STAT 3690 Lecture 30

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Apr 11, 2022

-

Testing the uncorrelatedness of canonical variates

- LRT for $H_0 : \mathbf{\Sigma_{YX}} = 0$ vs. H_1 : otherwise
  - LRT statistic $\lambda = \prod_{k=1}^{p}(1-\hat\rho_k^2)^{n/2}$
    - $\hat\rho_k$: the kth sample canonical correlation
    - Under H_0, $-2\ln\lambda = -n\sum_{k=1}^{p}\ln(1-\hat\rho_k^2) \approx \chi^2(pq)$

$\Sigma = \begin{bmatrix}\Sigma_Y & \Sigma_{YX}\\ \Sigma_{XY} & \Sigma_X\end{bmatrix}$, $\mu = \begin{bmatrix}\mu_Y\\ \mu_X\end{bmatrix}$

Let $\Theta = \{(\mu,\Sigma) \mid \Sigma > 0\}$

$\Theta_0 = \{(\mu,\Sigma) \mid \Sigma_{YX} = 0\}$

Known: $\begin{bmatrix}Y_i\\ X_i\end{bmatrix} \overset{iid}{\sim} MVN_{p+q}(\mu, \Sigma)$, $i = 1, \dots, n$

If $(\mu,\Sigma)\in\Theta$, then $\hat\mu = \begin{bmatrix}\bar Y\\ \bar X\end{bmatrix}$, $\hat\Sigma = n^{-1}\sum_{i=1}^n\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)^T$

$\therefore$ max log likelihood $= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma) - \frac12\sum_{i=1}^n\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)^T\hat\Sigma^{-1}\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma) - \frac12\sum_{i=1}^n \mathrm{tr}\left\{\hat\Sigma^{-1}\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)^T\right\}$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma) - \frac{n}{2}\mathrm{tr}\left[\hat\Sigma^{-1}\underbrace{\left\{\frac1n\sum_{i=1}^n\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)\left(\begin{bmatrix}Y_i\\ X_i\end{bmatrix}-\hat\mu\right)^T\right\}}_{=\hat\Sigma}\right]$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma) - \frac{n}{2}\mathrm{tr}(I_{p+q})$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma) - \frac{n(p+q)}{2}$ ①

If $(\mu,\Sigma)\in\Theta_0$, then $Y \perp\!\!\!\perp X$, i.e.,

samples $Y_i \sim MVN_p(\mu_Y, \Sigma_Y)$ and $X_i \sim MVN_q(\mu_X, \Sigma_X)$ are independent

$\therefore \hat\mu_Y = \bar Y$, $\hat\Sigma_Y = n^{-1}\sum_{i=1}^n(Y_i-\hat\mu_Y)(Y_i-\hat\mu_Y)^T$

$\hat\mu_X = \bar X$, $\hat\Sigma_X = n^{-1}\sum_{i=1}^n(X_i-\hat\mu_X)(Y_i-\hat\mu_X)^T$

That is, MLEs are $\hat\mu = \begin{bmatrix}\hat\mu_Y\\ \hat\mu_X\end{bmatrix} = \begin{bmatrix}\bar Y\\ \bar X\end{bmatrix}$ and $\hat\Sigma_0 = \begin{bmatrix}\hat\Sigma_Y & \\ & \hat\Sigma_X\end{bmatrix}$

$\therefore$ max log likelihood $= -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma_Y) - \frac12\sum_{i=1}^n(Y_i-\hat\mu_Y)^T\hat\Sigma_Y^{-1}(Y_i-\hat\mu_Y)$

$-\frac{nq}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma_X) - \frac12\sum_{i=1}^n(X_i-\hat\mu_X)^T\hat\Sigma_X^{-1}(X_i-\hat\mu_X)$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma_Y) - \frac{n}{2}\ln\det(\hat\Sigma_X)$

$-\frac12\sum_{i=1}^n\mathrm{tr}\{\hat\Sigma_Y^{-1}(Y_i-\hat\mu_Y)(Y_i-\hat\mu_Y)^T\}$

$-\frac12\sum_{i=1}^n\mathrm{tr}\{\hat\Sigma_X^{-1}(X_i-\hat\mu_X)(X_i-\hat\mu_X)^T\}$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma_Y) - \frac{n}{2}\ln\det(\hat\Sigma_X) - \frac{n}{2}\mathrm{tr}(I_p) - \frac{n}{2}\mathrm{tr}(I_q)$

$= -\frac{n(p+q)}{2}\ln(2\pi) - \frac{n}{2}\ln\det(\hat\Sigma_Y) - \frac{n}{2}\ln\det(\hat\Sigma_X) - \frac{n(p+q)}{2}$ ②

$\therefore$ $\lambda = \exp(② - ①) = \exp\left\{-\frac{n}{2}\ln\frac{\det(S_Y)\det(S_X)}{\det(S)}\right\}$ $\left(\because S = \frac{n}{n-1}\hat\Sigma, S_X = \frac{n}{n-1}\hat\Sigma_X, S_Y = \frac{n}{n-1}\hat\Sigma_Y\right)$

$= \left(\frac{\det(S)}{\det(S_Y)\det(S_X)}\right)^{\frac n2}$

$= \left(\frac{\cancel{\det(S_Y)}\det(S_X - S_{XY}S_Y^{-1}S_{YX})}{\cancel{\det(S_Y)}\det(S_X)}\right)^{\frac n2}$ $\left(\because S = \begin{bmatrix}S_Y & S_{YX}\\ S_{XY} & S_X\end{bmatrix}\right)$

$= \frac{\det\{S_X^{\frac12}(I - S_X^{-\frac12}S_{XY}S_Y^{-1}S_{YX}S_X^{-\frac12})S^{\frac12}\}}{\det(S_X^{\frac12}\cdot S_X^{\frac12})}$

$= \frac{\cancel{\{\det(S_X^{\frac12})\}^2}\det(I - \hat M^T\hat M)}{\cancel{\{\det(S_X^{\frac12})\}^2}}$ (let $\hat M = S_Y^{-\frac12}S_{YX}S_X^{-\frac12}$)

$= \det\{V(I - \Lambda^T\Lambda)V^T\}$ (svd of $\hat M$ is $\hat M = V\Lambda V^T$)

$= \det(I - \Lambda^T\Lambda)$ $(\because \det(V)\det(V^T) = \det(VV^T) = \det(I) = 1)$

$= \prod_{k=1}^{p}(1 - \hat\rho_k^2)$

- Sequential inference
  - Determining r, the number of pairs of canonical variates to retain
  - Note that $\mathbf{\Sigma_{YX}} = 0 \Leftrightarrow \rho_1 = \cdots = \rho_p = 0 \Leftrightarrow \rho_1 = 0$
    - Since $\rho_1 \geq \cdots \geq \rho_p$
  - Consider a sequence of p pairs of hypotheses: $H_{0,k} : \rho_{k-1} > 0, \rho_k = 0$ vs. $H_{1,k} : \rho_k > 0$
    - LRT statistic $\lambda_k = \prod_{\ell=k}^{p}(1 - \hat{\rho}_\ell^2)^{n/2}$
      - Under $H_{0,k}$, $-2\ln\lambda_k = -n\sum_{\ell=k}^{p}\ln(1 - \hat{\rho}_\ell^2) \approx \chi^2((p-k+1)(q-k+1))$
  - Different targets to control Type I errors
    - Family-wise error rate (FWER) $= \Pr(V \geq 1)$: the probability of at least one Type I error
      - V: the number of Type I errors
    - False discovery rate (FDR) $= \mathrm{E}(V/R \mid R > 0)\Pr(R > 0)$: the expected proportion of Type I errors among the rejected hypotheses
      - R: the number of rejected hypotheses
      - Less conservative and more powerful than FWER control at a cost of increased likelihood of Type I errors
  - Stopping rules
    - p_k: the p-value associated with the testing on $H_{0,k}$ vs. $H_{1,k}$
    - $p_{(k)}$: the kth smallest value among $\{p_1, \ldots, p_p\}$
    - Holm-Bonferroni procedure (Holm (1979), *Scandinavian Journal of Statistics*, 6, 65–70): if $p_{(k)} < \alpha/(p+1-k)$, reject $H_{0,(k)}$ and proceed to larger p-values; otherwise EXIT.
      - Control FWER at level α
    - B-H procedure (Benjamini & Hochberg (1995), *Journal of the Royal Statistical Society*, Series B., 57, 289–300): control FDR at level α
      1. For a given level α, find $k^* = \max\{k \in \{1, \ldots, p\} \mid p_{(k)} \leq k\alpha/p\}$
      2. Reject $H_{0,(k)}$ for $k = 1, \ldots, k^*$

```r
options(digits=4)
Y = as.matrix(dslabs::olive[,3:6])
X = as.matrix(dslabs::olive[,7:10])
p = ncol(Y)
q = ncol(X)

S_Y = cov(Y)
S_X = cov(X)
S_YX = cov(Y, X)
S_Y_sqrt = expm::sqrtm(S_Y)
S_X_sqrt = expm::sqrtm(S_X)
M = solve(S_Y_sqrt) %*% S_YX %*% solve(S_X_sqrt)
decomp1 = svd(M)

alpha = .05
n = nrow(Y)
rhos = decomp1$d
(test.stats = rev(-n*cumsum(rev(log(1-rhos^2)))))
pvals = numeric(length(test.stats))
for (k in 1:length(test.stats)){
  pvals[k] = 1-pchisq(test.stats[k], df=(p-k+1)*(q-k+1))
}
pvals.sort = sort(pvals)
# Holm-Bonferroni procedure
pvals.sort < alpha/(p+1-(1:p))
```

```r
# B-H procedure
pvals.sort <= (1:p)*alpha/p
```

Summary of CCA

- Dimension reduction method
    - Maximize correlation
    - Treat $\mathbf{Y}$ and $\mathbf{X}$ equally/reduce the dimension of both $\mathbf{Y}$ and $\mathbf{X}$ simultaneously
- Limitation: in need of invertible $\mathbf{\Sigma_Y}$ and $\mathbf{\Sigma_X}$