概率论与数理统计

复　习

考试范围：第1章到第9章，以下不作考试要求

1. 所有教材及课件中Excel内容。
2. （P31）（四）中的三个分布密度不要求记。
3. （1）(P67)数学期望公式（4.1.3）~ (4.1.5)；
   （2）(P73) *(四)条件数学期望；
   （3）(P78)变异系数；
   （4）(P85)其他数字特征；
   （5）(P87)协方差矩阵和二元正态概率密度不作要求，
   但多元正态变量的性质作为考试要求。
4. (P101) 李雅普诺夫中心极限定理。
5. (P119) （三）贝叶斯法；
   (P130)(7.4.8)(7.4.9)
   (P132)7.5节
6. (P147)8.3节中(8.3.3)-(8.3.6) ；
   (P156)8.5节（二）柯尔莫哥洛夫检验
   (P158)（三）正态W检验。
7. 第九章多因素方差分析、相关系数、多元回归分析及回归诊断。

第一章　随机事件与概率

1.交换律、结合律、分配律、德g摩根律

2.概率的性质：

$$P\left(B-A\right)=P\left(B\right)-P\left(AB\right)$$

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)$$
$$-P(AB)-P(AC)-P(BC)+P(ABC)$$

3. 古典概型： $P(A)=\dfrac{A\text{所包含的样本点数}}{\text{基本事件总数}}=\dfrac{m}{n}$

4. 抽签原理——跟先后顺序无关

5. 条件概率： $P(B|A)=\dfrac{P(AB)}{P(A)}$

6. 乘法公式： $P(A_1A_2\cdots A_n)=$

$$P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1})$$

7. 全概率： $P(A)=P(AS)$

$$=P(A(B_1\cup B_2\cup\cdots\cup B_n))\overset{\text{乘法公式}}{=}\sum_{j=1}^{n}P(B_j)\cdot P(A|B_j)$$

8. 贝叶斯： $P(B_i|A)=\dfrac{P(B_iA)}{P(A)}=\dfrac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)}$

9. 相容通过事件定义，独立通过概率定义。

设有n个人为过节日互赠礼物，每人准备一件礼物，集中在一起，然后每人随机取一件礼物，求（1）至少有一人恰好取到自己所准备的礼物的概率（2）恰好取到自己所准备的礼物的人数Y的数学期望及方差。

解：设$A_i=\{$第i个人恰好取到自己所准备的礼物$\}$，则所求概率为：

$$P(A_1\cup A_2\cup\cdots\cup A_n)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_iA_j)+\sum_{1\le i<j<k\le n}P(A_iA_jA_k)$$

$$+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)$$

$$=C_n^1P(A_1)-C_n^2P(A_1A_2)+C_n^3P(A_1A_2A_3)-\cdots+(-1)^{n-1}C_n^nP(A_1A_2\cdots A_n)$$

$$=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{n-1}\frac{1}{n!}$$

其中：$P(A_1)=\dfrac{1}{n}$ ，$P(A_1A_2)=P(A_1)P(A_2\mid A_1)=\dfrac{1}{n}\cdot\dfrac{1}{n-1}$

$$P(A_1A_2A_3)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2)=\frac{1}{n}\cdot\frac{1}{n-1}\cdot\frac{1}{n-2}$$

$\cdots$

$$P(A_1A_2\cdots A_n)=1/n!$$

设有n个人为过节日互赠礼物，每人准备一件礼物，集中在一起，然后每人随机取一件礼物，求（1）至少有一人恰好取到自己所准备的礼物的概率（2）恰好取到自己所准备的礼物的人数Y的数学期望及方差。

解（2）设 $X_i=\begin{cases}1, & \text{第}i\text{个人}\quad\text{取到自己所准备的礼物}\\ 0, & \text{第}i\text{个人没取到自己所准备的礼物}\end{cases}$

则 $Y=X_1+X_2+\cdots X_n$（同分布，不独立）

$$E(Y)=E(X_1+X_2+\cdots X_n)=nE(X_i)=1$$

$$D(Y)=D(X_1+X_2+\cdots X_n)=\sum_{i=1}^{n}D(X_i)+2\sum_{1\le i<j\le n}Cov(X_i,X_j)$$

$$=nD(X_i)+2C_n^2Cov(X_i,X_j)=1$$

	0	1
P	×	√

$$P(X_i=1)=\frac{1}{n}\ ,\ P(X_iX_j=1)=\frac{1}{n}\times\frac{1}{n-1}$$

$$E(X_i)=\frac{1}{n}\ ,\ D(X_i)=\frac{n-1}{n^2}\ ,\ Cov(X_i,X_j)=\frac{1}{n^2(n-1)}$$

盒中有红、黑、白球数分别为2、3、5，用不放回抽样取3球，*X*、*Y*、*Z* 分别表示取到的红、黑、白球数，求 $P(X=1,Z=1)$　　$P(X=1|Z=0)$

解：$P(X=1,Z=1)=P(X=1,Y=1,Z=1)=\dfrac{C_2^1C_3^1C_5^1}{C_{10}^3}=\dfrac{1}{4}$

$$P(X=1|Z=0)=\frac{P(X=1,Z=0)}{P(Z=0)}=\frac{3}{5}\quad \text{或} =\frac{C_2^1C_3^2}{C_5^3}$$

其中：$P(X=1,Z=0)=\dfrac{C_2^1C_3^2C_5^0}{C_{10}^3}=\dfrac{1}{20}$

$$P(Z=0)=\frac{C_2^0C_3^3C_5^0+C_2^1C_3^2C_5^0+C_2^2C_3^1C_5^0}{C_{10}^3}=\frac{1}{12}$$

第二章　随机变量及其分布

1. 六大常用分布的分布律或密度函数需牢记，

应用重点是二项分布及正态分布。

$$P(X=k)=p^k(1-p)^{1-k},\ k=0,1$$

$$P(X=k)=C_n^k p^k(1-p)^{n-k}\text{，}\quad k=0,1,\cdots,n$$

$$P(X=k)=\frac{l^{\,k}}{k!}e^{-l},\ k=0,1,2,\cdots,\ l>0$$

$$f(x)=\begin{cases}\dfrac{1}{b-a} & x\in(a,b)\\ 0 & \text{其他}\end{cases}\qquad f(x)=\begin{cases}l\,e^{-l\,x} & x>0\\ 0 & x\le 0\end{cases},$$

$$f(x)=\frac{1}{\sqrt{2p}\,s}e^{-\frac{(x-m)^2}{2s^2}}\text{，}\quad -\infty<x<+\infty$$

2. 分布函数 $F(x) = P(X \le x)$

$$= \begin{cases} \sum\limits_{x_k \le x} P\{X = x_k\} = \sum\limits_{x_k \le x} p_k, & \text{离散} \\ \int_{-\infty}^{x} f(t)dt \ , & \text{连续} \end{cases}$$

1) $F(x)$是单调不减函数

2) $0 \le F(x) \le 1$，且$F(-\infty) = 0$，$F(+\infty) = 1$

3) $F(x)$右连续,即$F(x+0) = F(x)$.

4) $P(a < X \le b) = F(b) - F(a)$

$$P(X = x_0) = F(x_0 + 0) - F(x_0 - 0)$$

3. 概率密度 $F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)dt$

1) $f(x)\ge 0$

2) $\int_{-\infty}^{+\infty} f(x)dx=1$

3) 对于任意的实数 $a,b \quad (b>a)$

$$P\{a<X\le b\}=\int_a^b f(t)\,dt$$

连续性随机变量任一指定值的概率为**0**,

$$P(X=C)=0$$

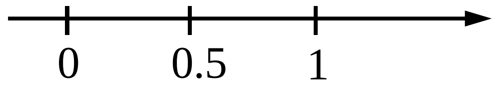
4) 在$f(x)$的连续点 x， $F'(x)=f(x)$

例：设随机变量X的密度函数为

$$f(x)=\begin{cases}2x, & 0\le x<0.5\\ 6-6x, & 0.5\le x\le 1\\ 0, & \text{其他}\end{cases}$$

，求分布函数$F(x)$。

解：参考$f(x)$的分段情况



$$F(x)=P\{X\le x\}=\int_{-\infty}^{x}f(t)dt=$$

$$\begin{cases}\int_{-\infty}^{0}0dt=0\ , & x<0\\ \int_{-\infty}^{0}0dt+\int_{0}^{x}2tdt=x^2, & 0\le x<0.5\\ \int_{-\infty}^{0}0dt+\int_{0}^{0.5}2tdt+\int_{0.5}^{x}(6-6t)dt=6x-3x^2-2, & 0.5\le x\le 1\\ 1, & x>1\end{cases}$$

4.随机变量函数的分布，$Y=g(X)=X^2$

离散型

X	-1	0	1
P_k	0.2	0.4	0.4
$Y=X^2$	1	0	1

$\Rightarrow$

Y	0	1
P_k	0.4	0.6

连续变量函数的分布，可用定义法先$F_Y(y)$后$f_Y(y)$

$$F_Y(y)=P(Y\le y)=P(g(X)\le y)=\ldots$$

$$f_Y(y)=F_Y'(y)=\ldots$$

*若$Y=g(X)$在$f_X(x)$非零区间上单调，则可用定理得到Y的概率密度：

$$f_Y(y)=\begin{cases} f_X(h(y))\cdot\left|h'(y)\right|, & a<y<b \\ 0, & \text{其他} \end{cases},$$

其中$x=h(y)$为$y=g(x)$的反函数。

2010年研究生入学数学题：

设$X \sim F(x) = \begin{cases} 0, & x<0 \\ 0.5, & 0 \le x < 1 \\ 1-e^{-x} & ,x \ge 1 \end{cases}$，则$P(X=1)=?$

解：$P(X=1) = P(X \le 1) - P(X < 1)$

$$= F(1+0) - F(1-0)$$

$$= 1 - e^{-1} - 0.5$$

$$= 0.5 - e^{-1}$$

设X,Y独立，且$X \sim N(0,1), P(Y=1)=1/3,$
$P(Y=2)=2/3,\quad Z=XY,\quad$ 求$f_Z(z)$

解: $F_Z(z)=P(XY\le z)$

$$=P(Y=1)P(XY\le z\,|\,Y=1)+P(Y=2)P(XY\le z\,|\,Y=2)$$

$$=P(Y=1)P(X\le z)+P(Y=2)P(X\le z/2)$$

$$=\frac{1}{3}\times\Phi(z)+\frac{2}{3}\Phi\left(\frac{z}{2}\right)$$

$$\therefore\quad f_Z(z)=F_Z'(z)=\frac{1}{3}\left[j\,(z)+j\left(\frac{z}{2}\right)\right]$$

$$=\frac{1}{3\sqrt{2p}}\left[e^{-\frac{z^2}{2}}+e^{-\frac{z^2}{8}}\right]\quad,\quad -\infty<z<\infty$$

第三章 二维随机变量及其概率分布

离散变量分布律 $P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j\mid X=x_i)$

离散/连续随机变量联合分布函数 $F(x,y)=P(X\le x,Y\le y)$

连续随机变量联合密度函数$f(x,y)$性质：

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=1$$

$$P((X,Y)\in G)=\iint_G f(x,y)dxdy$$

离散/连续随机变量的边缘分布函数, $F_X(x)=F(x,+\infty)$

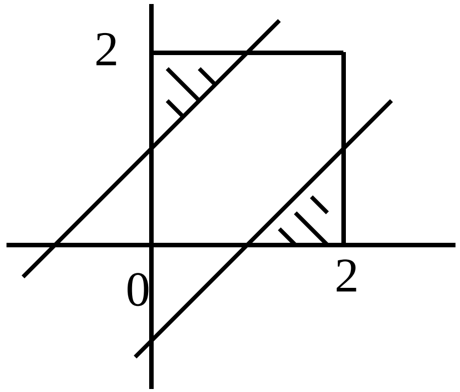
离散随机变量的边缘分布律 $P(X=x_i)=\sum_{j=1}^{+\infty}P(X=x_i,Y=y_j)$

连续随机变量的边缘密度函数 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy$

设X与Y独立均服从$U(0,2)$, $Z=|X-Y|$,求

（1）$P(Z>1)$　（2）$F_Z(z)$

解：$P(|X-Y|>1)=\iint\limits_{|x-y|>1} f(x,y)dxdy$



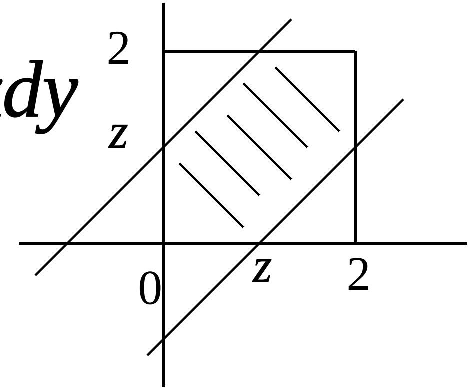
$$=\iint\limits_{\substack{x-y<-1\\0<x,y<2}} \frac{1}{2}\times\frac{1}{2}dxdy+\iint\limits_{\substack{x-y>1\\0<x,y<2}} \frac{1}{2}\times\frac{1}{2}dxdy$$

$$=\frac{1}{4}(\frac{1}{2}+\frac{1}{2})=\frac{1}{4}$$

设X与Y独立均服从$U(0,2)$, $Z=|X-Y|$,求

（1）$P(Z>1)$　（2）$F_Z(z)$

$$F_Z(z)=P(|X-Y|\le z)\overset{0<z\le 2}{=}\iint\limits_{|x-y|\le z} f(x,y)dxdy$$



$$=\iint\limits_{\substack{|x-y|\le z\\0<x,y<2}}\frac{1}{4}dxdy=\frac{1}{4}[4-\frac{(2-z)^2}{2}]$$

$$=0.5+0.5z-0.125z^2$$

$$\therefore F_Z(z)=\begin{cases}0, & z\le 0\\ 0.5+0.5z-0.125z^2, & 0<z\le 2\\ 1, & z>2\end{cases}$$

二维正态随机变量

$(X,Y)\sim N(m_1,\ m_2;s_1^2,\ s_2^2;r)$;

$$*f(x,y)=\frac{1}{2ps_1s_2\sqrt{1-r^2}}$$

$$\times\exp\left\{\frac{-1}{2(1-r^2)}\left[\frac{(x-m_1)^2}{s_1^2}-2r\frac{(x-m_1)(y-m_2)}{s_1s_2}+\frac{(y-m_2)^2}{s_2^2}\right]\right\}$$

$(X,Y)\sim N(m,\ m;s^2,\ s^2;0)$，求$E(XY^2)$ (2011研究生)

$$E(XY^2)\overset{独立}{=}E(X)E(Y^2)=E(X)[D(Y)+E^2(Y)]=m(s^2+m^2)$$

设$Z \sim N(1.5, 0.1^2), Z_i \sim N(1.5, 0.1^2), i = 1, 2, \mathbf{L}\, 30$

独立，求$P(Z_1 + Z_2 + \mathbf{L}\, Z_{15} \geq 23 \mid Z_1 + Z_2 + \mathbf{L}\, Z_{30} \geq 46)$

解：记$X = Z_1 + Z_2 + \mathbf{L} + Z_{15}$，$Y = Z_{16} + Z_{17} + \mathbf{L} + Z_{30}$

易知X与Y是独立同分布的，且$X \sim N(22.5, 0.15)$

$X + Y \sim N(45, 0.3)$

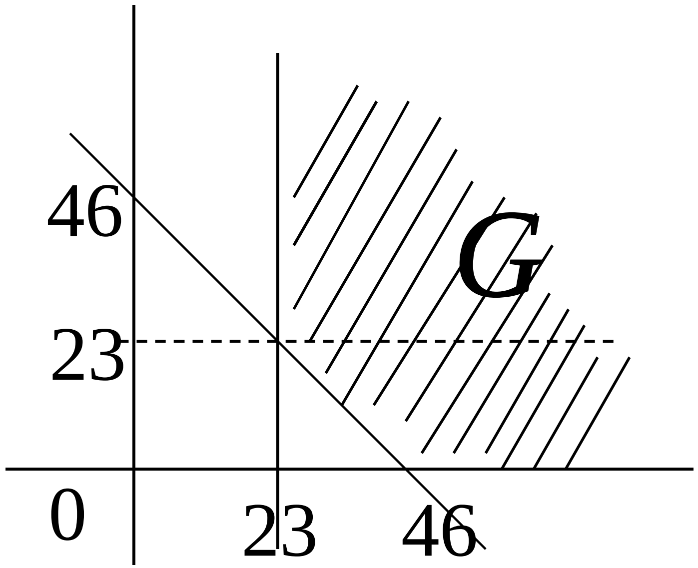
$$P(X \geq 23 \mid X + Y \geq 46) = \frac{P(X \geq 23, X + Y \geq 46)}{P(X + Y \geq 46)}$$

$$P(X + Y \geq 46) = 1 - \Phi\left(\frac{46 - 45}{\sqrt{0.3}}\right) = 0.291$$

$$P(X \geq 23, X + Y \geq 46) = ?$$

设$Z \sim N(1.5, 0.1^2), Z_i \sim N(1.5, 0.1^2), i = 1, 2, \cdots 30$

独立，求$P(Z_1 + Z_2 + \cdots Z_{15} \geq 23 | Z_1 + Z_2 + \cdots Z_{30} \geq 46)$



$$P(X \geq 23, X + Y \geq 46)$$

$$= \iint_G f(x, y) dx dy$$

$$= \iint_G f_X(x) f_Y(y) dx dy$$

$$= \int_{23}^{+\infty} dy \int_{23}^{+\infty} f_X(x) f_Y(y) dx + \int_0^{23} dy \int_{46-y}^{+\infty} f_X(x) f_Y(y) dx$$

$$= \cdots$$

条件分布

条件分布函数：

$$F_{X|Y}(x|y) = P(X \le x|Y = y) = \begin{cases} \sum\limits_{x_k \le x} P(X = x_k|Y = y) & ,\text{离散} \\ \int_{-\infty}^{x} f_{X|Y}(x|y)dx = \int_{-\infty}^{x} \dfrac{f(x,y)}{f_Y(y)}dx, & \text{连续} \end{cases}$$

离散型：$P(X = x_i|Y = y_j) = \dfrac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \dfrac{P_{ij}}{P_{\bullet j}}$

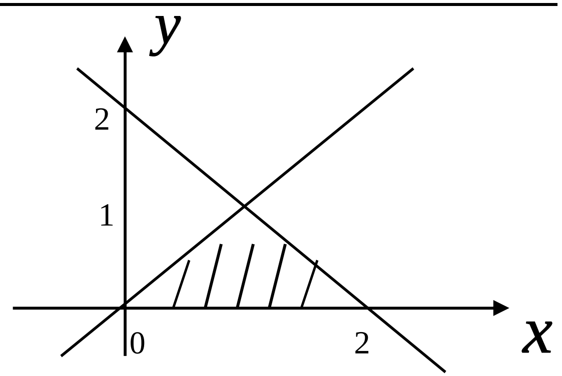
连续型：$f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$

$$P(a < X \le b|Y = y) = \int_a^b f_{X|Y}(x|y)dx$$

$$P(a < X \le b) = \int_a^b f(x)dx$$

2011年研究生入学数学题：

(X,Y)在G上服从均匀分布，G由$x - y = 0$, $x+y=2$与$y=0$围成。求$f_X(x)$, $f_{X|Y}(x|y)$



$$f(x,y)=\begin{cases}1,(x,y)\in G\\0,(x,y)\notin G\end{cases}$$

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy$$

$$=\begin{cases}\int_0^x 1dy = x, & 0<x<1\\ \int_0^{2-x}1dy=2-x, & 1\le x<2\\ 0, & \text{其它}\end{cases}$$

相互独立的随机变量

一　般：$F(x,y)=F_X(x)F_Y(y)$

在平面上每一点均成立。

离散型：$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$

对任意x_i,y_j均成立。

$\forall i,j$　均有$P_{ij}=P_{ig}*P_{gj}$

连续型：$f(x,y)=f_X(x)f_Y(y)$

在平面上每一点几乎处处成立。

例： 随机变量$X\sim N(0,1)$，令$U=|X|$，$V=\begin{cases}1, X>0\\0, X\le 0\end{cases}$. 求

(1) U的分布函数$F_U(u)$ (2) V的分布函数$F_V(v)$ (3)U、V独立性

$$(1)F_U(u)=P(U\le u)=P(|X|\le u)=\begin{cases}2\Phi(u)-1, & u\ge 0\\ 0, & u<0\end{cases}$$

$$(2)P(V=1)=P(X>0)=0.5$$

$$P(V=0)=P(X\le 0)=0.5$$

$$\Rightarrow F_V(v)=P(V\le v)=\begin{cases}0, & v<0\\ 0.5, & 0\le v<1\\ 1, & v\ge 1\end{cases}$$

$$(3)F_U(u)F_V(v)=\begin{cases}2\Phi(u)-1, & u\ge 0, v\ge 1\\ \Phi(u)-0.5, & u\ge 0, 0\le v<1\\ 0, & u<0\text{或}v<0\end{cases}$$

例： 随机变量$X \sim N(0,1)$，令$U=|X|$，$V=\begin{cases}1, X>0\\0, X\le 0\end{cases}$. 求

(1) U的分布函数$F_U(u)$ (2) V的分布函数$F_V(v)$ (3)U、V独立性

---

(3) 前面已得：$F_U(u)F_V(v)=\begin{cases}2\Phi(u)-1 \ , \ u\ge 0, v\ge 1\\ \Phi(u)-0.5 \ , u\ge 0, 0\le v<1\\ 0 \ , \ u<0\text{或}v<0\end{cases}$

当$u<0$或$v<0$时，$F(u,v)=P(U\le u, V\le v)$

$$=P(|X|\le u, V\le v)=0$$

当$u\ge 0, 0\le v<1$时，$F(u,v)=P(|X|\le u, X\le 0)$

$$=P(-u<X\le 0)=\Phi(u)-0.5$$

当$u\ge 0, v\ge 1$时，$F(u,v)=P(|X|\le u)=2\Phi(u)-1$

综上所述，$F(u,v)=F_U(u)F_V(v)$ ，即独立

判断以下命题的真伪

1. 若$D(X+Y)=D(X)+D(Y)$，则X,Y独立

2. 若$E(XY)=E(X)E(Y)$，则X,Y独立

√ 3. 若$E(XY)\neq E(X)E(Y)$，则X,Y不独立

√ 4. 存在点(x_0,y_0)使$F(x_0,y_0)\neq F_X(x_0)F_Y(y_0)$，则$X,Y$不独立

5. 存在点(x_0,y_0)使$F(x_0,y_0)=F_X(x_0)F_Y(y_0)$，则$X,Y$独立

√ 6. 若各连续函数有$f(x_0,y_0)\neq f_X(x_0)f_Y(y_0)$，则$X,Y$不独立

二元随机变量函数的分布

➤ 设二元离散型随机变量(X,Y)具有概率分布

$$P(X=x_i,Y=y_j)=p_{ij},\qquad i,j=1,2,\ldots$$

问题 (1)若$U=g(X,Y)$,则U的分布律是什么?

(2)若$U=u(X,Y),V=v(X,Y)$,则(U,V)的分布律是什么?

方法 对于(1),先确定U的取值u_i ,$i=1,2,\ldots$

再找出$(U=u_i)=\{(X,Y)\in D\}$,从而计算出分布律.

方法 对于(2),先确定(U,V)的取值 (u_i,v_j) $i,j=1,2,\ldots$

再找出$(U=u_i,V=v_j)=\{(X,Y)\in D\}$,从而计算出分布律;

例1：设X与Y的联合分布律为：

$X\backslash Y$	1	2
1	0.2	0.1
2	0.3	0.4

令$U = X + Y, V = \max(X, Y)$，

求U及(U, V)的分布律。

解：U的取值范围为2,3,4

$$P(U=2)=P(X+Y=2)=P(X=1,Y=1)=0.2$$

$$P(U=3)=P(X+Y=3)=P(\{X=1,Y=2\}\cup\{X=2,Y=1\})$$

$$=P(\{X=1,Y=2\})+P(\{X=2,Y=1\})=0.1+0.3=0.4$$

$$P(U=4)=P(X+Y=4)=P(X=2,Y=2)=0.4$$

$\therefore U$的分布律为：

U	2	3	4
P_k	0.2	0.4	0.4

例1：设X与Y的联合分布律为：

$X \backslash Y$	1	2
1	0.2	0.1
2	0.3	0.4

令$U=X+Y, V=\max(X,Y)$，

求U及(U,V)的分布律。

解：U的取值范围为2,3,4；　V的取值范围为1, 2

$$P(U=2,V=1)=P(X+Y=2,\max(X,Y)=1)=P(X=1,Y=1)=0.2$$

$$P(U=3,V=1)=P(X+Y=3,\max(X,Y)=1)=0$$

$$P(U=4,V=1)=P(X+Y=4,\max(X,Y)=1)=0$$

$$P(U=2,V=2)=P(X+Y=2,\max(X,Y)=2)=0$$

$$P(U=3,V=2)=P(X+Y=3,\max(X,Y)=2)$$

$$=P(\{X=1,Y=2\}\cup\{X=2,Y=1\})$$

$$=P(X=1,Y=2)+P(X=2,Y=1)=0.1+0.3=0.4$$

$$P(U=4,V=2)=P(X+Y=4,\max(X,Y)=2)$$

$$=P(X=2,Y=2)=0.4$$

$U \backslash V$	1	2
2	0.2	0
3	0	0.4
4	0	0.4

- 设二元连续型随机变量(X,Y)具有概率分布$f(x,y)$，Z是X,Y的函数，$Z=g(X,Y)$.

$\boxed{\text{问题}}$ Z的概率分布或密度函数是什么？

$\boxed{\text{方法}}$ 先求Z的分布函数再求导得到密度函数.

$$F_Z(z)=P(Z\le z)=P(g(X,Y)\le z)$$

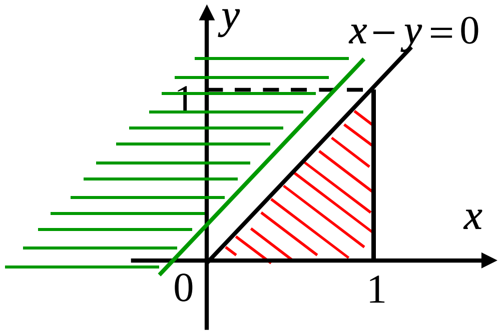
$$=\iint\limits_{g(x,y)\le z} f(x,y)dxdy$$

$$f_Z(z)=F_Z'(z)$$

和的分布(卷积公式)、最大、最小分布，

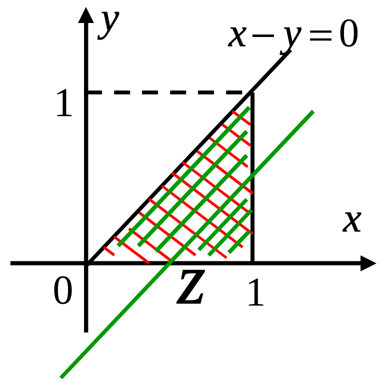
例3: 设(X,Y)的密度函数为$f(x,y)=\begin{cases}3x, 0<x<1, 0<y<x \\ 0, \quad 其他\end{cases}$，

求$Z=X-Y$的密度函数$f_Z(z)$。



解: $F_Z(z)=P(Z\le z)=P(X-Y\le z)=\iint\limits_{x-y\le z} f(x,y)dxdy$

当$z\le 0$时，画$x-y\le z$区域图，可见，不与$f(x,y)$非零区域相交，所以 $F_Z(z)=0.$



当$0<z<1$时，根据画$x-y\le z$区域图，得:

$$F_Z(z)=\iint\limits_{x-y\le z} f(x,y)dxdy=1-\iint\limits_{x-y>z} f(x,y)dxdy$$

$$=1-\int_z^1 dx\int_0^{x-z} 3xdy=\frac{3}{2}z-\frac{1}{2}z^3$$

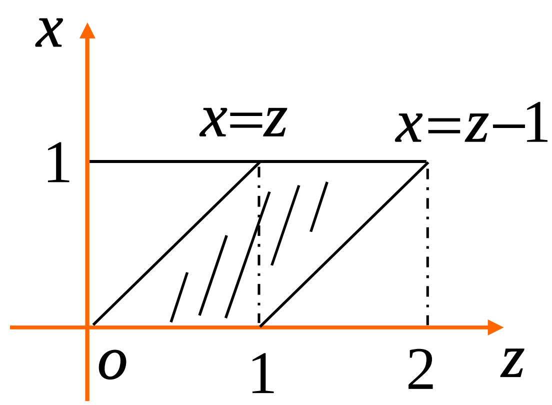
当$z\ge 1$时，$F_Z(z)=1$ $\therefore f_Z(z)=F_Z'(z)=\begin{cases}3(1-z^2)/2, 0<z<1 \\ 0, \quad 其他\end{cases}$

例：X,Y相互独立，同时服从[0,1]上的均匀分布，求$Z=X+Y$的概率密度。

解：根据卷积公式： $$f_Z(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx$$

易知仅当

$$\begin{cases} 0\le x\le 1 \\ 0\le z-x\le 1 \end{cases} \quad \text{即} \quad \begin{cases} 0\le x\le 1 \\ z-1\le x\le z \end{cases}$$

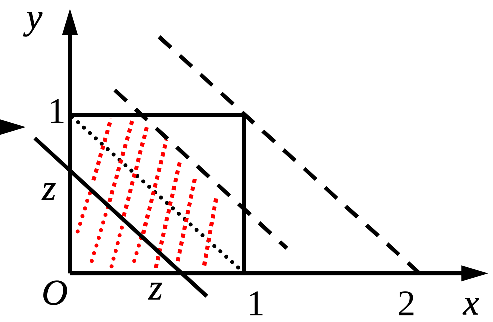


时，上述积分的被积函数不等于零。

参考图得：

$$f_Z(z)=\begin{cases} \int_0^z dx = z & 0\le z\le 1 \\ \int_{z-1}^1 dx = 2-z & 1\le z\le 2 \\ 0 & \text{其他} \end{cases}$$

上例的另外解法，“先F后f”



$$\text{解}: F_Z(z)=P(Z\le z)=P(X+Y\le z)$$

$$=\iint\limits_{x+y\le z} f(x,y)dxdy=\iint\limits_{x+y\le z} f_X(x)f_Y(y)dxdy$$

当$z<0$时，$F_Z(z)=0$

当$0\le z<1$时，$F_Z(z)=\iint\limits_{\substack{x+y\le z\\0<x,y<1}} 1\times 1dxdy=\text{三角形面积}=\frac{1}{2}z^2$

当$1\le z<2$时，$F_Z(z)=\text{长方形和梯形面积}=-\frac{z^2}{2}+2z-1$

当$z\ge 2$时，$F_Z(z)=1$

$$\therefore F_Z(z)=\begin{cases}0, & z<1\\ 0.5z^2, & 0\le z<1\\ -0.5z^2+2z-1, & 1\le z<2\\ 1, & z\ge 2\end{cases}\qquad f_Z(z)=F_Z{}'(z)=\begin{cases}z & ,0\le z\le 1\\ 2-z & ,1\le z\le 2\\ 0 & ,\text{其他}\end{cases}$$

2008年研究生入学数学题：

设X,Y独立，$P(X=i)=\dfrac{1}{3},(i=-1,0,1)$. $f_Y(y)=\begin{cases}1, 0\le y\le 1\\ 0, \quad 其它\end{cases}$

记$X+Y=Z$。求(1)$P(Z\le 0.5\,|\,X=0)$ (2) $f_Z(z)$

$$(1)P(Z\le 0.5\,|\,X=0)=P(X+Y\le 0.5\,|\,X=0)$$

$$=P(Y\le 0.5)=\int_{-\infty}^{0.5} f_Y(y)dy=\int_0^{0.5}1dy=0.5$$

(2)Z的取值范围为$-1\to 2$

$\therefore z\ge 2$时，$F(z)=P(Z\le z)=1,$

$z<-1$时，$F(z)=0$

2008年研究生入学数学题：

设X,Y独立，$P(X=i)=\dfrac{1}{3},(i=-1,0,1).$ $f_Y(y)=\begin{cases}1, 0\le y\le 1\\ 0, \quad 其它\end{cases}$

记$X+Y=Z$。求(1)$P(Z\le 0.5\mid X=0)$ (2)$f_Z(z)$

当-1$\le z<2$时，$F_Z(z)=P(Z\le z)=P(X+Y\le z)$

$$=\sum_{i=0}^{2}P(X=i-1)P(X+Y\le z\mid X=i-1)$$

$$=[P(Y\le z+1)+P(Y\le z)+P(Y\le z-1)]/3$$

分别在[-1,0) [0,1) [1,2)讨论

$$=\begin{cases}0, & z<-1\\ \dfrac{z+1}{3}, & -1\le z<2\\ 1, & z\ge 2\end{cases}$$

已知X,Y独立且均服从均值为**10**的指数分布，求

$P(\max(X,Y)\ge 10), P(\max(X,Y)\ge 10, \min(X,Y)\le 10)$

$E(\max(X,Y))$

解：$f(x)=\begin{cases}0.1e^{-0.1x}, x>0\\ 0 \quad ,\text{其他}\end{cases}, F(x)=\begin{cases}1-e^{-0.1x}, x>0\\ 0 \quad , x\le 0\end{cases}$

$$P(\max(X,Y)\ge 10)=1-P(\max(X,Y)<10)$$

$$=1-P(X<10, Y<10)=1-\{P(X<10)\}^2$$

$$=1-\{F(10)\}^2=1-(1-e^{-1})^2$$

$$P(\max \geq 10, \min \leq 10)$$

$$= 1 - P(\{\max \leq 10\} \cup \{\min \geq 10\})$$

$$= 1 - (P(\max \leq 10) + P(\min \geq 10) - 0)$$

$$= 1 - \{P(X \leq 10)\}^2 - \{P(X \geq 10)\}^2$$

$$= 1 - \{F(10)\}^2 - \{1 - F(10)\}^2$$

$$= 1 - \{1 - e^{-1}\}^2 - \{e^{-1}\}^2$$

$$= 2e^{-1} - 2e^{-2}$$

$$\because f(x)=\begin{cases}0.1e^{-0.1x}, & x>0\\ 0, & \text{其他}\end{cases}, F(x)=\begin{cases}1-e^{-0.1x}, & x>0\\ 0, & x\le 0\end{cases}$$

设$M=\max(X,Y)$

$$f_M(z)=2F(z)f(z)=\begin{cases}0.2(1-e^{-0.1z})e^{-0.1z}, & z>0\\ 0, & z\le 0\end{cases}$$

$$or \quad F_Z(z)=[F(z)]^2=\begin{cases}(1-e^{-0.1z})^2, & z>0\\ 0, & z\le 0\end{cases}$$

$$E(M)=\int_0^{\infty} z\cdot 0.2(1-e^{-0.1z})e^{-0.1z}dz=15$$

第四章 随机变量的数字特征

- 数学期望的六个计算式
- 数学期望、方差的性质；拆分法计算期望
- 六个分布的分布、期望及方差需牢记
- 协方差及相关系数定义、计算式
- 独立性与相关性的判断
- 正态分布的有关结论

$$E(X)=\sum_{i=1}^{\infty} x_i p_i \qquad E(X)=\int_{-\infty}^{+\infty} x f(x)dx$$

$$E[g(X)]=\sum_{i=1}^{\infty} g(x_i)p_i \qquad E(g(X))=\int_{-\infty}^{+\infty} g(x)f(x)dx$$

$$E[h(X,Y)]=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} h(x_i,y_j)p_{ij}$$

$$E(h(X,Y))=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)f(x,y)dxdy$$

C_n^2个Cov

$$E(XY)=r_{XY}\sqrt{D(X)D(Y)}+E(X)E(Y)$$

$$D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$$

$$D(X_1+X_2+\mathbf{L}+X_n)=\sum_{i=1}^{n} D(X_i)+2\sum_{1\le i<j\le n} Cov(X_i,X_j)$$

一盒中有5个球，其中有3个编不同号的红球，现有放回一个一个取球，直到三红球都取到，用*X*表示以上试验取球总次数，求*E*(*X*)

解：设X_1表示其中一只红球被取到时的取球次数

X_2：前1只红球取到后，为得第2只不同号红球的取球次数

X_3：前2只不同号红球取到后，为得第3只红球的取球次数

则 $X=X_1+X_2+X_3$

X_1	1	2	3	$\cdots$
P_k	$\frac{3}{5}$	$\frac{2}{5}\times\frac{3}{5}$	$\left(\frac{2}{5}\right)^2\times\frac{3}{5}$	

$$E(X_1)=\sum_{k=1}^{\infty}k\left(\frac{2}{5}\right)^{k-1}\frac{3}{5}=\frac{5}{3}$$

一盒中有5个球，其中有3个编不同号的红球，现有放回一个一个取球，直到三红球都取到，用*X*表示以上试验取球总次数，求*E*(*X*)

X_2	1	2	3	$\cdots$
P_k	$\frac{2}{5}$	$\frac{3}{5}\times\frac{2}{5}$	$\left(\frac{3}{5}\right)^2\times\frac{2}{5}$	

$$E(X_2)=\sum_{k=1}^{\infty}k\left(\frac{3}{5}\right)^{k-1}\frac{2}{5}=\frac{5}{2}$$

X_3	1	2	3	$\cdots$
P_k	$\frac{1}{5}$	$\frac{4}{5}\times\frac{1}{5}$	$\left(\frac{4}{5}\right)^2\times\frac{1}{5}$	

$$E(X_3)=\sum_{k=1}^{\infty}k\left(\frac{4}{5}\right)^{k-1}\frac{1}{5}=\frac{5}{1}$$

$$E(X)=E(X_1)+E(X_2)+E(X_3)=\frac{55}{6}$$

设$X\ \mathbf{:}\ N(m, s^{2}), (X_1, X_2\ \mathbf{L}, X_n)$是一样本，求$E(X_1\overline{X}), r_{X_1\overline{X}}$

解：$E(X_1\overline{X}) = \dfrac{1}{n}E(X_1^2 + X_1X_2 + \mathbf{L} + X_1X_n)$

$$= \frac{1}{n}[E(X_1^2) + E(X_1X_2) + \mathbf{L} + E(X_1X_n)]$$

$$= \frac{1}{n}[D(X_1) + E^2(X_1) + E(X_1)E(X_2) + \mathbf{L} + E(X_1)E(X_n)]$$

$$= \frac{1}{n}[s^{2} + m^{2} + (n-1)m^{2}] \quad = \frac{s^{2}}{n} + m^{2}$$

$$Cov(X_1, \overline{X}) = E(X_1\overline{X}) - E(X_1)E(\overline{X}) \quad = \frac{s^{2}}{n}$$

$$r_{X_1\overline{X}} = \frac{Cov(X_1, \overline{X})}{\sqrt{D(X_1)D(\overline{X})}} = \frac{s^{2}/n}{\sqrt{s^{2}s^{2}/n}} \quad = \frac{\sqrt{n}}{n}$$

$$Cov(X_1, \overline{X}) = Cov(X_1, \frac{X_1 + X_2 + \cdots + X_n}{n})$$

$$= \frac{1}{n} Cov(X_1, X_1 + X_2 + \cdots + X_n)$$

$$= \frac{1}{n}\{Cov(X_1, X_1) + Cov(X_1, X_2) + \cdots + Cov(X_1, X_n)\}$$

$$\overset{\text{独立}}{=} \frac{1}{n}\{Cov(X_1, X_1) + 0 + \cdots + 0\}$$

$$= \frac{1}{n} D(X_1) \quad = \frac{\sigma^2}{n}$$

2009年研究生入学数学题

设$X \sim F(x)=0.3\Phi\left(x\right)+0.7\Phi\left(\dfrac{x-1}{2}\right)$，则$E(X)=?$

解：$f(x)=0.3j\left(x\right)+0.35j\left(\dfrac{x-1}{2}\right)$

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{-\infty}^{+\infty}\left(0.3xj\left(x\right)+0.35xj\left(\dfrac{x-1}{2}\right)\right)dx$$

$$=\int_{-\infty}^{+\infty}0.35xj\left(\dfrac{x-1}{2}\right)dx \overset{u=(x-1)/2}{=\!=\!=\!=\!=}\int_{-\infty}^{+\infty}0.35(2u+1)j(u)\cdot 2du$$

$$=\int_{-\infty}^{+\infty}\left(1.4uj(u)+0.7j(u)\right)du \quad =\int_{-\infty}^{+\infty}0.7j(u)du$$

$$=0.7\int_{-\infty}^{+\infty}j(u)du=0.7$$

注意：$\displaystyle\int_{-\infty}^{+\infty}xj\left(x\right)dx=0$

设随机变量X与Y的概率分布分别为：

X	0	1
P	$\frac{1}{3}$	$\frac{2}{3}$

Y	-1	0	1
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

且$P(X^2=Y^2)=1$

(1)求二维随机变量(X,Y)的概率分布；

(2)求$Z=XY$的概率分布；

(3)求X与Y的相关系数r_{XY}.(2011年硕士研究生入学题)

$X \backslash Y$	-1	0	1	$P_{i\cdot}$
0	0	1/3	0	1/3
1	1/3	0	1/3	2/3
$P_{\cdot j}$	1/3	1/3	1/3	1

$$P(XY=-1)=P(X=1,Y=-1)=1/3$$

$$P(XY=1)=P(X=1,Y=1)=1/3$$

$$P(XY=0)=1-1/3-1/3=1/3$$

$$E(XY)=-1*1/3+0*1/3+1*1/3=0$$

$$E(X)=2/3,\quad E(Y)=0$$

$$COV(X,Y)=E(XY)-E(X)E(Y)=0,\ \ r_{XY}=\frac{COV(X,Y)}{\sqrt{D(X)D(Y)}}=0$$

第五章 大数定律和中心极限定理

1、契比雪夫不等式 $P\{|X-E(X)|\geq e\}\leq \frac{s^2}{e^2}$

2、随机变量序列$\{X_i, i\geq 1\}$，若满足以下任一条件：

(1) $\lim\limits_{n\to+\infty} D(\bar{X}) = \lim\limits_{n\to+\infty} D(\frac{1}{n}\sum_{i=1}^{n} X_i) = 0$, 独立不作要求；

(2) $X_1, X_2, \mathbf{L}, X_n, \mathbf{L}$ 独立, 方差存在；

(3) $X_1, X_2, \mathbf{L}, X_n, \mathbf{L}$ 独立同分布, 方差不作要求；

即有 $\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \xrightarrow{P} E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$

变量　　　　常数

第五章 大数定律和中心极限定理

3、变量若由大量的相互独立的变量的综合影响形成的，而其中每个个别的因素作用都很小，这种随机变量往往服从或近似服从正态分布.

- 当n足够大时， 各X_i独立, $\sum_{i=1}^{n} X_i \overset{(\text{近似})}{\sim} N(\sum_{i=1}^{n} E(X_i), \sum_{i=1}^{n} D(X_i))$
- 当n足够大时, 若$n_A \sim B(n,p)$, 则 $n_A \overset{\text{近似}}{\sim} N(np, np(1-p))$

如果样本容量足够大，则各种估计量往往近似正态分布。若总体$X \sim U(0,q)$中的q矩估计:

$$\hat{q} = 2\bar{X} \overset{\text{近似}}{\sim} N(q, \frac{q^2}{3n})$$

某种型号的器件的寿命X(以小时计)具有以下的概率密度：

$$f(x)=\begin{cases}\dfrac{1000}{x^2}, x>1000\\ 0 \quad ，其它\end{cases}$$

，现有一大批此种器件（设各器件损坏与否相互独立），任取n只，Y_n表示n只中寿命大于**3000**小时的器件个数，则当$n\to\infty$时，$\dfrac{Y_n}{n}\xrightarrow{P}?$ ，及$P(Y_{50}\le 20)=?$

解：$p=P(X>3000)=\int_{3000}^{\infty}\dfrac{1000}{x^2}dx=\dfrac{1}{3}$

$$Y_n\sim B(n,1/3) \qquad Y_n/n\xrightarrow{P}p=1/3$$

$$Y_{50}\overset{近似}{\sim} N(50*(1/3),50*(1/3)*(2/3))$$

$$P(Y_{50}\le 20)=\Phi\left(\frac{20-50*(1/3)}{\sqrt{50*(1/3)*(2/3)}}\right)\quad\approx\Phi(1)$$

设总体X的概率密度$f(x;q)=\begin{cases}\dfrac{\sqrt[q]{x}}{qx},0<x<1\\ \quad 0,其他\end{cases}$，$q>0$未知，$X_1,\mathbf{L},X_n$为来自$X$的简单随机样本，求$q$的矩估计量$\hat{q}$，并判断$\hat{q}$是否为$q$的相合估计量.

解：$m_1=E(X)=\int_0^1 x\cdot\dfrac{\sqrt[q]{x}}{qx}dx=\dfrac{1}{q+1}$ $,q=\dfrac{1-m_1}{m_1}$, $\hat{q}=\dfrac{1-\overline{X}}{\overline{X}}$

若用切比雪夫不等式判断$\lim\limits_{n\to\infty}P(|\hat{q}_1-q|<e)=1$是否成立就复杂了！

性质：若当$n\to\infty$时$,X_n\xrightarrow{P}a,Y_n\xrightarrow{P}b,$且函数$g(x,y)$

在点(a,b)连续，则 $g(X_n,Y_n)\xrightarrow{P}g(a,b)$

由辛钦大数定律,$\overline{X}\xrightarrow{P}E(X)=\dfrac{1}{q+1}$

则$\overline{X}$的连续函数：$\hat{q}=\dfrac{1-\overline{X}}{\overline{X}}\xrightarrow{P}\dfrac{1-\dfrac{1}{q+1}}{\dfrac{1}{q+1}}=q$

$\therefore\hat{q}_1$是q的相合性估计。

3. $X_1,\dots,X_n$ 为独立同分布的随机变量序列，均服从均匀分布 $U(0,1)$，(1) 当 n=64 时，$\frac{1}{8}\left(\sum_{i=1}^{64}X_i-32\right)$ 的近似分布为____________（要求写出参数），若以 Y 表示这 64 个 X_i 中取值大于 0.25 的个数，则 $P(Y\le 50)\approx$__________；(2) 由大数定律知，当 $n\to\infty$ 时，$\frac{2}{n}\sum_{i=1}^{n}(3X_i^2-X_i)$ 依概率收敛到____________。

$\frac{2}{n}\sum_{i=1}^{n}(3X_i^2-X_i)$依概率收敛到：

$$E(2(3X_i^2-X_i))=2(3E(X_i^2)-E(X_i))$$

$$=2(3(D(X_i)+E^2(X_i))-E(X_i))$$

$$=2(3(\frac{1}{12}+\frac{1}{4})-\frac{1}{2})=1$$

第六章 数理统计的基本概念

- 统计量　　$\overline{X},\ S^2, A_k, B_k$
- 三大分布　　$c^2,\quad t,\quad F$
  - 构成
  - 性质
- 抽样分布定理

$$\frac{\bar{X}-m}{s\ /\sqrt{n}} \sim N(0,1)$$

$$\frac{(\bar{X}-m)}{S/\sqrt{n}} \sim t(n-1)$$

$$\frac{(\bar{X}-\bar{Y})-(m_1-m_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

$$\frac{(n-1)S^2}{s^2} \sim c^2(n-1)$$

$$\frac{S_1^2}{S_2^2}\Big/\frac{s_1^2}{s_2^2} \sim F(n_1-1, n_2-1)$$

$$E(\overline{X})=E(X),\qquad D(\overline{X})=\frac{D(X)}{\text{样本容量}}$$

$$E(S^2)=D(X),\qquad \text{总体为正态时，}D(S^2)=\frac{2[D(X)]^2}{\text{样本容量}-1}$$

设$c^2 \sim c^2(n)$，则有 $E\left(c^2\right)=n,\quad D\left(c^2\right)=2n$

例：设$X \sim N(0,1),(X_1,X_2,\mathbf{L}\ X_n)$是样本,求$D(\overline{X}^2)$

解：$\overline{X} \sim N(0,\frac{1}{n})\qquad \frac{\overline{X}-0}{1/\sqrt{n}}=\sqrt{n}\ \overline{X} \sim N(0,1)$

$$n\overline{X}^2 \sim c^2(1)\qquad D\left(n\overline{X}^2\right)=2\qquad D\left(\overline{X}^2\right)=\frac{2}{n^2}$$

如何求 $\boldsymbol{D(\overline{X}+\sum_{i=1}^{n}(X_i-\overline{X})^2)}$

$$\boldsymbol{=D(\overline{X}+(n-1)S^2)=D(\overline{X})+D((n-1)S^2)}$$

***设总体X ：$N(0, s^2), (X_1, X_2)$是其样本，求$P\left(\left(\frac{X_1+X_2}{X_1-X_2}\right)^2 < 4\right)$**

解: **Q** (X_1, X_2)是样本， 由P88,(X_1, X_2)是二维正态随机变量

$\therefore (X_1+X_2, X_1-X_2)$也是二维正态随机变量

Q $Cov(X_1+X_2, X_1-X_2) = D(X_1) - D(X_2) = 0$

$\therefore X_1+X_2$与X_1-X_2是不相关的，所以它们也是相互独立的

$$Y = \left(\frac{X_1+X_2}{X_1-X_2}\right)^2 = \left(\frac{X_1+X_2-0}{\sqrt{2}s} \Big/ \frac{X_1-X_2-0}{\sqrt{2}s}\right)^2 \sim F(1,1)$$

由$P108$, 当$n_1 = 1, n_2 = 1$时，$f_Y(y) = \begin{cases} \dfrac{1}{p(1+y)\sqrt{y}}, & y > 0 \\ 0, & y \le 0 \end{cases}$

$$P(Y<4) = \int_0^4 \frac{1}{p(1+y)\sqrt{y}} dy = 2\int_0^4 \frac{1}{p(1+y)} d\sqrt{y} \;\; = \frac{2}{p}\arctan 2 = 0.70$$

2008年研究生入学数学题：

设X_1, L X_n是总体$N(m, s^2)$的样本，统计量$T=\overline{X}^2-\dfrac{S^2}{n}$

(1)证明T是m^2的无偏估计　　(2)当$m=0$，$s^2=1$时，求$D(T)$

（1）证明：$E(T)=E(\overline{X}^2-\dfrac{S^2}{n})=E(\overline{X}^2)-E(\dfrac{S^2}{n})$

$$=D(\overline{X})+E^2(\overline{X})-\frac{1}{n}E(S^2)=\frac{s^2}{n}+m^2-\frac{1}{n}s^2=m^2$$

（2）$n\overline{X}^2 \sim c^2(1)$，$D(\overline{X}^2)=\dfrac{2}{n^2}, D(S^2)=\dfrac{2s^2}{n-1}$

$$D(T)=D(\overline{X}^2-\frac{S^2}{n})\overset{\text{独立}}{=}D(\overline{X}^2)+\frac{1}{n^2}D(S^2)$$

$$=\frac{2}{n^2}+\frac{1}{n^2}\frac{2}{n-1}=\frac{2}{n(n-1)}$$

2010年研究生入学数学1&数学3题：

设$X_1, \mathbf{L}\ X_n$是总体$N(m, s^2)$的样本，统计量$T=\dfrac{1}{n}\sum_{i=1}^{n} X_i^2$,则$E(T)=?$

解：$$E(T)=\frac{1}{n}\sum_{i=1}^{n} E(X_i^2)$$

$$=\frac{1}{n}\sum_{i=1}^{n}(D(X_i)+E^2(X_i))$$

$$=\frac{1}{n}\sum_{i=1}^{n}(s^2+m^2)$$

$$=s^2+m^2$$

设$X \sim N(m, s^{2}), (X_1, \mathbf{L}, X_n)$是样本，$\overline{X}$、$S^2$是样本均值及方差，若再取$X_{n+1}$，证明统计量：

$$\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\overline{X}}{S} \sim t(n-1) \qquad Page113,11题$$

证：X_{n+1}与$\overline{X}$独立，且$X_{n+1}-\overline{X} \sim N(0, \frac{n+1}{n}s^{2})$

$$U=\frac{X_{n+1}-\overline{X}}{\sqrt{\frac{n+1}{n}s^{2}}} \sim N(0,1) \qquad V=\frac{(n-1)S^2}{s^{2}} \sim c^{2}(n-1) \qquad \text{独立}$$

$$\frac{U}{\sqrt{V/(n-1)}}=\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\overline{X}}{S} \sim t(n-1)$$

$$\text{若} X \sim t(n), \text{则} (1) X^2 \sim F(1,n), (2)\, t_{\alpha/2}^2(n) = F_\alpha(1,n)$$

$$(2) P(X > t_{\alpha/2}(n)) = \frac{\alpha}{2}$$

$$P(|X| > t_{\alpha/2}(n)) = \alpha$$

$$P(X^2 > t_{\alpha/2}^2(n)) = \alpha$$

$$\because X^2 \sim F(1,n)$$

$$P(X^2 > F_\alpha(1,n)) = \alpha$$

$$\therefore t_{\alpha/2}^2(n) = F_\alpha(1,n)$$

第七章 参数估计

矩估计法
极大似然估计法
估计量的评选标准
区间估计

(一) 矩估计法：

设总体X的分布函数为$F(x;q_1,q_2,\mathbf{L},q_k)$，$(q_1,q_2,\mathbf{L},q_k)$是待估计的未知参数，假定总体$X$的$k$阶原点矩$E(X^k)$存在且有未知数（否则用下一阶矩），则：

$$\begin{cases} m_1=E(X^1)=m_1(q_1,q_2,\mathbf{L}\, q_k) \\ m_2=E(X^2)=m_2(q_1,q_2,\mathbf{L}\, q_k) \\ \qquad \mathbf{L\ L\ L\ L\ L} \\ m_k=E(X^k)=m_k(q_1,q_2,\mathbf{L}\, q_k) \end{cases} \Rightarrow \begin{cases} q_1=q_1(m_1,m_2,\mathbf{L},m_k) \\ q_2=q_2(m_1,m_2,\mathbf{L},m_k) \\ \qquad \mathbf{L\ L\ L\ L\ L} \\ q_k=q_k(m_1,m_2,\mathbf{L},m_k) \end{cases}$$

$\mathbf{Q}\ A_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j \xrightarrow{P} m_j \quad (j=1,2,\mathbf{L}\,k)$， $\therefore$可用A_j作为m_j的估计，即得：

$$(q_1,q_2,\mathbf{L},q_k)\text{的一个矩估计量}\begin{cases} \hat{q}_1=q_1(A_1,A_2,\mathbf{L},A_k) \\ \hat{q}_2=q_2(A_1,A_2,\mathbf{L},A_k) \\ \qquad \mathbf{L} \\ \hat{q}_k=q_k(A_1,A_2,\mathbf{L},A_k) \end{cases}$$

（二）极(最)大似然估计法：

样本$\left(X_1, X_2, \mathbf{L}, X_n\right)$的观察值$\left(x_1, x_2, \mathbf{L}, x_n\right)$，

似然函数:

连续：$L\left(q\right)=\prod_{i=1}^{n} f\left(x_i\right)$

离散：$L(q)=\prod_{i=1}^{n} P(X_1=x_i)$

$$Ln(L(q))=.....$$

$$\frac{dLn(L(q))}{dq}=0,$$ 若不为零，则从定义出发求得。

$$\hat{q}_L=......$$

常用分布中参数的矩估计和最大似然估计

分布名称	分布	未知参数	矩估计量	极大似然估计量
0-1分布	$X \sim B(1, p)$	p	$\overline{X}$	$\overline{X}$
二项分布	$X \sim B(m, p)$	p	$\overline{X} / m$	$\overline{X} / m$
泊松分布	$X \sim p(\lambda)$	λ	$\overline{X}$	$\overline{X}$
均匀分布	$X \sim U[a,b]$	a,b	$\begin{cases} \hat{a} = \overline{X} - \sqrt{3B_2} \\ \hat{b} = \overline{X} + \sqrt{3B_2} \end{cases}$	$\begin{cases} \hat{a} = \min(X_1, \cdots, X_n) \\ \hat{b} = \max(X_1, \cdots, X_n) \end{cases}$
指数分布	$X \sim E(\lambda)$	λ	$1 / \overline{X}$	$1 / \overline{X}$
正态分布	$X \sim N(\mu, \sigma^2)$	μ, σ^2	$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = B_2 \end{cases}$	$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = B_2 \end{cases}$

估计量的评价准则

1. 满足$E\left(\hat{q}\right)=q$,则称$\hat{q}$是q的一个无偏估计量。

2. 设$\hat{q}_1,\hat{q}_2$是q的两个无偏估计，如果

   $D\left(\hat{q}_1\right)\le D\left(\hat{q}_2\right)$,则称$\hat{q}_1$比$\hat{q}_2$有效。

3. 设$\hat{q}$是参数q的点估计，方差存在，则称$E(\hat{q}-q)^2$

   是估计量的均方误差，记为$Mse(\hat{q})$.

4. $\lim\limits_{n\to+\infty} P\left\{\left|\hat{q}-q\right|<e\right\}=1$即当$n\to+\infty$时,$\hat{q}$依概率收敛

   于q，则称$\hat{q}$为q的相合估计量或一致估计量.

2010年研究生入学数学1

设X的分布为：

X	1	2	3
P_k	$1-q$	$q-q^2$	q^2

其中$q\in(0,1)$未知，以N_i来表示来自总体X的样本（容量为n）中等于i的个数（$i=1,2,3$）。求常数a_1,a_2,a_3使$T=\sum_{i=1}^{3}a_iN_i$为q的无偏估计量，并求$D(T)$.

解：N_i是随机变量，并且N_i服从$B(n,p_i)$，其中$p_i=P(X=i)$

$$E(T)=\sum_{i=1}^{3}a_iE(N_i)=a_1n(1-q)+a_2nq(1-q)+a_3q^2$$

$$=na_1+(na_2-na_1)q+(na_3-na_2)q^2\overset{\text{题设}}{\equiv}q$$

$$\Rightarrow a_1=0,a_2=a_3=1/n$$

$$D(T)=\sum_{i=1}^{3}a_i^2D(N_i)=\mathbf{L}=\frac{q(1-q)}{n}$$

2011年研究生入学数学

设$X_1, X_2, \mathbf{L}, X_n$为来自正态总体$N(m_0, s^2)$的简单随机样本，其中$m_\mathbf{0}$已知，$s^2>0$未知，$\overline{X}$和$S^2$分别表示样本均值和样本方差。

(1)求参数s^2的最大似然估计$\hat{s}^2$；

(2)计算$E(\hat{s}^2), D(\hat{s}^2)$.

$$(1)L(s^2)=\prod_{i=1}^{n} f(x_i)=(2ps^2)^{-\frac{n}{2}} e^{-\frac{1}{2s^2}\sum_{i=1}^{n}(x_i-m_0)^2}$$

$$\ln L(s^2)=-\frac{n}{2}\ln(2ps^2)-\frac{1}{2s^2}\sum_{i=1}^{n}(x_i-m_0)^2$$

$$\frac{\ln L(s^2)}{ds^2}=-\frac{n}{2s^2}+\frac{1}{2s^4}\sum_{i=1}^{n}(x_i-m_0)^2\overset{令}{=}0$$

$$\hat{s}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-m_0)^2$$

2011年研究生入学数学

设$X_1, X_2, \mathbf{L}, X_n$为来自正态总体$N(m_0, s^2)$的简单随机样本，其中$m_{\mathbf{0}}$已知，$s^{\mathbf{2}} > 0$未知，$\overline{X}$和$S^2$分别表示样本均值和样本方差。

(1)求参数s^2的最大似然估计$\hat{s}^2$；

(2)计算$E(\hat{s}^2), D(\hat{s}^2)$.

$$(2)E(\hat{s}^2)=E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - m_0)^2\right) = \frac{1}{n}\sum_{i=1}^{n}E(X_i - m_0)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}D(X_i) = \frac{1}{n}\sum_{i=1}^{n}s^2 = s^2 \quad \text{或：}$$

$$\frac{n\hat{s}^2}{s^2} = \sum_{i=1}^{n}\left(\frac{X_i - m_0}{s}\right)^2 \sim c^2(n), \quad E\left(\frac{n\hat{s}^2}{s^2}\right) = n$$

2011年研究生入学数学

设$X_1, X_2, \mathbf{L}, X_n$为来自正态总体$N(m_0, s^2)$的简单随机样本，其中$m_{\mathbf{0}}$已知，$s^{\mathbf{2}} > 0$未知，$\overline{X}$和$S^2$分别表示样本均值和样本方差。

(1)求参数s^2的最大似然估计$\hat{s}^2$;

(2)计算$E(\hat{s}^2), D(\hat{s}^2)$.

$$\frac{n\hat{s}^2}{s^2} = \sum_{i=1}^{n}\left(\frac{X_i - m_0}{s}\right)^2 \sim c^2(n)$$

$$D(\frac{n\hat{s}^2}{s^2}) = 2n$$

$$D(\hat{s}^2) = \frac{2s^4}{n}$$

六．（12 分）设总体 $X \sim N(0,\sigma^2)$，$X_1,\ldots,X_n(n>1)$ 是来自总体的简单随机样本，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$；下面给出 σ^2 的三个估计量，$\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，$\hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$，$\hat{\sigma}_3^2=n\bar{X}^2$；(1)求这些估计量的均值，并判断其中哪些是 σ^2 的无偏估计量；(2)求这三个估计量的方差，并在无偏估计中判断哪个是有效估计。

(1). $E(\hat{\sigma}_1^2)=E(S^2)=\sigma^2$， $\hat{\sigma}_1^2$ 是无偏估计.

$E(\hat{\sigma}_2^2)=E(X^2)=\sigma^2$， $\hat{\sigma}_2^2$ 〃

$E(\hat{\sigma}_3^2)=nD(\bar{X})=\sigma^2$， $\hat{\sigma}_3^2$ 〃

六．（12 分）设总体 $X \sim N(0,\sigma^2)$，$X_1,\ldots,X_n (n>1)$ 是来自总体的简单随机样本，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$；下面给出 σ^2 的三个估计量，$\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，$\hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$，$\hat{\sigma}_3^2=n\bar{X}^2$；(1)求这些估计量的均值，并判断其中哪些是 σ^2 的无偏估计量；(2)求这三个估计量的方差，并在无偏估计中判断哪个是有效估计。

(2). $\frac{(n-1)\hat{\sigma}_1^2}{\sigma^2} \sim \chi^2(n-1)$. $\Rightarrow D(\hat{\sigma}_1^2)=\frac{2\sigma^4}{n-1}$

$\frac{n\hat{\sigma}_2^2}{\sigma^2} \sim \chi^2(n)$. $\Rightarrow D(\hat{\sigma}_2^2)=\frac{2\sigma^4}{n}$

$\frac{\hat{\sigma}_3^2}{\sigma^2} \sim \chi^2(1)$. $\Rightarrow D(\hat{\sigma}_3^2)=2\sigma^4$

$D(\hat{\sigma}_2^2)<D(\hat{\sigma}_1^2)<D(\hat{\sigma}_3^2)$

故 $\hat{\sigma}_2^2$ 是有效估计.

区间估计

引言：点估计是由样本求出未知参数θ的一个估计值$\hat{\theta}$，

由于其随机性，$\hat{\theta}$总是不会恰好等于θ，它仅仅是θ的参考值，没有反映这个近似值的误差范围。

而区间估计则要由样本给出参数θ的一个估计范围，并指出该区间包含θ的可靠程度。

假设$(X_1,\cdots,X_n)$是总体X的一个样本，区间估计的方法是给出两个统计量

$$\hat{\theta}_1=g_1(X_1,\cdots,X_n),\quad \hat{\theta}_2=g_2(X_1,\cdots,X_n)$$

使区间$\left[\hat{\theta}_1,\hat{\theta}_2\right]$以一定的可靠程度盖住$\theta$。

求未知参数θ的置信区间（置信限）方法：

1. 根据得到的样本构造函数（枢轴量）$G(X_1,\cdots,X_n;\theta)$，要求(1)包含待估计θ；(2)包含θ的点估计(如无偏估计等)；(3)包含总体已知的信息；(4)不含除θ外的其它未知参数；(5)分布已知.

2. 对于给定的置信度$1-\alpha$，确定尽可能大的a，尽可能小的b，使得 $P\{a<G(\theta)<b\}\geq 1-\alpha$；

3. 若能从 $a<G(\theta)<b$ 得到等价的不等式

$$\hat{\theta}_1(X_1,\cdots,X_n)<\theta<\hat{\theta}_2(X_1,\cdots,X_n)$$

那么$(\hat{\theta}_1,\hat{\theta}_2)$就是$\theta$的置信度为$1-\alpha$的双侧置信区间。

注:若要求单侧置信限，只要将“2.”中的 $P\{a<G(\theta)<b\}\geq 1-\alpha$

改为 $P\{a<G(\theta)\}\geq 1-\alpha$ 或 $P\{G(\theta)<b\}\geq 1-\alpha$ 即可。

正态总体下常见的枢轴量

(1)单个正态总体$N(m, s^2)$情形

s^2已知，求m的区间估计：$$\frac{\bar{X}-m}{s/\sqrt{n}} \sim N(0,1)$$

s^2未知，求m的区间估计：$$\frac{\bar{X}-m}{S/\sqrt{n}} \sim t(n-1)$$

m未知，求s^2的区间估计：$$\frac{(n-1)S^2}{s^2} \sim c^2(n-1)$$

m已知，求s^2的区间估计：$$\sum_{i=1}^{n}\left(\frac{X_i-m}{s}\right)^2 \sim c^2(n)$$

正态总体下常见的枢轴量

(2)二个正态总体$N(m_1, s_1^{\ 2}), N(m_2, s_2^{\ 2})$情形

$s_1^{\ 2}, s_2^{\ 2}$已知，求$m_1 - m_2$的区间估计：$\dfrac{(\bar{X} - \bar{Y}) - (m_1 - m_2)}{\sqrt{\dfrac{s_1^{\ 2}}{n_1} + \dfrac{s_2^{\ 2}}{n_2}}} \sim N(0,1)$

$s_1^{\ 2} = s_2^{\ 2}$未知，求$m_1 - m_2$的区间估计：$\dfrac{(\bar{X} - \bar{Y}) - (m_1 - m_2)}{S_w \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$

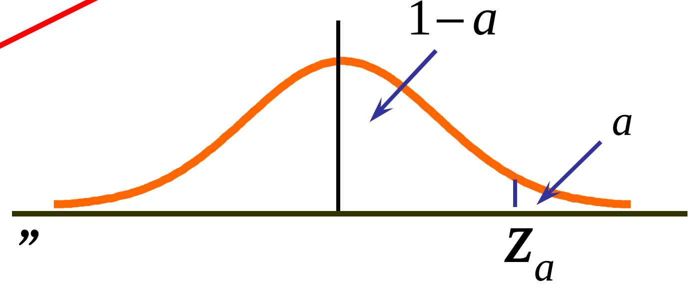
m_1, m_2未知，求$\dfrac{s_1^{\ 2}}{s_2^{\ 2}}$的区间估计：$\dfrac{S_1^2}{S_2^2} \dfrac{s_2^{\ 2}}{s_1^{\ 2}} \sim F(n_1 - 1, n_2 - 1)$

m_1, m_2已知，求$\dfrac{s_1^{\ 2}}{s_2^{\ 2}}$的区间估计：$\dfrac{\sum_{i=1}^{n_1} (X_i - m_1)^2 \Big/ n_1}{\sum_{i=1}^{n_2} (Y_i - m_2)^2 \Big/ n_2} \dfrac{s_2^{\ 2}}{s_1^{\ 2}} \sim F\left(n_1, n_2\right)$

单个正态总体$X \sim N\left(m, s^{2}\right)$中，$s^{2}$已知时,

求均值m的置信度$1-a$的单侧置信下限

同样取 $G=\dfrac{\bar{X}-m}{s/\sqrt{n}} \sim N(0,1)$



为求形如 $P(m>?)=1-a$ 中的“？”

设 $P\left(\dfrac{\bar{X}-m}{s/\sqrt{n}}<?\right)=1-a$

$$\Rightarrow P\left(\frac{\bar{X}-m}{s/\sqrt{n}}<z_{a}\right)=1-a$$

即$P\left\{m>\bar{X}-\dfrac{s}{\sqrt{n}}z_{a}\right\}=1-a$

置信区间为: $\left(\bar{X}-\dfrac{s}{\sqrt{n}}z_{a}\ ,\ +\infty\right)$

思考题:

均值m的置信度 $1-a$的置信上限是什么呢?

答案: $\bar{X}+\dfrac{s}{\sqrt{n}}z_{a}$

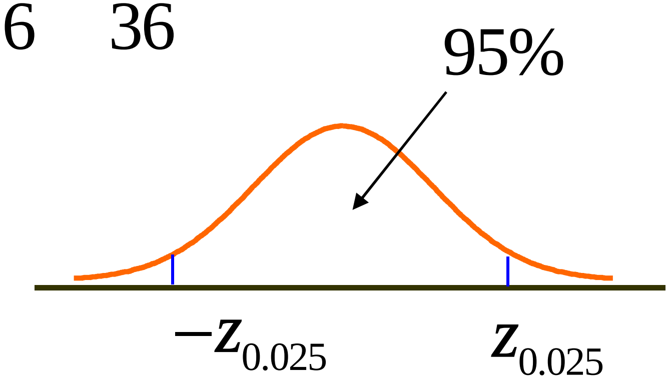
例：两个独立总体 $X \sim N(m_1,1), Y \sim N(m_2,4)$， m_1, m_2未知，分别从总体X、Y中抽取简单随机样本$X_1, X_2, \mathbf{L}\ X_9$和$Y_1, Y_2, \mathbf{L}\ , Y_{16}$， 求 $2m_1 - m_2$ 的95%的置信区间。

解：$2\overline{X} - \overline{Y}$ 是 $2m_1 - m_2$ 的无偏估计，

$$D(2\overline{X} - \overline{Y}) = 4D(\overline{X}) + D(\overline{Y}) = 4 * \frac{1}{9} + \frac{4}{16} = \frac{25}{36}$$

$$\therefore\ 2\overline{X} - \overline{Y} \sim N(\ 2m_1 - m_2, \frac{25}{36})$$

取 $G = \dfrac{(2\overline{X} - \overline{Y}) - (2m_1 - m_2)}{5/6} \sim N(0,1)$



设 $P(-z_{0.025} < G < z_{0.025}) = 95\%$ ，$z_{0.025} = 1.96$

得 $2m_1 - m_2$ 的95%的置信区间为:

$$2\overline{X} - \overline{Y} \pm \frac{5}{6} z_{0.025} = 2\overline{X} - \overline{Y} \pm 1.633$$

（三）成对数据情形

例：为考察某种降压药的降压效果，测试了n个高血压病人在服药前后的血压（收缩压）分别为

(X_1, Y_1)，$(X_2, Y_2), \cdots, (X_n, Y_n)$.

由于个人体质的差异，$X_1, \cdots, X_n$不能看成来自同一个正态总体的样本，即$X_1, \cdots, X_n$是相互独立但不同分布的样本，$Y_1, \cdots, Y_n$也是.

另外对同一个个体，X_i和Y_i也是不独立的.

$P136:18(1)$ 记$Y = X_{(1)} = \min(X_1, X_2, \cdots, X_n)$

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_m^x e^{-(t-m)}dt = 1 - e^{m-x}, & x \ge m \\ 0, & x < m \end{cases}$$

$$F_Y(y) = P(Y \le y) = 1 - P(Y > y) = 1 - [P(X > y)]^n = 1 - [1 - F_X(y)]^n$$

$$= \begin{cases} 0, & y < m \\ 1 - e^{nm-ny}, & y \ge m \end{cases}$$

$$f_Y(y) = F_Y'(y) = \begin{cases} 0, & y < m \\ ne^{nm-ny}, & y \ge m \end{cases}$$

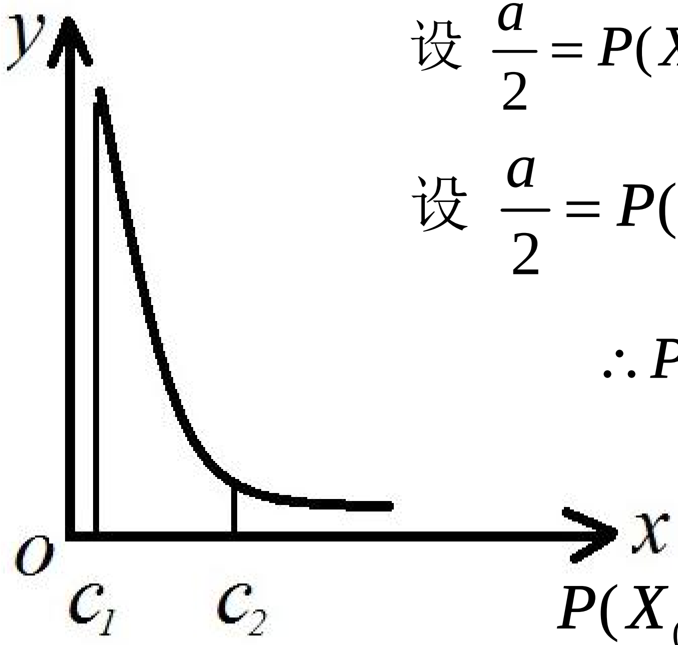
设$Z = X_{(1)} - m = Y - m$

$z=y-m, z'=1>0$单调增函数，反函数：$y=z+m$, 反函数导数：1

$$f_Z(z)=f_Y(z+m)*1=f_Y(z+m)=\begin{cases} ne^{-nz}, z\geq 0 \\ 0 \quad\quad, z<0 \end{cases}$$

由Z的密度函数可知，Z服从指数分布。

Z包含唯一未知数m，分布已知，所以可作为区间估计的枢轴量。



设 $\dfrac{a}{2}=P(X<c_1)=\int_0^{c_1} ne^{-nz}dz=1-e^{-nc_1}$， $c_1=-\dfrac{1}{n}\ln(1-\dfrac{a}{2})$

设 $\dfrac{a}{2}=P(X>c_2)=\int_{c_2}^{\infty} ne^{-nz}dz=e^{-nc_2}$， $c_2=-\dfrac{1}{n}\ln\dfrac{a}{2}$

$$\therefore P(c_1<X_{(1)}-m<c_2)=1-a$$

$$P(X_{(1)}+\frac{1}{n}\ln\frac{a}{2}<m<X_{(1)}+\frac{1}{n}\ln(1-\frac{a}{2}))=1-a$$

第八章 假设检验

假设检验

正态总体参数的假设检验

拟合优度检验

➢ 参数的假设检验问题处理步骤

- 1．根据实际问题的要求，提出原假设H_0 和备择假设H_1；特别注意H_0与H_1的不平等性。
- 2．根据已知条件选取检验统计量(选取方法同区间估计中的枢轴量)，画出统计量密度函数草图；
- 3．按照“在原假设H_0成立时，拒绝原假设的概率不大于显著性水平a ”这一原则，画出统计量分布的分位数图，(左边检验左边留a，右边检验右边留a，两边检验两边各留a/2)确定H_0拒绝域；
- 4．查分位数表、用样本观测值数据代入公式进行计算；根据样本数据是否落在H_0拒绝域内，作出拒绝原假设还是接受原假设的决策。

（$3'$）按“3”确定拒绝域，计算检验统计量的观测值与$P_$值；

（$4'$）根据给定的显著水平a，作出判断.

正态总体下常见的检验统计量

(1)单个正态总体$N(m, s^2)$情形

s^2已知，对m的检验：$$\frac{\bar{X} - m_0}{s / \sqrt{n}} \sim N(0,1)$$

s^2未知，对m的检验：$$\frac{\bar{X} - m_0}{S / \sqrt{n}} \sim t(n-1)$$

m未知，对s^2的检验：$$\frac{(n-1)S^2}{s_0^2} \sim c^2(n-1)$$

m已知，对s^2的检验：$$\sum_{i=1}^{n} \left(\frac{X_i - m}{s_0} \right)^2 \sim c^2(n)$$

正态总体下常见的检验统计量

(2)二个正态总体$N(m_1, s_1^2), N(m_2, s_2^2)$情形

s_1^2, s_2^2已知，对$m_1 - m_2$的检验：$\dfrac{(\bar{X}-\bar{Y})-d}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}} \sim N(0,1)$

$s_1^2 = s_2^2$未知，对$m_1 - m_2$的检验：$\dfrac{(\bar{X}-\bar{Y})-d}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2)$

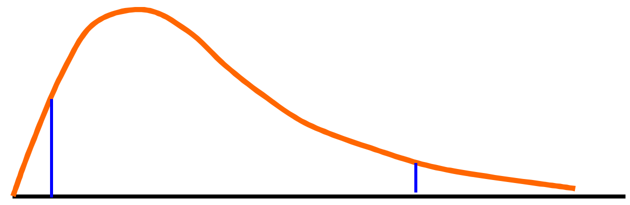
m_1, m_2未知，对s_1^2与s_2^2是否相同的检验：$\dfrac{S_1^2}{S_2^2} \sim F(n_1-1, n_2-1)$

m_1, m_2已知，对s_1^2与s_2^2是否相同的检验：$\dfrac{\sum_{i=1}^{n_1}(X_i-m_1)^2 \Big/ n_1}{\sum_{i=1}^{n_2}(Y_i-m_2)^2 \Big/ n_2} \sim F(n_1, n_2)$

例：自两个独立总体$N(m_1, s_1^2)$与$N(m_2, s_2^2)$中抽取了n_1、n_2个数据，并已计算得：$\bar{x}, s_1$；$\bar{y}, s_2$，在水平a下，检验 $H_0: s_1 = 2s_2 \quad H_1: s_1 \neq 2s_2$

$$\frac{\dfrac{(n_1-1)S_1^2}{s_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{s_2^2}\Big/(n_1-1)} = \frac{\dfrac{(n_1-1)S_1^2}{s_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)(2S_2)^2}{(2s_2)^2}\Big/(n_1-1)} = \frac{S_1^2}{(2S_2)^2}\frac{(2s_2)^2}{s_1^2} \sim F(n_1-1, n_2-1)$$

$$\therefore \text{取检验统计量：} F = \frac{S_1^2}{4S_2^2}$$



$$H_0\text{拒绝域为} F = \frac{S_1^2}{4S_2^2} \leq F_{1-a/2}(n_1-1, n_2-1)$$

$$\text{或} F = \frac{S_1^2}{4S_2^2} \geq F_{a/2}(n_1-1, n_2-1)$$

正态总体中均值、方差的双侧置信区间估计与双边假设检验比较

	待估参数	原假设 H_0	枢轴量 G	检验统计量	分布	置信区间	拒绝域
一个正态总体	m （s^2已知）	$m=m_0$ （s^2已知）	$\frac{\bar{X}-m}{s/\sqrt{n}}$	$\frac{\bar{X}-m_0}{s/\sqrt{n}}$	$N(0,1)$	$\frac{\lvert\bar{X}-m\rvert}{s/\sqrt{n}}<z_{a/2}$	$\frac{\lvert\bar{X}-m_0\rvert}{s/\sqrt{n}}\geq z_{a/2}$
	m （s^2未知）	$m=m_0$ （s^2未知）	$\frac{\bar{X}-m}{S/\sqrt{n}}$	$\frac{\bar{X}-m_0}{S/\sqrt{n}}$	$t(n-1)$	$\frac{\lvert\bar{X}-m\rvert}{S/\sqrt{n}}<t_{a/2}(n-1)$	$\frac{\lvert\bar{X}-m_0\rvert}{S/\sqrt{n}}\geq t_{a/2}(n-1)$
	s^2 (m未知)	$s^2=s_0^2$ (m未知)	$\frac{(n-1)S^2}{s^2}$	$\frac{(n-1)S^2}{s_0^2}$	$c^2(n-1)$	$c^2_{1-a/2}(n-1)$ $<(n-1)S^2/s^2$ $<c^2_{a/2}(n-1)$	$\frac{(n-1)S^2}{s_0^2}\leq c^2_{1-a/2}(n-1)$ 或 $\frac{(n-1)S^2}{s_0^2}\geq c^2_{a/2}(n-1)$
两个正态总体	m_1-m_2 $(s_1^2=s_2^2=s^2)$	$m_1=m_2$ $(s_1^2=s_2^2=s^2)$	$\frac{(\bar{X}-\bar{Y})-(m_1-m_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$\frac{\bar{X}-\bar{Y}}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$	$t(n_1+n_2-2)$	$\frac{\lvert(\bar{X}-\bar{Y})-(m_1-m_2)\rvert}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $<t_{a/2}(n_1+n_2-2)$	$\frac{\lvert\bar{X}-\bar{Y}\rvert}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $\geq t_{a/2}(n_1+n_2-2)$
	$\frac{s_1^2}{s_2^2}$	$s_1^2=s_2^2$	$\frac{S_1^2}{S_2^2}\Big/\frac{s_1^2}{s_2^2}$	$\frac{S_1^2}{S_2^2}$	$F(n_1-1,n_2-1)$	$F_{1-a/2}(n_1-1,n_2-1)$ $<\frac{S_1^2}{S_2^2}\Big/\frac{s_1^2}{s_2^2}<$ $F_{a/2}(n_1-1,n_2-1)$	$\frac{S_1^2}{S_2^2}\leq F_{1-a/2}(n_1-1,n_2-1)$ 或 $\frac{S_1^2}{S_2^2}\geq F_{a/2}(n_1-1,n_2-1)$

犯2类错误概率计算

设总体$X \sim N(m, 4)$,从中取容量25的样本，求

（1）$a = 0.05$下检验$H_0 : m \leq 1$,$H_1 : m > 1$的拒绝域

（2）当$m = 0.86$时，犯第一类错误的概率

（3）当$m = 2$时，犯第二类错误的概率

---

(1)H_0拒绝域为:$\dfrac{\overline{X} - m_0}{s / \sqrt{n}} = \dfrac{\overline{X} - 1}{2/5} \geq z_{0.05}$,即$\overline{X} \geq 1.66$

(2)第一类错误的概率=$P($拒绝$H_0 \mid H_0$为真$)$

$$= P_{m=0.86}(\overline{X} \geq 1.66) = 1 - \Phi\left(\frac{1.66 - 0.86}{0.4}\right) = 1 - \Phi(2) = 0.0228$$

$$\neq a$$

$$\overline{X} \sim N(0.86, 4/25)$$

犯2类错误概率计算

设总体$X \sim N(m,4)$,从中取容量25的样本，求

（1）$a = 0.05$下检验$H_0: m \le 1$,$H_1: m > 1$的拒绝域

（2）当$m = 0.86$时，犯第一类错误的概率

（3）当$m = 2$时，犯第二类错误的概率

---

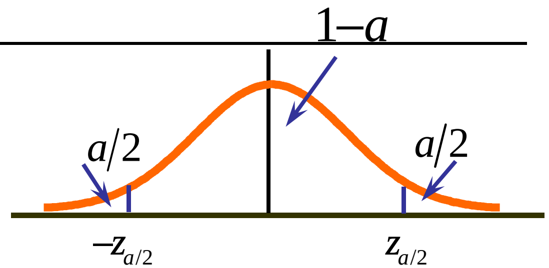
(3)第二类错误的概率=P(接受$H_0 \mid H_0$为假)

$$= P_{m=2}(\overline{X} < 1.66) \qquad \overline{X} \sim N(2, 4/25)$$

$$= \Phi\left(\frac{1.66-2}{0.4}\right) = 1 - \Phi(0.85) = 0.1977$$

P_值与统计显著性

P_值： 当原假设成立时，检验统计量取比观察到的结果更为极端的数值的概率.



- H_0的拒绝域 $|Z| = \left|\dfrac{\bar{X}-m_0}{s/\sqrt{n}}\right| \geq z_{a/2}$,

$$P\_ = P_{H_0}\{|Z| \geq |z_0|\} = 2P_{H_0}\{Z \geq |z_0|\} = 2(1-\Phi(|z_0|))$$

- H_0的拒绝域为： $t = \dfrac{\bar{X}-m_0}{S/\sqrt{n}} \geq t_a(n-1)$,

$$P\_ = \sup_{m \leq m_0}\{t \geq t_0\} = P\{t(n-1) \geq t_0\}.$$

- H_0的拒绝域为： $c^2 = \dfrac{(n-1)S^2}{s_0^2} \leq c_{1-a}^2(n-1)$,

$$P\_ = P(c^2(n-1) \leq c_0^2)$$

（二）成对数据的t检验

成对数据在7.4节中(区间估计)已作过介绍.

成对样本设为 $(X_1, Y_1), \cdots, (X_n, Y_n)$,

差值为 $D_i = X_i - Y_i, \quad i = 1, \cdots, n.$

可以看成来自正态总体$N(\mu_d, \sigma_d^2)$的样本。

为比较两总体均值是否有显著差异，可考虑假设问题

$$H_0: \mu_d = 0, \qquad H_1: \mu_d \neq 0$$

转化为单个正态总体的均值的假设检验。

分布拟合检验

定理. 若n充分大($n \geq 50$)，则当H_0为真时，统计量

$$c^2 = \sum_{i=1}^{k} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{n_i^2}{np_i} - n$$

近似服从$c^2(k-1)$分布。因此检验拒绝域为

$$c^2 \geq c_a^2(k-r-1).$$

注：c^2拟合检验使用时必须注意n要足够大，

np_i(或$n\hat{p}_i$)不能太小。根据实践，要求$n \geq 50$，

np_i (或 $n\hat{p}_i$) ≥ 5，否则应适当合并A_i，以满足要求。

A_i	n_i	$\hat{p}_i$或$\hat{p}_i$	np_i或$n\hat{p}_i$
A_1			
A_2			
A_3			
A_k			

计算 $c^2 = \sum_{i=1}^{k} \frac{n_i^2}{np_i} - n =$

查表 $c_a^2(k-r-1) =$

若$c^2 \geq c_a^2(k-r-1)$则拒绝H_0

第九章 方差分析及回归分析

单因素方差分析

一元线性回归

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ij}-\bar{X}\right)^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ij}-\bar{X}_{i\bullet}+\bar{X}_{i\bullet}-\bar{X}\right)^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(X_{ij}-\bar{X}_{i\bullet}\right)^2+\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(\bar{X}_{i\bullet}-\bar{X}\right)^2$$

$$= \quad S_E \quad + \quad S_A$$

(1) S_A与S_E相互独立； (2) $\dfrac{S_E}{s^2}\sim c^2(n-r)$；

(3)当H_0为真时，$\dfrac{S_A}{s^2}\sim c^2(r-1)$。

（4）当H_0为真时，$F=\dfrac{S_A/(r-1)}{S_E/(n-r)}\overset{\Delta}{=}\dfrac{MS_A}{MS_E}\sim F(r-1,n-r)$

单因素试验方差分析表

方差来源	平方和	自由度	均方	F比
因素A	S_A	$r-1$	$MS_A = S_A/(r-1)$	$\dfrac{MS_A}{MS_E}$
误差	S_E	$n-r$	$MS_E = S_E/(n-r)$	
总和	S_T	$n-1$		

$H_0: \mu_1 = \mu_2 = \cdots = \mu_r = 0,\ H_1: \mu_1, \mu_2, \cdots, \mu_r$不全相等

或，$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0,\ H_1: \alpha_1, \alpha_2, \cdots, \alpha_r$不全为零。

H_0的检验统计量及α水平的拒绝域为：

$$\boxed{F = \frac{S_A/(r-1)}{S_E/(n-r)} \geq F_\alpha(r-1, n-r)}$$

σ^2的估计$\hat{\sigma}^2 = \dfrac{S_E}{n-r} = MS_E$

当拒绝H_0时，求$m_i - m_j$的置信水平为$1-a$的置信区间

$$取\ G = \frac{(\bar{X}_{i\bullet} - \bar{X}_{j\bullet}) - (m_i - m_j)}{s\sqrt{(1/n_i + 1/n_j)}} \Bigg/ \sqrt{\frac{S_E}{s^2}\Big/(n-r)}$$

$$= \frac{(\bar{X}_{i\bullet} - \bar{X}_{j\bullet}) - (m_i - m_j)}{\sqrt{MS_E}\sqrt{\dfrac{1}{n_i} + \dfrac{1}{n_j}}} \sim t(n-r)$$

当拒绝H_0时，求m_i的置信水平为$1-a$的置信区间

$$取函数 G = \frac{\dfrac{\bar{X}_{ig} - m_i}{s / \sqrt{n_i}}}{\sqrt{\dfrac{S_E}{s^2}/(n-r)}} = \frac{\bar{X}_{ig} - m_i}{\sqrt{MS_E}/\sqrt{n_i}} \sim t(n-r)$$

当拒绝 "$H_0: m_1 = m_2 = \mathbf{L} = m_r = 0$"时，作假设：

$$H_0: m_i = m_j \quad , \quad H_1: m_i \neq m_j$$

取 $$t_{ij} = \frac{\overline{X}_{i\bullet} - \overline{X}_{j\bullet}}{\sqrt{MS_E}\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \sim t(n-r)$$

$$H_0: m_i = m_0 \quad , \quad H_1: m_i \neq m_0$$

取 $$t_i = \frac{\frac{\overline{X}_{ig} - m_0}{s / \sqrt{n_i}}}{\sqrt{\frac{S_E}{s^2} / (n-r)}} = \frac{\overline{X}_{ig} - m_0}{\sqrt{MS_E} / \sqrt{n_i}} \sim t(n-r)$$

一元线性回归

$$Y \sim N(a+bx, s^2) \qquad Y = a+bx+e \qquad e \sim N(0, s^2)$$

$$Y_i \sim N(a+bx_i, s^2) \qquad Y_i = a+bx_i+e_i \qquad e_i \sim N(0, s^2)$$

其中未知参数a, b, s^2都不依赖于x的常数！

$$S_{xy} = \sum_i (x_i - \overline{x})(y_i - \overline{y}),\ S_{xx} = \sum_i (x_i - \overline{x})^2,\ S_{yy} = \sum_i (y_i - \overline{y})^2.$$

a, b的最小二乘估计：$\begin{cases} \hat{b} = S_{xy}/S_{xx} \\ \hat{a} = \overline{y} - \overline{x}\hat{b} \end{cases}$

Y关于x的（经验）回归方程：$\hat{y} = \hat{a} + \hat{b}x$

平方和分解：$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i+\hat{y}_i-\bar{y})^2$

$$=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$S_{yy}=SST=SSE+SSR$$

$$SSE=S_{yy}-\hat{b}S_{xy} \qquad \frac{SSE}{s^2}\sim c^2(n-2)$$

s^2 的估计用 $\hat{s}^2=\frac{SSE}{n-2}\overset{定义}{=}S^2$

$$\hat{b}\sim N(b,\frac{s^2}{S_{xx}})$$

线性假设的显著性检验

检验假设 $H_0: b = 0, H_1: b \neq 0,$

当H_0为真即$b = 0$时，取检验统计量

$$t = \frac{\hat{b} - b}{\sqrt{s^2 / S_{xx}}} \Big/ \sqrt{\frac{SSE}{s^2} \Big/ (n-2)} = \frac{\hat{b}}{S / \sqrt{S_{xx}}} \sim t(n-2),$$

回归系数b的置信区间

取函数 $G = \dfrac{\hat{b} - b}{S / \sqrt{S_{xx}}} \sim t(n-2)$

可化为一元线性回归

考试注意事项

1. 考试时间、地点
   时间：2015年1月26日　10:30～12:30
   地点：西2-303*,313,315,415,417
2. 考前答疑时间、地点：
   1月25日 8:30~11:30,13:30~16:30
   东1A－207
3. 考试时按座位号就坐，并要签名，请同时把“作业本编号”抄到试卷的左上角。

考 场：紫 金 港 西 2-205

座位号	学号	姓名	性别	类(专业)	NO	作业本号	签名
2	3120100335	葛佳俊	男	应用生物科学类	1	P097	
4	3120000121	康路夷	男	混合班	2	A097	
6	3120100630	王博雅	男	工科试验班（工学）	3	P011	
8	3120100510	李峥	男	工科试验班（工学）	4	P009	
9	3120101969	石丹妮	女	工科试验班（信息）	5	A043	
11	3110103106	陈云龙	男	机械工程及其自动化	6	P064	
13	3120102020	王一川	男	工科试验班（信息）	7	A050	
15	3120101910	郑来文	女	工科试验班（信息）	8	A042	
18	3110100307	王震宇	男	动物科学	9	P007	
20	3120100462	袁焕杰	男	理科试验班类	10	A111	
22	3120101038	谢欣成	女	工科试验班（工学）	11	P012	
24	3110101254	杜峰百	男	信息工程（光电系）	12	P087	

考试注意事项

- 可带计算器，之前要学会计算器的使用
- 统一阅卷、统一评分标准（注意必要过程）
- 考题类型：一般为填空与计算题
- 平时成绩一般20%～30%