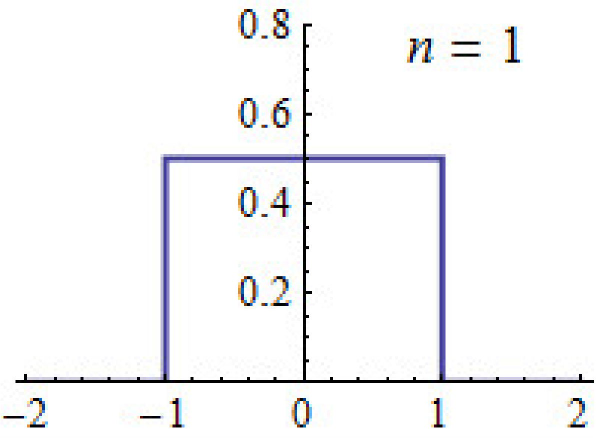
Convergence in distribution is often denoted by adding the letters d over an arrow indicating convergence:

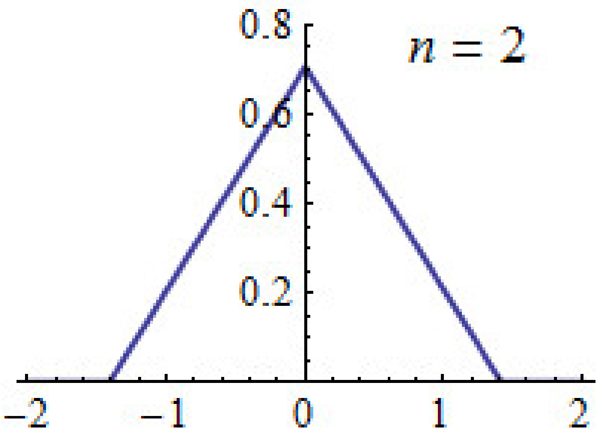
$$X_n \xrightarrow{d} X$$

Graphic example

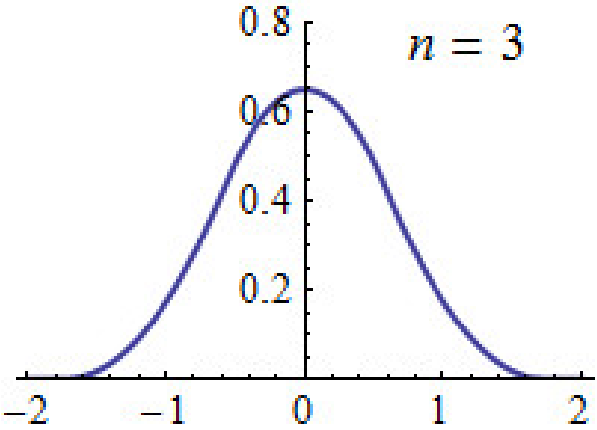
Suppose $\{X_i\}$ is an iid sequence of uniform $U(-1,1)$ random variables. Let $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ be their (normalized) sums. Then according to the central limit theorem, the distribution of Z_n approaches the normal $N(0, \frac{1}{3})$ distribution.

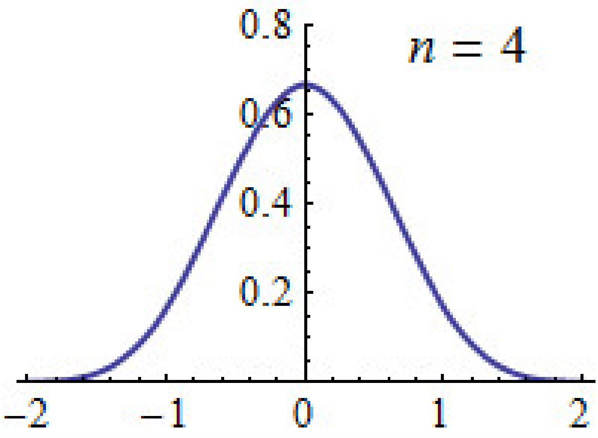
This convergence is shown in the picture: as n grows larger, the shape of the pdf function gets closer and closer to the Gaussian curve.

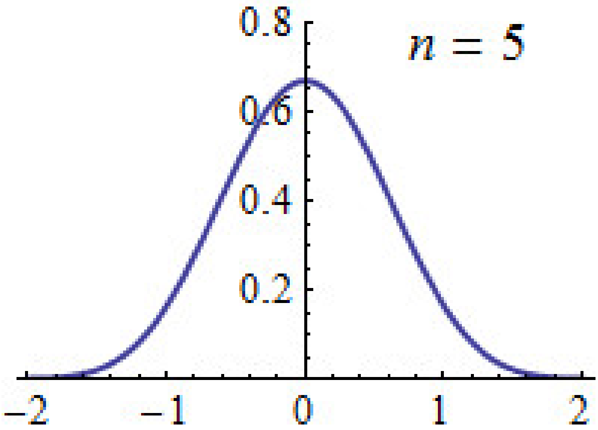

0.8
0.6
0.4
0.2
n = 1
−2
−1
0
1
2

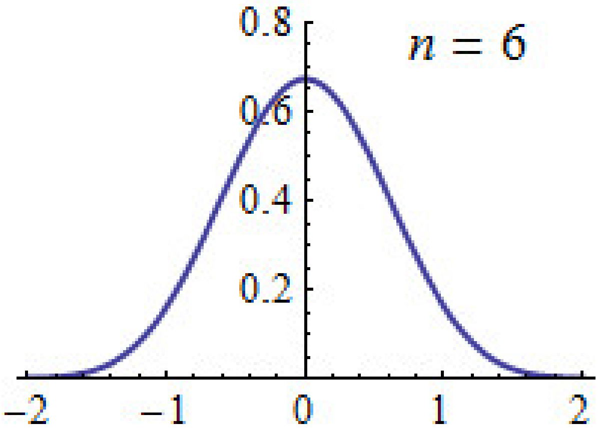


0.8
0.6
0.4
0.2
$n = 2$
−2
−1
0
1
2

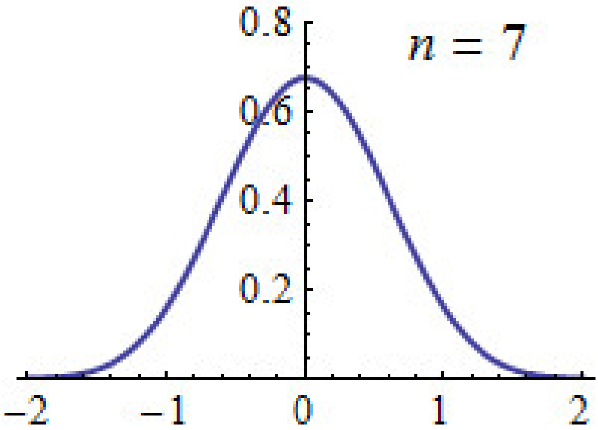

0.8
0.6
0.4
0.2
$n = 3$
−2
−1
0
1
2

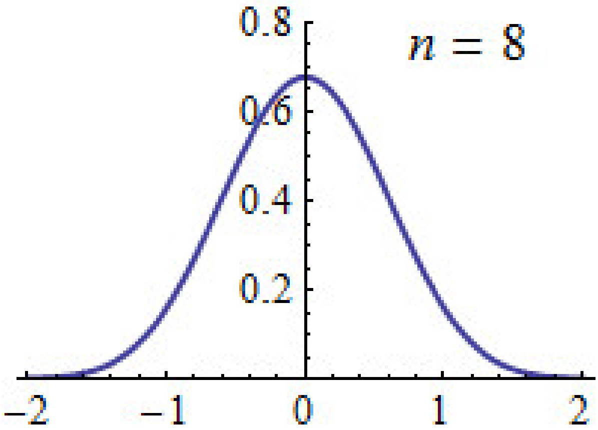

0.8
0.6
0.4
0.2
$n = 4$
−2
−1
0
1
2

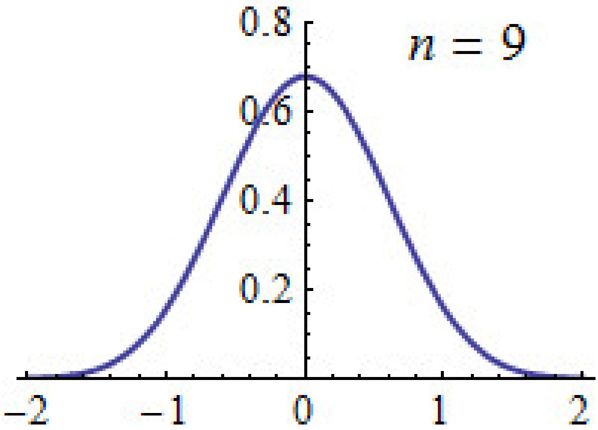

0.8
0.6
0.4
0.2
$n = 5$
−2
−1
0
1
2



0.8
0.6
0.4
0.2
$n = 6$
−2
−1
0
1
2


0.8
0.6
0.4
0.2
$n = 7$
−2
−1
0
1
2


0.8
0.6
0.4
0.2
$n = 8$
−2
−1
0
1
2



0.8
0.6
0.4
0.2
$n = 9$
−2
−1
0
1
2

Properties/性质

1. If $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$, then $X = Y$ almost surely.
2. If $X_n \xrightarrow{a.s.} X$ and $X_n \xrightarrow{a.s.} Y$, then $X = Y$ almost surely.
3. If $X_n \xrightarrow{L^r} X$ and $X_n \xrightarrow{L^r} Y$, then $X = Y$ almost surely.
4. If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$, then $aX_n + bY_n \xrightarrow{p} aX + bY$ (for any real numbers a and b) and $X_nY_n \xrightarrow{p} XY$.
5. If $X_n \xrightarrow{a.s.} X$ and $Y_n \xrightarrow{a.s.} Y$, then $aX_n + bY_n \xrightarrow{a.s.} aX + bY$ (for any real numbers a and b) and $X_nY_n \xrightarrow{a.s.} XY$.
6. If $X_n \xrightarrow{L^r} X$ and $Y_n \xrightarrow{L^r} Y$, then $aX_n + bY_n \xrightarrow{L^r} aX + bY$ (for any real numbers a and b).

The chain of implications between the various notions of convergence are noted in their respective sections. They are, using the arrow notation:

$$\begin{array}{ccccc} \xrightarrow{L^s} & \underset{s>r\geq 1}{\Rightarrow} & \xrightarrow{L^r} & & \\ & & \Downarrow & & \\ \xrightarrow{a.s.} & \Rightarrow & \xrightarrow{p} & \Rightarrow & \xrightarrow{d} \end{array}$$

- Almost sure convergence implies convergence in probability:

$$X_n \xrightarrow{as} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$$

- Convergence in probability implies there exists a sub-sequence (k_n) which almost surely converges:

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad X_{k_n} \xrightarrow{as} X$$

- Convergence in probability implies convergence in distribution:

$$X_n \xrightarrow{p} X \quad \Rightarrow \quad X_n \xrightarrow{d} X$$

- Convergence in r-th order mean implies convergence in probability:

$$X_n \xrightarrow{L^r} X \quad \Rightarrow \quad X_n \xrightarrow{p} X$$

- Convergence in r-th order mean implies convergence in lower order mean, assuming that both orders are greater than or equal to one:

$$X_n \xrightarrow{L^r} X \quad \Rightarrow \quad X_n \xrightarrow{L^s} X, provided\ r \geqslant s \geqslant 1.$$

- If X_n converges in distribution to a constant c, then X_n converges in probability to c:

$$X_n \xrightarrow{d} c \quad \Rightarrow \quad X_n \xrightarrow{p} c, \qquad c \text{ is a constant.}$$

- If X_n converges in distribution to X and the difference between X_n and Y_n converges in probability to zero, then Y_n also converges in distribution to X:

$$X_n \xrightarrow{d} X, \ |X_n - Y_n| \xrightarrow{p} 0 \quad \Rightarrow \quad Y_n \xrightarrow{d} X$$

- If X_n converges in distribution to X and Y_n converges in distribution to a constant c, then the joint vector (X_n, Y_n) converges in distribution to (X, c):

$$X_n \xrightarrow{d} X, \ Y_n \xrightarrow{d} c \quad \Rightarrow \quad (X_n, Y_n) \xrightarrow{d} (X, c), \qquad c \text{ is a constant.}$$

- If X_n converges in probability to X and Y_n converges in probability to Y, then the joint vector (X_n, Y_n) converges in probability to (X, Y):

$$X_n \xrightarrow{p} X, \ Y_n \xrightarrow{p} Y \quad \Rightarrow \quad (X_n, Y_n) \xrightarrow{p} (X, Y)$$

Markov inequality

Proposition

Let X an integrable random variable we have

$$P(|X| > \epsilon) \leq \frac{1}{\epsilon} E(|X|)$$

Proof

Comes from the splitting of the integral into two parts:

$$\begin{aligned} E(|X|) &= \int_{\{\omega||X(\omega)|\leq\epsilon\}} X(\omega)dP(\omega) + \int_{\{\omega||X(\omega)|>\epsilon\}} X(\omega)dP(\omega) \\ &\geq \int_{\{\omega||X(\omega)|>\epsilon\}} X(\omega)dP(\omega) \\ &\geq \int_{\{\omega||X(\omega)|>\epsilon\}} \epsilon dP(\omega) \\ &= P(|X| > \epsilon) \end{aligned}$$

Bienayme-Tchebychev inequality

Corollary

Bienayme-Tchebychev inequality: Let X a second order random variable we have

$$P(|X - E(X)| > \epsilon) \geq \frac{1}{\epsilon^2} Var(X)$$

Law of large numbers

Theorem

Let $\{X_n\}$ be an iid (independent identically distributed) sequence of random variables, then the empirical mean $M_n = \frac{X_1+\cdots+X_n}{n} \to E$ in the following senses:

- Mean-square convergence,
- Convergence in probability,
- Weak convergence and convergence of characteristic functions.

Proof We have

$$\begin{aligned} E\{(M_n - E(X))^2\} &= \frac{1}{n^2} E[(X_1 - E(X))^2 + \cdots \\ &\quad +(X_n - E(X))^2] \\ &\to 0 \end{aligned}$$

Central limit theorem

Theorem

Let $\{X_n\}$ be an iid (independent identically distributed) sequence of random variables, then $U_n = \frac{(X_1 - E(X)) + \cdots + (X_n - E(X))}{\sqrt{n}}$ converges in law towards a centered Gaussian law with variance $Var(X)$

Proof Let ϕ_X the characteristic function of the common law of the (X_n). Let ϕ_{U_n} the characteristic function of U_n. We have

$$\begin{aligned}\phi_{U_n}(t) &= \int \cdots \int \exp\{j \frac{t}{\sqrt{n}}[(x_1 - E(X)) + \cdots \\ &\quad +(x_n - E(X))]\} dP(x_1) \cdots dP(x_n)\end{aligned}$$

Let $\phi_X(t) = \int \exp\{jtx\} dP(x)$, so
$\phi_{U_n}(t) = \exp\{-jtE(X)\sqrt{n}\} \left\{\phi_X\left(\frac{t}{\sqrt{n}}\right)\right\}^n$

$$\log[\phi_{U_n}(t)] = -jtE(X)\sqrt{n} + n \log[\phi_X\left(\frac{t}{\sqrt{n}}\right)]$$

End of the proof of the central limit theorem

We have the following second order Taylor development of ϕ_X

$$\phi_x(t) = 1 + jtE(X) - \frac{t^2}{2}(Var(X) + E(X)^2) + o(t^2)$$

$$log[\phi_X(t)] = jtE(X) - \frac{t^2}{2}[Var(X) + E(X)^2] - \frac{t^2}{2}E(X)^2 + o(t^2)$$

$$log[\phi_X(t)] = jtE(X) - \frac{t^2}{2}[Var(X) + E(X)^2] - \frac{t^2}{2}E(X)^2 + o(t^2)$$

$$log[\phi_{U_n}(t)] = -jt\sqrt{n}E(X) + n\frac{jt}{\sqrt{n}}E(X) - n\frac{t^2}{2n}Var(X) + o(1)$$

$$\phi_{U_n}(t) \to exp\left(\frac{-t^2}{2}Var(X)\right)$$