5—2 中心极限定理

中心极限定理的客观背景

在实际问题中许多随机变量是由相互独立随机因素的综合（或和)影响所形成的.

例如：炮弹射击的落点与目标的偏差，就受着许多随机因素（如瞄准，空气阻力，炮弹或炮身结构等）综合影响的.每个随机因素的对弹着点（随机变量和）所起的作用都是很小的.那么弹着点服从怎样分布哪 ？

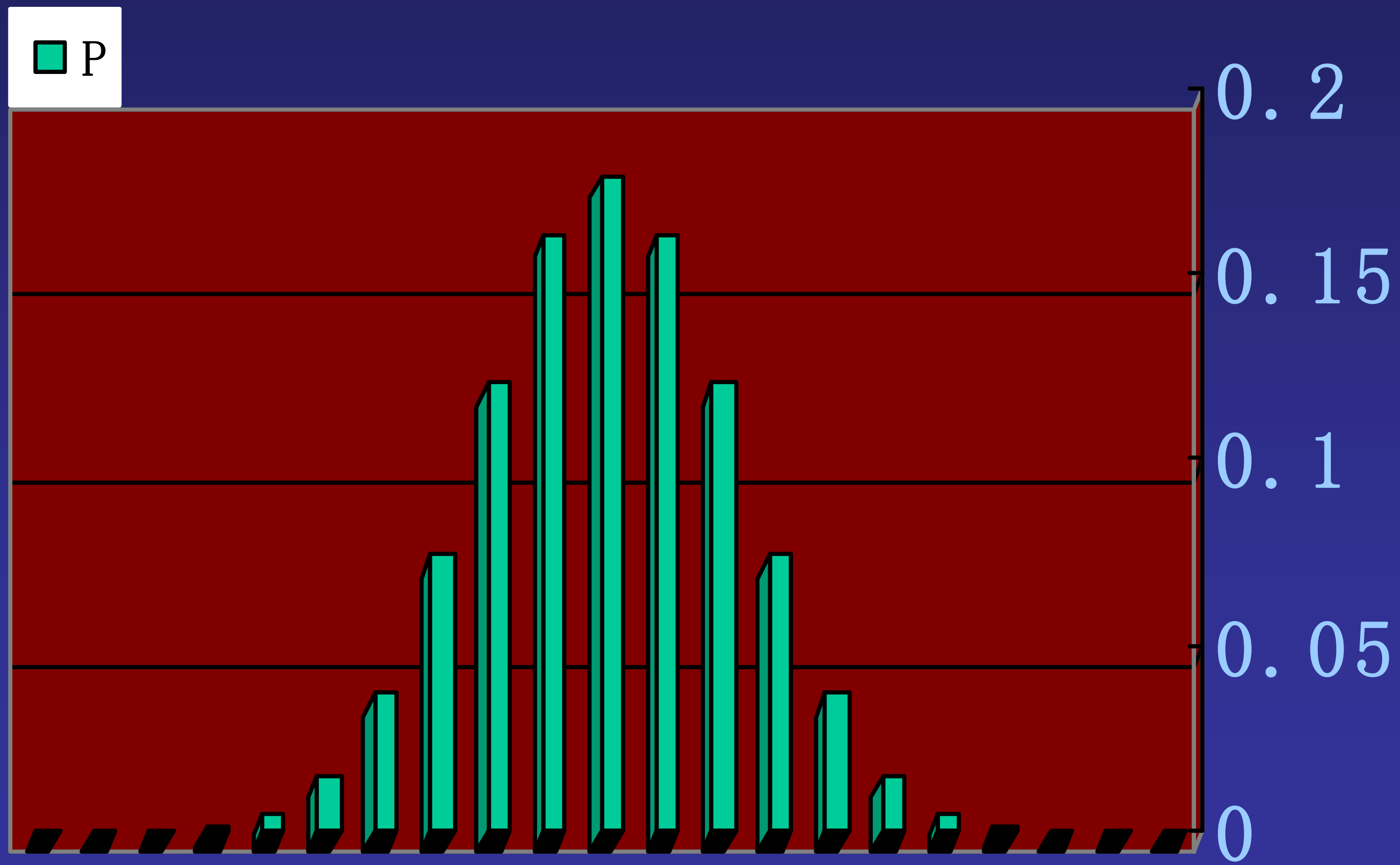
自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见.

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.

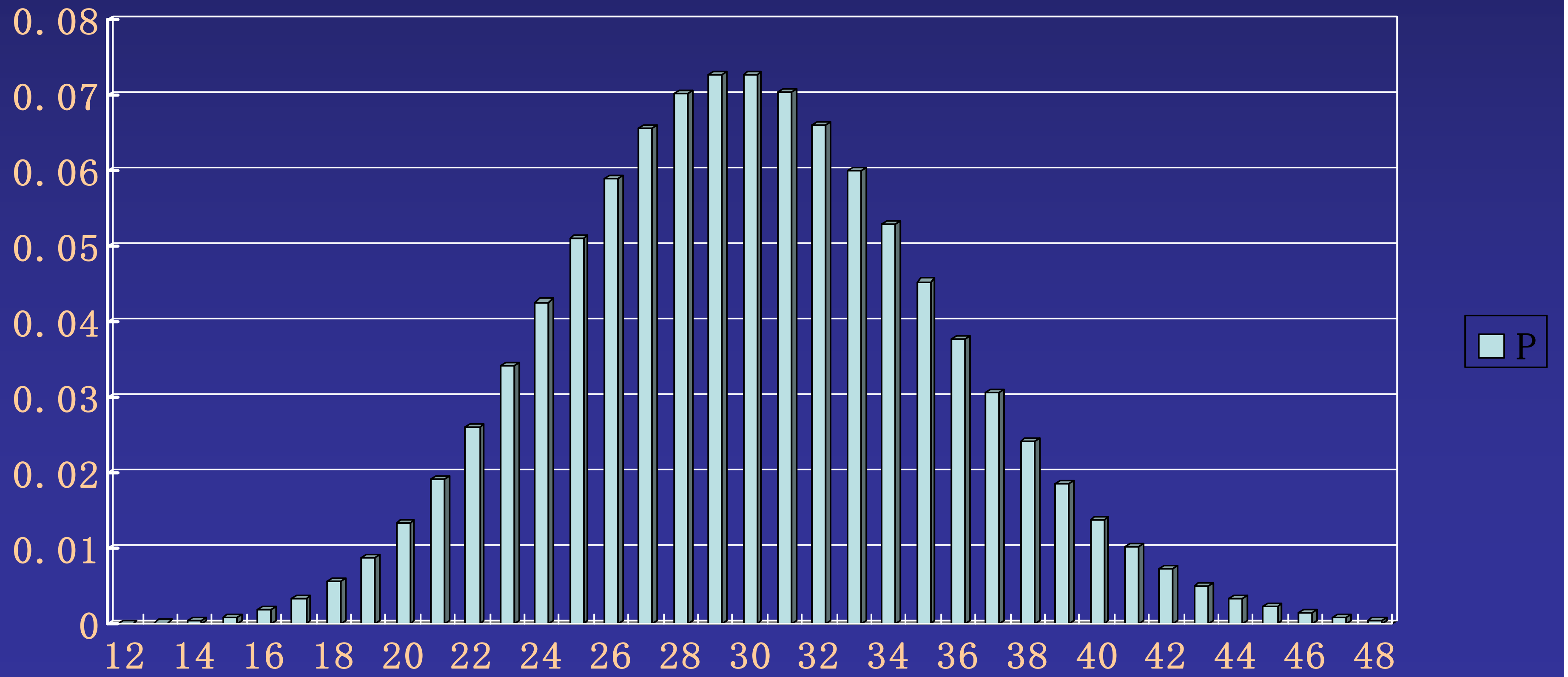
现在我们就来研究独立随机变量之和所特有的规律性问题.

当n无限增大时，这个和的极限分布是什么呢？

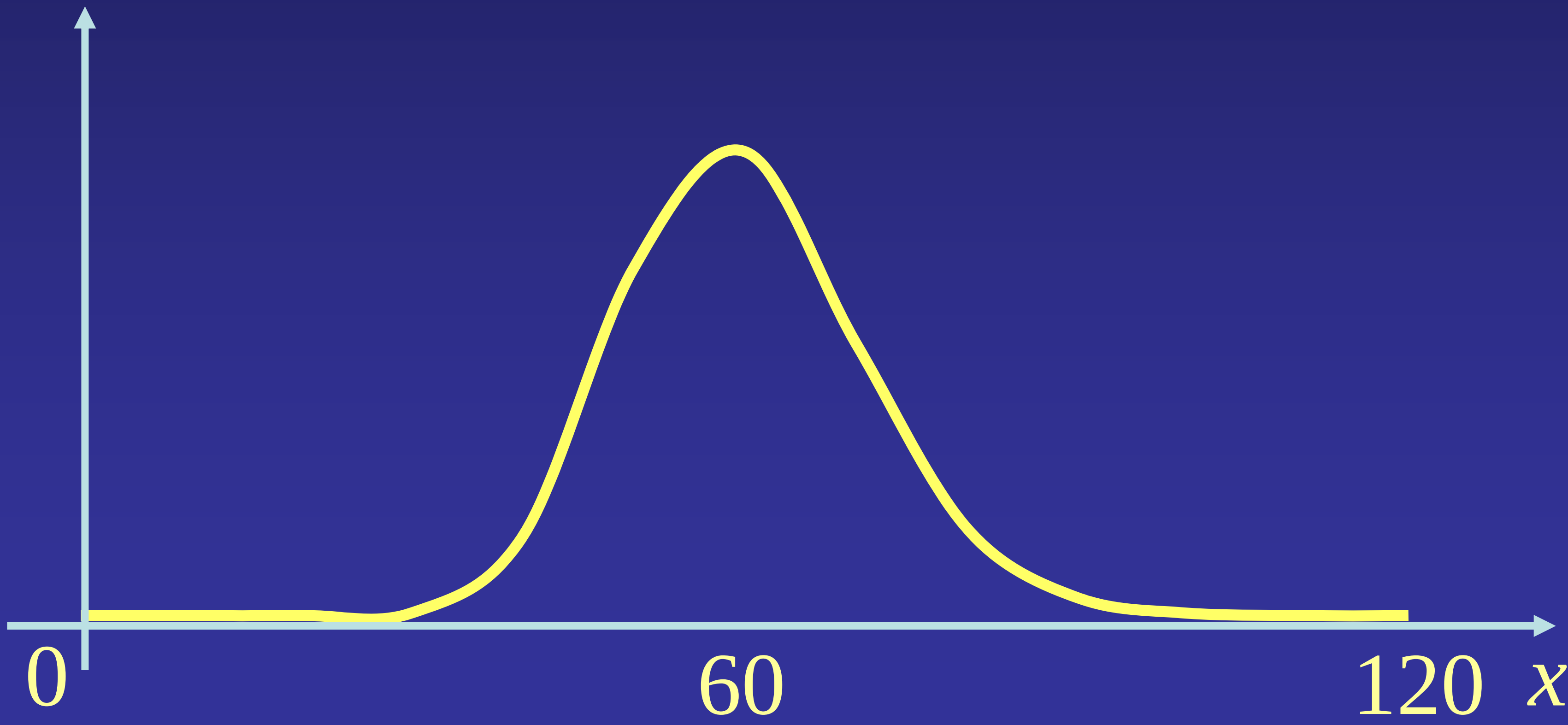
二项分布的随机变量可看作许多相互独立的**0-1**分布的随机变量之和, 下面是当$\xi \sim B(20,0.5)$时, ξ的概率分布图



泊松分布相当于二项分布中*p*很小*n*很大的分布, 因此, 参数$\lambda=np$当很大时也相当于*n*特别大, 这个时候泊松分布也近似服从正态分布, 下面是$\lambda=30$时的泊松概率分布图.



在χ^2(n)分布中, 如果自由度n很大, 也可以认为是多个自由度为1的相互独立的χ^2(1)分布的随机变量的和, 因此也近似服从正态分布. 下面是χ^2(60)的概率密度曲线.



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$X_k\ (k=1,\cdots n)$的和$\sum_{k=1}^{n} X_k$

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

讨论Y_n的极限分布是否为标准正态分布

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.

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：$E(X_k) = \mu, D(X_k) = \sigma^2 \neq 0$ $(k = 1, 2, \cdots)$，则对于任意x，有

$$\lim_{n\to\infty} P\left\{\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

令$Y_n = \dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} = \dfrac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\sigma/\sqrt{n}} = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$,

则$\lim\limits_{n\to\infty} P\{Y_n \leq x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$.

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

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

2、独立同分布中心极限定理的另一种形式可写为

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

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.

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：

$$E(X_k) = \mu_k, \qquad D(X_k) = \sigma_k^2 \neq 0, (k = 1, 2, \cdots)$$

记 $B_n^2 = \sum_{k=1}^{n} \sigma_k^2$.　　若存在正数$\delta$，使得当$n \to \infty$时，

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

则随机变量$Z_n = \dfrac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} = \dfrac{\sum_{k=1}^{n}(X_k - \mu_k)}{B_n}$

的分布函数$F_n(x)$对于任意x，满足

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

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

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

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.

定理6(棣莫弗－拉普拉斯（De Moivre-Laplace定理）

设随机变量 $Y_n \sim B(n, p)$, 其中 $0 < p < 1, n = 1, 2, \ldots$ ，则对任意*x*，有

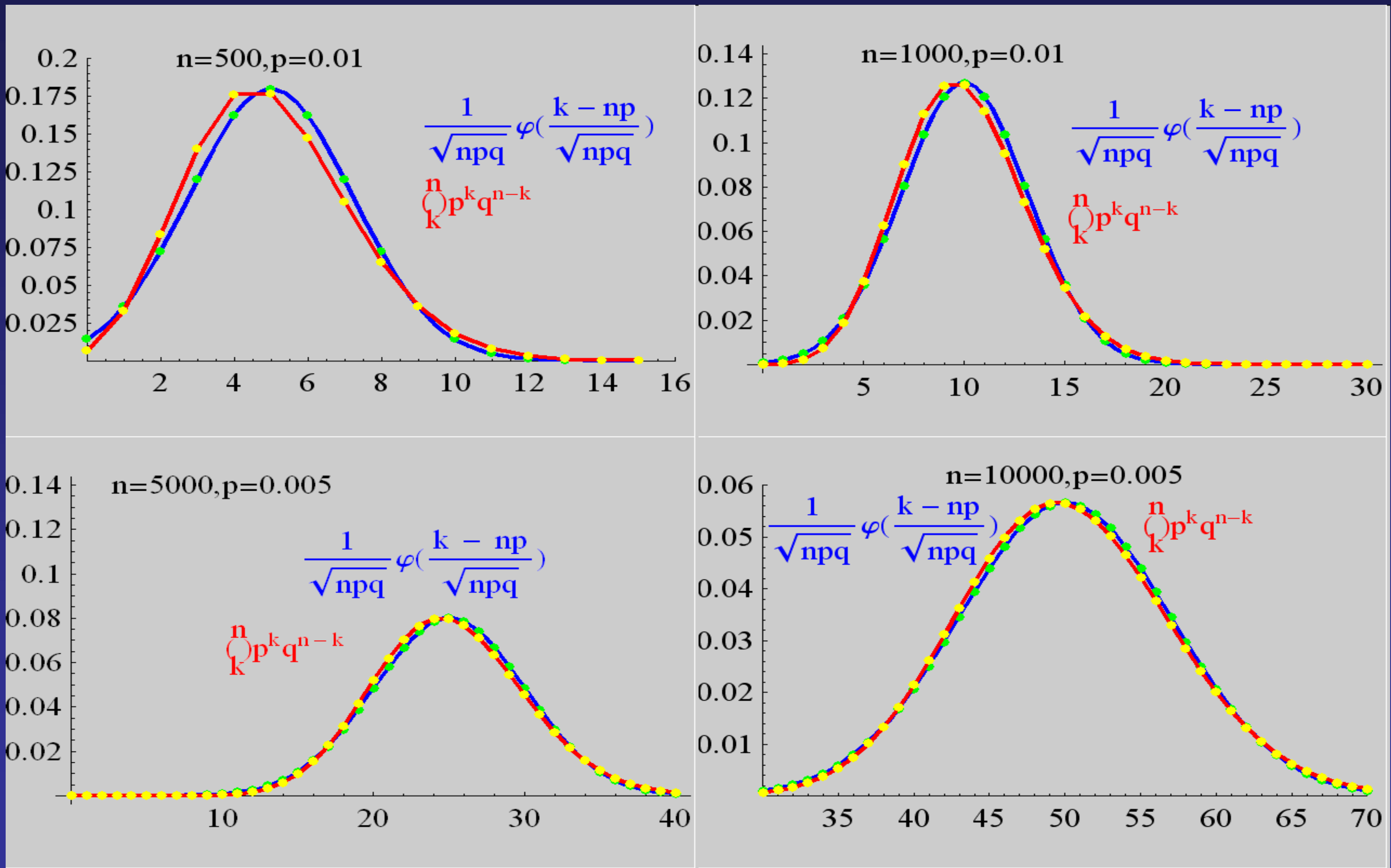
$$\lim_{n\to\infty} P\{\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad = \Phi(x)$$

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 Y_n 的分布近似正态分布 ***N*(*np*,*np*(1-*p*))**.

即

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

下面的图形表明:正态分布是二项分布的逼近.



例： 加法器同时收到20个噪声电压$V_k(k=1,2,\cdots n)$，设它们是相互独立的随机变量，且都在区间(0, 10)上服从均匀分布.记$V=\sum_{k=1}^{n}V_k$，求$P\{V>105\}$的近似值.

例：设某单位电话总机有**2000**个分机，每个分机有**5%**的时间使用外线通话，各分机使用外线相互独立，问总机至少要设多少条外线才能保证每个分机使用外线时不占线的概率大于**90%**。

例　一船舶在某海区航行, 已知每遭受一次波浪的冲击, 纵摇角大于3º的概率为p=1/3, 若船舶遭受了90000次波浪冲击, 问其中有29500~30500次纵摇角大于3º的概率是多少?

例　对于一个学生而言, 来参加家长会的家长人数是一个随机变量, 设一个学生无家长, 1名家长, 2名家长来参加会议的概率分别为0.05, 0.8, 0.15. 若学校共有400名学生, 设各学生参加会议的家长数相互独立, 且服从同一分布. (1)求参加会议的家长数X超过450的概率; (2) 求有一名家长来参加会议的学生数不多于340的概率.