# 期末复习

## 课程难点

- Ch3：条件密度、连续性随机变量的卷积公式
- Ch4：条件期望及其性质
- Ch5：强大数定律，以概率1收敛
- Ch7：联合区间估计，非正态
- Ch8：单边检验，非正态

考试时间：6月29日

考试须知：

1. 记得填写学号，座位号
2. 不能带计算器。
   因此，分位数不用写具体数值（除非另有提示）
   e.g. $\left[\frac{40}{F_{0.975}(7,9)}, \frac{40}{F_{0.025}(7,9)}\right]$
3. 题型：选择 × 12，大题 × 6

古典概率：$C_n^m := \frac{m!(n-m)!}{n!}$ ：n中(无关先后次序)取m个的取法.

例. 带编号1~10的10个球，随机取3个球

问：1. 最小号码为4的概率

2. 大 . . . . 4 . . .

3. 中间 . . . . 4 . . .

解：样本空间含样本数为$C_{10}^3$

1. 事件A = "最小..为4"

= "一个取4" & "另外两个取5,...,10"

$\therefore P(A) = \frac{C_6^2}{C_{10}^3}$

2. 事件B = "最大为4" = . . . . . . . 1,2,3

$P(B) = \frac{C_3^2}{C_{10}^3}$

3. 事件C = "中间为4" = "一个取4" & "一个1,2,3" & "一个5~10"

$P(C) = \frac{C_3^1 \cdot C_6^1}{C_{10}^3}$

条件概率

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}$$

$$= \frac{P(B|A) \cdot P(A)}{\sum_C P(B|C) \cdot P(C)}$$

例：题设同上.

问：1. 已知"最大为4"的条件下，"含有5"的概率

2. 小

3. 中间

解：1. 记 A = "最大为4"，D = "含有5"

可见 A与D互斥，得 $P(D|A) = \frac{P(A \cdot D)}{P(A)} = \frac{0}{P(A)} = 0$

2. B·D = "一个4" & "一个5" & "一个6,...,10"

$P(D|B) = \frac{P(B \cdot D)}{P(B)} = \frac{C_5^1}{P(B)}$

3. C·D = "一个4" & "一个5" & "一个1,2,3"

$P(D|C) = \frac{P(C \cdot D)}{P(C)} = \frac{C_3^1}{P(C)}$

## 1 (10 分)

将 A, B, C 三个字母之一输入信道, 输出为原字母的概率为 $\alpha$, 而输出为其他一个字母的概率为 $(1-\alpha)/2$。

今将字母串 AAAA，BBBB，CCCC 之一输入信道，记输入 AAAA，BBBB，CCCC 的概率分别为 $p_1$, $p_2$, $p_3$，其中 $p_1 + p_2 + p_3 = 1$。

已知输出信号为 $ABCA$，问输入的字母串为 AAAA 的概率是多少?

解：记 $X_i$ 为 ith 输入信号，$Y_i$ 为 ith 输出信号

$$X = X_1X_2X_3X_4, \quad Y = Y_1Y_2Y_3Y_4$$

要求 $P[X=AAAA \mid Y=ABCA]$

$$= \frac{P[X=AAAA, Y=ABCA]}{P[Y=ABCA]} \quad (\text{条件概率定义})$$

$$= \frac{P[Y=ABCA \mid X=AAAA]\cdot P[X=AAAA]}{\sum\limits_{*\in\{A,B,C\}} P[Y=ABCA \mid X=****]\cdot P[X=****]}.$$

计算 $P[Y=y \mid X=****]$

$$= \prod_{i=1}^{4} P[Y_i=y_i \mid X=****] \quad (\because Y_i \text{ 与 } Y_j \text{ 独立})$$

$$= \prod_i P[Y_i=y_i \mid X_i=x_i] \quad (\because Y_i \text{ 与 } X_j \text{ 独立}, j\neq i)$$

$$= \alpha^{\#\{i \mid x_i=y_i\}}\cdot\beta^{\#\{i \mid x_i\neq y_i\}} \quad (\text{where } \beta = 1-\alpha)$$

$$\therefore \begin{cases} P[Y=ABCA \mid X=AAAA] = \alpha^2\cdot\beta^2 \\ P[\quad\cdot\quad\cdot\quad \mid X=BBBB] = \alpha\cdot\beta^3 \\ P[\qquad \mid X=CCCC] = \alpha\cdot\beta^3 \end{cases}$$

$$\therefore \text{原} = \frac{\alpha^2\cdot\beta^2\cdot P_1}{\alpha^2\cdot\beta^2\cdot P_1 + \alpha\cdot\beta^3\cdot P_2 + \alpha\beta^3\cdot P_3} \quad (\text{Def of } P_i)$$

$$= \frac{\alpha\cdot P_1}{\alpha\cdot P_1 + (1-\alpha)\cdot(1-P_1)} \quad (\because P_2+P_3 = 1-P_1)$$

分布函数 $F_X(x) = P(X \le x)$ 满足 $\lim_{x\to-\infty} F(x) = 0$，$\lim_{x\to+\infty} F(x) = 1$

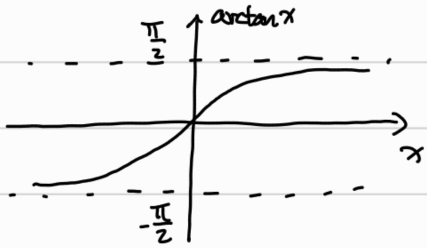
概率密度 $f_X(x_0) = \frac{d}{dx}F_X(x)\Big|_{x=x_0}$ $F_X$ 在 $x_0$ 处连续

例： $X \sim F(x) = A + B\cdot \arctan x$ $(-\infty < x < +\infty)$

问： 1. $A, B = ?$

2. $P[-1 \le X \le 1] = ?$

3. $f_X(x) = ?$




解：① 分布函数满足 $\begin{cases} \lim_{x\to-\infty} F(x) = 0 \\ \lim_{x\to+\infty} F(x) = 1 \end{cases}$

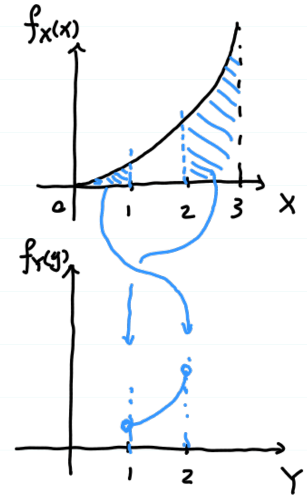
$\Rightarrow \begin{cases} A + B\times(-\frac{\pi}{2}) = 0 \\ A + B\times(\frac{\pi}{2}) = 1 \end{cases} \Rightarrow \begin{cases} A = \frac{1}{2} \\ B = \frac{1}{\pi} \end{cases}$

② $P[-1 \le X \le 1] = F(1) - F(-1)$

$= \left(\frac{1}{2} + \frac{1}{\pi}\arctan 1\right) - \left(\frac{1}{2} + \frac{1}{\pi}\arctan(-1)\right) = \frac{1}{2}$

③ $f_X(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(\frac{1}{2} + \frac{1}{\pi}\arctan x\right) = \frac{1}{\pi(1+x^2)}$ #

## 2 (15 分)




设随机变量 $X$ 的概率密度为

$$f_X(x) := \begin{cases} \frac{1}{9}x^2, & 0 < x < 3 \\ 0, & \text{else} \end{cases}$$

记随机变量

$$Y := \begin{cases} 2, & X \leq 1 \\ X, & 1 < X < 2 \\ 1, & X \geq 2 \end{cases}$$

1. 求 $Y$ 的分布函数 $F_Y$
2. 求 $Y$ 的分布密度函数 $f_Y$
3. 求概率 $P[X \leq Y]$

解：1. 当 $y<1$ 时，$F_Y(y)=0$ （∵ Y 取值范围为 $[1,2]$）

当 $1\le y<2$ 时，$F_Y(y) = P[Y\le y] = P[Y=1] + P[1<Y\le y]$

其中 $P[Y=1] = P[X\ge 2] = \int_2^{+\infty} f_X(x)dx = \int_2^3 \frac{1}{9}x^2 dx = \frac{19}{27}$

$P[1<Y\le y] = P[1<X\le y] = \int_1^y \frac{1}{9}x^2dx = \frac{1}{27}(y^3-1)$

∴此时，$F_Y(y) = \frac{1}{27}(y^3+18)$

当 $y\ge 2$ 时 $F_Y(y)=1$.

综上所述，$F_Y(y) = \begin{cases} 0, & y<1 \\ \frac{1}{27}(y^3+18), & 1\le y<2 \\ 1, & 2\le y \end{cases}$

2. $f_Y(y) = \begin{cases} 0, & y<1 \text{ 或 } y>2 \\ \frac{1}{9}y^2, & 1<y<2. \\ \frac{19}{27}\cdot\delta(0), & y=1 \\ \frac{1}{27}\cdot\delta(0), & y=2 \end{cases}$ （$\delta(x)$ : Dirac $\delta$ function）

3. $P[X\le Y] = P[X\le Y, X\le 1] + P[X\le Y, 1<X<2] + P[X\le Y, X\ge 2]$

其中 $P[X\le Y, X\le 1] = P[X\le 2, X\le 1] = P[X\le 1]$

$P[\ldots, 1<X<2] = P[X\le X, 1<X<2] = P[1<X<2]$

$P[\ldots, X\ge 2] = P[X\le 1, X\ge 2] = 0$

$\therefore P[X\le Y] = P[X\le 1] + P[1<X<2] = F_X(2)$

$= \int_0^2 \frac{1}{9}x^2dx = \frac{8}{27}$ #

# 数字特征

- $\mathbb{E}[X] = \int x f(x) dx$.
  - 线性：$\mathbb{E}[aX+bY] = a\cdot\mathbb{E}[X] + b\cdot\mathbb{E}[Y]$
  - $X, Y$ 独立 $\Rightarrow \mathbb{E}[X\cdot Y] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$.

- $Var[X] = \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
  - $Var[aX] = a^2\cdot Var[X]$
  - $X, Y$ 独立 $\Rightarrow Var[X\pm Y] = Var[X] + Var[Y]$

例. $(X,Y)$ 在 $D=\{(x,y) \mid 0<x<1, |y|<x\}$ 上服从均匀分布

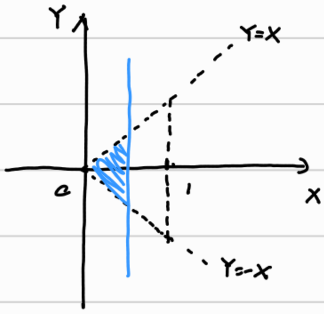
$Z = 2X+1$

问：1. $F_Z$

2. $f_Z$

3. $\mathbb{E}[Z]$

4. $Var[Z]$




解：D 的面积 $\int_{(0,1)}\int_{|y|<x} dy\,dx$

$= \int_0^1\int_{-x}^{x} dy\,dx = 1$

由 $(X,Y)$ 在 D 上服从均匀分布得

$$f_{X,Y}(x,y) = \begin{cases} 1, & (x,y)\in D \\ 0, & \text{else.} \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x<0 \\ \int_0^x\int_{-x}^{x} dy\,dx = x^2, & 0\le x<1 \\ 1, & 1\le x \end{cases}$$

$F_Z(z) = P[Z\le z] = P[2X+1\le z] = P[X\le \frac{1}{2}(z-1)] = F_X(\frac{1}{2}(z-1))$

$$= F_X(\tfrac{1}{2}(z-1)) = \begin{cases} 0, & z<1 \ (\because \frac{1}{2}(z-1)<0) \\ (\frac{1}{2}(z-1))^2 = \frac{1}{4}(z-1)^2, & 1\le z<3 \\ 1, & 3\le z \end{cases}$$

$$f_Z(z) = \begin{cases} \frac{1}{2}(z-1), & 1\le z<3 \\ 0, & \text{else} \end{cases}$$

$$\mathbb{E}[Z] = \int z f_Z(z)\,dz = \int_1^3 z\cdot\frac{1}{2}(z-1)\,dz = \frac{7}{3}$$

$$\mathbb{E}[Z^2] = \int z^2 f_Z(z)\,dz = \int_1^3 z^2\cdot\frac{1}{2}(z-1)\,dz = \frac{17}{3}$$

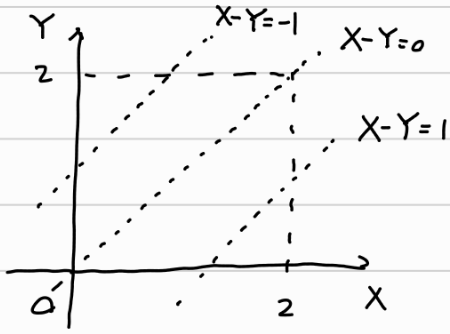
$$Var[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{2}{9}$$

例: $(X,Y)$ 在 $D=\{(x,y)\mid 0\le x\le 2,\ 0\le y\le 2\}$ 服从均匀分布

$Z=|X-Y|$

问: 1. $f_Z(z)$

2. $\mathbb{E}[Z]$ & $Var[Z]$.



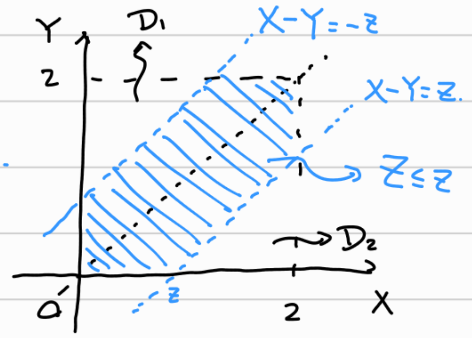

解: $(X,Y)$ 概率密度

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{4}, & (x,y)\in D \\ 0, & else \end{cases}$$

记 $D'(z) = \{(x,y)\in D \mid |x-y|\le z\}$

$$\begin{aligned} Area\,D'(z) &= Area\,D - Area\,D_1 - Area\,D_2 \\ &= 4 - \frac{1}{2}(2-z)^2 - \frac{1}{2}(2-z)^2 \\ &= 4z - z^2 \end{aligned}$$

$(z\ge 0)$ $-z\le X-Y\le z.$

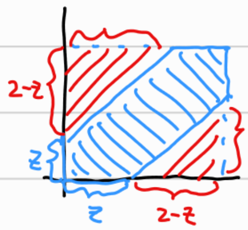



$$\therefore F_Z(z) = P(Z\le z) = P(|X-Y|\le z)$$

$$= \int_{|X-Y|\le z} f_{X,Y}(x,y)\,dxdy$$

$$= \int_{D'(z)} \frac{1}{4}\,dxdy = \frac{1}{4}\int_{D'(z)} dxdy = \frac{1}{4}Area\,D'(z) = -\frac{1}{4}z^2 + z.$$

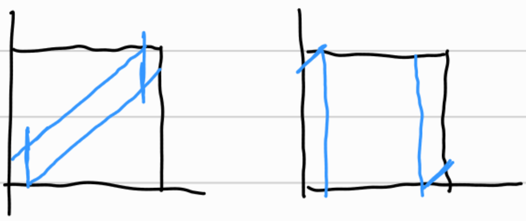
$$\therefore f_Z(z) = \begin{cases} \frac{d}{dz}F_Z(z) = -\frac{1}{2}z+1, & 0\le z<2 \\ 0 & . \quad else \end{cases}$$




$$\mathbb{E}[Z] = \int_0^2 z\cdot\left(-\frac{1}{2}z+1\right)dz = \frac{2}{3}$$

$$\mathbb{E}[Z^2] = \int_0^2 z^2\cdot\left(-\frac{1}{2}z+1\right)dz = \frac{2}{3}$$

$$Var[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{4}{9}$$



(λ≠5!)

## 4 (15 分)

两台相同型号的自动记录仪，每台无故障工作的时间服从指数分布 Exp($\lambda$)。首先开动其中一台，当其发生故障时停用，并自动开动另外一台。记随机变量 $T$ 为两台记录仪无故障工作的总时间。求：

- $T$ 的概率密度 $f_T$
- 数学期望 $\mathbb{E}[T]$ 和方差 $\mathrm{Var}[T]$

解：① $X$：第一台记录仪的工作时间，$Y$：另一台 . . .

$\Rightarrow T = X + Y$

$f_T(t) = \int_0^t f_X(s) \cdot f_{Y|X}(t-s \mid s)\, ds$ （∵ 全概率公式）

$= \int_0^t f_X(s) \cdot f_Y(t-s)\, ds$ （∵ $X, Y$ 独立）

$= \int_0^t \lambda \cdot e^{-\lambda s} \cdot \lambda \cdot e^{-\lambda(t-s)} ds$

$= \lambda^2 t \cdot e^{-\lambda t}$

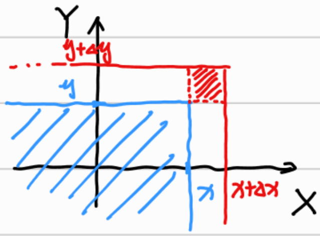
$$\therefore f_T(t) = \begin{cases} \lambda^2 t \cdot e^{-\lambda t}, & t > 0 \\ 0, & \text{else} \end{cases}$$

② $\mathbb{E}[T] = \mathbb{E}[X] + \mathbb{E}[Y] = 2\lambda^{-1}$ （∵ $X, Y \sim Exp(\lambda)$ ∴ $\mathbb{E}X = \mathbb{E}Y = \lambda^{-1}$）

$\mathrm{Var}[T] = \mathrm{Var}[X] + \mathrm{Var}[Y] = 2\lambda^{-2}$ （$\mathrm{Var}X = \mathrm{Var}Y = \lambda^{-2}$）

#

# 独立 vs 不相关




- $X, Y$ 相互独立.

  $\Leftrightarrow$ 分布: $F_{X,Y}(x,y) = F_X(x) \cdot F_Y(y)$.

  $\Leftrightarrow$ 密度: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$.

- $X, Y$ 相关

  $\Leftrightarrow$ 相关系数 $r(X,Y) \neq 0$

- $X, Y$ 不相关

  $\Leftrightarrow$ $r(X,Y) = 0$ (验证 $\mathrm{Cov}(X,Y) = 0$ 即可).

  $r(X,Y) = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X} \cdot \sqrt{\mathrm{Var}Y}}$, $\quad \mathrm{Cov}(X,Y) = E[XY] - E[X] \cdot E[Y]$

- $X, Y$ 具有线性关系: $\exists\, a \neq 0, b$ s.t. $Y = aX + b$ for almost $\omega \in \Omega$ w.r.t. $P$.

  (i.e. $P[Y = aX + b] = 1$)

  $\Leftrightarrow$ $|r(X,Y)| = 1$

$F_{X,Y} = F_X \cdot F_Y$ $\leftrightarrow$ 独立 $\leftrightarrow$ $f_{X,Y} = f_X \cdot f_Y$

独立 $\Rightarrow$ 不相关 $\longleftrightarrow$ $\mathrm{Cov}(X,Y) = 0$

不相关 $\overset{?}{\Rightarrow}$ 独立

- 不成立: 例 4.2.1.
- 成立: 只有正态分布

  (i.e. $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$)

## 3 (15 分)

设二维随机变量 $(X,Y)$ 在区域 $D=\{(x,y)|0<x<1, x^2<y<\sqrt{x}\}$ 上服从均匀分布，令

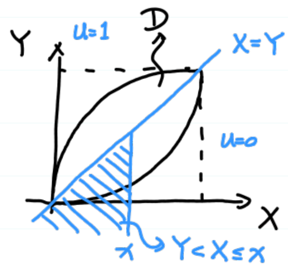
$$U=\begin{cases}1, & X\le Y,\\ 0, & X>Y\end{cases}$$

1. 求 $(X,Y)$ 的概率密度 $f_{X,Y}$
2. 说明 $U$ 与 $X$ 是否相互独立？是否相关？
3. 求 $Z=U+X$ 的分布函数 $F_Z(z)$

解：1. D的面积 $=\int_0^1\int_{x^2}^{\sqrt{x}} dy\,dx = \frac{1}{3}$

由 (X,Y) 服从均匀分布得

$$f_{X,Y}(x,y)=\begin{cases}3, & (x,y)\in D\\ 0, & \text{else.}\end{cases}$$




2. 独立性：验证 $F_{U,X}(u,x) \overset{?}{=} F_U(u)\cdot F_X(x)$

由于 U 取值为 $\{0,1\}$，只考察 $u=0$

$F_{U,X}(0,x) = P[U\le 0, X\le x] = P[Y<X\le x]$

$= \int_0^x\int_{x^2}^{x} 3\cdot dy\,dx = x^2(\frac{3}{2}-x)$

$F_U(0) = P[Y<X] = \frac{1}{2}$

$F_X(x) = \int_0^x\int_{x^2}^{\sqrt{x}} 3\cdot dy\,dx = 2\cdot x^{\frac{3}{2}} - x^3$

$\Rightarrow F_{U,X}(0,x) \ne F_U(0)\cdot F_X(x) \Rightarrow$ 不独立

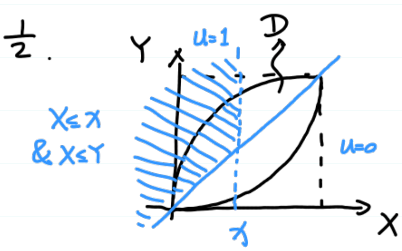
相关性：$P[U=0]=P[U=1]=\frac{1}{2}$. $\Rightarrow E[U]=\frac{1}{2}$.

$f_X(x) = 3\sqrt{x}-3x^2 \Rightarrow E[X]=\frac{9}{20}$

$f_{U,X}(1,x) = \frac{d}{dx}P[U=1, X\le x]$

$= \frac{d}{dx}\int_0^x\int_x^{\sqrt{x}} 3\,dy\,dx$

$= 3\sqrt{x}-3x$




$\therefore E[U\cdot X] = \int_{\{0,1\}\times[0,1]} ux\cdot f_{U,X}(u,x)\,du\,dx = \int_{\{1\}\times[0,1]} ux\, f_{U,X}(u,x)\,du\,dx$

$= \int_{[0,1]} x\, f_{U,X}(1,x)\,dx = \int_0^1 x(3\sqrt{x}-3x)\,dx = \frac{1}{5}$

$\therefore E[U\cdot X] - EU\cdot EX = \frac{1}{5}-\frac{1}{2}\times\frac{9}{20} = -\frac{1}{40} < 0 \Rightarrow$ 负相关

3. $F_Z(z) = P(U+X\le z) = P(U+X\le z, U=0) + P(U+X\le z, U=1)$

where $P(U+X\le z, U=0) = P(X\le z, X>Y) = \int_0^z\int_{x^2}^{x} 3\,dy\,dx = \frac{3}{2}z^2 - z^3$

$P(U+X\le z, U=1) = P(1+X\le z, X\le Y) = \int_0^{z-1}\int_x^{\sqrt{x}} 3\,dy\,dx = 2(z-1)^{\frac{3}{2}} - \frac{3}{2}(z-1)^2$

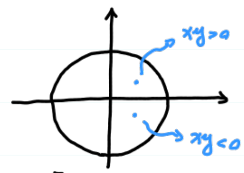
| $z$ | $(-\infty,0)$ | $[0,1)$ | $[1,2)$ | $[2,+\infty)$ |
|---|---|---|---|---|
| $F_Z(z)$ | 0 | $\frac{3}{2}z^2-z^3$ | $\frac{1}{2}+2(z-1)^{\frac{3}{2}}-\frac{3}{2}(z-1)^2$ | 1 |

## 5 (15 分)

设二维随机变量 $(X, Y)$ 的概率密度为

$$f(x,y)=\begin{cases}\frac{1}{\pi}, & x^2+y^2\le 1\\ 0, & \text{else}\end{cases}$$

说明 $X$ 与 $Y$ 的相关性和独立性。




解：相关性：$E[XY] = \iint_{\mathbb{R}^2} xy\, f_{X,Y}(x,y)\,dxdy$

$= \int_0^{+\infty}\left[x\cdot\int_{-\infty}^{+\infty} y\, f_{X,Y}(x,y)\,dy + (-x)\cdot\int_{-\infty}^{+\infty} y\, f_{X,Y}(-x,y)\,dy\right]dx$

$= \int_0^{+\infty} x\cdot\left(\int_{-\infty}^{+\infty} y\, f_{X,Y}(x,y)\,dy - \int_{-\infty}^{+\infty} y\, f_{X,Y}(x,y)\,dy\right)dx$ $\quad(\because f_{X,Y}(-x,y)=f_{X,Y}(x,y))$

$= 0$

$E[X] = \int_{-\infty}^{+\infty} x\cdot\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy\,dx = \int_{-1}^{1} x\cdot\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}\,dy\,dx$

$= \left(\int_{-1}^{0}+\int_{0}^{1}\right) x\cdot\int \cdots dy\,dx$

$= \int_0^1 (x+(-x))\int \cdots dy\,dx = 0$

$E[Y] = 0$

$\therefore E[X\cdot Y] - E[X]\cdot E[Y] = 0$

$\Rightarrow$ $X, Y$ 不相关

独立性：对 $x\in(-1,1)$，$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}\,dy = \frac{2}{\pi}\sqrt{1-x^2}$

对 $y\in(-1,1)$，$f_Y(y) = \frac{2}{\pi}\sqrt{1-y^2}$

则 $f_X(x)\cdot f_Y(y) = \frac{4}{\pi^2}\sqrt{(1-x^2)(1-y^2)}$

$\neq \frac{1}{\pi} = f_{X,Y}(x,y)$ for $x^2+y^2\le 1$.

#

例. 已知 $X \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ , $Y \sim \begin{pmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$ , $P[X=Y]=\frac{1}{4}$

求 相关系数 $\rho_{X,Y} = ?$

解: 由 $X=Y \iff X=1 \ \& \ Y=1$ 得

$P[X=1, Y=1] = P[X=Y] = \frac{1}{4}$

$\Rightarrow$

| Y \ X | −1 | 1 |
|---|---|---|
| 0 | 0 | $\frac{1}{4}$ $(= P(X=1) - P(X=1,Y=1))$ |
| 1 | $\frac{1}{2}$ $(= P(Y=1) - P(X=1,Y=1))$ | $\frac{1}{4}$ |

$\therefore \ \mathbb{E}[X]=0$ , $Var[X]=1$ , $\mathbb{E}[Y]=\frac{3}{4}$ , $Var[Y]=\frac{3}{16}$

$\mathbb{E}[XY] = -\frac{1}{4}$ , $Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = -\frac{1}{4}$

$\therefore \ \rho_{XY} = \dfrac{Cov(X,Y)}{\sqrt{Var[X]} \cdot \sqrt{Var[Y]}} = -\frac{1}{\sqrt{3}}$ #

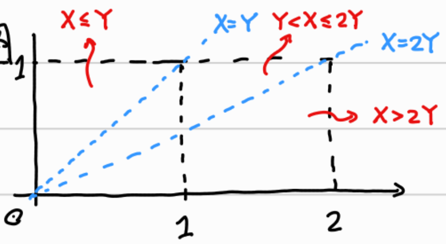
例. $(X,Y)$ 在 $D = \{(x,y) \mid 0 \le x \le 2, 0 \le y \le 1\}$ 服从均匀分布

取 $U = \begin{cases} 0, & \text{if } X \le Y \\ 1, & \text{else} \end{cases}$ $\qquad V = \begin{cases} 0, & \text{if } X \le 2Y \\ 1, & \text{else} \end{cases}$

求: 1. 联合分布 $f_{UV}$

2. 相关系数 $\rho_{UV}$

解: 画图




$P[X \le Y] = \frac{1}{4}$

$P[X > 2Y] = \frac{1}{2}$

$\Rightarrow P[Y < X \le 2Y] = \frac{1}{4}$

$\therefore$

| V \ U | 0 | 1 |
|---|---|---|
| 0 | $\frac{1}{4}$ $(= P[X<Y])$ | $\frac{1}{4}$ $(= P[Y<X\le 2Y])$ |
| 1 | $0$ $(= P[X\le Y, X>2Y])$ | $\frac{1}{2}$ $(= P[X>2Y])$ |

$\therefore \ U \cdot V \sim \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ , $U \sim \begin{bmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{bmatrix}$ , $V \sim \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

$\therefore \ \mathbb{E}[U] = \frac{3}{4}$ , $Var[U] = \frac{3}{16}$ , $\mathbb{E}[V] = \frac{1}{2}$ , $Var[V] = \frac{1}{4}$

$\mathbb{E}[UV] = \frac{1}{2}$ , $Cov(U,V) = \mathbb{E}[UV] - \mathbb{E}[U] \cdot \mathbb{E}[V] = \frac{1}{8}$

$\therefore \ \rho_{UV} = \dfrac{Cov(U,V)}{\sqrt{VarU} \cdot \sqrt{VarV}} = \frac{1}{\sqrt{3}}$ #

## 6 (15 分)

已知 $X$ 的概率密度为

$$f(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

其中 $\theta > 0$ 为未知参数。设 $X_1, \ldots, X_n$ 为取自 $X$ 的一个样本。

1. 求 $\theta$ 的矩法估计量 $\hat{\theta}_M$
2. 求 $\theta$ 的最大似然估计量 $\hat{\theta}_L$
3. 说明 $\hat{\theta}_M$ 和 $\hat{\theta}_L$ 的无偏性

解: ① $\mathbb{E}[X] = \int_0^{+\infty} x \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \theta \int_0^{+\infty} y e^{-y} dy = \theta$

$\therefore$ 由 $\theta = \mathbb{E}[X]$ 得 矩法估计量 $\hat{\theta}_M = \overline{X}$

② 似然函数 $\ln P[X_1, \ldots, X_n; \theta] = \ln \prod_i P[X_i; \theta]$

$= \ln\left( \frac{1}{\theta^n} \cdot e^{-\frac{1}{\theta}\sum x_i} \right)$

$= -n \cdot \ln\theta - \frac{1}{\theta} \cdot \sum_i x_i$

要使似然函数最大的必要条件

$\frac{d}{d\theta} \ln P[\ldots] = -\frac{n}{\theta} + \frac{1}{\theta^2} \cdot \sum x_i = 0 \Rightarrow \theta^* = \frac{1}{n}\sum_i x_i = \overline{X}$

验证 此时最大:

$\frac{d^2}{d\theta^2} \ln P[\ldots] \Big|_{\theta=\theta^*} = \frac{n}{\theta^2} - \frac{2}{\theta^3}\sum_i x_i \Big|_{\theta=\theta^*} = \frac{n}{\overline{X}^2} - \frac{2n}{\overline{X}^3} \cdot \overline{X} = -\frac{n}{\overline{X}^2} < 0$

$\therefore$ 最大似然估计量 $\hat{\theta}_L = \overline{X}$

③ 记 $\hat{\theta} = \overline{X}$，验证 $\hat{\theta}$ 的无偏性.

$\mathbb{E}_{X_1, \ldots, X_n}[\hat{\theta}] = \mathbb{E}_{X_1 \ldots X_n}[\overline{X}] = \mathbb{E}_{X_1 \ldots X_n}[\frac{1}{n}(X_1 + \ldots + X_n)]$

$= \frac{1}{n} \cdot n \cdot \mathbb{E}[X] = \theta$

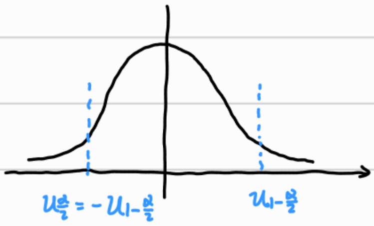
$\therefore$ $\hat{\theta}_M$ 和 $\hat{\theta}_L$ 的无偏估计量 #

## 常见统计量

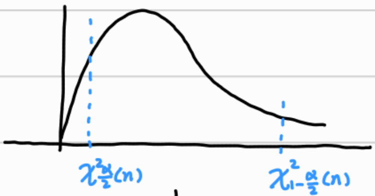
- 样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- 样本方差 $S_n^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})$
- 修正样本方差 $S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})$

标准.

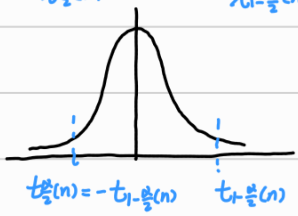
- 正态分布：$X \sim N(0,1)$



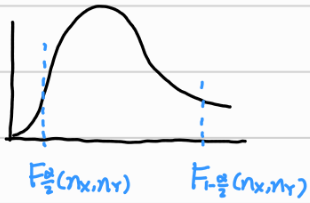

- $\chi^2$分布：$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$




- t分布：$\dfrac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}} \sim t(n)$




- F分布：$\dfrac{\frac{1}{n_X}\cdot\sum_{i=1}^{n_X} X_i^2}{\frac{1}{n_Y}\cdot\sum_{j=1}^{n_Y} Y_j^2} \sim F(n_X, n_Y)$




- 常用统计量的分布
  - 单个 $X \sim N(\mu, \sigma^2)$

    $\mu, \sigma$ ∘ $\bar{X} \sim N(\mu, \frac{1}{n}\sigma^2)$, $\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

    $\mu$ ∘ $\dfrac{\bar{X}-\mu}{\sqrt{S_n^2/(n-1)}} = \dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \Big/ \sqrt{\dfrac{n\cdot S_n^2}{\sigma^2}/(n-1)} \sim t(n-1)$

    $\mu, \sigma$ ∘ $\sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$

    $\sigma$ ∘ $\dfrac{n\cdot S_n^2}{\sigma^2} \sim \chi^2(n-1)$

  - 两个 $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$

    $\mu_X - \mu_Y$
    $\sigma_X^2, \sigma_Y^2$ ∘ $\dfrac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{\sigma_X^2}{n_X} + \dfrac{\sigma_Y^2}{n_Y}}} \sim N(0,1)$

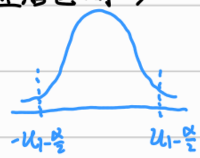
    $\mu_X - \mu_Y$
    $\sigma_X^2 = \sigma_Y^2 = ?$ ∘ $\dfrac{(\bar{X}-\bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\dfrac{1}{n_X} + \dfrac{1}{n_Y}} \cdot \sqrt{\dfrac{n_X S_{n_X}^2 + n_Y S_{n_Y}^2}{n_X + n_Y - 2}}} \sim t(n_X + n_Y - 2)$

    $\sigma_X^2 / \sigma_Y^2$
    $\mu_X, \mu_Y$ ∘ $\dfrac{\frac{1}{\sigma_X^2}\sum_i (X_i - \mu_X)^2 / n_X}{\frac{1}{\sigma_Y^2}\sum_j (Y_j - \mu_Y)^2 / n_Y} \sim F(n_X, n_Y)$

    $\sigma_X, \sigma_Y$ ∘ $\dfrac{n_X S_{n_X}^2 / (n_X - 1)}{n_Y S_{n_Y}^2 / (n_Y - 1)} \sim F(n_X - 1, n_Y - 1)$

# 区间估计（求置信度为 1-α 的置信区间）

- 假设 $X \sim N(\mu, \sigma^2)$
  - 求 $\mu$




$$-u_{1-\frac{\alpha}{2}} \le \frac{\bar X-\mu}{\sqrt{\sigma_0^2/n}} \le u_{1-\frac{\alpha}{2}}$$

    - $\sigma^2=\sigma_0^2$

$$\frac{\bar X-\mu}{\sqrt{\sigma_0^2/n}} \sim N(0,1) \Rightarrow \bar X \pm \sqrt{\frac{\sigma_0^2}{n}}\cdot u_{1-\frac{\alpha}{2}}$$

$\sigma_0^2 \leftarrow \frac{1}{n-1}\sum_i (X_i-\bar X)^2 = \frac{n\cdot S_n^2}{n-1}$ i.e. $\frac{\sigma_0^2}{n} \leftarrow \frac{S_n^2}{n-1}$

$N(0,1) \leftarrow t(n-1)$

    - $\sigma^2 = ?$

$$\frac{\bar X-\mu}{\sqrt{S_n^2/(n-1)}} \sim t(n-1) \Rightarrow \bar X \pm \sqrt{\frac{S_n^2}{n-1}}\cdot t_{1-\frac{\alpha}{2}}(n-1)$$

  - 求 $\sigma$
    - $\mu=\mu_0$:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} (X_i-\mu_0)^2 \sim \chi^2(n) \Rightarrow \frac{1}{\chi^2_{1-\frac{\alpha}{2}}(n)}\times\sum(X_i-\mu_0)^2 \le \sigma^2 \le \frac{1}{\chi^2_{\frac{\alpha}{2}}(n)}\times\sum_i(\cdot\cdot)^2$$

$\mu_0 \leftarrow \bar X \Rightarrow \sum(X_i-\mu_0)^2 \leftarrow n\cdot S_n^2$

$\chi^2(n) \leftarrow \chi^2(n-1)$

    - $\mu = ?$:

$$\frac{1}{\sigma^2}\cdot nS_n^2 \sim \chi^2(n-1) \Rightarrow \frac{1}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\cdot nS_n^2 \le \sigma \le \cdots$$

- 假设 $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$
  - 求 $\mu_X-\mu_Y$
    - $\sigma_X^2=\sigma_Y^2=\sigma_0^2$:

$$\frac{(\bar X-\bar Y)-(\mu_X-\mu_Y)}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sigma_0} \sim N(0,1) \Rightarrow (\bar X-\bar Y)\pm u_{1-\frac{\alpha}{2}}\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sigma_0$$

$\sigma_0^2 \leftarrow \frac{\sum(X_i-\bar X)^2+\sum(Y_j-\bar Y)^2}{n_X+n_Y-2} = \frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}$

$N(0,1) \leftarrow t(n_X+n_Y-2)$

    - $\sigma_X^2=\sigma_Y^2=?$

$$\frac{(\bar X-\bar Y)-(\mu_X-\mu_Y)}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sqrt{\frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}}} \sim t(n_X+n_Y-2) \Rightarrow (\bar X-\bar Y) \pm t_{1-\frac{\alpha}{2}}(n_X+n_Y-2)\cdot\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sqrt{\frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}}$$

  - 求 $\frac{\sigma_1^2}{\sigma_2^2}$
    - $\mu_X, \mu_Y$ ✓

$$\frac{\frac{1}{\sigma_X^2}\cdot\sum_i(X_i-\mu_X)^2/n_X}{\frac{1}{\sigma_Y^2}\cdot\sum_j(Y_j-\mu_Y)^2/n_Y} \sim F(n_X,n_Y) \Rightarrow \left[\frac{1}{F_{1-\frac{\alpha}{2}}(n_X,n_Y)}\cdot\frac{n_Y}{n_X}\cdot\frac{\sum(X_i-\mu_X)^2}{\sum(Y_j-\mu_Y)^2}, \frac{1}{F_{\frac{\alpha}{2}}(n_X,n_Y)}\cdot\cdots\cdots\right]$$

$\mu_\cdot \leftarrow \bar{\cdot} \Rightarrow \sum(\cdot_i-\mu_\cdot)^2 \leftarrow n_\cdot\cdot S_{n_\cdot}^2$

$F(n_X,n_Y) \leftarrow F(n_X-1,n_Y-1)$

    - $\mu_X, \mu_Y$ ?

$$\frac{\frac{1}{\sigma_X^2}\cdot n_X S_{n_X}^2/(n_X-1)}{\frac{1}{\sigma_Y^2}\cdot n_Y S_{n_Y}^2/(n_Y-1)} \sim F(n_X-1,n_Y-1) \Rightarrow \left[\frac{1}{F_{1-\frac{\alpha}{2}}(n_X-1,n_Y-1)}\cdot\frac{n_Y-1}{n_X-1}\cdot\frac{n_X S_{n_X}^2}{n_Y S_{n_Y}^2}, \frac{1}{F_{\frac{\alpha}{2}}(n_X-1,n_Y-1)}\cdot\cdots\right]$$

- 井
  - 单 → 求
    - $\mu$: $\dfrac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$ $\xrightarrow{\sigma^2 \leftarrow \frac{1}{n-1}\cdot n S_n^2}$ $\dfrac{\bar{X}-\mu}{\sqrt{S_n^2/(n-1)}} \sim t(n-1)$
    - $\sigma$: $\dfrac{1}{\sigma^2}\sum_i (X_i-\mu)^2 \sim \chi^2(n)$ $\xrightarrow{\mu \leftarrow \bar{X}}$ $\dfrac{1}{\sigma^2}\cdot n S_n^2 \sim \chi^2(n-1)$
  - 双 → 求
    - $\mu_1-\mu_2$: $\dfrac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sigma_0} \sim N(0,1)$ $\xrightarrow{\sigma_0 \leftarrow \frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}}$ $\dfrac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sqrt{\frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}}} \sim t(n_X+n_Y-2)$
    - $\dfrac{\sigma_1^2}{\sigma_2^2}$: $\dfrac{\frac{1}{\sigma_X^2}\sum(X_i-\mu_X)^2/n_X}{\frac{1}{\sigma_Y^2}\sum(Y_j-\mu_Y)^2/n_Y} \sim F(n_X,n_Y)$ $\xrightarrow{\mu_\cdot \leftarrow \bar{\cdot}}$ $\dfrac{\frac{1}{\sigma_X^2}\cdot n_X S_{n_X}^2/(n_X-1)}{\frac{1}{\sigma_Y^2}\cdot n_Y S_{n_Y}^2/(n_Y-1)} \sim F(n_X-1,n_Y-1)$

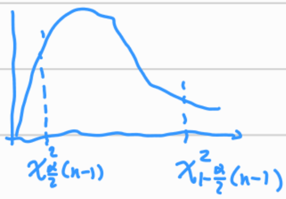
例. $X \sim N(\mu,\sigma^2)$，其中$\mu,\sigma^2$未知，容量为n的样本 $X_1,\dots,X_n$，分别求 $\mu$ 与 $\sigma^2$ 置信度为 $1-\alpha$ 的置信区间.

解:

∘ $\dfrac{\bar{X}-\mu}{\sqrt{S_n^2/(n-1)}} \sim t(n-1) \Rightarrow -t_{1-\frac{\alpha}{2}}(n-1) \le \dfrac{\bar{X}-\mu}{\sqrt{S_n^2/(n-1)}} \le t_{1-\frac{\alpha}{2}}(n-1)$

$\therefore \mu$ ... 置信区间: $\left[\bar{X}-\sqrt{\frac{S_n^2}{n-1}}\cdot t_{1-\frac{\alpha}{2}}(n-1),\ \bar{X}+\sqrt{\frac{S_n^2}{n-1}}\cdot t_{1-\frac{\alpha}{2}}(n-1)\right]$

∘ $\dfrac{1}{\sigma^2}\cdot n\cdot S_n^2 \sim \chi^2(n-1) \Rightarrow \chi^2_{\frac{\alpha}{2}}(n-1) \le \dfrac{1}{\sigma^2}\cdot n S_n^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)$




$\therefore \sigma$ .... 置信区间: $\left[\dfrac{1}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\times n S_n^2,\ \dfrac{1}{\chi^2_{\frac{\alpha}{2}}(n-1)}\times n S_n^2\right]$ #

假设检验 ⇝ 验证是否落在接受域（置信区间）

\# → 单 → ?

- $\mu=\mu_0$: $\dfrac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}} \sim N(0,1) \xrightarrow{\sigma^2 \leftarrow \frac{1}{n-1}\cdot nS_n^2} \dfrac{\bar{X}-\mu_0}{\sqrt{S_n^2/(n-1)}} \sim t(n-1)$
- $\sigma=\sigma_0$: $\dfrac{1}{\sigma_0^2}\sum_i (X_i-\mu)^2 \sim \chi^2(n) \xrightarrow{\mu \leftarrow \bar{X}} \dfrac{1}{\sigma_0^2}\cdot nS_n^2 \sim \chi^2(n-1)$

\# → 双 → ?

- $\mu_X-\mu_Y=\delta$: $\dfrac{(\bar{X}-\bar{Y})-\delta}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sigma_0} \sim N(0,1) \xrightarrow{\sigma_0 \leftarrow \frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}} \dfrac{(\bar{X}-\bar{Y})-\delta}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sqrt{\frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}}} \sim t(n_X+n_Y-2)$
- $\dfrac{\sigma_X^2}{\sigma_Y^2}=k$: $\dfrac{1}{k}\cdot\dfrac{\sum(X_i-\mu_X)^2/n_X}{\sum(Y_j-\mu_Y)^2/n_Y} \sim F(n_X,n_Y) \xrightarrow{\mu_\cdot \leftarrow \bar{\cdot}} \dfrac{1}{k}\cdot\dfrac{n_X S_{n_X}^2/(n_X-1)}{n_Y S_{n_Y}^2/(n_Y-1)} \sim F(n_X-1,n_Y-1)$

例. $X \sim N(\mu_X,\sigma_X^2)$, $Y \sim N(\mu_Y,\sigma_Y^2)$,

样本 $X_1,\dots,X_{n_X}$, $Y_1,\dots,Y_{n_Y}$.

在显著性水平 $\alpha$ 下检验假设.

1). 当 $\sigma_X^2=\sigma_Y^2$ 时, $H_0: \mu_X-\mu_Y=\delta$; $H_1: \mu_X-\mu_Y\neq\delta$ （Q: 没有假设 $\sigma_X^2=\sigma_Y^2$ 时如何计算？）

2). $H_0: \sigma_X^2=\sigma_Y^2$, $H_1: \sigma_X^2\neq\sigma_Y^2$

解: 1) 当 $H_0$ 为真 & $\sigma_X^2=\sigma_Y^2$ 时,

$$\frac{\bar{X}-\bar{Y}-\delta}{\sqrt{\frac{1}{n_X}+\frac{1}{n_Y}}\cdot\sqrt{\frac{n_X S_{n_X}^2+n_Y S_{n_Y}^2}{n_X+n_Y-2}}} \sim t(n_X+n_Y-2)$$

得 ----- 的显著性水平 $\alpha$ 的接受域为 $[-t_{1-\frac{\alpha}{2}}(n_X+n_Y-2),\ t_{1-\frac{\alpha}{2}}(n_X+n_Y-2)]$

2) 当 $H_0$ 为真时, $\dfrac{n_X S_{n_X}^2/(n_X-1)}{n_Y S_{n_Y}^2/(n_Y-1)} \sim F(n_X-1,n_Y-1)$

得 . . . . . . . $[F_{\frac{\alpha}{2}}(n_X-1,n_Y-1),\ F_{1-\frac{\alpha}{2}}(n_X-1,n_Y-1)]$

## 7 (15 分)

为比较不同季节出生的女婴体重的方差，从某年 12 月和 6 月出生的女婴中分别随机地选取 6 名及 7 名，测其体重（单位：g）如下表所示

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 12 月 $X$ | 3520 | 2960 | 2560 | 2960 | 3260 | 3960 | |
| 6 月 $Y$ | 3220 | 3220 | 3760 | 3000 | 2920 | 3740 | 3060 |

假定 $X$ 和 $Y$ 分别服从正态分布 $N(\mu_X, \sigma_X^2)$ 和 $N(\mu_Y, \sigma_Y^2)$。利用双侧检验，在显著性水平 $\alpha = 0.05$ 下，检验以下假设：

假设 $\sigma_X^2 = \sigma_Y^2$

1. $H_0 : \mu_1 = \mu_2$，$H_1 : \mu_1 \neq \mu_2$
2. $H_0 : \sigma_1^2 = \sigma_2^2$，$H_1 : \sigma_1^2 \neq \sigma_2^2$

解：① 假设 $\sigma^2 = \sigma_X^2 = \sigma_Y^2$

选取 $$T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n_X-1)S_X^{*2} + (n_Y-1)S_Y^{*2}}{n_X+n_Y-2}} \cdot \sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

当 $H_0$ 为真时，$T \sim t(n_X+n_Y-2)$

$\Rightarrow$ 显著性水平为 $\alpha$ 的拒绝域为 $|T| > t_{\frac{\alpha}{2}}(n_X+n_Y-2)$

$\Rightarrow$ 代入计算 $T$

并比较 $|T|$ 与 $t_{\frac{\alpha}{2}}(n_X+n_Y-2)$ 关系（略）

② 选取 $$F = \frac{S_X^{*2}}{S_Y^{*2}}$$

当 $H_0$ 为真时，$F \sim F(n_X-1, n_Y-1)$

$\Rightarrow$ 显著性水平为 $\alpha$ 的拒绝域为 $F < F_{\frac{\alpha}{2}}(n_X-1, n_Y-1) \cup F > F_{1-\frac{\alpha}{2}}(n_X-1, n_Y-1)$

$\Rightarrow$ 代入计算 $F$

并比较 $F$，$F_{\frac{\alpha}{2}}(n_X-1, n_Y-1)$，$F_{1-\frac{\alpha}{2}}(n_X-1, n_Y-1)$ 的大小（略） #