第四章

第四节

中心极限定理

关于随机变量的计算

—— 确定其分布！

$X_1, X_2, \ldots \ldots, X_n \;\Rightarrow X = \sum_{i=1}^{n} X_i \sim F_X(x) = ?$

在许多情况下，和的极限分布就是正态分布——借用极限的思想

在什么情况下，随机变量和的极限分布是正态分布？—— 中心极限定理

一、中心极限定理

设随机变量序列$\{X_n\}$的部分和为$S_n = \sum_{k=1}^{n} X_k$，若$\forall x \in R$,

$$\lim_{n\to\infty} P\left\{\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \le x\right\} = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

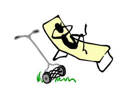
即

$$Y_n^* = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} \xrightarrow{n\to\infty} N(0,1)$$

则称$X_1, X_2, \cdots$服从中心极限定理。

二、常用的中心极限定理

- 独立同分布下的中心极限定理
  - 棣莫弗-拉普拉斯CLT
  - 林德伯格-莱维CLT
- 独立不同分布下的中心极限定理
  - 林德伯格CLT
  - 李雅普诺夫CLT



1、独立同分布下的中心极限定理

（1）．棣莫弗-拉普拉斯CLT

若r. v. $X_1, X_2, \cdots$ i. i. d. (0−1)分布，即

$P(X_i = 1) = p$，$P(X_i = 0) = 1 - p \quad i = 1, 2, \ldots$

记$S_n = \sum_{i=1}^{n} X_i$，则

$$Y_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} \overset{近似}{\sim} N(0,1)$$

即$\forall x \in R$，有 $\lim_{n\to\infty} P\left\{\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right\} = \Phi(x)$

1、独立同分布下的中心极限定理

Remark.

二项分布是离散分布，而正态分布是连续分布，所以用正态分布作为二项分布的近似时，可作如下修正:

$$P(k_1 \le \mu_n \le k_2) = P(k_1 - 0.5 < \mu_n < k_2 + 0.5)$$

$$\approx \Phi\left(\frac{k_2 + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - 0.5 - np}{\sqrt{npq}}\right)$$

1、独立同分布下的中心极限定理

中心极限定理的应用有三大类：

i) 已知 n 和 y，求概率；

ii) 已知 n 和概率，求 y；

iii) 已知 y 和概率，求 n .

1、独立同分布下的中心极限定理

例1. 100个独立工作(工作的概率为0.9)的部件组成一个系统，求系统中至少有85个部件工作的概率.

例2. 有200台独立工作(工作的概率为0.7)的机床，每台机床工作时需15kw电力. 问共需多少电力，才可有95%的可能性保证正常生产?

例3. 用调查对象中的收看比例k/n 作为某电视节目的收视率 p 的估计。要有 90% 的把握，使k/n与 p的差异不大于0.05，问至少要调查多少对象？

1、独立同分布下的中心极限定理

例4 设每颗炮弹命中目标的概率为0.01，求500发炮弹中命中 5 发的概率.

解：设 X 表示命中的炮弹数，则 $X \sim b(500, 0.01)$

(1) $P(X=5) = C_{500}^{5} \times 0.01^5 \times 0.99^{495} = 0.17635$

(2) 应用正态逼近:

$$P(X=5) = P(4.5 < X < 5.5) \approx \Phi\left(\frac{5.5-5}{\sqrt{4.95}}\right) - \Phi\left(\frac{4.5-5}{\sqrt{4.95}}\right)$$

$$= 0.1742$$

1、独立同分布下的中心极限定理

(2). 林德伯格-莱维CLT

设随机变量序列$\{X_n\}$i. i. d.，且$E(X_i) = \mu$,

$Var(X_i) = \sigma^2 > 0$, 记$S_n = \sum_{k=1}^{n} X_k$, 则

$$Y_n^* = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\text{近似}} N(0,1)$$

即 $\forall x \in R$, $\lim_{n\to\infty} P\{Y_n^* \le x\} = \Phi(x)$

应用：正态随机数的产生；误差分析

证明: 连续性定理可知，只需证明 Y_n^* 的特征函数收敛于标准正态分布的特征函数

因为$X_1, X_2, \ldots, X_n$独立且同分布,

故设 $X_i - \mu$ 的特征函数为 $\varphi(t)$

则 $Y_n^* = \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma\sqrt{n}}$ 的特征函数为

$$\varphi_{Y_n^*}(t) = \left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$$

$$\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = \varphi(0) + \varphi'(0)\frac{t}{\sigma\sqrt{n}} + \frac{\varphi''(0)}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\left(\frac{t}{\sigma\sqrt{n}}\right)^2\right)$$

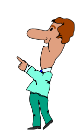
而 $E(X_i - \mu) = 0$, $Var(X_i - \mu) = \sigma^2$ $i = 1, 2, \cdots, n$

故 $\varphi(0) = 1$, $\varphi'(0) = 0$, $\varphi''(0) = -\sigma^2$

于是 $\varphi\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)$

$$\lim_{n\to+\infty} \varphi_{Y_n^*}(t) = e^{-\frac{t^2}{2}}$$

$$\boxed{\lim_{x\to 0}(1+x)^{\frac{1}{x}} = e}$$



1、独立同分布下的中心极限定理

例5　每袋味精的净重为随机变量，平均重量为100克，标准差为10克. 一箱内装200袋味精，求一箱味精的净重大于20500克的概率?

课堂练习. 一食品店有三种蛋糕出售，由于售出哪一种蛋糕是随机的，因而售出一只蛋糕的价格是一个随机变量，它取5（元），7（元），10（元）的概率分别为0.3, 0.2, 0.5. 某天售出300只蛋糕。试求这天的收入至少是2000（元）的概率。

1、独立同分布下的中心极限定理

例6　某药厂生产的某种药品，声称对某疾病的治愈率为80%，现为了检验此治愈率，任意抽取100个此种病患者进行临床试验。如果有多于75人治愈,则此药通过检验。试在以下两种情况下，分别计算此药通过检验的可能性

(1) 此药的实际治愈率为80%;

(2) 此药的实际治愈率为70%.

解: 记Y_n=100个此种病患者中治愈的人数，则

(1) $Y_n \sim B(100, 0.8) \Rightarrow E(Y_n) = 80 \quad Var(Y_n) = 16$

$$P\{Y_n \ge 75\} \approx 1 - \Phi\left(\frac{75 - 0.5 - 80}{4}\right) = 1 - \Phi\left(-\frac{5.5}{4}\right)$$

$$= \Phi(1.375) = 0.9155$$

此药通过检验的可能性是很大的

(2) $Y_n \sim B(100, 0.7)$

$$\Rightarrow E(Y_n) = 70 \quad Var(Y_n) = 21$$

$$P\{Y_n \ge 75\} \approx 1 - \Phi\left(\frac{75 - 0.5 - 70}{\sqrt{21}}\right) = 1 - \Phi(0.982)$$

$$= 1 - 0.8370 = 0.163$$

可见，此药通过检验的可能性是很小的

2、独立不同分布下的中心极限定理

(1). 林德伯格CLT

设$\{X_n\}$为独立随机变量序列，若任对 $\tau > 0$，有

$$\lim_{n\to\infty} \frac{1}{\tau^2 B_n^2} \sum_{i=1}^{n} \int_{|x-\mu_i| > \tau B_n} (x - \mu_i)^2 p_i(x) \mathrm{d}x = 0$$ 林德贝格条件

则 $$\lim_{n\to\infty} P\left\{\frac{1}{B_n} \sum_{i=1}^{n} (X_i - \mu_i) \le y\right\} = \Phi(y)$$

林德贝格条件较难验证

2、独立不同分布下的中心极限定理

(2). 李雅普诺夫CLT

设$\{X_n\}$是相互独立的随机变量序列，且$E(X_i) = \mu_i$，

$D(X_i) = \sigma_i^2 \ne 0, \ i = 1, 2, \cdots$, 记$B_n^2 = \sum_{i=1}^{n} \sigma_i^2$,

若$\exists \delta > 0$, 使

$$\frac{1}{B_n^{2+\delta}} \sum_{i=1}^{n} E\left\{|X_i - \mu_i|^{2+\delta}\right\} \xrightarrow{n\to\infty} 0$$

则 $$\lim_{n\to\infty} P\left\{\frac{1}{B_n} \sum_{i=1}^{n} (X_i - \mu_i) \le y\right\} = \Phi(y)$$

继　续

思 考 题

一份考卷由99个题目组成，按由易到难顺序排列，某学生答对第一题的概率为0.99；答对第二题的概率为0.98；一般地，答对第i题的概率为$1-i/100$，$i=1,2,\ldots$；假如该学生回答各题目是相互独立的，并且要正确回答其中60个题目以上才算通过考试。试计算该学生通过考试的可能性有多大?

解: 设 $X_i=\begin{cases}1 & \text{若学生答对第}i\text{道题}\\ 0 & \text{若学生答错第}i\text{道题}\end{cases}\quad i=1,2,\cdots,99$

则X_i服从不同的两点分布:

$$P(X_i=1)=p_i=1-\frac{i}{100}$$

$$P(X_i=0)=1-p_i=\frac{i}{100}$$

$$E(X_i)=p_i,\quad Var(X_i)=p_i(1-p_i)<+\infty$$

为使用中心极限定理，设从X_{100}开始的随机变量都与X_{99}同分布，且相互独立 令δ=1，验证李雅普诺夫条件:

$$B_n^2=\sum_{i=1}^{n}Var(X_i)=\sum_{i=1}^{n}p_i(1-p_i)\xrightarrow{n\to+\infty}\infty$$

$$E\left(|X_i-p_i|^3\right)=(1-p_i)^3p_i+(1-p_i)p_i^3\le(1-p_i)p_i$$

于是 $$\frac{1}{B_n^3}\sum_{k=1}^{n}E\left(|X_k-\mu_k|^3\right)\le\frac{1}{\left[\sum_{i=1}^{n}p_i(1-p_i)\right]^{1/2}}\to 0$$

即$\{X_i\}$满足李雅普诺夫条件，可以使用CLT

$$E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}\left(1-\frac{i}{100}\right)=49.5$$

$$Var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}p_i(1-p_i)=\sum_{i=1}^{n}\left(1-\frac{i}{100}\right)\left(\frac{i}{100}\right)=16.665$$

故 $$P\left(\sum_{i=1}^{n}X_i\ge 60\right)=P\left\{\frac{\sum_{i=1}^{n}X_i-49.5}{\sqrt{16.665}}\ge\frac{60-49.5}{\sqrt{16.665}}\right\}$$

$$\approx 1-\Phi(2.5735)=0.005$$

由此看出: 该学生通过考试的可能性很小.

课 堂 练 习

假设每个人路过某报亭时购买当日商报的概率是0.2，而每个人是否买报是相互独立的，试问（1）该报亭卖出第200份报纸时路过的人数超过1000人的概率是多少？（2）路过第1000人时卖出的报纸超过200份的概率是多少？