应用统计

第3讲　　多元正态，条件分布与统计量

二元正态分布

$$(X,Y)\sim N\left(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho\right)$$

$$p(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\frac{-1}{2\left(1-\rho^2\right)}\left[\frac{\left(x-\mu_1\right)^2}{\sigma_1^2}-2\rho\frac{\left(x-\mu_1\right)\left(y-\mu_2\right)}{\sigma_1\sigma_2}+\frac{\left(y-\mu_2\right)^2}{\sigma_2^2}\right]\right\}$$

令： $\Sigma=\begin{pmatrix}\sigma_1{}^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2{}^2\end{pmatrix}$， $\mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}$， $x=\begin{pmatrix}x_1\\ x_2\end{pmatrix}$

则有 $p(x_1,x_2)=\dfrac{1}{2\pi\left|\Sigma\right|^{\frac{1}{2}}}\exp\left\{-\dfrac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$。

多元正态分布 $\eta \sim N_n(\theta, \Sigma)$

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{pmatrix} \qquad \theta = \begin{pmatrix} E(\eta_1) \\ E(\eta_2) \\ \vdots \\ E(\eta_n) \end{pmatrix} \qquad \Sigma = \begin{pmatrix} Var(\eta_1) & Cov(\eta_1,\eta_2) & \cdots & Cov(\eta_1,\eta_n) \\ Cov(\eta_2,\eta_1) & Var(\eta_2) & \cdots & Cov(\eta_2,\eta_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\eta_n,\eta_1) & Cov(\eta_n,\eta_2) & \cdots & Var(\eta_n) \end{pmatrix}$$

$$f_\eta(y_1,\cdots,y_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(y-\theta)'\Sigma^{-1}(y-\theta)\right\}$$

定理 $\eta \sim N_n(\theta, \Sigma)$，则$\eta_1,\cdots,\eta_n$相互独立的充分必要条件是$\eta_1,\cdots,\eta_n$两两不相关。

多元正态分布 $\eta \sim N_n(\theta,\Sigma)$

$$\Sigma \overset{\text{正交对角化}}{=} T\Lambda T' \qquad TT'=I_n \qquad \Lambda = diag(\lambda_1,\cdots,\lambda_n) \qquad -\frac{1}{2}x'\Lambda^{-1}x = -\frac{1}{2}\left(\frac{x_1^2}{\lambda_1}+\cdots+\frac{x_n^2}{\lambda_n}\right)$$

$$\iint\limits_{y\in R^n} f_\eta(y_1,\cdots,y_n)dy = \iint\limits_{y\in R^n} \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y-\theta)'\Sigma^{-1}(y-\theta)\right\}dy$$

$$= \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}(y-\theta)'(T\Lambda T')^{-1}(y-\theta)\right\}dy_1\cdots dy_n$$

$$\overset{x=T'(y-\theta)}{=} \frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}x'\Lambda^{-1}x\right\}dx_1\cdots dx_n$$

$$= \frac{1}{(2\pi)^{n/2}(\lambda_1\cdots\lambda_n)^{\frac{1}{2}}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}\frac{x_1^2}{\lambda_1}\right\}dx_1\cdots\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}\frac{x_n^2}{\lambda_n}\right\}dx_n = \prod_{k=1}^{n}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi\lambda_k}}\exp\left\{-\frac{1}{2}\frac{x_k^2}{\lambda_k}\right\}dx_k = 1$$

引理 1.2.1 设 $\eta_1,\cdots,\eta_n$ 为相互独立同服从正态分布 $N(0,1)$ 的 n 个随机变量，T 为 n 阶正交矩阵，$\zeta=T'\eta$，其中 $\zeta=(\zeta_1,\cdots,\zeta_n)'$，$\eta=(\eta_1,\cdots,\eta_n)'$. 则 $\zeta_1,\cdots,\zeta_n$ 也为相互独立同服从正态分布 $N(0,1)$ 的 n 个随机变量，即 $\zeta\sim N_n(0,I_n)$，其中 I_n 为 n 阶单位矩阵.

证明 因为 $f_\eta(y_1,\cdots,y_n)=(2\pi)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2}y'y\right\}$，其中 $y=(y_1,y_2,\cdots,y_n)'$. 设 $x=(x_1,x_2,\cdots,x_n)'$，又因为变换 $y=Tx$ 的雅可比行列式的绝对值为 $|J|=\|T\|=1$，且 $y'y=x'T'Tx=x'x$. 所以由文献[21]中定理 2.7.5 知，ζ 的密度函数为

$$f_\zeta(x_1,\cdots,x_n)=(2\pi)^{-n/2}\exp\left\{-\frac{1}{2}x'TT'x\right\}|J|$$

$$=(2\pi)^{-n/2}\exp\left\{-\frac{1}{2}x'x\right\},$$

故

$$\zeta\sim N_n(0,I_n).$$

推论 1.2.1 设 $\eta\sim N_n(\theta,\sigma^2I_n)$，$T$ 为 n 阶正交矩阵，则 $\xi=T'\left(\frac{\eta-\theta}{\sigma}\right)\sim N_n(0,I_n)$.

$$Y=(Y_1,Y_2,\cdots,Y_n)^T=(g_1(X_1,\cdots,X_n),\cdots,g_n(X_1,\cdots,X_n))^T$$

$$f_Y(y_1,\cdots,y_n)=\begin{cases} f_X\left(x_1(y_1,\cdots,y_n),\cdots,x_n(y_1,\cdots,y_n)\right)\cdot|J|, & \text{当}\bigcap_{i=1}^{n}x_i(y_1,\cdots,y_n)\neq\Phi \\ 0, & \text{其他}\end{cases}$$

$$J=\det\left(\begin{bmatrix}\frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n}\end{bmatrix}\right)$$

正态分布随机变量的线性变换

性质 1：（正态分布在可逆仿射变换下仍是正态分布） 设 n 维随机向量 $X \sim N(\mu,\Sigma)$，A 是 n 阶实数可逆方阵，$b \in \mathrm{R}^n$。则 $Y = AX + b \sim N(A\mu + b, A\Sigma A')$。

性质 1'：（正态分布在非退化仿射变换下的不变性，性质 1 的一般形式）

设 n 维随机向量 $X \sim N(\mu,\Sigma)$，A 是 $m \times n$ 阶实数方阵，$\mathrm{rank}A = m$（即 A 的行向量是线性无关的），$b \in \mathrm{R}^m$。则 $Y = AX + b \sim N(A\mu + b, A\Sigma A')$。

一个常用的结论（性质 1'的特例）

设 $X_1, X_2, \ldots, X_n$ 独立，$X_i \sim N(\mu_i, \sigma_i^2)$，$i = 1,2,\ldots,n$，$a_1, a_2, \ldots, a_n, b \in \mathrm{R}$，其中 $a_1, a_2, \ldots, a_n$ 不全为零。则

$$a_1X_1 + \cdots a_nX_n + b \sim N(a_1\mu_1 + \cdots a_n\mu_n + b, a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2)。$$

$$Y=(Y_1,Y_2,\cdots,Y_n)^T=(g_1(X_1,\cdots,X_n),\cdots,g_n(X_1,\cdots,X_n))^T$$

$$f_Y(y_1,\cdots,y_n)=\begin{cases} f_X(x_1(y_1,\cdots,y_n),\cdots,x_n(y_1,\cdots,y_n))\cdot|J|, & \text{当}\bigcap_{i=1}^{n}x_i(y_1,\cdots,y_n)\neq\Phi \\ 0, & \text{其他}\end{cases}$$

$$J=\det\left(\begin{bmatrix}\frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n}\\ \vdots & \ddots & \vdots\\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n}\end{bmatrix}\right)$$

$$X\sim N(\mu,\Sigma),\quad Y=AX+b,\quad |X|\neq 0$$

$$f_Y(y)=f_Y\left(x_1(y_1,\cdots,y_n),\cdots,x_n(y_1,\cdots,y_n)\right)\left|A^{-1}\right|$$

$$=\frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\left|A^{-1}\right|\exp\left\{-\frac{1}{2}\left[A^{-1}(y-b)-\mu\right]'\Sigma^{-1}\left[A^{-1}(y-b)-\mu\right]\right\}$$

$$=\frac{1}{(2\pi)^{n/2}|\Sigma|^{\frac{1}{2}}}\left|A^{-1}\right|\exp\left\{-\frac{1}{2}(y-b-A\mu)'A'^{-1}\Sigma^{-1}A^{-1}(y-b-A\mu)\right\}$$

$$=\frac{1}{(2\pi)^{n/2}|A\Sigma A'|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(y-A\mu-b)'(A\Sigma A')^{-1}(y-A\mu-b)\right\}$$

$$Y\sim N(A\mu+b,A\Sigma A')$$

定理 1.2.1 设 $\eta \sim N_n(0, I_n)$，A 为 n 阶对称幂等(即 $A^2 = A$)矩阵. 则

$$\eta' A \eta \sim \chi^2(\mathrm{tr}(A)),$$

即 $\eta' A \eta$ 服从自由度为 $\mathrm{tr}(A)$ 的卡方分布，其中 $\mathrm{tr}(A)$ 表示 A 的迹.

$$Ax = \lambda x, \quad A^2 x = AAx = A\lambda x = \lambda Ax = \lambda^2 x, \qquad \lambda^2 x = \lambda x \Rightarrow \lambda = 0 \text{或} 1$$

$$A \overset{\text{正交对角化}}{=} T\Lambda T', \qquad TT' = I_n, \qquad \Lambda = diag(1, \cdots 1, 0, \cdots, 0)$$

$$\eta' A \eta = \eta' T \Lambda T' \eta$$

$$\zeta = T'\eta, \quad \zeta \sim N_n(0, I_n)$$

$$\eta' A \eta = \eta' T \Lambda T' \eta = (\zeta_1, \cdots, \zeta_n) \Lambda (\zeta_1, \cdots, \zeta_n)' = \sum_{k=1}^{tr(A)} \zeta_k^2 \sim \chi^2(tr(A))$$

（下侧）α 分位数

X 为一连续分布随机变量，如果 $P(X \le a) = \alpha$，a 称为该分布的α 分位数,也称为下侧α 分位数

标准正态分布的α 分位点记为 u_α

n 个自由度的 χ^2 分布的α 分位点记为 $\chi^2_\alpha(n)$

n 个自由度的 t 分布的α 分位点记为 $t_\alpha(n)$

(m,n) 自由度的 F 分布的α 分位点记为 $F_\alpha(m,n)$

（上侧）α 分位数

X 为一连续分布随机变量，如果 $P(X \geq a) = \alpha,$
a 称为该分布的α 分位数,也称为上侧α 分位数

标准正态分布的α 分位点记为 u_α

n 个自由度的χ^2 分布的α 分位点记为 $\chi^2_\alpha(n)$

n 个自由度的 t 分布的α 分位点记为 $t_\alpha(n)$

(m,n) 自由度的 F 分布的α 分位点记为 $F_\alpha(m,n)$

(下侧)分位数练习

例1. $X \sim N(0,1)$，则 $P\left(|X| < u_{0.975}\right) =$

例2. 设随机变量 X 服从正态总体 $N(0,1)$ ，对给定的 $\alpha(0<\alpha<1)$，数 u_α 满足 $P(X \le u_\alpha) = \alpha$，若 $P(|X| < x) = \alpha$，则 $x =$

(A) $u_{\frac{1+\alpha}{2}}$ (B) $u_{1-\frac{\alpha}{2}}$ (C) $u_{\frac{1-\alpha}{2}}$ (D) $u_{1-\alpha}$

条件数学期望

$$E(X\mid Y=y)=\begin{cases}\sum_i x_i P(X=x_i\mid Y=y), & (X,Y)\text{为离散随机变量}\\ \int_{-\infty}^{+\infty} xp(x\mid y)dx, & (X,Y)\text{为连续随机变量}\end{cases}$$

重期望公式（全期望公式） $E(X)=E(E(X\mid Y))$

证明： 以连续型为例

$$E(X)=\int_{-\infty}^{+\infty} xp_X(x)dx=\int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} p(x,y)dydx=\int_{-\infty}^{+\infty} x\int_{-\infty}^{+\infty} p_Y(y)p(x\mid y)dydx$$

$$=\int_{-\infty}^{+\infty} p_Y(y)dy\int_{-\infty}^{+\infty} xp(x\mid y)dx=\int_{-\infty}^{+\infty} E(X\mid Y=y)p_Y(y)dy$$

$$=E(E(X\mid Y))$$

例 3. 口袋里有编号$1,2,\cdots,n$的n个球，任取 1 个，若为 1 号则得 1 分停止，若为i $(i\geq 2)$号，则得到i分，放回继续摸球，求总得分的期望。

解：设X为总得分，Y为第一次抽到的号码

则 $P(Y=k)=\dfrac{1}{n}$ $(k=1,\cdots,n)$,

$$E(X\mid Y=1)=1, \quad E(X\mid Y=k,\ k\neq 1)=k+E(X)$$

$$E(X)=E(E(X\mid Y))=\sum_{k=1}^{n}E(X\mid Y=k)P(Y=k)=\frac{1}{n}\sum_{k=1}^{n}[k+E(X)]$$

$$\Rightarrow E(X)=\frac{n(n+1)}{2}。$$

例 4. 投掷一个公平的硬币，直至首次出现相继的两个正面停止，求投掷次数的期望。

解：设 X 为投掷次数， Y 定义为 $\begin{array}{cccc} Y & T & HH & HT \\ P & 1/2 & 1/4 & 1/4 \end{array}$，其中

T 表示第一次掷出反面， HH 表示前两次依次掷出正、正，

HT 表示前两次依次掷出正、反。

$$E(X) = E(E(X \mid Y))$$
$$= P(Y = T)E(X \mid Y = T) + P(Y = HH)E(X \mid Y = HH) + P(Y = HT)E(X \mid Y = HT)$$
$$= \frac{1}{2}(1 + E(X)) + \frac{1}{4} \cdot 2 + \frac{1}{4}(2 + E(X))$$

$\Rightarrow E(X) = 6$。

思考： X 的方差如何计算？

例 5. $(X,Y)\sim N\left(0,0,\sigma_1^2,\sigma_2^2,\rho\right)$，证明：$E(XY)=\rho\sigma_1\sigma_2$。

证明：$E(XY)=E\left[E(XY\mid X)\right]=E\left[X\cdot E(Y\mid X)\right]$

$$p(y\mid x)=\frac{p(x,y)}{p_X(x)}=\frac{\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{\dfrac{-1}{2\left(1-\rho^2\right)}\left[\dfrac{x^2}{\sigma_1^2}-2\rho\dfrac{xy}{\sigma_1\sigma_2}+\dfrac{y^2}{\sigma_2^2}\right]\right\}}{\dfrac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\dfrac{x^2}{2\sigma_1^2}\right\}}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2\sigma_2^2\left(1-\rho^2\right)}\left(y-\rho\frac{\sigma_2}{\sigma_1}x\right)^2\right\}\sim N\left(\rho\frac{\sigma_2}{\sigma_1}x,\sigma_2{}^2\left(1-\rho^2\right)\right)$$

$$E(Y\mid X)=\rho\frac{\sigma_2}{\sigma_1}X\quad\Rightarrow$$

$$E(XY)=E\left[E(XY\mid X)\right]=E\left[X\cdot E(Y\mid X)\right]=E\left(\rho\frac{\sigma_2}{\sigma_1}X^2\right)=\rho\frac{\sigma_2}{\sigma_1}\sigma_1{}^2=\rho\sigma_1\sigma_2\text{。}$$

瓦尔德（Wald）方程

例 6. $X_1, X_2, \cdots, X_n$ 为一列独立同分布随机变量，N 为一个取整数值的随机变量，且 N 与 $\{X_k\}$ 相互独立。证明 $E\left(\sum_{k=1}^{N} X_k\right) = E(X_1)E(N)$。

证明：

$$E\left(\sum_{k=1}^{N} X_k\right) = E\left(E\left(\sum_{k=1}^{N} X_k \middle| N\right)\right) = \sum_{n=1}^{\infty} P(N=n)\cdot E\left(\sum_{k=1}^{N} X_k \middle| N=n\right)$$

$$= \sum_{n=1}^{\infty} P(N=n)\cdot n \cdot E(X_1) = E(X_1)\cdot E(N)。$$

多层模型，混合分布

- 一只昆虫产下大量的卵，已知每颗卵的成活率是p，问平均多少颗卵可以成活？
- X：成活的卵的数量，　Y：昆虫产下的卵的数量$\sim P(\lambda)$
- $X|Y\sim B(Y,p)$，　　X?

$$P(X=x)=\sum_{y=0}^{+\infty}P(X=x,Y=y)=\sum_{y=0}^{+\infty}P(X=x\mid Y=y)\,P(Y=y)$$

$$=\sum_{y=x}^{+\infty}\binom{y}{x}p^{x}(1-p)^{y-x}\frac{e^{-\lambda}\lambda^{y}}{y!}=\frac{(\lambda p)^{x}e^{-\lambda}}{x!}\sum_{y=x}^{+\infty}\frac{\big((1-p)\lambda\big)^{y-x}}{(y-x)!}$$

$$=\frac{(\lambda p)^{x}e^{-\lambda}}{x!}\sum_{t=0}^{+\infty}\frac{\big((1-p)\lambda\big)^{t}}{t!}=\frac{(\lambda p)^{x}e^{-\lambda}}{x!}e^{(1-p)\lambda}=\frac{(\lambda p)^{x}}{x!}e^{-\lambda p}$$

多层模型，混合分布

- 一只昆虫产下大量的卵，已知每颗卵的成活率是p，问平均多少颗卵可以成活？

- X：成活的卵的数量，Y：昆虫产下的卵的数量$\sim P(\lambda)$
- $X|Y \sim B(Y,p)$，　$E(X)=E\big(E(X|Y)\big)=E(pY)=\lambda p$

从一个区域内昆虫中任取一只产卵，问成活卵的平均数量

Z：昆虫的产卵能力$\sim Exp(\beta)$，　$Y|Z \sim P(Z)$

$X|Y \sim B(Y,p)$，　X：成活的卵的数量

$$E(X)=E\big(E(X|Y)\big)=E(pY)=pE\big(E(Y|Z)\big)=pE(Z)=p\beta$$

统计量

- 总体：　一个统计问题研究对象的全体。构成总体的每个成员称为个体。简单说总体即为分布。
- 设想你参加了一次考试，在知道自己得到了78分后，希望了解自己的成绩在班上处于什么水平。你会怎样做？
- 你对自己未来工作收入的预期是什么？
- 设 $X_1, X_2, \cdots, X_n$ 为取自某总体的样本，不含有任何未知参数的样本函数 $T = T(X_1, X_2, \cdots, X_n)$，即称为**统计量**。

统计量

- 强国知十三数：境内仓口之数，壮男壮女之数，老弱之数，官士之数，以言说取食者之数，利民之数，马牛刍藁之数。欲强国，不知国十三数，地虽利，民虽众，国愈弱至削。国无怨民曰强国。兴兵而伐，则武爵武任，必胜；按兵而农，粟爵粟任，则国富。兵起而胜敌，按兵而国富者，王。（秦·商鞅《商君书》）

商鞅（前390～前338年），卫国人。战国时期政治家，思想家，著名法家代表人物。应秦孝公求贤令入秦，说服秦孝公变法图强。孝公死后，受到贵族诬害以及秦惠文王的猜忌，车裂而死。其在秦执政二十余年，秦国大治，史称“商鞅变法”。

常用统计量：样本均值

样本均值 $\bar{x}=\dfrac{x_1+x_2+\cdots x_n}{n}=\dfrac{1}{n}\sum_{i=1}^{n}x_i$， $x_1,x_2,\cdots,x_n$ 为取自某总体的样本

定理 1. 若把样本中数据与样本均值之差称为偏差，则样本所有偏差之和为 0，即 $\sum_{i=1}^{n}\left(x_i-\bar{x}\right)=0$ 。

定理 2. 数据观察值与均值的偏差平方和最小，即对所有的 $c\in R$， $c=\bar{x}$，使得 $\sum_{i=1}^{n}\left(x_i-c\right)^2$ 最小。

常用统计量：样本均值

定理 3. 设 $x_1, x_2, \cdots, x_n$ 是来自某个总体的样本，$\bar{x}$ 为样本均值，

（1）若总体分布为 $N(\mu, \sigma^2)$，则 $\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$；

（2）若总体分布不是正态分布或根本未知，$E(X) = \mu$，$Var(X) = \sigma^2$，

则 n 较大时，$\bar{x}$ 的渐近分布为 $N\left(\mu, \frac{\sigma^2}{n}\right)$，常记为 $\bar{x} \dot{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$。

（中心极限定理）

常用统计量：样本方差

样本方差 $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

定理. 设总体 X 具有二阶矩，即 $E(X) = \mu$， $Var(X) = \sigma^2 < +\infty$，

$x_1, x_2, \cdots, x_n$ 为从该总体得到的样本， $\bar{x}$ 和 s^2 分别是样本均值和

样本方差，则 $E(\bar{x}) = \mu$， $Var(\bar{x}) = \dfrac{\sigma^2}{n}$， $E(s^2) = \sigma^2$。 无偏估计

样本 k 阶原点矩 $a_k = \dfrac{1}{n}\sum_{i=1}^{n} x_i^k$， 样本 k 阶中心矩 $b_k = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$

$$E\left(s^2\right)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2\right]=\sigma^2$$

样本均值和样本方差的数字特征

$$E\left(\bar{x}\right)=E\left(\frac{x_1+x_2+\cdots x_n}{n}\right)=\frac{1}{n}E\left(x_1+x_2+\cdots x_n\right)=\mu$$

$$Var\left(\bar{x}\right)=Var\left(\frac{x_1+x_2+\cdots x_n}{n}\right)=\frac{1}{n^2}Var\left(x_1+x_2+\cdots x_n\right)=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n}$$

$$\sum_{i=1}^{n}\left[\left(x_i-E(x_i)\right)-\left(\bar{x}-E(\bar{x})\right)\right]^2$$

$$=\sum_{i=1}^{n}E\left(x_i-E\left(x_i\right)\right)^2+\sum_{i=1}^{n}E\left(\bar{x}-E\left(\bar{x}\right)\right)^2-2\cdot E\sum_{i=1}^{n}\left[\left(x_i-E\left(x_i\right)\right)\cdot\left(\bar{x}-E\left(\bar{x}\right)\right)\right]$$

$$=\sum_{i=1}^{n}Var(x_i)+\sum_{i=1}^{n}Var(\bar{x})-2\cdot E\left[\left(\sum_{i=1}^{n}x_i-nE(x_i)\right)\cdot\left(\bar{x}-E(\bar{x})\right)\right]$$

$$=n\sigma^2+n\cdot Var(\bar{x})-2\cdot E\left[\left(n\bar{x}-nE(\bar{x})\right)\cdot\left(\bar{x}-E(\bar{x})\right)\right]$$

$$=n\sigma^2-n\cdot Var(\bar{x})=(n-1)\sigma^2$$

三大统计分布，卡方、t、F

I.　$X_1, X_2, \cdots, X_n$ 独立同分布，服从 $N(0,1)$ 则

$Y = X_1^{\ 2} + \cdots + X_n^{\ 2} \sim \chi_n^{\ 2}$ 或 $\chi^2(n)$，称为自由度为 n 的 χ^2 分布。

II.　t 分布　$X_1 \sim N(0,1)$，$X_2 \sim \chi_n^{\ 2}$，X_1, X_2 相互独立

$$t = \frac{X_1}{\sqrt{X_2 / n}} \sim t(n)$$，称为自由度为 n 的 t 分布。

III.　F 分布　X_1, X_2 相互独立，$X_1 \sim \chi_m^{\ 2}$，$X_2 \sim \chi_n^{\ 2}$

$$F = \frac{X_1 / m}{X_2 / n} \sim F(m, n)$$，称为自由度为 m 与 n 的 F 分布。

统计抽样定理

设 $x_1, x_2, \cdots, x_n$ 是来自正态总体$N\left(\mu, \sigma^2\right)$的样本，其样本均值和样本方差分别为：

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \text{ 和 } s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2, \text{ 则有}$$

(1) $\overline{x}$ 与 s^2 相互独立

(2) $\overline{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

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

哪一组显得更随机一些

A

0 0 0 0 1 0 1 0 1 1 1 0 0 1 0 1 0 0 0 0 1 0 1 1 1
0 0 0 0 1 0 0 1 0 0 1 1 1 1 0 0 0 0 0 0 1 0 1 1 0
1 0 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0 0 0 0 1 0 0 1 0
0 1 1 0 0 0 0 1 0 0 0 0 0 1 0 0 1 1 0 1 0 0 1 1 0
1 1 1 0 0 0 0 0 0 0 1 0 1 0 1 1 0 1 1 0 0 1 0 1 0
0 1 1 0 1 1 0 0 0 0 1 0 1 1 1 1 1 0 1 0 1 1 1 1 1
1 0 0 0 1 0 1 0 1 1 1 0 0 1 0 0 1 1 1 1 1 0 0 0 0
0 0 1 1 1 1 0 1 0 0 0 0 0 0 1 1 0 0 1 0 1 1 0 1 0

B

1 1 1 1 0 0 1 0 0 1 0 0 1 0 1 0 0 0 1 0 0 1 1 1 1
0 0 1 0 1 0 0 0 1 0 0 1 1 0 1 0 1 1 1 0 0 1 1 1 0
0 1 1 1 1 0 1 1 0 1 0 1 1 1 1 0 1 1 0 1 0 0 1 0 0
0 1 1 1 0 0 1 1 0 1 0 1 1 0 1 0 1 0 1 1 0 1 0 0 1
1 0 1 0 1 0 0 1 1 1 0 1 1 0 1 0 0 1 0 0 0 0 0 1 1
0 1 0 1 1 0 1 1 0 0 0 0 1 1 1 0 1 0 1 1 0 0 1 1 0
0 1 1 1 0 0 1 1 1 1 0 1 1 0 0 0 1 1 1 0 0 1 1 0 1
0 1 1 1 0 0 1 1 0 0 0 1 1 1 1 1 0 1 0 1 0 0 1 1 1

掷硬币的随机性鉴别

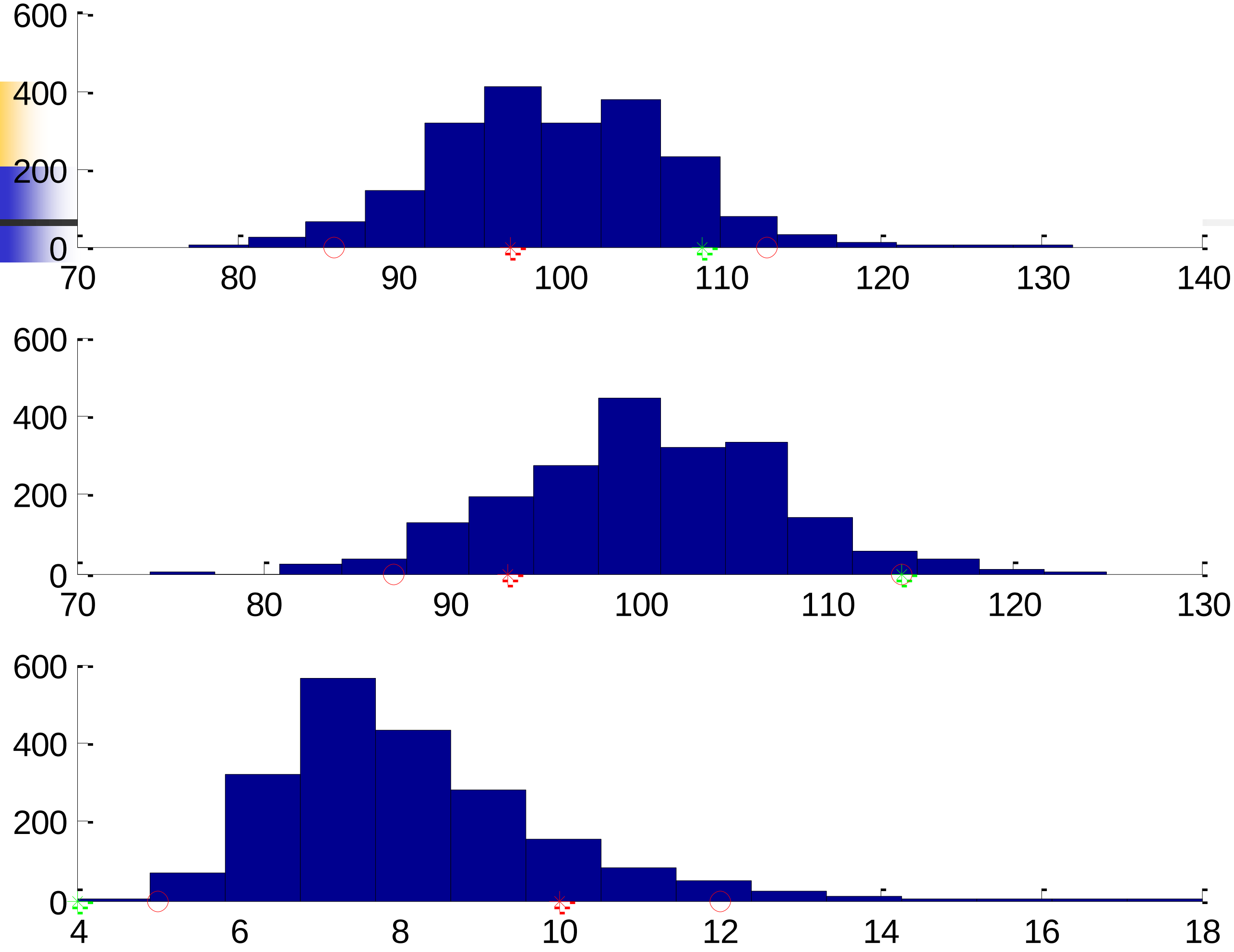
1. 定义关于 200 位 0-1 序列的统计量；找到公平抛掷(即 b(1, 0.5))假设下统计量的分布，统计量的分布可能可以解析地算出，也可用直方图等近似；

2. 提供几个对这个问题不一定是很好的统计量供大家参考

   (1) 正面的个数；
   (2) 最长0或1串的长度；
   (3) 0-1变化次数，比如01001的切换次数为4， 0-1-0-1 ；

3. 争取提出更多适合的统计量进行判断。

统计量的分布与经验分布函数



nn=2000; % sample size

作业

习题一. 17,18,20

附加题. 假设某医生每天门诊挂号的病人数为 N，是服从参数为 a 的泊松分布随机变量。又假设每位病人门诊看病的时间也为随机的，均服从参数为 b 的指数分布随机变量，且相互独立。这名医生总的门诊看病时间记为 T，求 T 的期望和方差。