概率论与数理统计

第二十八讲（下）

中心极限定理

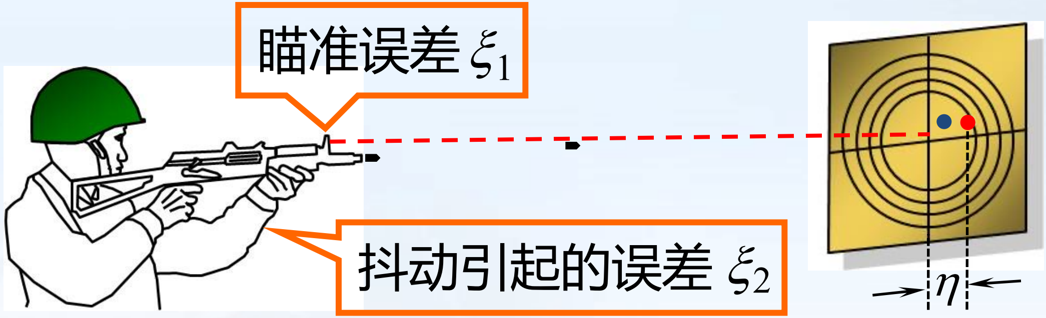
第28讲（下） 中心极限定理

- **问题背景:** 自然界许多随机指标均服从或近似服从正态分布

  例 子弹和炮弹的弹着点
  测量误差
  一个班级的课程考试成绩
  人的身高和体重
  一个城市的日平均耗电量
  一个地区的家庭年收入
  农作物的产量
  海浪的高度
  ……

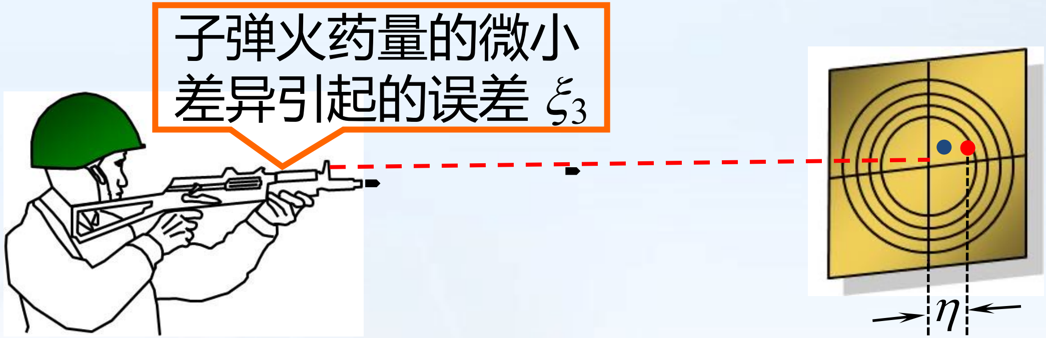
- **问题** 产生这一现象的原因是什么?

例 步枪射击时,子弹落点的横向偏差η服从正态分布.



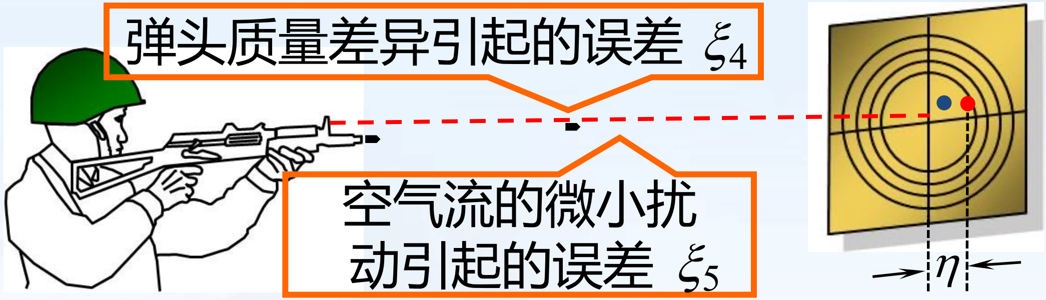
$$\eta=\xi_1+\xi_2$$

例 步枪射击时,子弹落点的横向偏差η服从正态分布.



$$\eta=\xi_1+\xi_2+\xi_3$$

例 步枪射击时,子弹落点的横向偏差η服从正态分布.

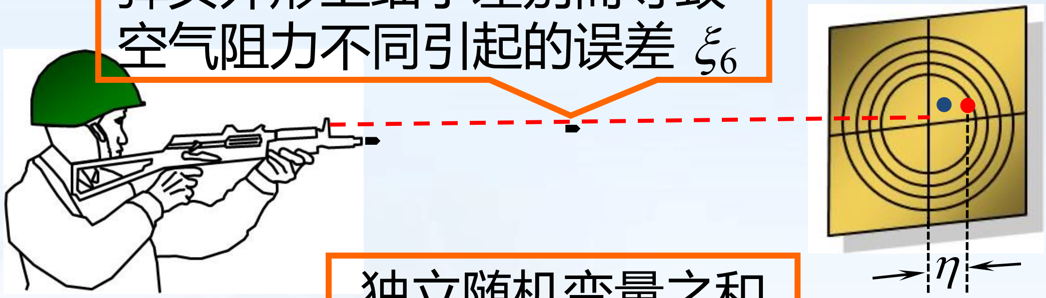


$$\eta=\xi_1+\xi_2+\xi_3+\xi_4+\xi_5$$

第28讲（下） 中心极限定理

例 步枪射击时,子弹落点的横向偏差 η 服从正态分布.

弹头外形上细小差别而导致空气阻力不同引起的误差 ξ_6



独立随机变量之和

$$\eta=\xi_1+\xi_2+\xi_3+\xi_4+\xi_5+\xi_6+\cdots+\xi_n$$

- **问题** 当 $n\to+\infty$ 时,在什么情况下

  $\eta_n=\sum_{i=1}^{n}\xi_i$ 的极限分布是正态分布？

- **标准化** 设$\{\xi_n\}$为均值,方差存在的随机变量列.令

$$\eta_n=\frac{\sum_{k=1}^{n}\xi_k-E(\sum_{k=1}^{n}\xi_k)}{\sqrt{D(\sum_{k=1}^{n}\xi_k)}}\quad(n=1,2,\cdots)$$

即有 $E(\eta_n)=0,D(\eta_n)=1\quad(n=1,2,\cdots).$

定义 设$\{\xi_n\}$为均值、方差存在的随机变量列.若$\{\xi_n\}$的**部分和标准化随机变量列**$\{\eta_n\}$满足$\forall x\in(-\infty,+\infty)$有

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

则称$\{\xi_n\}$**服从中心极限定理**.

问题 $\{\xi_n\}$在什么条件下服从中心极限定理？

假设(1) $\{\xi_n\}$是独立随机变量列；(2)均值和方差都存在

$$E(\xi_n)=\mu_n, D(\xi_n)=\sigma_n^2 \quad (n=1,2,\cdots)$$

结论成立否？

例 设X为均值、方差存在的任一随机变量,令

$$\xi_1=X, \xi_n=0 \quad (n=2,3,\cdots)$$

则$\{\xi_n\}$是独立随机变量列,且均值和方差都存在.

$$\eta_n=\frac{\sum_{k=1}^{n}\xi_k-E(\sum_{k=1}^{n}\xi_k)}{\sqrt{D(\sum_{k=1}^{n}\xi_k)}}=\frac{X-E(X)}{\sqrt{D(X)}} \quad (n=1,2,\cdots)$$

除非X服从正态分布,否则结论不成立.

定理(独立同分布中心极限定理) 设$\{\xi_n\}$为独立同分布随机变量列,其数学期望和方差均存在,记

$$E(\xi_n)=\mu,\ D(\xi_n)=\sigma^2 \quad (n=1,2,\cdots)$$

则$\{\xi_n\}$服从中心极限定理. 即$\forall x\in(-\infty,\infty)$有

$$\lim_{n\to\infty}P\left\{\frac{\sum_{k=1}^{n}\xi_k-E(\sum_{k=1}^{n}\xi_k)}{\sqrt{D(\sum_{k=1}^{n}\xi_k)}}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t$$

(证略)

第28讲（下） 中心极限定理

定理(独立同分布中心极限定理) 设$\{\xi_n\}$为独立同分布随机变量列,其数学期望和方差均存在,记

$$E(\xi_n)=\mu,\ D(\xi_n)=\sigma^2 \quad (n=1,2,\cdots)$$

则$\{\xi_n\}$服从中心极限定理.

定理(棣莫弗-拉普拉斯中心极限定理) 设$\eta_n \sim B(n,p)$, $(n=1,2,\cdots;0<p<1)$. 则对任意实数x有

$$\lim_{n\to\infty} P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t.$$

- **如何证明？**
- **回顾二项分布产生的背景**

中心极限定理的意义

设$\{\xi_n\}$为独立同分布随机变量列,期望和方差均存在

$$E(\xi_n)=\mu,\ D(\xi_n)=\sigma^2 \quad (n=1,2,\cdots)$$

则

$$\frac{\sum_{k=1}^{n}\xi_k-n\mu}{\sqrt{n}\sigma} \underset{\sim}{\text{近似}} N(0,1)$$

$$\xi_1+\xi_2+\cdots+\xi_n \underset{\sim}{\text{近似}} N(n\mu,n\sigma^2)$$

故在实际问题中,如果某随机指标满足

(1)该指标是由大量相互独立的随机因素迭加而成

(2)每个因素作用都是微小的,且没有一个因素起到突出的作用

则这个随机指标近似地服从正态分布.

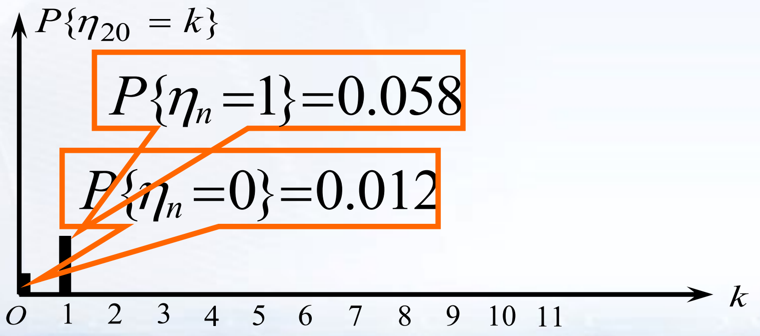
棣莫弗-拉普拉斯中心极限定理的应用

若随机变量列$\eta_n \sim B(n,p)$ $(n=1,2,\cdots)$, 则有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$$

故当 n 充分大时, 可近似认为 $\eta_n \sim N(np, np(1-p))$.

例 $\eta_{20} \sim B(20, 0.2)$的图形为



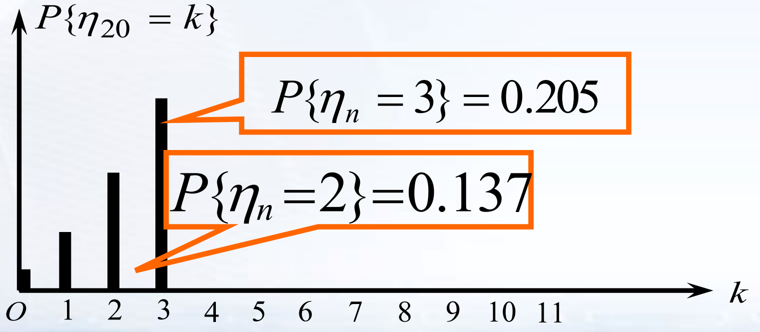
棣莫弗-拉普拉斯中心极限定理的应用

若随机变量列$\eta_n \sim B(n,p)\quad (n=1,2,\cdots)$，则有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$$

故当 n 充分大时，可近似认为 $\eta_n \sim N(np, np(1-p))$.

例 $\eta_{20} \sim B(20, 0.2)$的图形为



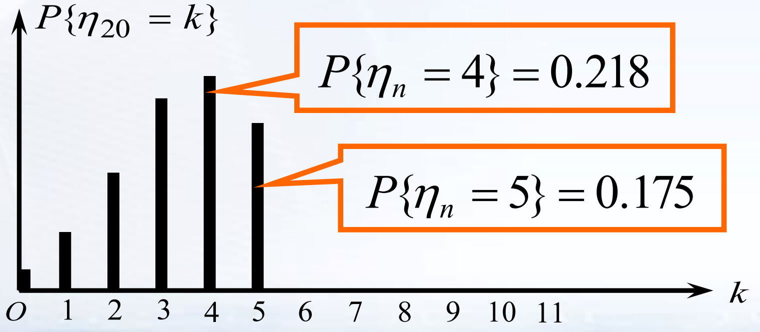
棣莫弗-拉普拉斯中心极限定理的应用

若随机变量列$\eta_n \sim B(n,p)$ $(n=1,2,\cdots)$, 则有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$$

故当 n 充分大时, 可近似认为 $\eta_n \sim N(np, np(1-p))$.

例 $\eta_{20} \sim B(20, 0.2)$ 的图形为



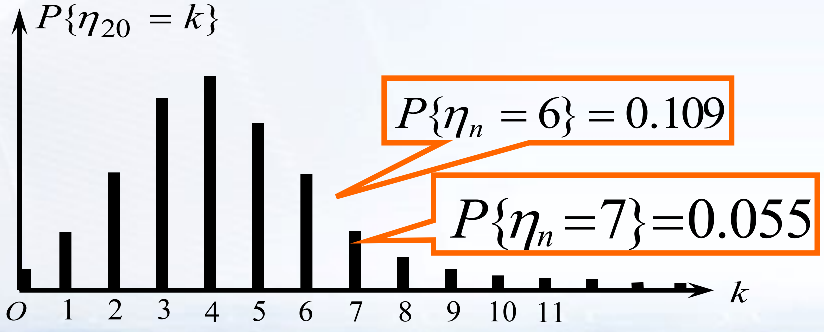
棣莫弗-拉普拉斯中心极限定理的应用

若随机变量列$\eta_n \sim B(n,p)$ $(n=1,2,\cdots)$，则有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$$

故当 n 充分大时，可近似认为 $\eta_n \sim N(np, np(1-p))$.

例 $\eta_{20} \sim B(20, 0.2)$ 的图形为



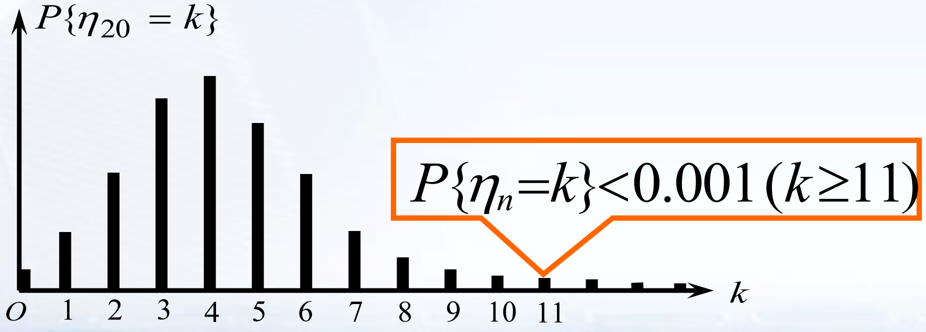
棣莫弗-拉普拉斯中心极限定理的应用

若随机变量列$\eta_n \sim B(n,p)$ $(n=1,2,\cdots)$, 则有

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$$

故当 n 充分大时, 可近似认为 $\eta_n \sim N(np, np(1-p))$.

例 $\eta_{20} \sim B(20, 0.2)$ 的图形为



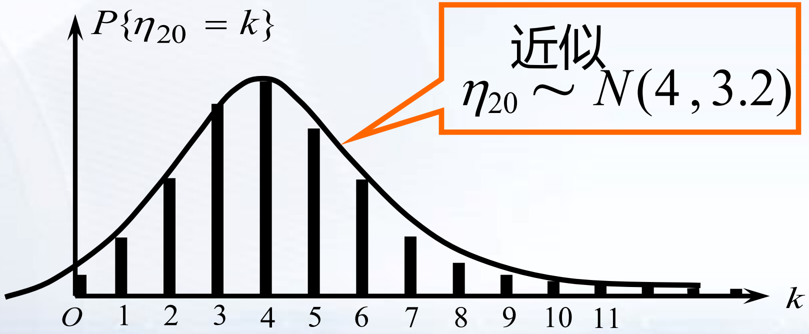
棣莫弗-拉普拉斯中心极限定理的应用

若随机变量列 $\eta_n \sim B(n,p)\quad(n=1,2,\cdots)$, 则有

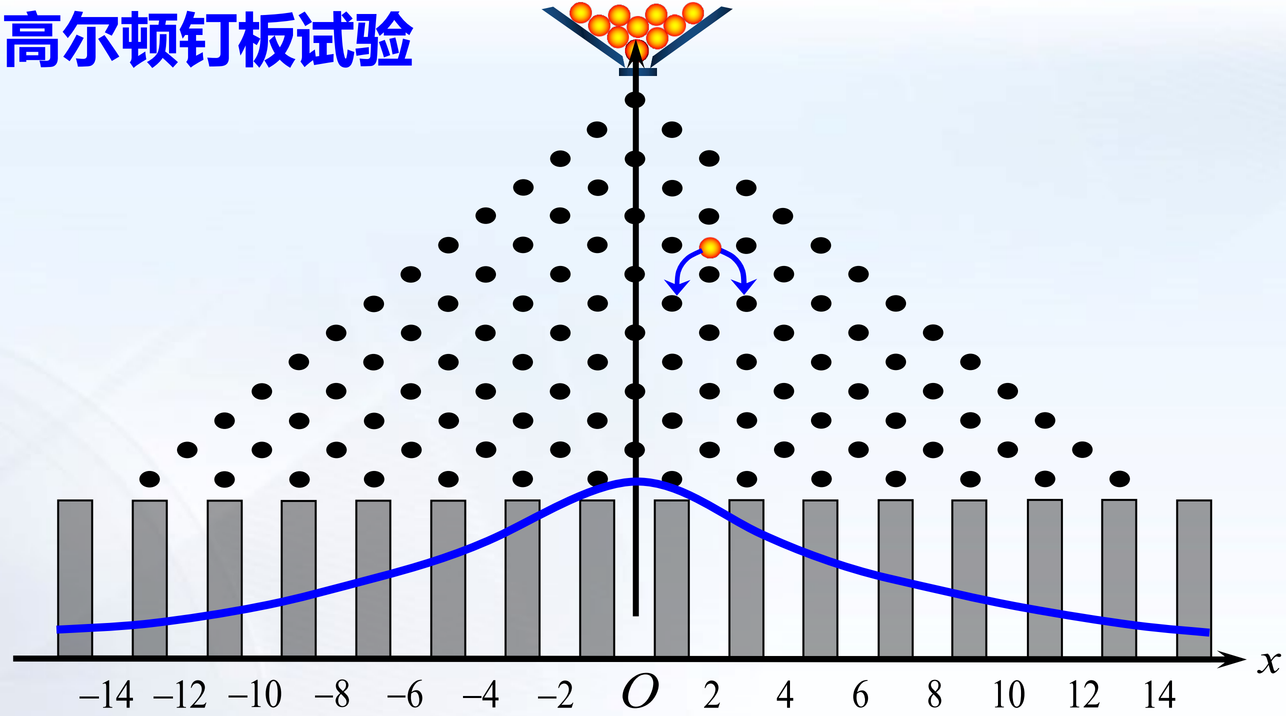
$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1)$$

故当 n 充分大时, 可近似认为 $\eta_n \sim N(np, np(1-p))$.

例 $\eta_{20} \sim B(20, 0.2)$ 的图形为

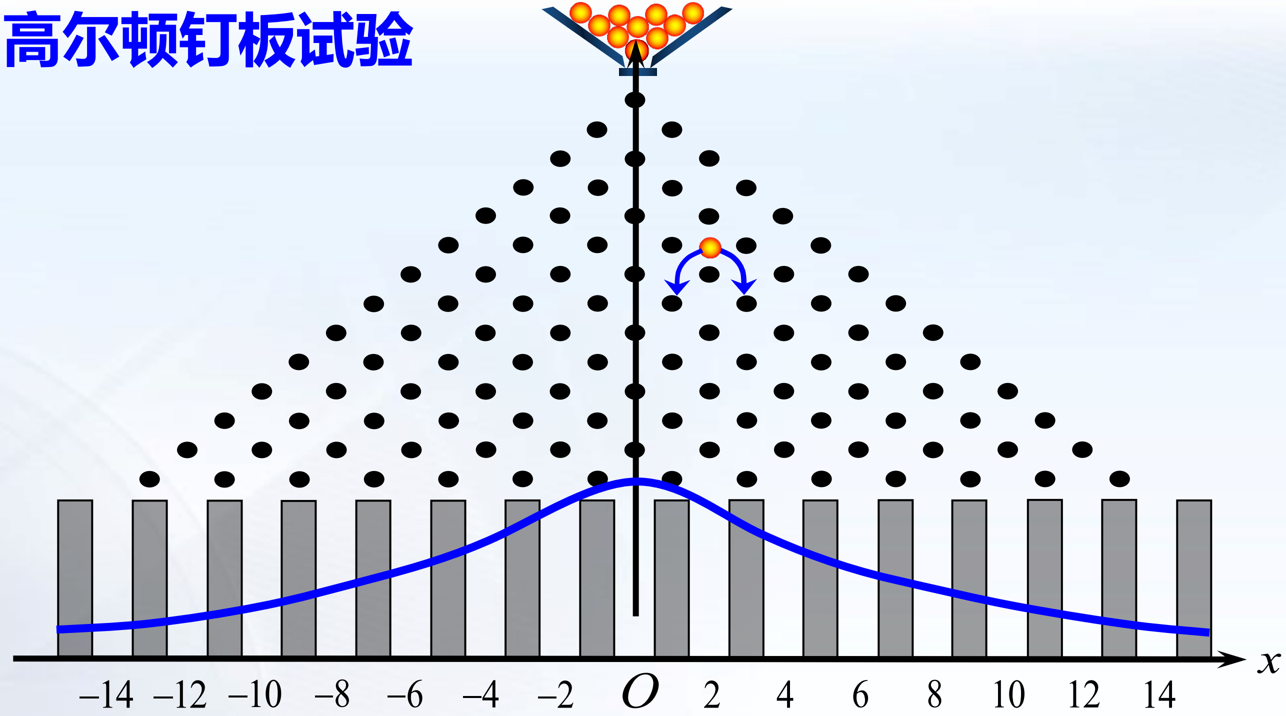


高尔顿钉板试验



$$\sum_{k=1}^{14} X_k \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2) \quad (n=14, \mu=0, \sigma^2=1)$$

高尔顿钉板试验



$$\sum_{k=1}^{14} X_k \overset{\text{近似}}{\sim} N(0, 14) \quad (n=14, \mu=0, \sigma^2=1)$$

例 某单位电话交换机接有500部电话,在所有通话中有96%次通话是在各分机内进行的.假定每部分机是否需要打外线是相互独立的,问需要配备多少条外线才能以95%的概率保证每个分机要用外线时不必等候？

解 记 η_{500} 表示500台分机中同时打外线电话的台数.

则有 $\eta_{500} \sim B(500,0.04)$. 设满足需要的外线条数为 N ,则应有

$$P\{\eta_{500} \le N\} = P\{\frac{\eta_{500}-500\times0.04}{\sqrt{500\times0.04\times0.96}} \le \frac{N-500\times0.04}{\sqrt{500\times0.04\times0.96}}\}$$

$$\approx \Phi(\frac{N-20}{\sqrt{19.2}}) \ge 95\%$$

查正态分布表得 $\Phi(1.65)=0.95$,故有

$$\frac{N-20}{\sqrt{19.2}} \ge 1.65 \quad \Rightarrow \quad N \ge 27.23$$

故至少应配备28条外线才能满足要求.

本讲结束 谢谢大家