Topic 11: Limit Theorems

Outline

- Markov Inequality
- Chebyshev Inequality
- Weak Law of Large Number (LLN) 大數法則
- The Central Limit Theorem (CLT) 中央極限定理
- Convergence of a Sequence of RVs 隨機序列的收斂性

Reading:

Textbook Section 5.1~5.5

Introduction

- Why study limit theorems?
  - Study the *asymptotic behavior* of sequences of RVs. We'd like to know whether a **sequence of RVs *converges***
    (如鐵穹防禦系統要能鎖定並持續追蹤直到擊中目標)
  - **大數法則(Law of Large Numbers)** 陳述sample mean $\frac{1}{n}\sum_{i=1}^{n} X_i$在樣本數很大時會***收斂***到真實的期望值。
    應用：可估計群體中某現象存在個數（某事件發生次數）。在群體樣本非常大時，此估計愈準確。
  - 我們經常對一連串i.i.d.隨機變數之總和感興趣，如計算個數、多次賽局輸贏總合、長隊伍等待時間等。**中央極限定理(Central Limit Theorem)**宣告我們可借用**常態分佈**來***近似***此總和的PDF

Review of Sample Mean

- Law of large numbers and Central Limit Theorem are related to the sample mean. Let's have a quick review.

- $X_1, X_2, \ldots X_n$ are ***i.i.d.*** RVs with $E[X_i]=\mu$, $\mathrm{var}(X_i)=\sigma^2$

  Sample mean: $M_n = (X_1+X_2+\ldots+X_n)/n$

  Then,

$$E[M_n] = \mu; \quad \mathrm{var}(M_n) = \frac{\sigma^2}{n}$$

Markov Inequality

If a random variable X can only take nonnegative values, then we must have the following Markov inequality

$$P(X \geq a) \leq \frac{E[X]}{a} \quad \text{for all } a > 0.$$

Remark: 僅需知道期望值即可!

Markov inequality provides a loose bound, but is *easy to calculate*!

Example 5.1: Uniform in [0,4] (p.266)

$$\text{Markov: } \Pr(X \geq 3) \leq \tfrac{2}{3} = 0.67; \quad \text{Exact: } \Pr(X \geq 3) = 0.25$$

Chebyshev Inequality

If X is a random variable with mean μ and variance σ^2, then

$$P(|X-\mu| \geq c) \leq \frac{E[(X-\mu)^2]}{c^2} \quad \text{for all } c > 0.$$

Remarks:

- *Markov* and *Chebyshev* inequalities allow us to derive bounds on probabilities when only the mean, or both the mean and the variance are available

- In general, *Markov and Chebyshev inequality* are NOT very tight upper bounds, as the required information is only the mean and the variance. (Examples 5.1 and 5.2)

Chebyshev Inequality

$$P(|X-\mu| \geq c) \leq \frac{E[(X-\mu)^2]}{c^2} \quad \text{for all } c > 0.$$

In general, *Markov and Chebyshev inequality* are NOT very tight upper bound.

Example 5.2:

Let X be uniform in [0,4]. Use the Chebyshev inequality to bound the probability of $|X-2| \geq 1$.

Chebyshev Inequality Example 1: Weak Law of Large Numbers (LLN)

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) RVs with mean μ and **finite** variance σ^2. For every $\epsilon > 0$ (no matter how small it is), we have the following (weak) law of large numbers

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \to 0, \quad \text{as } n \to \infty$$

where $M_n = (X_1 + X_2 + \ldots + X_n)/n$ is the sample mean.

The weak LLN (大數法則) says that the *sample mean* will be very close to *the true mean*, when the sample size is sufficiently large.

(Proof) Use Chebyshev inequality!

$$\mathrm{P}(|M_n - \mu| \geq \epsilon) \;\leq\; \frac{\mathrm{Var}(M_n)}{\epsilon^2} \;=\; \frac{\sigma^2}{n\epsilon^2}$$

Chebyshev Inequality Example: The Pollster’s Problem

The Pollster’s Problem (民意調查問題)

You may often see something like this, particularly during elections:

「這次選前民意調查，XX的支持率約為52%。此結果有 95% 信心水準，抽樣誤差於正負 3% 以內。」

Questions:

1) 直覺上，抽樣樣本數越多，則估計出來的支持率越接近真實。但，最少需抽樣幾份，才能保證有上述之抽樣誤差與信心水準?

2) 另外一個問法是，民調公司預算僅能允許抽樣1000份。若欲保證預估的支持率可有95%信心水準，那抽樣誤差要如何修正?

The Pollster's Problem

Example 5.5: Accuracy of Polling (p.270) (民調數據的準確率)

- The ith person' polled: $X_i = \begin{cases} 1 & \text{If "Yes"} \\ 0 & \text{If "No"} \end{cases}$

- Fraction of "Yes" : $M_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$

◆ p: (unknown) fraction of population voting "Yes": $\mu_x = p$

◆ The question here is:

95% 信心水準，抽樣誤差於正負 1% 以內，則需要抽樣幾份?

The Pollster's Problem

- Accuracy of Polling (p.270) (民調數據的準確率)

  95% 信心水準，抽樣誤差於正負 1% 以內

$$\Pr(|M_n - \mu_X| \geq 0.01) \leq 0.05$$

◆ 使用 Chebyshev: $\Pr(|M_n - \mu_x| \geq \varepsilon) \leq \dfrac{\sigma_x^2}{n\varepsilon^2}$ $\quad \mu_x = p$

➢ In fact, X_i is i.i.d. Bernoulli RV: $\mu_x = p; \quad \sigma_x^2 = p(1-p) \leq 1/4$

$$\Pr(|M_n - p| \geq 0.01) \leq \frac{1}{0.0004n} \quad \Rightarrow n > 50{,}000$$

➢ *Remarks:*

1. Chebyshev 算出的 n 可保證不等式成立，but...
2. CLT can also do the tricks, and give a much smaller n (i.e., fewer samples. See page 14.)

The Central Limit Theorem (CLT)

Central Limit Theorem (中央極限定理)

Let $X_1, X_2, \ldots, X_n$ be a sequence of **i.i.d.** RVs with mean μ and variance σ^2, and define

$$Z_n \triangleq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then, the CDF of Z_n converges to the standard normal CDF $\Phi(z)$ in the sense that

$$\lim_{n\to\infty} P(Z_n \leq z) = \Phi(z) \quad \text{for every } z.$$

Remarks:

- Note that $X_1, X_2, \ldots, X_n$ can be of **any distributions** (this is why CLT is so powerful), and the CLT is also true.

不論 $X_1, X_2, \ldots, X_n$ 原本的統計行為如何，Z_n 在n很大時會趨近於常態分佈

The Central Limit Theorem (CLT)

Central Limit Theorem (中央極限定理)

Let $X_1, X_2, \ldots, X_n$ be a sequence of **i.i.d.** RVs with mean μ and variance σ^2, and define

$$Z_n \triangleq \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

Then, the CDF of Z_n converges to the standard normal

- *該如何去理解上式Z_n的結構? 減去 $n\mu$ 的意義，分母的物理意義各為何?*

- CLT well models many practical real-world problems. Sum of i.i.d. *RVs* with *arbitrary distributions* having finite mean and variance can be approximated by Gaussian
  - ✓ The number of occurrence of a certain event, the sample mean
  - ✓ Noise, channel effects in the received signal of your smartphone

Proof of CLT (p.290)

- Sum of indep. RVs → use moment generating function

- We have:

$$M_{Z_n}(s) = \mathbf{E}[e^{sZ_n}] = \mathbf{E}\left[e^{(s/\sqrt{n})(X_1+\cdots+X_n)}\right]$$

$$\mathbf{E}[e^{sX/\sqrt{n}}] \approx 1 + \frac{s}{\sqrt{n}}\mathbf{E}[X] + \frac{s^2}{2n}\mathbf{E}[X^2]$$

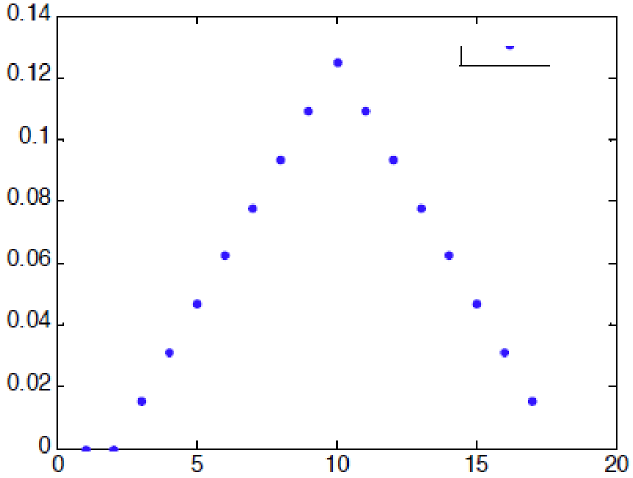
$$M_{Z_n}(s) = (\mathbf{E}[e^{sX/\sqrt{n}}])^n \approx \left(1 + \frac{s^2}{2n}\right)^n \longrightarrow e^{s^2/2}$$

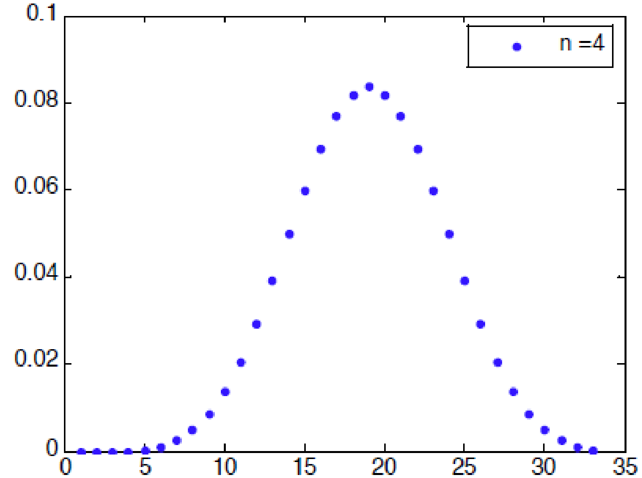
which is the transform of the standard normal.

“N” in CLT

- To make a “good” Gaussian, how large is n?

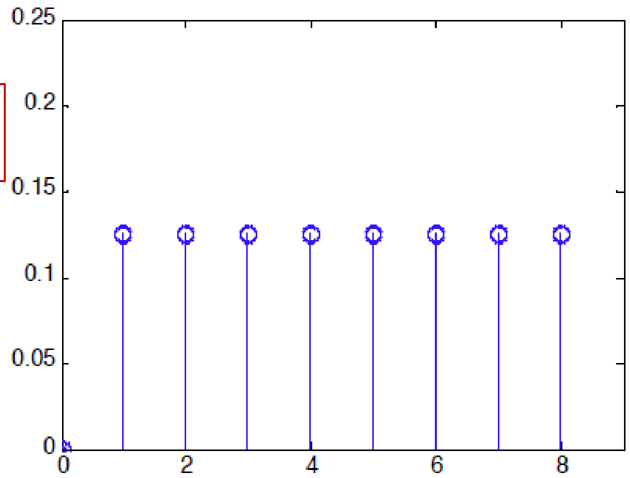
n=2

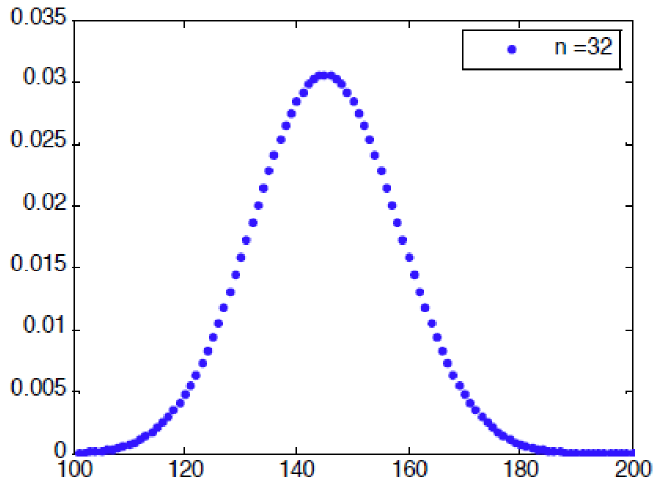




n=4

n=1

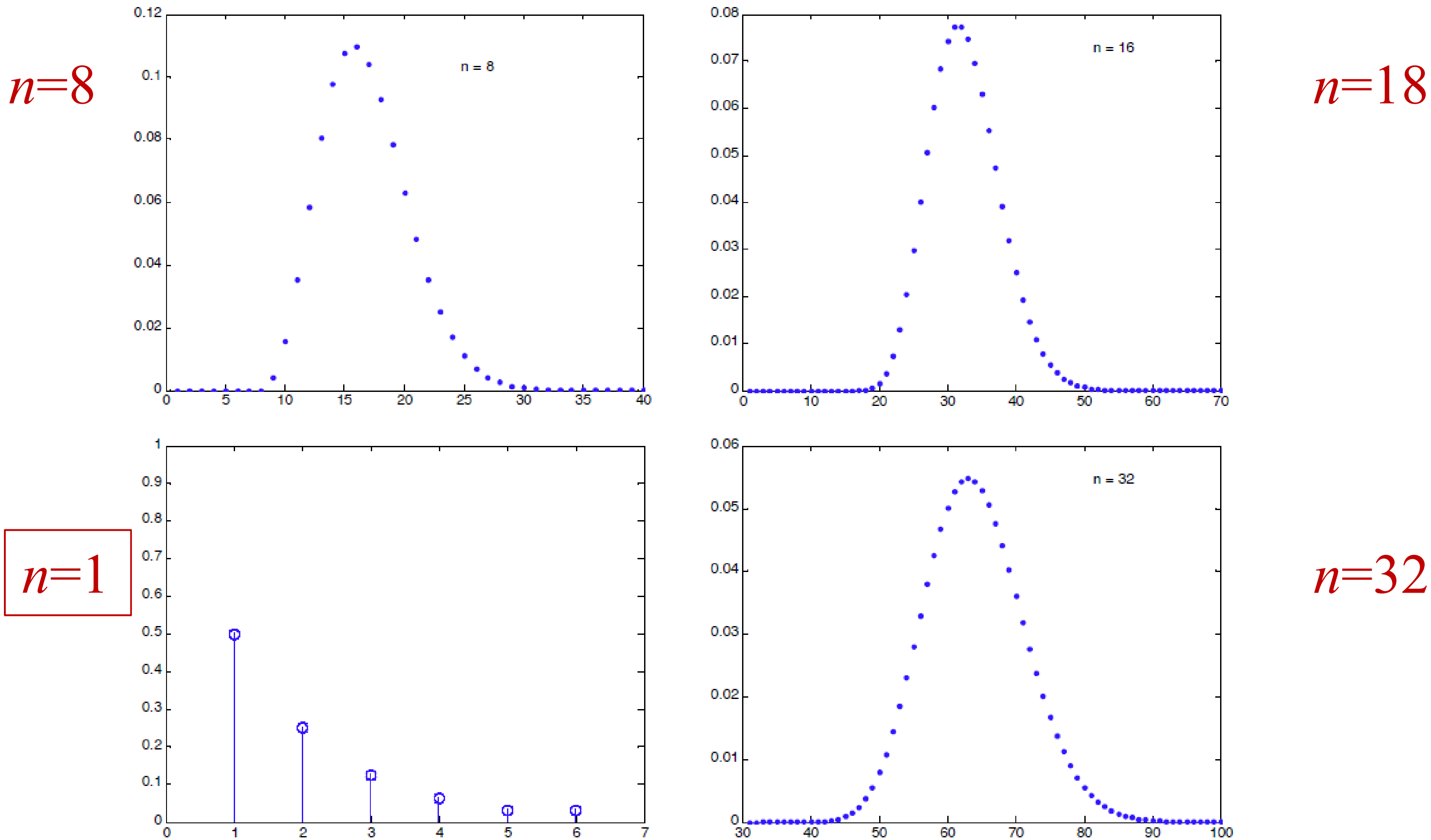




n=32

“N” in CLT (2)

- The shape of PDF of X has “some” impact on n.



The Pollster's Problem (p.277)

- For the ith person: $X_i = \begin{cases} 1 & \text{If "Yes"} \\ 0 & \text{If "No"} \end{cases}$

- Fraction of "Yes": $M_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ $\qquad \sigma_x^2 = p(1-p) \le ¼$

◆ p: fraction of population voting "Yes" = μ_X

◆ Suppose we want: 95% 信心水準，抽樣誤差於正負 1% 以內

$$\Pr(\left|M_n - \mu_X\right| \ge 0.01) \le 0.05$$

◆ CLT facilitates calculations of the probability:

$$\left|\frac{X_1 + X_2 + \cdots + X_n - np}{n}\right| \ge 0.01 \rightarrow \left|\frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{n}\sigma}\right| \ge 0.01\frac{\sqrt{n}}{\sigma}$$

$$\Pr(\left|M_n - \mu_X\right| \ge 0.01) \approx \Pr(|Z| \ge 0.02\sqrt{n}) \le 0.05$$

$$2 \cdot 0.01 \cdot \sqrt{n} \ge 1.96 \Rightarrow n > 9604$$

CLT Applied to Binomial

- Bernoulli: $\mu_x = p; \quad \sigma_x^2 = p(1-p)$
- Binomial: $S_n = X_1 + X_2 + \cdots + X_n$

$$\mu_S = np; \quad \sigma_S^2 = np(1-p)$$

- Approximation:

$$\text{CDF of } \frac{S - np}{\sqrt{np(1-p)}} \to \text{standard normal}$$

- *Example*: n=36, p=0.5; find $\Pr(S_n=19)$

➢ Exact answer:

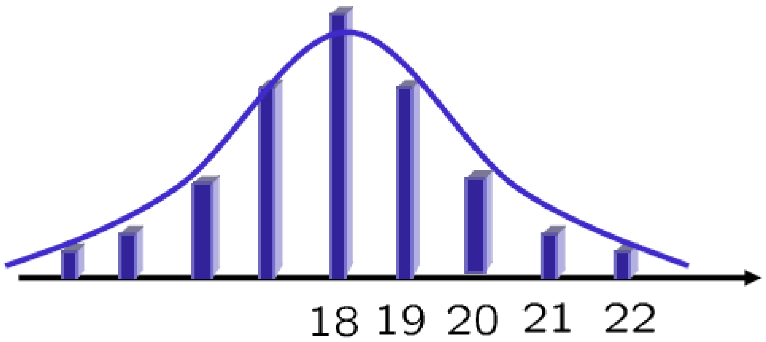
$$\binom{36}{19}\left(\frac{1}{2}\right)^{36} = 0.1251$$

De Moivre-Laplace Approximation to Binomial

- **The ½ Correction for Binomial Approximation**

  Let S_n be normalized to zero mean and unit variance:

  $$\Pr(k \le S_n \le l) \approx \Phi(l+0.5) - \Phi(k-0.5)$$



- *Ex*: $n=36$, $p=0.5$; $\Pr(S_n=19) = \Pr(18.5 \le S_n \le 19.5)$

➢

$$18.5 \le S_n \le 19.5 \iff$$

$$\frac{18.5-18}{3} \le \frac{S_n-18}{3} \le \frac{19.5-18}{3} \iff$$

$$0.17 \le Z_n \le 0.5$$

$$\begin{aligned} P(S_n = 19) &\approx P(0.17 \le Z \le 0.5) \\ &= P(Z \le 0.5) - P(Z \le 0.17) \\ &= 0.6915 - 0.5675 \\ &= 0.124 \end{aligned}$$

Convergence

- **Deterministic limit**: a sequence → a number
  - a_n converges to a: $\lim_{n\to\infty} a_n = a$

    a_n eventually gets and stays (arbitrarily) close to a
  - Def: For every $\varepsilon>0$ there exists n_0, such that for all

    $n \geq n_0, \quad |a_n - a| \leq \varepsilon.$

- **Stochastic limits**: there exists several definitions
  - Convergence "in probability"
  - Convergence "almost surely" or "with probability 1"

Convergence “in Probability”

- $Y_1, Y_2, \ldots$ is a sequence of RVs, and let a be a real number. We say Y_n converges to a **in probability**, if for every $\varepsilon>0$, we have

$$\lim_{n\to\infty} \Pr(|Y_n - a| \geq \varepsilon) = 0$$

- (Almost) all of the PMF/PDF of Y_n eventually gets concentrated (arbitrarily) close to a

- *例子: (大數法則)* The weak law of large numbers → the sample mean *converges in prob* to the true mean.

- **Accuracy level:** ε, **Confidence** level: δ;

$$\Pr(|Y_n - a| \geq \varepsilon) \leq \delta, \quad \text{for all } n \geq n_0$$

Example (p.272)

Example 5.6:

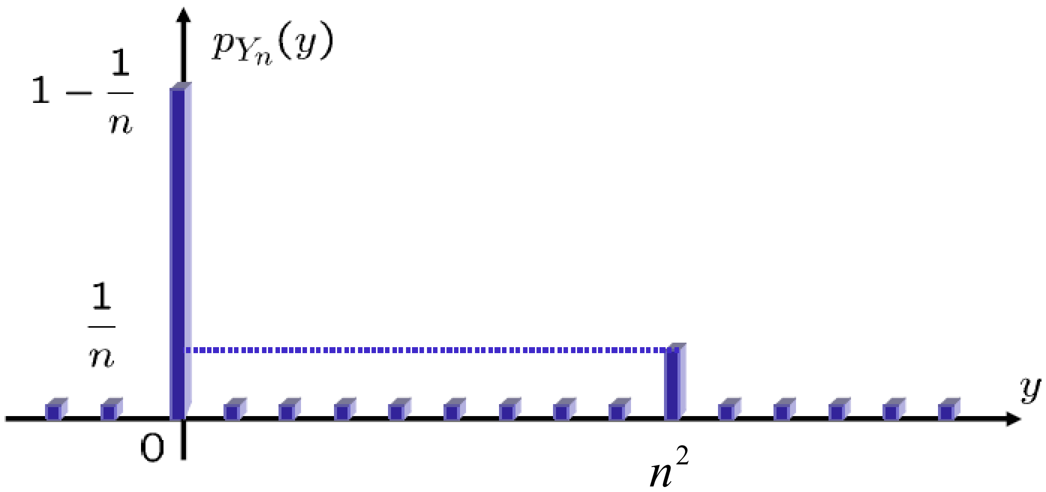
Consider a sequence of independent random variables X_n that are uniform in [0,1]. Let

$$Y_n = \min\{X_1, \dots, X_n\}$$

Does Y_n converge to   in probability?

Example (p.272)

- Consider a sequence of RVs Y_n with the following PMFs.



- Does Y_n converge? Yes.

$$\lim_{n\to\infty} \Pr(|Y_n - 0| \geq \varepsilon) = \lim_{n\to\infty} \frac{1}{n} = 0$$

- What is $E[Y_n]$? "*n*"

$$0 \cdot \left(1 - \frac{1}{n}\right) + n^2 \frac{1}{n} = n$$

Convergence "with Probability 1"

- $Y_1, Y_2, \ldots$ is a sequence of RVs, and let c be a real number. We say Y_n converges to c **with probability 1**, if

$$\Pr(\lim_{n\to\infty} Y_n = c) = 1$$

- *Remarks*:
  - $\left\{\lim_{n\to\infty} Y_n = c\right\}$ 寫得精確一些應為 $\left\{\omega\colon \lim_{n\to\infty} Y_n(\omega) = c\right\}$

    大括號表示一個事件，此事件發生機率為 1 時稱 Y_n converges to c **with probability 1**

- 隨機變數所構成的序列，其收斂的不同定義會在研究所 "隨機過程"課程中有更詳細解說

Example -- Convergence with Probability 1

Example

Consider a sample space $\Omega = [0,1]$, the closed interval between 0 and 1. Assume that each sample point $\epsilon \in \Omega$ follows continuous uniform distribution. Then, the random sequence $X_n(\epsilon) = \exp(-n^2(\epsilon - 1/n))$ converges to 0 with probability 1.

Example

Example 5.14

Consider a sequence of independent random variables X_n that are uniform in [0,1]. Let

$$Y_n = \min\{X_1, \dots, X_n\}$$

Show that Y_n converge to 0 with probability 1.

Convergence "with Probability 1"

- $Y_1, Y_2, \ldots$ is a sequence of RVs, and let c be a real number. We say Y_n converges to c **with probability 1**, if

$$\Pr(\lim_{n\to\infty} Y_n = c) = 1$$

- *Remarks*
  - Convergence with probability 1 定義中有傳統數列收斂的概念(see p.19)。樣本空間中若*幾乎*每一個sample point 對應到的數列都會收斂，僅部分的sample points對應到的數列不收斂，若那些對應到不收斂數列的sample points所成集合發生機率是零，即為此處所謂收斂 with probability 1 或稱為 convergence almost surely
  - Convergence *with probability 1* implies convergence *in probability*. The converse is not necessarily true. (Problem 15*, p.292)

    Example 5.15 (p.282) says a random sequence may converge *in probability,* but not *with probability 1*

Example

Example 5.15

Consider a discrete time arrival process. The set of times is $\mathbf{I}_k = \{2^k, 2^k + 1, \ldots, 2^{k+1} - 1\}$ for $k=1,2,\ldots$. During each $\mathbf{I}_k$, there is exactly one arrival, and all times within an interval are equally likely. The arrival times in different intervals are assumed to be independent. Define $Y_n=1$ if there is an arrival at time n, and $Y_n=0$ if there is no arrival.

Show that Y_n converge to 0 in probability, but NOT with probability 1

The Strong Law of Large Numbers

- Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with mean μ.

  Let $M_n = \left(X_1 + X_2 + \cdots + X_n\right)/n$,

$$\boxed{\Pr(\lim_{n\to\infty} M_n = \mu) = 1}$$

- Sample space: set *A* = collections of $(x_1, x_2, \ldots)$

$$\omega = (x_1, x_2, \cdots) \text{ such that } \left[\lim_{n\to\infty} \frac{x_1 + x_2 + \cdots + x_n}{n} = \mu\right]$$

➔ Prob of *A* = 1.

- *Weak Law*: $\lim_{n\to\infty}\left\{\Pr(\left|M_n - \mu\right| \geq \varepsilon)\right\} = 0$

  這是convergence in probability的概念

- *Strong Law:* For every ε>0 there exists n_0,

$$\Pr(\left\{\omega_n : \left|M_n - \mu\right| < \varepsilon, n \geq n_0\right\}) = 1$$

這是convergence with probability 1 的概念