第三章　概率和理论分布

- 3.1　事件与概率
- 3.2　二项分布
- 3.3　多项分布
- 3.4 泊松分布
- 3.5　正态分布
- 3.6 二项分布的正态近似

3.1 事件与概率

3.1.1 事件和概率的定义

事件（event）：自然界中一种事物常存在几种可能出现的情况，每一种可能出现的情况称为事件。

概率（probability）：每一个事件出现的可能性。

必然事件：在一组条件下，必然要发生的事件，其概率 P=1。

不可能事件：在一组条件下，不可能发生的事件，其概率 P=0。

随机事件：某特定事件只是可能发生的几种事件中的一种，其概率 $0 \le P \le 1$。

3.1 事件与概率

3.1.2 事件的相互关系

（1）和事件

事件 A和 B至少有一个发生而构成的新事件称为事件 A和 B的和事件，记为 $A+B$，读作“或 A发生，或 B发生”。

（2）积事件

事件 A和 B同时发生所构成的新事件，称为事件 A和 B的积事件，记作 AB，读作“ A和 B同时发生或相继发生”。

3.1 事件与概率

3.1.2 事件的相互关系

（3）互斥事件

事件 A和 B不可能同时发生，即 AB为不可能事件，称事件 A和 B互斥。

（4）对立事件

事件 A和 B不可能同时发生，但必发生其一。即 $A+B$为必然事件，AB为不可能事件，则称事件 B为事件 A的对立事件，并记 $B=\overline{A}$。

3.1 事件与概率

3.1.2 事件的相互关系

（5）完全事件系

若事件 A_1、A_2、….、A_n两两互斥，而且每次必发生其一，则称 n个事件构成完全事件系。

3.1 事件与概率

3.1.3 计算事件发生概率的法则

（1）加法定理

互斥事件 A和 B的和事件的概率等于事件 A和事件 B概率之和，

即$P(A+B) = P(A) + P(B)$ 。

推论1：完全事件系的概率为“1”

推论2：对立事件 $\bar{A}$ 的概率$P(\bar{A})=1- P(A)$

3.1 事件与概率

3.1.3 计算事件发生概率的法则

例题 已知一颗6面骰子（6面点数分别是1 2 3 4 5 6），

求（1）掷得骰子点数大于等于5的概率（2）骰子点数小于 5的概率

（1） $P(A) = P(y=5) = \frac{1}{6}$　　$P(B) = P(y=6) = \frac{1}{6}$

$$P(A+B) = P(A) + P(B) = \frac{1}{3}$$

（2） $P(C) = P(y\geq5) = \frac{1}{3}$　　$P(\overline{C})=1- P(C)=\frac{2}{3}$

3.1 事件与概率

3.1.3 计算事件发生概率的法则

（2）乘法定理

事件 A 和事件 B 同时发生的概率等于事件 A 的概率乘以在事件 A 的发生下事件 B 的概率，或者等于事件 B 的概率乘以在事件 B 的发生下事件 A 的概率。

即 $P(AB)=P(A)P(B/A)$ 或 $P(AB)=P(B)P(A/B)$

3.1 事件与概率

3.1.3 计算事件发生概率的法则

例题 在数字1、2、3、4、5、6、7、8、9中，试求抽得一个既是偶数又能被3整除的数的概率。

设：抽得偶数为事件 A，抽得被3整除的数为事件 B,

则 $P(A)=\frac{4}{9}$ ， $P(B)=\frac{3}{9}$

$$P(AB)=P(A)*P(B/A)=\frac{4}{9}*\frac{1}{4}=\frac{1}{9}$$

或 $P(AB)=P(B)*P(A/B)=\frac{3}{9}*\frac{1}{3}=\frac{1}{9}$

3.1 事件与概率

3.1.3 计算事件发生概率的法则

推论1：若事件 A、B相互独立，那么 A、B同时发生的概率为各自概率之积，即$P(AB) = P(A) * P(B)$

推论2：若事件 A_1、A_2、....、A_n皆彼此独立，那么

$$P(A_1 * A_2 *....A_n)= P(A_1) * P(A_2) *.... P(A_n)$$

3.1 事件与概率

3.1.3 计算事件发生概率的法则

例题 同时掷两个6面骰子，求掷得都是6点的概率。

设：掷得第一个为6点事件为 A，掷得第二个为6点事件为 B

则 $P(A)=\frac{1}{6}$　　$P(B)=\frac{1}{6}$

$$P(AB)=P(A)*P(B)=\frac{1}{6}*\frac{1}{6}=\frac{1}{36}$$

第三章 概率和理论分布

扬州大学 YANGZHOU UNIVERSITY

- 3.1 事件与概率
- **3.2 二项分布**
- 3.3 多项分布
- 3.4 泊松分布
- 3.5 正态分布
- 3.6 二项分布的正态近似

3.2 二项分布

3.2.1 二项总体和二项分布的意义

二项总体（binomial population）： 由非此即彼事件构成的总体称为二项总体。为了便于研究，通常给“此”事件以数值“1”，具概率 p；给“彼”事件以数值“0”，具概率 q，其中$p+q=1$。因而二项总体又称0、1总体。

二项分布（binomial distribution）： 二项总体中有 n+1种变量及其各自概率组成的分布称为二项分布。二项分布是间断性变数的一种最重要的理论分布。

3.2 二项分布

3.2.2 计算二项分布概率的方法

例题1：假设大袋子中分别放100粒黑豆，100粒黄豆，采取复置抽样两次，分别求抽得两黑豆、一黑豆一黄豆、两黄豆的概率

设：抽得黑豆概率为此事件“1”即 p=0.5

抽得黄豆概率为彼事件“0”即 q=0.5

则 抽得两黑豆可能存在情况（1，1）

$P(Y=2)= p*p= p^2=0.5^2=0.25$

3.2 二项分布

3.2.2 计算二项分布概率的方法

抽得一黑豆一黄豆可能存在情况（1，0）、（0，1）

$$P(Y=1)=p*q+p*q=2pq=2*0.5*0.5=0.5$$

抽得两黄豆可能存在的情况（0，0）

$$P(Y=0)=q*q=q^2=0.5*0.5=0.25$$

上述各项概率正好是二项式$(p+q)^2$在$p=q=0.5$时展开后的各项：

$$(p+q)^2=p^2+2pq+q^2=1$$

3.2 二项分布

3.2.2 计算二项分布概率的方法

例题2： 根据遗传学原理，豌豆的红花纯合基因型和白花纯合基因型杂交后，在F_2代红花植株的出现概率 p=0.75，白花植株的出现概率 q=0.25。若每次观察 n=4株，问得红花为4株、3株、2株、1株和0株的概率

设：红花事件1，白花事件0

根据概率乘法定理，红花4株可能事件（1，1，1，1）

$P(Y=4)= p*p*p*p= p^4=0.75^4=0.3164$

3.2 二项分布

3.2.2 计算二项分布概率的方法

红花3株（3红1白）由互斥事件（1，1，1，0）、（1，1，0，1）、（1，0，1，1）、（0，1，1，1）

$P(Y=3)=4p^3*q=4*0.75^3*0.25=0.4219$

红花2株（2红2白）由互斥事件（1，1，0，0）、（1，0，1，0）、（1，0，0，1）、（0，1，1，0）、（0，1，0，1）、（0，0，1，1）

$P(Y=2)=6p^2*q^2=6*0.75^2*0.25^2=0.2109$

3.2 二项分布

3.2.2 计算二项分布概率的方法

红花1株（1红3白）由互斥事件（1，0，0，0）、（0，1，0，0）、（0，0，1，0）、（0，0，0，1）

$P(Y=1)=4p*q^3=4*0.75*0.25^3=0.0469$

红花0株（4白）仅有事件（0，0，0，0）

$P(Y=0)= q*q*q*q= q^4=0.25^4=0.0039$

3.2 二项分布

3.2.2 计算二项分布概率的方法

上述各项概率也正好是$(p+q)^4$在 p=0.75、q=0.25时展开后的各项：

$$(p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

以上二项式展开后各项的系数，正是在 n个事物中抽得k个事物（$0 \le k \le n$）

的组合数C_n^k $\quad C_n^k = \frac{n!}{k!(n-k)!}$

由此我们可以得到计算二项分布中任何一项概率的通式，

$$P(Y=k) = C_n^k \, p^k q^{n-k}$$

3.2 二项分布

3.2.2 计算二项分布概率的方法

例题3：已知高杆香稻和矮杆非香稻杂交后，在F_2代出现目标植株矮杆香稻的概率p=0.0625，出现非矮杆香稻的概率q=0.9375，试求（1）若F_2代种植30株，则至少获得1株矮杆香稻的概率是多少？（2）如希望有0.95的概率获得至少1株矮杆香稻，则F_2应种植多少株？

（1）根据二项分布概率计算公式，得0株矮秆香稻的概率为：

$$P\{Y=0\}=C_{30}^{0}p^{0}q^{30}=q^{30}=0.9375^{30}=0.1443$$

由于$Y=0$和$Y=1$，2，3，…30是互斥事件，且构成完全事件系，

因此至少获得1株矮秆香稻的概率为：

$$P\{Y\geq 1\}=1-P\{Y=0\}=1-0.1443=0.8557$$

3.2 二项分布

3.2.2 计算二项分布概率的方法

（2）F_2应种植的株数 n 需满足 $P\{Y=0\}=1-0.95=0.05$，即满足 $q^n=0.05$ 。所以：

$$n=\frac{\lg 0.05}{\lg q}=\frac{-1.3010}{-0.0280}=46.5\approx 47 \quad \text{（株）}$$

这就是说，如希望有95%的把握至少获得1株矮秆香稻，F_2需种植47株。

3.2 二项分布

3.2.3 二项分布的平均数和标准差

二项分布的总体平均数

$$\mu=\sum_i Y_i P\{Y_i\}=np\sum_0^{n-1}C_{n-1}^k p^k q^{n-k-1}=np(q+p)^{n-1}=np$$

二项分布总体的标准差

$$\sigma=\sqrt{\sum_i (Y_i-\mu)^2 P\{Y_i\}}=\sqrt{\sum_i Y_i^2 P\{Y_i\}-\left[\sum_i Y_i P\{Y_i\}\right]^2}$$

$$=\sqrt{n(n-1)p^2+np-(np)^2}=\sqrt{np(1-p)}\quad=\sqrt{npq}$$

第三章 概率和理论分布

3.4 泊松分布

3.4.1 泊松分布

二项分布当 $n\to\infty$ 而 $p\to 0$ 时，可导得一个分布为：

$$e^{-m}e^{m}=e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\frac{m^4}{4!}+\ldots\right)$$

该分布称为**泊松分布**（poisson distribution），其中 $m=np$，$e=2.71828\ldots$；

其在y=k时的概率为

$$P=(y=k)=e^{-m}\frac{m^k}{k!}\quad(k=0,1,2\ldots)$$

3.4 泊松分布

3.4.1 泊松分布

且当 $m<1$时，$p（y=0）=max$；当$1<m<2$时，$p（y=1）=max$；

当$2<m<3$时，$p（y=2）=max$…..

泊松分布的平均数和标准差

$$\mu=np=m \qquad \sigma=\sqrt{m}=\sqrt{np}$$

3.4 泊松分布

3.4.1 泊松分布

泊松分布主要用途

① 农业生产上本身存在很多小概率事件，直接利用泊松分布解决

② 当二项分布的 $p<0.1$或 $np<5$,可用泊松分布近似

第三章 概率和理论分布

3.4 泊松分布

3.4.1 泊松分布

二项分布当 $n\to\infty$而 $p\to 0$时，可导得一个分布为：

$$e^{-m}e^{m}=e^{-m}\left(1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\frac{m^4}{4!}+\ldots\right)$$

该分布称为泊松分布(poisson distribution)，其中 $m= np$，e=2.71828…；

其在 $y=k$ 时的概率为

$$P=\{y=k\}=e^{-m}\frac{m^k}{k!}\quad(k=0，1，2\ldots)$$

3.4 泊松分布

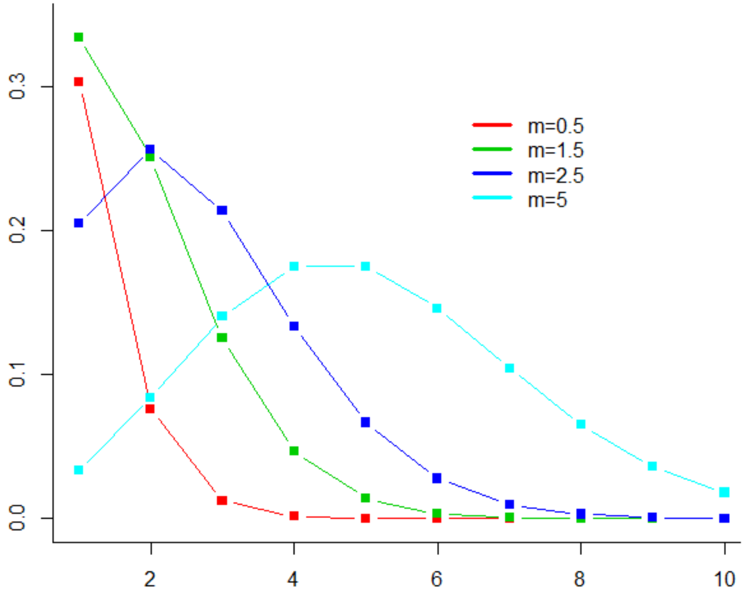
3.4.1 泊松分布

泊松分布仅具有参数$m = np$。因其是二项分布在 $n\to\infty$而 $p\to 0$时的极限分布，故根据二项分布总体平均数和标准差的计算公式可以导出泊松分布的平均数和标准差为

$$\mu = np = m \qquad \sigma = \sqrt{np} = \sqrt{m}$$

泊松分布的形状取决于m。

- 当 $m < 1$时，$p\{y = 0\}=\max$；
- 当$1< m <2$时，$p\{y = 1\}=\max$；
- 当$2< m <3$时，$p\{y = 2\}=\max$…..



不同m值的泊松分布

3.4 泊松分布

3.4.1 泊松分布

泊松分布主要用途

① 一是在农业上有好多小概率事件，其发生概率p往往<0.1，甚至<0.01。对于这些小概率事件，往往都可用泊松分布描述其概率分布，从而作出需要的频率预期。

② 由于泊松分布是描述小概率事件，当二项分布的 $p<0.1$或 $np<5$，可用泊松分布近似

3.4 泊松分布

3.4.1 泊松分布

例题4： **【例题3 】资料矮秆香稻的出现概率p=0.0625，可看成是小概率事件。试以泊松分布解决【例题3 】所提的两个问题。**

（1）求得 $m = np = 30 \times 0.0625 = 1.875$， 故根据泊松分布概率计算公式有：

$$P\{y=0\} = e^{-1.875} = 0.1534$$

所以，$P\{y \geq 1\} = 1 - 0.1534 = 0.8466$

（2）F_2种植的株数n应满足：

$$P(y=0) = e^{-m} = e^{-np} = 0.05$$

故：$n = \dfrac{-\lg 0.05}{p \lg e} = \dfrac{1.3010}{0.0625 \times 0.43429} = 47.9 \approx 48$（株）

第三章 概率和理论分布

3.5 正态分布

3.5.1 正态分布及其性质

正态分布是连续性变量的一种最重要而常见的理论分布。

① 客观世界中许多现象的数据是服从正态分布的

② 在适当条件下，可以用做二项分布等间断性变数分布及其他连续性变数分布的近似分布

③ 虽然有些总体不服从正态分布，但从总体中抽出的样本平均数及其他统计数是服从正态分布的

3.5 正态分布

3.5.1 正态分布及其性质

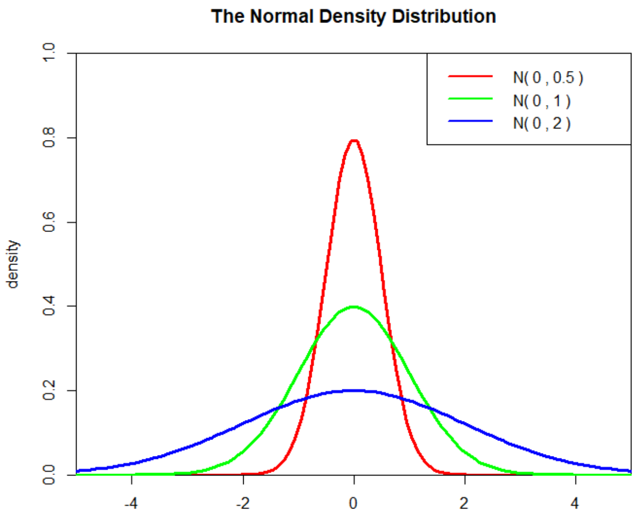
设有 n 个独立的随机因素，每个因素都可能使观测值 y 产生一个偏离总体平均数 μ 的微小误差，这个误差可正可负，且取正取负的概率相等。这样，误差 ε_i=（y_i-μ）的分布相当于 $p = q = 0.5$ 时的 $(p + q)^n$ 展开。如果 $n\to\infty$，则可导得一个表示各 y_i 出现的概率密度方程为

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

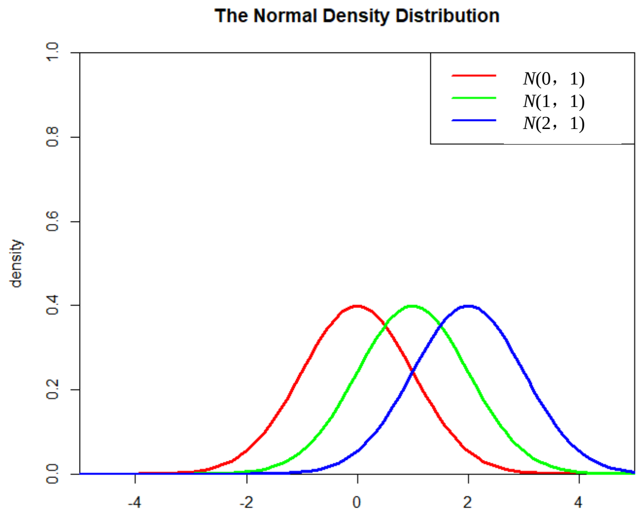
3.5 正态分布

3.5.1 正态分布及其性质

上式是二项分布的极限，称为正态误差曲线方程，或简称正态分布方程。$f(y)$ 为正态分布下y_i的概率密度函数（Probability density function），即分布的高度，μ 和 σ 分别为总体的平均数和标准差，记为 $N(\mu, \sigma^2)$，读作“具平均数为 μ、方差为 σ^2 的正态分布”。



平均数相同标准差不同的正态分布



标准差相同平均数不同的正态分布

μ决定分布中心，

σ^2决定分布的变异度，

且 σ^2越大，变异度越大。

3.5 正态分布

3.5.1 正态分布及其性质

① y的取值区间为 $-\infty \sim +\infty$，$y=\mu$ 时，$f(y)$ 值最大，曲线最高；且在 $y-\mu$ 的绝对值相等时，$f(y)$ 值也相等，故正态分布是以 μ 为中心左右对称的。在正态分布下，平均数、中位数和众数，三者等值。

② $\frac{y-\mu}{\sigma}$ 的绝对值愈大，$f(y)$ 值愈小，但 $f(y)$ 值恒非负。故正态曲线以横轴为渐近线（与横轴不相交）。

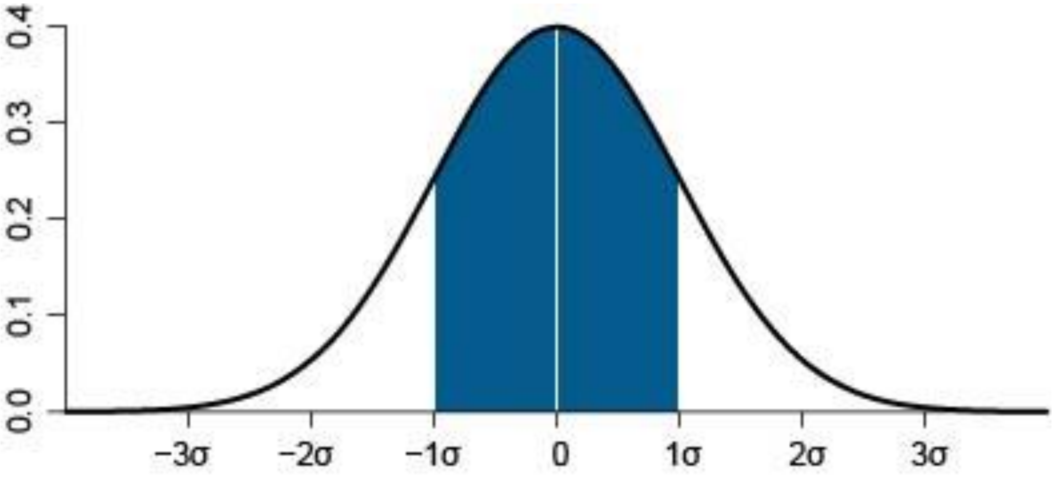
$$f(y)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

3.5 正态分布

3.5.1 正态分布及其性质

③ y 在 $-\infty \sim +\infty$ 间可处处取值构成了 y 取值的完全事件系，故正态曲线与横轴所围成的全部面积必等于1，而变量 y 出现在任两个定值 y_1 到 y_2（$y_1 \neq y_2$）之间的概率即为这两个定值间的面积占总面积的成数。

④ 对 $f(y)$ 取一阶和二阶导数可得：在 $y < \mu-\sigma$ 和 $y > \mu+\sigma$ 上，曲线凹向上；在 $\mu-\sigma < y < \mu+\sigma$上，曲线凸向上；在$y = \mu \pm 1\sigma$上，曲线各有一个拐点。



3.5 正态分布

3.5.2 标准正态分布及其累积函数

正态分布标准化，假设 $y \sim N(\mu, \sigma^2)$令

$$u = \frac{y-\mu}{\sigma}$$

$$u_1 = \frac{y_1-\mu}{\sigma}, \quad u_2 = \frac{y_2-\mu}{\sigma}, \quad u_3 = \frac{y_3-\mu}{\sigma} \quad \cdots$$

则 $u \sim N(0, 1)$，并具有概率密度函数

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2}$$

称为标准正态分布或 u 分布，而 u 称为标准正态离差

3.5 正态分布

3.5.2 标准正态分布及其累积函数

由于 u 分布仅是若干正态分布中的一个特例，故正态分布的基本性质 u 分布同样具备，但由于 u 分布的 $\mu = 0$和 $\sigma^2 = 1$，故 u 分布的性质可以相应总结如下：

① $u = 0$ 时，$\varphi(u)$值最大，曲线最高。故是以 0 为中心而左右对称分布

② u 分布仍以横轴为渐近线，且变量 u 出现在任两个定值u_1到u_2（$u_1 \neq u_2$)之间的概率为这两个定值间的面积占总面积的成数

③ u 分布的两个拐点分别在$u = -1$ 和 $u = 1$上

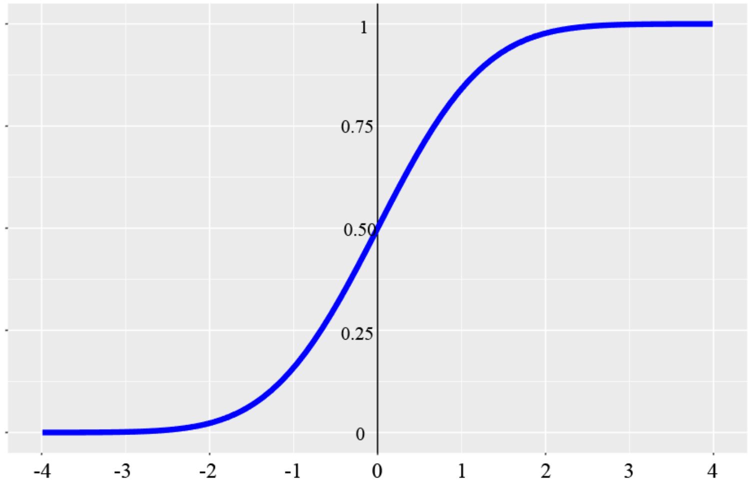
3.5 正态分布

3.5.2 标准正态分布及其累积函数

为了便于概率计算，引出了u分布的累积函数（cumulative function）

$F(u)$：变量u小于某一定值u_i的概率，即算从-∞到u_i的定积分:

$$F(u_i)=P\,(u\le u_i)=\int_{-\infty}^{u_i}\varphi(u)\,\mathrm{d}u$$



u分布的累积函数分布

横坐标为正态标准离差u，纵坐标为$F(u)$

3.5 正态分布

3.5.2 标准正态分布及其累积函数

由于u分布累积函数应用的广泛性，前人已算好实际需要的各个$F(u)$ 值，见附表1。

附表1：标准正态分布累积函数表

$F(u_i) = P(u \le u_i)$

$= \int_{-\infty}^{u_i} \varphi(u)\mathrm{d}u$

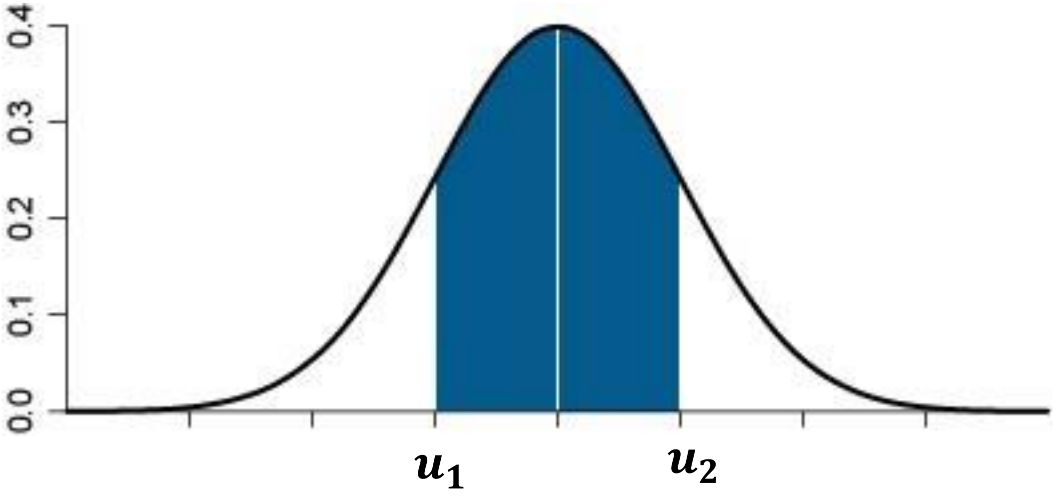
u_i	0.00	−0.01	−0.02	−0.03	−0.04	−0.05	−0.06	−0.07	−0.08	−0.09	u_i
−3.0	0.0^{2}13 50	0.0^{2}13 06	0.0^{2}12 64	0.0^{2}12 23	0.0^{2}11 83	0.0^{2}11 44	0.0^{2}11 07	0.0^{2}10 70	0.0^{2}10 35	0.0^{2}10 01	−3.0
−2.9	0.0^{2}18 66	0.0^{2}18 07	0.0^{2}17 50	0.0^{2}16 95	0.0^{2}16 41	0.0^{2}15 89	0.0^{2}15 38	0.0^{2}14 89	0.0^{2}14 41	0.0^{2}13 95	−2.9
−2.8	0.0^{2}25 55	0.0^{2}24 77	0.0^{2}24 01	0.0^{2}23 27	0.0^{2}22 56	0.0^{2}21 86	0.0^{2}21 18	0.0^{2}20 52	0.0^{2}19 88	0.0^{2}19 26	−2.8
−2.7	0.0^{2}34 67	0.0^{2}33 64	0.0^{2}32 64	0.0^{2}31 67	0.0^{2}30 72	0.0^{2}29 80	0.0^{2}28 90	0.0^{2}28 03	0.0^{2}27 18	0.0^{2}26 35	−2.7
−2.6	0.0^{2}46 61	0.0^{2}45 27	0.0^{2}43 96	0.0^{2}42 69	0.0^{2}41 45	0.0^{2}40 25	0.0^{2}39 07	0.0^{2}37 93	0.0^{2}36 81	0.0^{2}35 73	−2.6
−2.5	0.0^{2}62 10	0.0^{2}60 37	0.0^{2}58 68	0.0^{2}57 03	0.0^{2}55 43	0.0^{2}53 86	0.0^{2}52 34	0.0^{2}50 85	0.0^{2}49 40	0.0^{2}47 99	−2.5
−2.4	0.0^{2}81 98	0.0^{2}79 76	0.0^{2}77 60	0.0^{2}75 49	0.0^{2}73 44	0.0^{2}71 43	0.0^{2}69 47	0.0^{2}67 56	0.0^{2}65 69	0.0^{2}638 7	−2.4
−2.3	0.010 72	0.010 44	0.010 17	0.0^{2}990 3	0.0^{2}964 2	0.0^{2}93 87	0.0^{2}91 37	0.0^{2}88 94	0.0^{2}86 56	0.0^{2}84 24	−2.3
−2.2	0.013 90	0.013 55	0.013 21	0.012 87	0.012 55	0.012 22	0.011 91	0.011 60	0.011 30	0.011 01	−2.2
−2.1	0.017 86	0.017 43	0.017 00	0.016 59	0.016 18	0.015 78	0.015 39	0.015 00	0.014 63	0.014 26	−2.1
−2.0	0.022 75	0.022 22	0.021 69	0.021 18	0.020 68	0.020 18	0.019 70	0.019 23	0.018 76	0.018 31	−2.0
−1.9	0.028 72	0.028 07	0.027 43	0.026 80	0.026 19	0.025 59	0.025 00	0.024 42	0.023 85	0.023 30	−1.9
−1.8	0.035 93	0.035 15	0.034 38	0.033 62	0.032 88	0.032 16	0.031 44	0.030 74	0.030 05	0.029 38	−1.8
−1.7	0.044 57	0.043 63	0.042 72	0.041 82	0.040 93	0.040 06	0.039 20	0.038 36	0.037 54	0.036 73	−1.7
−1.6	0.054 80	0.053 70	0.052 62	0.051 55	0.050 50	0.049 47	0.048 46	0.047 46	0.046 48	0.045 51	−1.6
−1.5	0.066 81	0.065 52	0.064 26	0.063 01	0.061 78	0.060 57	0.059 38	0.058 21	0.057 05	0.055 92	−1.5
−1.4	0.080 76	0.079 27	0.077 80	0.076 36	0.074 93	0.073 53	0.072 15	0.070 78	0.069 44	0.068 11	−1.4
−1.3	0.096 80	0.095 10	0.093 42	0.091 76	0.090 12	0.088 51	0.086 91	0.085 34	0.083 79	0.082 26	−1.3
−1.2	0.115 1	0.113 1	0.111 2	0.109 3	0.107 5	0.105 6	0.103 8	0.102 0	0.100 3	0.098 53	−1.2

3.5 正态分布

3.5.3 利用正态分布计算概率的方法

- 设标准正态离差u_1、u_2为实数，且 $u_1 < u_2$，现需要计算 u 在 u_1 到 u_2 区间内的出现概率 $P(u_1 \leq u \leq u_2)$ 。

$$P(u_1 \leq u \leq u_2) = P(u \leq u_2) - P(u \leq u_1) = F(u_2) - F(u_1)$$



$$P(u \geq u_2) = 1 - P(u \leq u_2) = 1 - F(u_2)$$

3.5 正态分布

3.5.3 利用正态分布计算概率的方法

- 一般正态分布概率的计算方法：若 $y \sim N(\mu, \sigma^2)$，计算 y 在 y_1 到 y_2 区间内的概率，等价于计算 u 在相应的 u_1 到 u_2 区间内的概率，即

$$P(y_1 \le y \le y_2) \quad \Leftrightarrow \quad P(u_1 \le u \le u_2)$$

其中，

$$u_1 = \frac{y_1 - \mu}{\sigma} \qquad u_2 = \frac{y_2 - \mu}{\sigma}$$

值得注意的是，由于连续性变量分布任一点的概率为0，

因此$P(y_1 \le y \le y_2)$与$P(y_1 < y < y_2)$相等。

3.5 正态分布

3.5.3 利用正态分布计算概率的方法

例题1 计算标准正态分布，变量 u 在（-1，1）的概率

$$P(-1 \le u \le 1) = P(u \le 1) - P(u \le -1) = F(1) - F(-1)$$

查附表1 得 $F(-1) = 0.1587$ $F(1) = 0.8413$

$$P(-1 \le u \le 1) = F(1) - F(-1) = 0.6836$$

3.5 正态分布

3.5.3 利用正态分布计算概率的方法

例题2 计算变量 u 在(-∞，-1.96) 和 (1.96，+∞)这两个区间概率之和

$P(u < -1.96) = F(-1.96)$

$P(u > 1.96) = 1 - F(1.96)$

查附表1得 $F(-1.96) = 0.025$　　$F(1.96) = 0.975$

$P(u < -1.96) = F(-1.96) = 0.025$

$P(u > 1.96) = 1 - F(1.96) = 1 - 0.975 = 0.025$

即　$P(|u| > 1.96) = 0.05$

扬州大学 YANGZHOU UNIVERSITY

3.5.3 利用正态分布计算概率的方法

例题3 计算变量 u 在$(-\infty$，$-2.58)$ 和 $(2.58$，$+\infty)$这两个区间概率之和

$$P(u < -2.58) = F(-2.58)$$

$$P(u > 2.58) = 1 - F(2.58) = F(-2.58)$$

查附表1得 $F(-2.58) = 0.005$

即 $P(|u| > 2.58) = 0.01$

3.5 正态分布

3.5.3 利用正态分布计算概率的方法

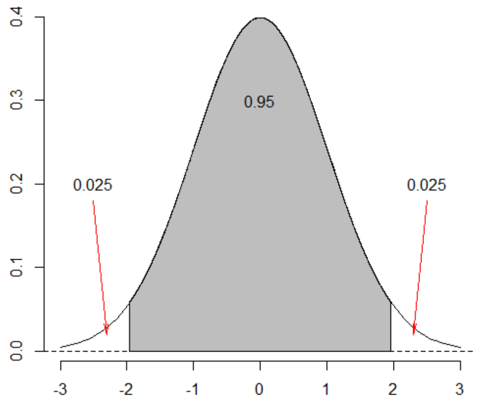
上述计算的几何意义： 虽然 u 分布的取值区间在（-∞， +∞），但实际上，

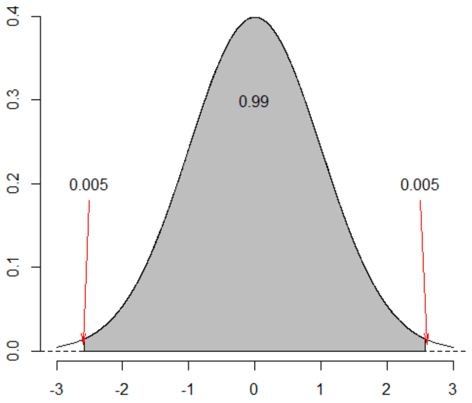
$|u| > 2.58$的概率只有0.01，$|u| > 1.96$的概率只有0.05。由于$y = \mu + u\sigma$，

所以也可以说，y 取值区间在（-∞， +∞），但实际上，

在 $\mu \pm 2.58\sigma$ 的概率只有0.01，在 $\mu \pm 1.96\sigma$的概率只有0.05。

1.96和2.58分别称为 u 分布的两尾临界值。该值与后面的假设测验有关，望能记住。





u 分布概率计算示意图

3.5 正态分布

3.5.3 利用正态分布计算概率的方法

例题4 假定随机变量 y 遵循正态分布，$y \sim N(20，16)$。

试计算（1）$P(y<16)$和（2）$P(y>28)$。

（1）首先将 y 转换为 u 值。

$$u=\frac{y-\mu}{\sigma}=\frac{16-20}{4}=-1$$

查附表1，当 $u=-1$时，$F(-1)=0.1587$，故$P(y<16)=0.1587$

（2） $$u=\frac{y-\mu}{\sigma}=\frac{28-20}{4}=2$$

查附表1，当$u=2$时，$F(2)=0.9773$，故$P(y<28)=0.9773$。

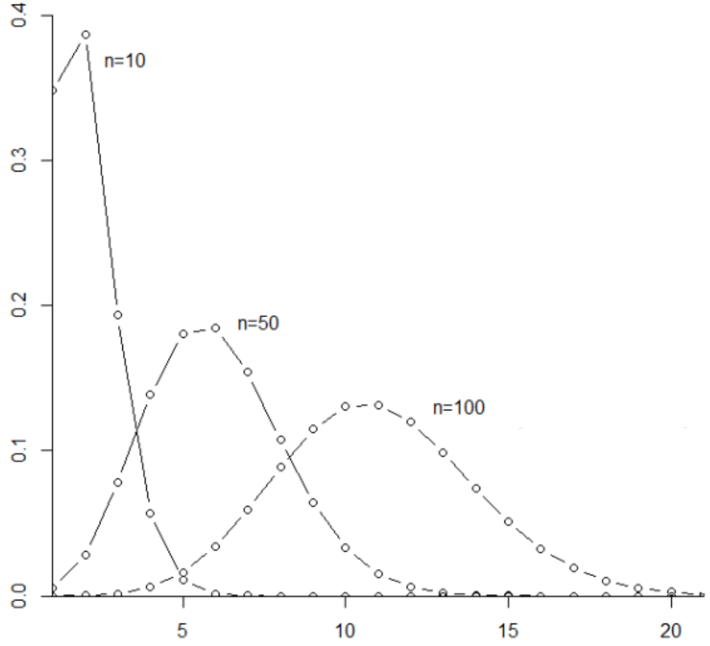
而$P(y>28)=1-P(y<28)=0.0227$

第三章 概率和理论分布

3.6 二项分布的正态近似

在二项分布中，给出 n 和 p，就能得到变量 y 的概率分布，故 n 和 p 是决定二项分布形状的两个参数。在 $p=q$ 时，不论 n 的大小，二项分布的多边形图必形成对称。如果 $p \neq q$ 时，当 n 小，二项分布将明显偏斜；当 n 增大到使 $np > 5$（$p \leq 0.5$时）或 $nq > 5$ （$p > 0.5$时），且p和q都不过大或过小（譬如> 0.1，< 0.9），则二项分布亦渐趋向对称，并近似正态分布。因此，具备上述条件时，可用正态近似法计算二项分布的概率。

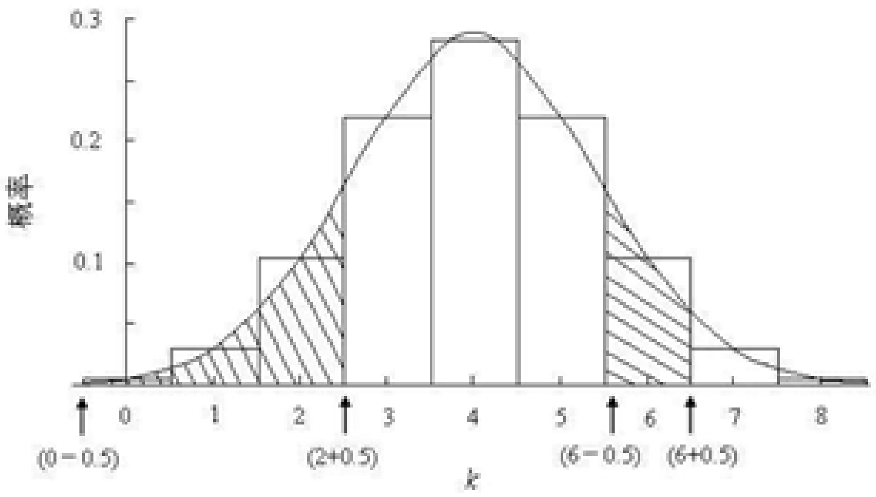
在 $p = 0.1$ 但 n 不同时的二项分布



3.6 二项分布的正态近似

扬州大学 YANGZHOU UNIVERSITY

但是正态分布是连续性分布，二项分布是间断性分布，它们表示概率的方式不一样。二项分布中 $y=k$ 的概率应为正态分布中 $y_1=(k-0.5)$ 到 $y_2=(k+0.5)$区间的概率；同样，二项分布中的 $P\{k_1 \le y \le k_2\}$ 相当于正态分布中的 $P\{(k_1-0.5)\le y <(k_2+0.5)\}$ 。



二项分布的正态近似图解

以正态概率近似二项概率，需应用“连续性矫正常数”0.5，其正态标准离差为：

$$u_c=\frac{(y\mp 0.5)-\mu}{\sigma}=\frac{(y-\mu)\mp 0.5}{\sigma}$$

$$u_c=\frac{(y-np)\mp 0.5}{\sqrt{npq}}$$

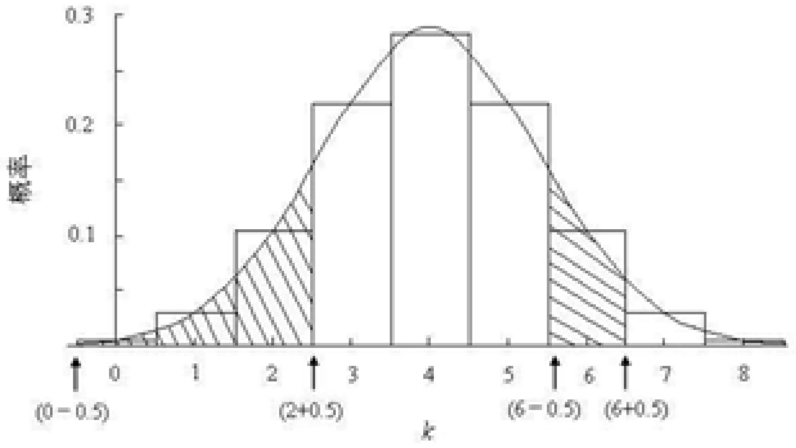
u_c叫做为适合连续性的矫正正态标准离差，简称矫正正态离差。

在 np和 nq皆 >30时，可以不作矫正，直接应用。

3.6 二项分布的正态近似

例题 一批棉花种子的发芽率 p=0.5，在每穴播8粒时，求：(1)每穴出苗为 $0 \le y \le 2$ 株的概率？(2)每穴出苗为6株的概率？

$$\mu = 8 \times 0.5 = 4 \qquad \sigma = \sqrt{8 \times 0.5 \times 0.5} = 1.414$$



二项分布的正态近似图解

（1）

$$u_{c1} = \frac{(0-4)-0.5}{1.414} = -3.18 \qquad u_{c2} = \frac{(2-4)+0.5}{1.414} = -1.06$$

查附表1得

$$F(-3.18) \approx 0.0007 \qquad F(-1.06) = 0.1446$$

$$P\{0 \le y \le 2\} = F(-1.06) - F(-3.18) = 0.1439$$

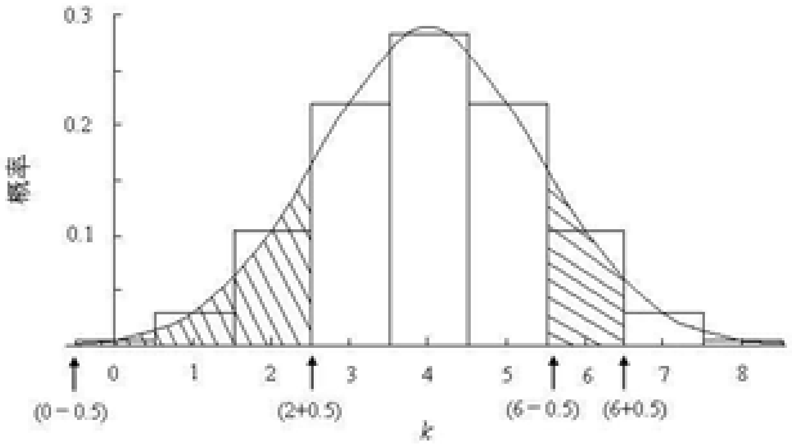
如按二项分布直接计算，则 $P\{0 \le y \le 2\} = P\{y=0\} + P\{y=1\} + P\{y=2\}$

$$= 0.0039 + 0.0313 + 0.1094 = 0.1446$$

3.6 二项分布的正态近似

例题 一批棉花种子的发芽率 p=0.5，在每穴播8粒时，求：(1)每穴出苗为 $0 \le y \le 2$ 株的概率？(2)每穴出苗为6株的概率？

$$\mu = 8 \times 0.5 = 4 \qquad \sigma = \sqrt{8 \times 0.5 \times 0.5} = 1.414$$



二项分布的正态近似图解

（2）

$$u_{c1} = \frac{(6-4)-0.5}{1.414} = 1.06 \qquad u_{c2} = \frac{(6-4)+0.5}{1.414} = 1.77$$

查附表1得

$$F(1.06) = 0.8554 \qquad F(1.77) = 0.9616$$

$$P\{y=6\} = F(1.77) - F(1.06) = 0.1062$$

如按二项分布直接计算，则 $P\{y=6\} = C_8^6 \times 0.5^6 \times 0.5^2 = 0.1094$