§2 中心极限定理

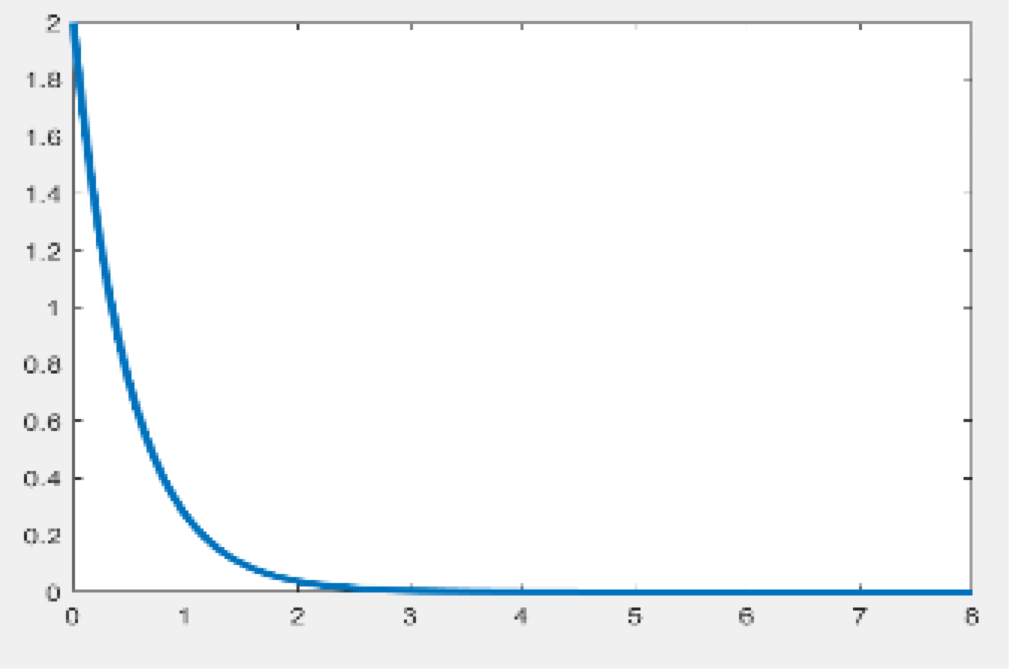
案例 某公司接待的客户的服务时间(单位：分钟)服从参数为$\lambda = 0.25$的指数分布。设第i个客户的服务时间为X_i。试估计一天8小时接待的客户数超过100的概率。

设Y表示服务100个客户的总时间，则

$$Y = X_1 + X_2 + \cdots + X_{100}$$

$$P(Y \le 480) = ?$$

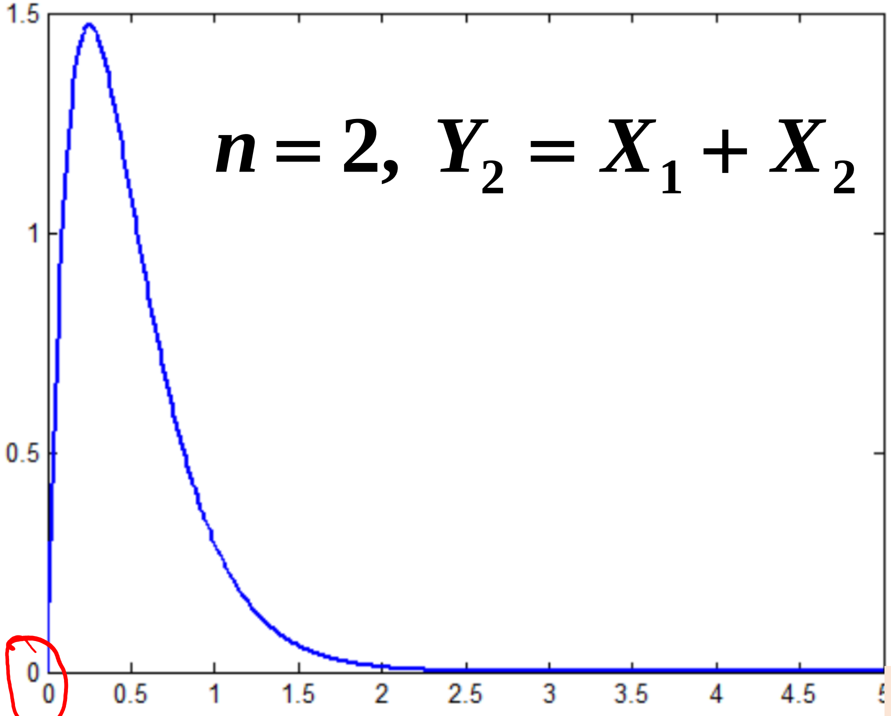
第i个客户的服务时间X_i（单位：分钟）服从参数为 $\lambda = 0.25$的指数分布，密度函数如下图



我们考查服务n个客户的总时间

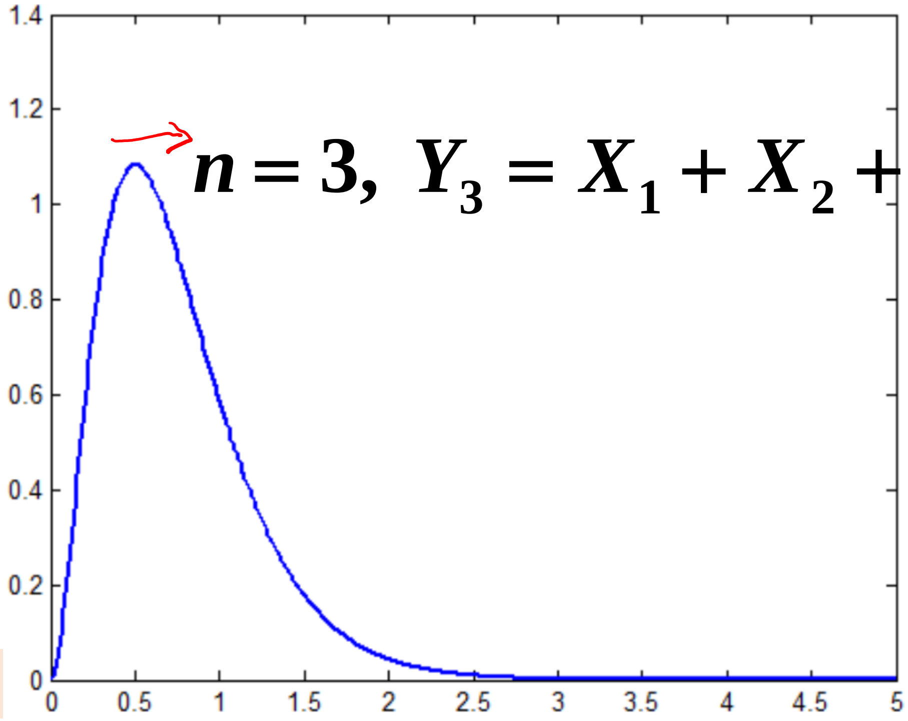
$$Y_n = \sum_{i=1}^{n} X_i$$

的分布。

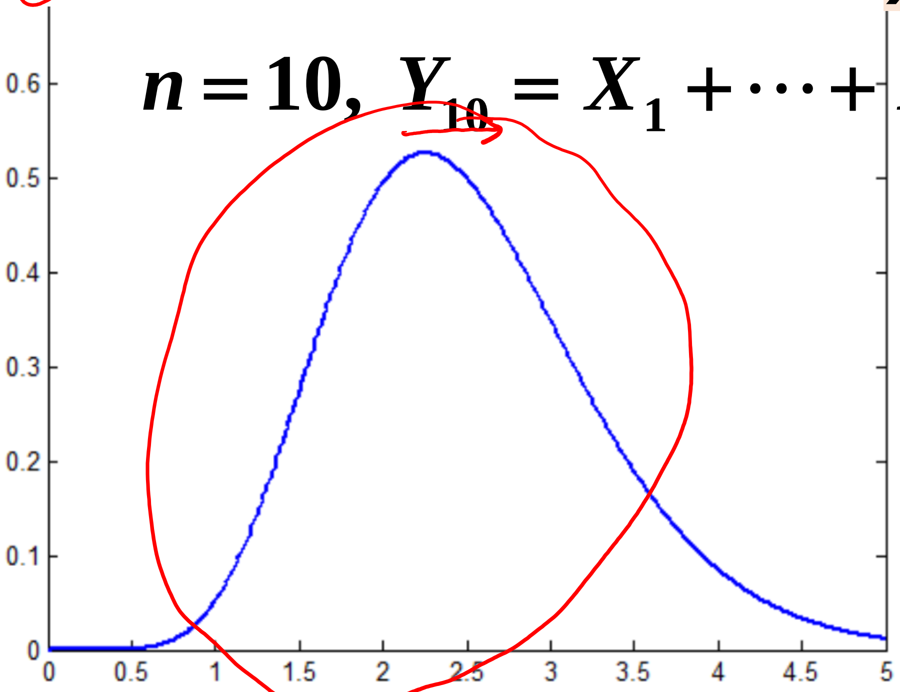


$$n=2,\ Y_2=X_1+X_2$$

$$\lambda=0.25$$



$$n=3,\ Y_3=X_1+X_2+X_3$$

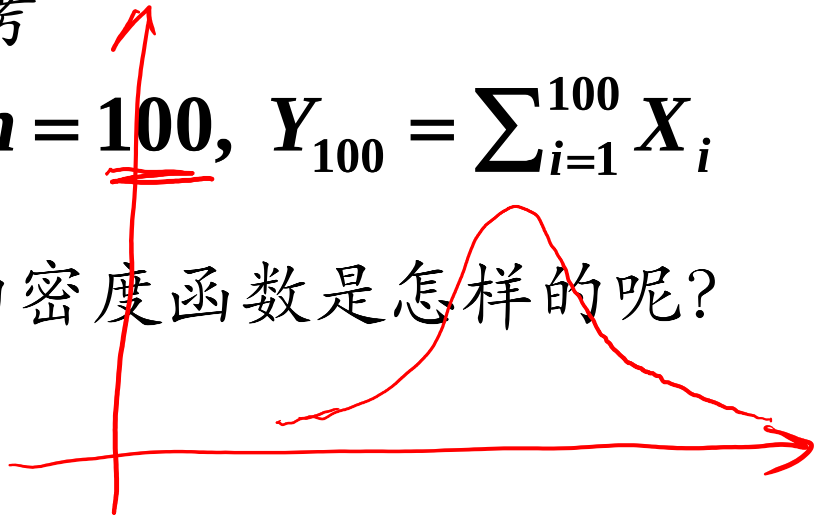


$$n=10,\ Y_{10}=X_1+\cdots+X_{10}$$

思考

$$n=100,\ Y_{100}=\sum_{i=1}^{100}X_i$$

的密度函数是怎样的呢?

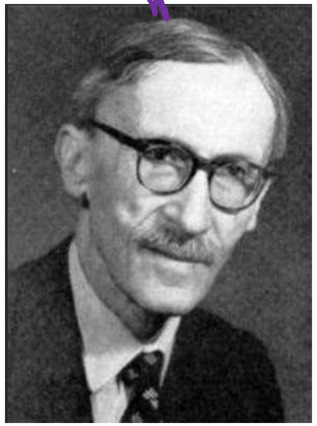


定理　独立同分布的中心极限定理　　Linderberg-Lévy中心极限定理

设随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 相互独立同分布（i.i.d.），且有期望和方差：

$$E(X_k) = \mu\,,\quad D(X_k) = \sigma^2 > 0\,,\quad k = 1, 2, \cdots$$

$< +\infty$



则对于任意实数 x，

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$$

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma} \xrightarrow{W} N(0,1)$$

依分布收敛

$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma}$ 的分布函数收敛于 $N(0,1)$ 的分布函数

$$\sum_{k=1}^{n} X_k \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2)$$

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

注 ❒ X_k 看成第 k 个随机因素，则

$\sum_{k=1}^{n} X_k$ 可看成 n个随机因素的叠加，中心极限定理表明

多个随机因素的叠加近似正态分布

身高为什么服从正态分布？

身高 $= \underbrace{X_0 + X_1 + X_2 + \cdots + X_n}$

体重为什么服从正态分布？

考试卷面成绩为什么服从正态分布？

- $X_k = \begin{cases} 1, & 第k次A发生; \\ 0, & o.w. \end{cases}$ $k = 1, \ldots, n$ $\quad X_1, X_2, \ldots, X_n$相互独立

成功率P

A发生的次数

$$X = \sum_{k=1}^{n} X_k \sim B(n, p)$$

中心极限定理表明----- $X = \sum_{k=1}^{n} X_k \xrightarrow{近似} N(np, np(1-p))$

De Moivre-Laplace中心极限定理

例 某公司接待的客户的服务时间(单位：分钟)服从参数为$\lambda = 0.25$的指数分布。设第i个客户的服务时间为X_i。试估计一天8小时接待的客户数超过100的概率。

解：$X_i \sim Exp(0.25)$，$EX_i = \frac{1}{0.25} = 4$ $DX_i = \frac{1}{0.25^2} = 16$，$X_1, \cdots, X_{100}$ i.i.d

$$Y = \sum_{i=1}^{100} X_i \overset{\text{近似}}{\sim} N(400, 1600)$$

$$P(Y \leq 480) = P\left(\frac{Y-400}{40} \leq \frac{480-400}{40}\right)$$

$$\approx \Phi(2) = 0.977$$

#

例 生产线生产的产品成箱包装，每箱的重量是随机的，每箱的平均重为 50kg，标准差为 5kg. 若用最大载重量为 5 吨的卡车承运，问每辆卡车最多可以装多少箱，才能保证不超载的概率大于0.977. （ $\Phi(2) = 0.977$ ）

解： X_i —— 第 i 箱重量，设最多可以装 n 箱， $X_1,\cdots,X_n$ i.i.d

$EX_i = 50$， $DX_i = 25$

$$T = \sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(50n, 25n)$$

$$P\left(\sum_{i=1}^{n} X_i \le 5000\right) = P\left(\frac{\sum_{i=1}^{n} X_i - 50n}{5\sqrt{n}} \le \frac{5000-50n}{5\sqrt{n}}\right)$$

$$\approx \Phi\left(\frac{5000-50n}{5\sqrt{n}}\right) \geq 0.977 = \Phi(2)$$

$$\frac{5000-50n}{5\sqrt{n}} \geq 2$$

$$\Rightarrow 5n + \sqrt{n} - 500 \le 0 \Rightarrow \sqrt{n} \le \frac{-1+\sqrt{10001}}{10} = 9.9 \Rightarrow n \le 9.9^2 = 98.01$$

最可装 98 箱 ※

作业 习题五

9, 11

补充题 设$X_k, k=1,2,\ldots$为独立同分布的随机变量，且$EX_1=0$, $E(X_1^2)=1$, $E(X_1^4)=\mu_4<+\infty$，令

$$S_n = X_1^2+\cdots+X_n^2$$

（1）证明：$\mu_4\geq 1$，等号成立的充要条件是 $X_1^2=1$, $a.s.$

(2) 计算 $E\left[\left(X_1^2-1\right)^2\right]$;

(3) 计算 $\lim\limits_{n\to\infty} P\left(S_n\leq n+\sqrt{n(\mu_4-1)}\right)$.