概率论与数理统计

中 心 极 限 定 理

主讲人：曾华琳

信息科学与技术学院

中心极限定理的客观背景

自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见。

Gauss

如果一个随机变量是由大量相互独立的随机因素的综合影响所造成，而每一个别因素对这种综合影响中所起的作用不大。则这种随机变量一般都服从或近似服从正态分布。

现在我们就来研究独立随机变量之和所特有的规律性问题。

当 n 无限增大时，这个和的极限分布是什么呢?

中心极限定理的客观背景

由于无穷个随机变量之和可能趋于∞，故我们不研究 n 个随机变量之和本身而考虑它的标准化的随机变量。

即考虑随机变量X_k $(k=1,\ldots n)$ 的和 $\sum_{k=1}^{n} X_k$

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}}$$

Y_n 讨论的极限分布是否为标准的正态分布

在概率论中，习惯于把和的分布收敛于正态分布这一类定理都叫做**中心极限定理**。

一、中心极限定理

定理1（独立同分布下的中心极限定理）

设随机变量$X_1, X_2, \cdots X_n, \cdots$相互独立，服从同一分布，且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots)$，则随机变量之和 $\sum_{k=1}^{n} X_k$的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$ 的分布函数$F_n(x)$对于任意x满足

$$\boxed{\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x \right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} \mathrm{d}t = \Phi(x)}$$

一、中心极限定理

注：

1. **定理表明，独立同分布的随机变量之和** $\sum_{k=1}^{n} X_k$ **，当** n **充分大时，随机变量之和与其标准化变量分别有**

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2); \qquad \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1).$$

一、中心极限定理

注：

2. **独立同分布中心极限定理的另一种形式可写为**

$$\bar{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n) \quad \text{或} \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1)$$

其中 $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

3. **虽然在一般情况下，我们很难求出** $\sum_{k=1}^{n} X_k$ **的分布的确切形式，但当** n **很大时，可以求出近似分布。**

一、中心极限定理

定理2（李雅普诺夫(Liapounov)定理)

设随机变量$X_1, X_2, \cdots, X_n \cdots$相互独立，它们具有数学期望和方差：$E(X_k) = \mu_k$,　$D(X_k) = \sigma_k^{\ 2}, (k = 1, 2, \cdots)$

记　$B_n^{\ 2} = \sum_{k=1}^{n} \sigma_k^{\ 2}$

若存在正数δ，使得当$n \to \infty$时，

$$\frac{1}{B_n^{\ 2+\delta}} \sum_{k=1}^{n} E\left\{ \left| X_k - \mu_k \right|^{2+\delta} \right\} \to 0$$

一、中心极限定理

则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量：

$$Z_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

的分布函数$F_n(x)$对于任意x，满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} \mathrm{d}t = \Phi(x)$$

一、中心极限定理

请注意：

1. 定理中随机变量之和$\sum_{k=1}^{n} X_k$及其标准化变量Z_n在n很大时，分别近似服从

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(\sum_{k=1}^{n} \mu_k, B_n^{\ 2}) \quad ; \qquad Z_n \overset{\text{近似地}}{\sim} N(0,1)$$

2. 随机变量X_k无论服从什么分布，只要满足定理条件，随即变量之和$\sum_{k=1}^{n} X_k$，当n很大时，就近似服从正态分布，这就是为什么正态分布在概率论中所占的重要地位的一个基本原因。

一、中心极限定理

定理6(棣莫佛－拉普拉斯（De Laplace定理)

设随机变量η_n(n=1,2,….)服从参数$n,p(0<p<1)$的二项分布，则对任意x，有

$$\lim_{n\to\infty} P\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

一、中心极限定理

定理6（棣莫佛－拉普拉斯（De Laplace定理）

证： 由第四章知识知可将 η_n 分解成为n 个互相独立、服从同一(0-1) 分布的诸随机变量 $X_1,X_2,\ldots X_n$ 之和，

即有

$$\eta_n = \sum_{k=1}^{n} X_k$$

其中X_k $(k=1,2,\ldots,n)$的分布律为

$$P\{X_k=i\}=p^i(1-p)^{1-i}, i=0,1$$

一、中心极限定理

由于 $E(X_k)=p$，$D(X_k)=p(1-p)$ ($k=1,2,\ldots,n$),

由定理4得

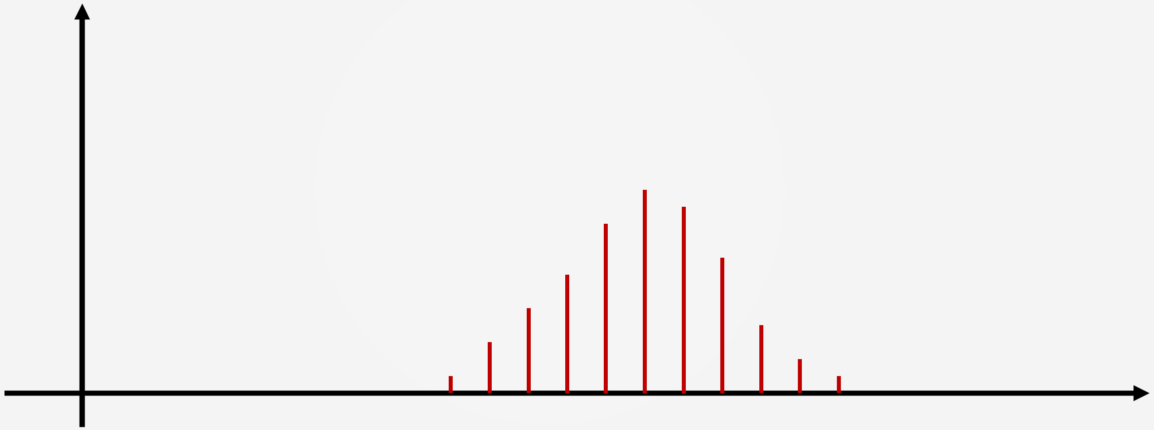
$$\lim_{n\to\infty}P\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\le x\}=\lim_{n\to\infty}P\{\frac{\sum_{k=1}^{n}X_k-np}{\sqrt{np(1-p)}}\le x\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)$$

定理表明，当 n 很大，$0<p<1$是一个定值时（或者说，$np(1-p)$ 也不太小时），二项变量η_n的分布近似正态分布 $N(np,np(1-p))$。

即 $$\eta_n \overset{\text{近似地}}{\sim} N(np,np(1-p))$$

一、中心极限定理

中心极限定理的客观背景



例：20个0-1分布的和的分布

谢 谢 大 家