清華大學

L1

Display & Modeling

$X_{n\times p} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & & \vdots \\ x_{n1} & \cdots & x_{np}\end{bmatrix} = \begin{bmatrix}\vec{x_1}^T \\ \vdots \\ \vec{x_n}^T\end{bmatrix}$

$\vec{x_1}\cdots\vec{x_n}$ i.i.d

$f(\vec{x_i}) = f(x_{i1}\cdots x_{ip})$

Descriptive Statistics

$\bar{x}_k = \frac{1}{n}\sum_{i=1}^n x_{ik}$ $k=1,2\cdots p$

$s_k^2 = \frac{1}{n}\sum_{i=1}^n (x_{ik}-\bar x_k)^2$ $k=1,2,\cdots p$

$s_{ik} = \frac{1}{n}\sum_{j=1}^n (x_{ji}-\bar x_i)(x_{jk}-\bar x_k)$ $i,k=1,2,\cdots p$

$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}}$ $r=0$ 非线性相关未必独立.

Mahalanobis Distance

$\vec x,\vec y$ of same distribution with cov matrix Σ

$d(x,y)=\sqrt{(\vec x-\vec y)^T\Sigma^{-1}(\vec x-\vec y)}$

Random Vector $\vec X = \begin{pmatrix} X_1 \\ \vdots \\ X_p\end{pmatrix}$

$\vec u = E(\vec X) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p)\end{pmatrix}$

$\Sigma = Cov(\vec X) = E[(\vec X-\vec u)(\vec X-\vec u)^T] = \begin{bmatrix}\sigma_{11} & \cdots & \sigma_{1p} \\ \sigma_{p1} & \cdots & \sigma_{pp}\end{bmatrix}$

$\sigma_{ij}=\sigma_{ji} = E[(X_i-u_i)(X_j-u_j)]$

Thm1. $E(\vec Z^T A\vec Z) = tr(A\Sigma)+u^TAu$

Thm2. $E(\|\vec x-\vec y\|^2) = 2\sum_{i=1}^p \sigma_{ii}$ 上述证明 $A=I$

$\Rightarrow E(\|\vec x-\vec y\|_\Sigma) = tr(\Sigma^{-1}\Sigma) = 2p$ 当 $\vec x,\vec y$ iid

L2

A is k×k symetric matrix 谱分解

$A = P\Lambda P^T$ $\Lambda = \begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_k\end{bmatrix}$ $P=[\vec e_1 \cdots \vec e_k]$

$\{\vec e_1\cdots\vec e_k\}$ is an orthogonal basis.

旋转→放缩→再旋转.

$A^{-1} = P\Lambda^{-1}P^T = \sum_{i=1}^k \frac{1}{\lambda_i}\vec e_i\vec e_i^T$

$A^{\frac12} = \sum_{i=1}^k \sqrt{\lambda_i}\,\vec e_i\vec e_i^T = P\Lambda^{\frac12}P^T$

$\vec x = P\Lambda^{\frac12} y$ $Var(y)=I$

$\Sigma = Var(x) = P\Lambda^{\frac12}\, Var(y)\, \Lambda^{\frac12}P^T = P\Lambda^2P^T$

Mahalanobis-distance

$\vec x^T S^{-1}\vec x$ $S = P\Lambda P^T$.

SVD. $A = U\Lambda V^T$

Cauchy-schwarz Inequalities.

$(\vec b^T\vec d)^2 \le (\vec b^T\vec b)(\vec d^T\vec d)$

or $\le (\vec b^T B\vec b)(\vec d^T B^{-1}\vec d)$ B正定.

equality if and only if $\vec b = cB^{-1}\vec d$

Maximization Lemma

$\max_{\vec x\ne 0}\frac{(\vec x^T\vec d)^2}{\vec x^TB\vec x} = \vec d^TB^{-1}\vec d$

$\vec x = cB^{-1}\vec d$ 时取等.

Maximization of Quadratic Form for Points On the Unit Sphere.

$\max_{x\ne 0}\frac{\vec x^TB\vec x}{\vec x^T\vec x} = \lambda_1$ (attain when $\vec x=\vec e_1$)

$\max_{\vec x\perp\vec e_1\cdots\vec e_k}\frac{\vec x^TB\vec x}{\vec x^T\vec x} = \lambda_{k+1}$ (attained when $\vec x=\vec e_{k+1}$)

Matrix Operation On Descriptive Statistics.

$\bar X = \frac{1}{n}X^T 1_n$ $nS_n = X^T(I-\frac{1}{n}1_n1_n^T)X$

Geomatry:

$L_{d_i}^2 = \vec d_i^T\vec d_i$, $\vec d_i = \begin{bmatrix}x_{1i} \\ \vdots \\ x_{ni}\end{bmatrix} - \bar x_i 1 = \sum_{j=1}^n (x_{ji}-\bar x_i)^2$ α Variance.

$\vec d_i^T\vec d_k = L_{d_i}L_{d_k}\cos(\theta_{ik})$

$r_{ik} = \frac{s_{ik}}{\sqrt{s_{ii}}\sqrt{s_{kk}}} = \cos(\theta_{ik})$ closer di,dj larger corr.

Generalize Variance.

$|S| = (volumn)^2/n^p$

Vector and Matrix calculus

$\frac{\partial}{\partial\vec x}A\vec x = A^T$ $\frac{\partial}{\partial\vec x}\vec x^TA = A$ $\frac{\partial}{\partial\vec x}\vec x^T\vec x = 2\vec x$

$\frac{\partial|A|}{\partial A} = |A|A^{-1}$ $\frac{\partial tr(AB)}{\partial A} = B^T$ $\frac{\partial tr(A^{-1}B)}{\partial A} = -A^{-1}BA^{-1}$

Theoretical results of Sample Statistics:

$E(\bar x) = u$ $E(\frac{n}{n-1}S_n) = \Sigma$.

mean $E(c^TX) = c^T\vec u$ var $Var(c^TX) = c^T\Sigma c$

$E(AX) = Au_x$ $Cov(AX) = A\Sigma_xA^T$

L3.

Multivariative Normal Distribution.

$f(\vec x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{\frac12}}\exp\{-\frac{(\vec x-\vec u)^T\Sigma^{-1}(\vec x-\vec u)}{2}\}$

$(\vec x-\vec u)^T\Sigma^{-1}(\vec x-\vec u) = c^2$ 构成一组等高线.

MLE.

Joint Distribution $= \frac{1}{(2\pi)^{\frac{np}{2}}|\Sigma|^{\frac n2}} e^{-\sum_{j=1}^n\frac{(x_j-u)^T\Sigma^{-1}(x_j-u)}{2}}$

$\hat u = \bar x$ $\hat\Sigma = \frac{1}{n}\sum_{j=1}^n(\vec x_j-\bar x)(\vec x_j-\bar x)^T$

Unbiased Estimator:

$\hat u = \bar X$ $S = \frac{1}{n-1}\sum_{j=1}^n(\vec x_j-\bar x)(\vec x_j-\bar x)^T = \frac{n}{n-1}\hat\Sigma$

$E(\bar X) = u$ $E(S) = \Sigma$.

Sufficient Statistics. $\bar X$ and S

$\sum_{j=1}^n(\vec x_j-\vec u)^T\Sigma^{-1}(\vec x_j-u)$

$= tr(\Sigma^{-1}[\sum_{j=1}^n(\vec x_j-\bar x)(\vec x_j-\bar x)^T + n(\bar x-u)(\bar x-u)^T])$

Properties.

$X\sim N_p(\vec u,\Sigma)$ $AX\sim N_q(Au, A\Sigma A^T)$

$X+\vec d\sim N_p(\vec u+\vec d,\Sigma)$

$\Rightarrow Z\sim N_p(\vec 0, I)$ $\vec X = \vec u + \Sigma^{\frac12}\vec Z\sim N_p(u,\Sigma)$

Conditional Distribution

$\begin{bmatrix}X_1\\X_2\end{bmatrix}\sim N_{q+q}\left(\begin{bmatrix}u_1\\u_2\end{bmatrix},\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$

given $X_2=x_2$ then X_1 is normal distribution

mean $= u_1+\Sigma_{12}\Sigma_{22}^{-1}(x_2-u_2)$

Cov $= \Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

$(X-\vec u)^T\Sigma^{-1}(\vec X-\vec u)\sim\chi^2_p$ ($\vec Z = \Sigma^{-\frac12}(\vec X-\vec u)$)

L4

Sample Distribution of Sufficient Statistics

$\begin{cases}\bar X\sim N_p(u,\frac{1}{n}\Sigma)\\ (n-1)S\sim W_p(n-1,\Sigma)\\ \bar X\perp\!\!\!\perp S\end{cases}$

Wishart Distribution:

$W_p(m,\Sigma) = \sum_{j=1}^m Z_jZ_j^T$ $Z_j\sim N_p(0,\Sigma)$

Test of one-sample Mean:

$H_0: u=u_0 \leftrightarrow H_1: u\ne u_0$

$T^2 = \sqrt n(\bar X-u_0)^T(S)^{-1}\sqrt n(\bar X-u_0)$

under H_0: $= N_p(0,\Sigma)^T(\frac{1}{n-1}W_p(n-1,\Sigma))^{-1}N_p(0,\Sigma)$

$\frac{T^2}{n-1}\sim\frac{p}{n-p}F_{p,n-p}$

Then we reject if $T^2 > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$

LRT:

$\Lambda = \frac{\max_{H_0}L(u_0,\Sigma)}{\max_{H_0\cup H_1}L(u,\Sigma)} = (1+\frac{T^2}{n-1})^{-\frac n2}$

又 $\Lambda = (\frac{|\hat\Sigma|}{|\hat\Sigma_0|})^{\frac n2}$.

$T^2 = \frac{(n-1)|\sum_{j=1}^n(x_j-u_0)(x_j-u_0)^T|}{|\sum_{j=1}^n(x_j-\bar x)(x_j-\bar x)^T|} - (n-1)$

更有效地计算 T.

Confidence Interval: 100(1-α)% region

$R(x) = \{u: n(\bar X-u)^TS^{-1}(\bar X-u)\le c^2\}$

$= \{u: T_u^2\le c^2\}$

$c^2 = T^2_{(\alpha)} = \frac{p(n-1)}{n-p}F_{p,n-p}(\alpha)$

Linear Combination.

$Z_i = a^TX_i\sim N(a^Tu, a^T\Sigma a)$

$\hat\mu = \bar z = \vec a^T\bar X$ $se(\hat\mu) = \frac{S_z}{\sqrt n}$

Thus 100(1-β)% CI for the mean is

$I_\psi(z) = \hat\psi\pm t_{n-1}(\frac{\beta}{2})se(\hat\psi)$

$= a^T\bar X\pm t_{n-1}(\frac{\beta}{2})\sqrt{\frac{a^TSa}{n}}$

$\frac{\partial}{\partial\vec x}\vec x^TA\vec x = A\vec x + A^T\vec x$ (分两部分)

$H_{f\circ g}(x) = H_g(x)H_f(g(x))$ (反过来)

Linear combination

$V_1 = \sum_{i=1}^n c_iX_i\sim N_p(\sum_{j=1}^n c_ju_j, \sum_{j=1}^n c_j^2\Sigma)$

$V_2 = \sum_{i=1}^n b_iX_i$ V_1,V_2 has joint distribution $\begin{bmatrix}\sum_{j=1}^n c_j^2\Sigma & (\vec b^T\vec c)\Sigma\\ (\vec b^T\vec c)\Sigma & \sum_{j=1}^n b_j^2\Sigma\end{bmatrix}$

$\therefore V_1\perp V_2$ if $\vec b^T\vec c = 0$

Bonferroni Intervals:

$I_{\varphi_k}(z) = \varphi_k \pm t_{n-1}(\frac{\alpha}{2m}) \hat{se}(\varphi_k)$

三者对比

$I_a(C) = a^T \bar{x} \pm C\sqrt{a^T \frac{S}{n} a}$ $\begin{cases} C = t_{n-1}(\frac{\alpha}{2}) \\ C = t_{n-1}(\frac{\alpha}{2m}) \\ C = \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p}(\alpha)} \end{cases}$

Bonferroni一定框得比t大

但和F dist关系未知.

F区域框到了R(x)边界.

L_5

双样本检验 $\Sigma_1 = \Sigma_2$

$H_0: u_1 - u_0 = \delta_0 \Leftrightarrow H_1: u_1 - u_0 \neq \delta_0$

$T^2 = (\frac{1}{n_1} + \frac{1}{n_2})^{-\frac{1}{2}} (\bar{x}_1 - \bar{x}_2 - \delta_0)^T S_{pooled}^{-1} (\frac{1}{n_1} + \frac{1}{n_2})^{-\frac{1}{2}} (\bar{x}_1 - \bar{x}_2 - \delta_0)$

$S_{pooled} = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} \sim \frac{W_p(n_1+n_2-2, \Sigma)}{n_1+n_2-2}$

reject if $T^2 > T^2(\alpha) = \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} F_{p, n_1+n_2-p-1}(\alpha)$

CR:

$(\bar{x}_1 - \bar{x}_2 - (u_1 - u_2)) [(\frac{1}{n_1} + \frac{1}{n_2}) S_{pooled}]^{-1} [\bar{x}_1 - \bar{x}_2 - (u_1 - u_2)] \leq T^2(\alpha)$

CI:

$\vec{a}^T(\bar{x}_1 - \bar{x}_2) \pm C\sqrt{\vec{a}^T (\frac{1}{n_1} + \frac{1}{n_2}) S_{pooled} \vec{a}}$

$C^2 = \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} F_{p, n_1+n_2-p-1}(\alpha)$

当 $\vec{a} = S_{pooled}^{-1}(\bar{x}_1 - \bar{x}_2)$ 时，$\vec{a}$ 样本差异最大 $(\arg\max |\frac{\vec{a}^T(\bar{x}_1 - \bar{x}_2)}{\sqrt{\vec{a}^T(\frac{1}{n_1} + \frac{1}{n_2}) S_{pooled} \vec{a}}}|)$

Large Sample

$\bar{X} \to u \quad S \to \Sigma$.

$T^2 = ((\bar{x}_1 - \bar{x}_2) - (u_1 - u_2))^T (\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2)^{-1} ((\bar{x}_1 - \bar{x}_2) - (u_1 - u_2)) \xrightarrow{d} \chi_p^2$

CR:

$(\bar{x}_1 - \bar{x}_2 - (u_1 - u_2))^T (\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2)^{-1} (\bar{x}_1 - \bar{x}_2 - (u_1 - u_2)) \leq \chi_p^2(\alpha)$

CI: $C^2 = \chi_p^2(\alpha)$

$\vec{a}^T(\bar{x}_1 - \bar{x}_2) \pm C\sqrt{\vec{a}^T (\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2) \vec{a}}$

L_6:

$X^T = [x_1, \cdots x_p] \quad \Sigma$

$\{(\lambda_i, e_i), i = 1, 2, \cdots p\}$ 为 Σ 特征值及特征向量

$Y_1 = e_1^T X \quad Y_2 = e_2^T X \cdots Y_p = e_p^T X$

为其主成分，$\sigma_{11} + \cdots \sigma_{pp} = \sum_{i=1}^p Var(X_i)$

$= \lambda_1 + \cdots + \lambda_p = \sum_{i=1}^p Var(Y_i)$

% of total population due to kth PC

$= \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_p}$.

$\rho_{Y_i X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \quad \sum_{i=1}^p \rho_{Y_i, X_k}^2 = 1$

标准化后: $Z_k = \frac{x_k - u_k}{\sqrt{\sigma_{kk}}} \quad \vec{z} = V^{-1}(\vec{x} - \vec{u})$

$V = \begin{bmatrix} \sigma_{11} & & \\ & \ddots & \\ & & \sigma_{pp} \end{bmatrix} \quad cov(z) = \rho$

$\vec{Y}_i = \vec{e}_i^T \vec{z} \quad \sum_{i=1}^p Var(Y_i) = \sum_{i=1}^p Var(Z_i) = p$

$\rho_{Y_i Z_k} = e_{ik}\sqrt{\lambda_i}$

SVD: $S = P\Lambda P^T = \frac{1}{n}(HX)^T(HX)$

$nS = \tilde{X}^T \tilde{X} = V\Lambda^T \Lambda V^T$

$Y = \tilde{X}P = (HX)P = U\Lambda$

$n \times p \quad n \times p \quad p \times p$

MDS: 只知相似度，降维 $D = [d_{ij}]$

任意两样本相似度 ① $D = D^T$ ② $d_{ij} \geq 0$ $d_{ii} = 0$

样本间相似 $B = \tilde{X}\tilde{X}^T = U\Lambda\Lambda^T U$

$d_{ij}^2 = b_{ii} + b_{jj} - 2b_{ij}$

Bonfearroni change with m

T^2 change with p.

LRT For Equal Variance

$H_0: \Sigma_1 = \Sigma_2$

$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)}$

$v - v_0 = \dim \Theta - \dim \Theta_0$

under H_0

$-2\ln\Lambda \sim \chi_{v-v_0}^2$