Displays for Statistics 5401

Lecture 33

November 23, 2005

Christopher Bingham, Instructor
612-625-1024

Class Web Page
http://www.stat.umn.edu/~kb/classes/5401

Copyright© Christopher Bingham 2005

You get canonical variables from the multistandardized $\tilde{\mathbf{x}}^{(1)} = \boldsymbol{\Sigma}_{11}^{-T/2}(\mathbf{x}^{(1)} - \boldsymbol{\mu}^{(1)})$ and $\tilde{\mathbf{x}}^{(2)} = \boldsymbol{\Sigma}_{22}^{-T/2}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})$ using left and right singular vectors $\boldsymbol{\ell}_j$ and $\mathbf{r}_j$ of

$$\tilde{\boldsymbol{\rho}}_{12} = \text{corr}[\tilde{\mathbf{x}}^{(1)},\tilde{\mathbf{x}}^{(2)}] = \boldsymbol{\Sigma}_{11}^{-T/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}.$$

How do you get canonical variables <u>directly</u> from $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$, rather than from $\tilde{\mathbf{x}}^{(1)}$ and $\tilde{\mathbf{x}}^{(2)}$?

- $z_j^{(1)} = \boldsymbol{\ell}_j^T\tilde{\mathbf{x}}^{(1)} = \boldsymbol{\ell}_j^T\boldsymbol{\Sigma}_{11}^{-T/2}(\mathbf{x}^{(1)} - \boldsymbol{\mu}^{(1)})$
  $= \mathbf{u}_j^T(\mathbf{x}^{(1)} - \boldsymbol{\mu}^{(1)})$
  $= \mathbf{u}_j^T\mathbf{x}^{(1)} - \mathbf{u}_j^T\boldsymbol{\mu}^{(1)}$, where $\mathbf{u}_j = \boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\ell}_j$
- $z_j^{(2)} = \mathbf{r}_j^T\tilde{\mathbf{x}}^{(2)} = \mathbf{r}_j^T\boldsymbol{\Sigma}_{22}^{-T/2}(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})$
  $= \mathbf{v}_j^T(\mathbf{x}^{(2)} - \boldsymbol{\mu}^{(2)})$
  $= \mathbf{v}_j^T\mathbf{x}^{(2)} - \mathbf{v}_j^T\boldsymbol{\mu}^{(2)}$, where $\mathbf{v}_j = \boldsymbol{\Sigma}_{22}^{-1/2}\mathbf{r}_j$

Thus you need to find $\mathbf{u}_j$ and $\mathbf{v}_j$. Although they are defined using $\boldsymbol{\ell}_j$ and $\mathbf{r}_j$, they can be computed directly from $\boldsymbol{\Sigma}$.

Facts (easily checkable):

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\mathbf{u}_j = \tau_j^2\Sigma_{11}\mathbf{u}_j = \theta_j\Sigma_{11}\mathbf{u}_j$$
$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathbf{v}_j = \tau_j^2\Sigma_{22}\mathbf{v}_j = \theta_j\Sigma_{22}\mathbf{v}_j$$

- Coefficient vector $\mathbf{u}_j$ for $z_j^{(1)}$ is a eigenvector of $\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ relative to Σ_{11}
- Coefficient vector $\mathbf{v}_j$ for $z_j^{(2)}$ is a eigenvector of $\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ relative to Σ_{22}

So you can find canonical variables by solving two relative eigenvalue/vector problems involving pieces of Σ.

Usually, the canonical variables are defined as

$$z_j^{(1)} = \mathbf{u}_j^T\mathbf{x}^{(1)} = (\Sigma_{11}^{-1/2}\boldsymbol{\ell}_j)^T\mathbf{x}^{(1)}$$
$$z_j^{(2)} = \mathbf{v}_j^T\mathbf{x}^{(1)} = (\Sigma_{22}^{-1/2}\mathbf{r}_j)^T\mathbf{x}^{(2)}$$

without subtracting means. These differ only by constants $\mathbf{u}_j^T\boldsymbol{\mu}^{(1)}$ and $\mathbf{v}_j^T\boldsymbol{\mu}^{(1)}$ from the previous definition, and

$$\mathbf{u}_j^T\boldsymbol{\mu}^{(1)} = E[z_j^{(1)}], \quad \mathbf{v}_j^T\boldsymbol{\mu}^{(1)} = E[z_j^{(2)}]$$

My examples have not, of course, had to do with population principal components, but rather with sample canonical correlations.

You define

- *sample* canonical correlations $\hat{\tau}_j$
- pairs of sample canonical variables $\hat{z}_j^{(1)}$ and $\hat{z}_j^{(2)}$

in a similar way, starting with $\mathbf{S}$ instead of Σ.

Continue with analysis of artificial data:

```
Cmd> s <- tabs(scores,covar:T)

Cmd> J1 <- run(3); J2 <- run(4,7) # selectors for variables

Cmd> s11 <- s[J1,J1]; s22 <- s[J2,J2]

Cmd> s12 <- s[J1,J2]; s21 <- s12'

Cmd> tauhatsq <- releigenvals(s21 %*% solve(s11) %*% s12, s22)

Cmd> tauhatsq # squared canonical correlations
(1)     0.83093    0.030001   0.0089408  6.5688e-18
```

Compute canonical correlations $\hat{\tau}_j$ from SVD of correlation matrix of multi-standardized data:

```
Cmd> tauhat <- svd(cor(scores[,J1] %*% solve(cholesky(s11)),\
            scores[,J2] %%*% solve(cholesky(s22)))[J1,J2])

Cmd> tauhat^2 # same as tauhatsq
(1)     0.83093    0.030001   0.0089408           0
```


There is a close relationship between sample canonical correlations and relative eigenvalues from the regression approach discussed on Monday.

If $\hat{\lambda}_j$ are the sample eigenvalues of **H** relative to **E** in either the multivariate regression of $\mathbf{x}^{(1)}$ on $\mathbf{x}^{(2)}$ or of $\mathbf{x}^{(2)}$ on $\mathbf{x}^{(1)}$, then

$$\hat{\tau}_i = \sqrt{\hat{\theta}_i} = \sqrt{\{\hat{\lambda}_i/(1 + \hat{\lambda}_i)\}}$$

```
Cmd> manova("x2 = x1_1 + x1_2 + x1_3",silent:T)

Cmd> h2 <- sum(SS[run(2,4),,]); e2 <- SS[5,,]

Cmd> lambdahat <- releigenvals(h2,e2)

Cmd> lambdahat
(1)     4.9149   0.030929  0.0090215 1.2698e-15

Cmd> lambdahat[run(3)]/(1 + lambdahat[run(3)])
(1)     0.83093    0.030001   0.0089408   thetahat = tauhat^2

Cmd> tauhatsq[run(3)]
(1)     0.83093    0.030001   0.0089408
```

The correlation canonical variables $\hat{z}_j^{(1)}$ and $\hat{z}_j^{(2)}$ are the same as the MANOVA canonical variables of regressions of $\mathbf{x}^{(1)}$ on $\mathbf{x}^{(2)}$ and of $\mathbf{x}^{(2)}$ on $\mathbf{x}^{(1)}$, except possibly for change of sign.

Alternative Approach: Find features or summaries of $x^{(1)}$ and $x^{(2)}$ that are highly correlated with each other.

This is the more traditional approach to canonical correlation.

We concentrate on linear features $\mathbf{u}^T\mathbf{x}^{(1)}$ and $\mathbf{v}^T\mathbf{x}^{(2)}$ and try to find $\mathbf{u}$ and $\mathbf{v}$ to maximize (make as large as possible)

$$\rho^2[\mathbf{u}^T\mathbf{x}^{(1)}, \mathbf{v}^T\mathbf{x}^{(2)}] = \frac{\mathrm{Cov}[\mathbf{u}^T\mathbf{x}^{(1)}, \mathbf{v}^T\mathbf{x}^{(2)}]^2}{V[\mathbf{u}^T\mathbf{x}^{(1)}]V[\mathbf{v}^T\mathbf{x}^{(2)}]} = \frac{(\mathbf{u}^T\mathbf{\Sigma}_{12}\mathbf{v})^2}{(\mathbf{u}^T\mathbf{\Sigma}_{11}\mathbf{u})(\mathbf{v}^T\mathbf{\Sigma}_{22}\mathbf{v})}$$

We work with ρ^2 because the sign of the correlation will be arbitrary.

I'll skip any derivation, but the solution can be stated using relative eigenvectors:

- $\mathbf{u} = \mathbf{u}_1$, where $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p$ are the ***relative eigenvectors*** of

  $$\mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{-1}\mathbf{\Sigma}_{21} \text{ relative to } \mathbf{\Sigma}_{11} \text{ (both } p\times p),$$

  with corresponding *relative eigen-values* $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_p$.

- $\mathbf{v} = \mathbf{v}_1$, where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ are the ***relative eigenvectors*** of

  $$\mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12} \text{ relative to } \mathbf{\Sigma}_{22} \text{ (both } q\times q),$$

  with corresponding *relative eigen-values* $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_q$.

Furthermore the maximized value (largest squared correlation) is $\theta_1 = \tau_1^2$.

These are the same coefficient vectors from the first approach to canonical correlation.

That is

$$\max_{\mathbf{u},\mathbf{v}} \rho^2[\mathbf{u}^T\mathbf{X}^{(1)}, \mathbf{v}^T\mathbf{X}^{(2)}] = \rho^2[\mathbf{u}_1{}^T\mathbf{X}^{(1)}, \mathbf{v}_1{}^T\mathbf{X}^{(2)}] = \theta_1$$

Note: These θ_j's are the same as before, that is $\theta_j = \tau_j{}^2$ where τ_j is a SV of $\tilde{\boldsymbol{\Sigma}}_{12}$.

With the usual normalization for $\mathbf{u}_1$,

$$V[\mathbf{u}_1{}^T\mathbf{X}^{(1)}] = \mathbf{u}_1{}^T\boldsymbol{\Sigma}_{11}\mathbf{u}_1 = 1$$

and

$$V[\mathbf{v}_1{}^T\mathbf{X}^{(2)}] = \mathbf{v}_1{}^T\boldsymbol{\Sigma}_{22}\mathbf{v}_1 = 1.$$

and

$$\mathrm{Cov}[\mathbf{u}_1{}^T\mathbf{X}^{(1)},\mathbf{v}_1{}^T\mathbf{X}^{(2)}] = \tau_1 = \sqrt{\theta_1}.$$

Similarly

$$z_j{}^{(1)} = \mathbf{u}_j{}^T\mathbf{X}^{(1)}, \quad z_j{}^{(2)} = \mathbf{v}_j{}^T\mathbf{X}^{(2)} \qquad j = 1, \ldots, \min(p,q)$$

have $\mathrm{Corr}[z_j{}^{(1)}, z_j{}^{(2)}] = \tau_j = \sqrt{\theta_j}$.

$z_j{}^{(1)}$ and $z_j{}^{(2)}$ have the largest squared correlation of any linear combinations uncorrelated with $z_k{}^{(1)}$ and $z_k{}^{(2)}$, $k < j$

Here is what the correlation matrix (and variance matrix) of standardized canonical variables looks like when p = 4 and q = 3.

$$V[\mathbf{z}] = \left[\begin{array}{ccc|cccc} 1 & 0 & 0 & \sqrt{\theta_1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \sqrt{\theta_2} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \sqrt{\theta_3} & 0 \\ \hline \sqrt{\theta_1} & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \sqrt{\theta_2} & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \sqrt{\theta_3} & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\mathbf{z} = [z_1{}^{(1)}, z_2{}^{(1)}, z_3{}^{(1)}, z_1{}^{(2)}, z_2{}^{(2)}, z_3{}^{(2)}, z_4{}^{(2)}]^T$$

There are only s = min(3,4) = 3 non-zero canonical correlations $\tau_1 = \sqrt{\theta_1}$, $\tau_2 = \sqrt{\theta_2}$ and $\tau_3 = \sqrt{\theta_3}$. Note that all correlations with $z_4{}^{(2)}$ are 0.

In general, there are s = min(p,q) pairs $(z_j^{(1)}, z_j^{(2)})$ of canonical variables.

All the correlation between $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ is "concentrated" in

$$\tau_i = \text{corr}[z_i^{(1)}, z_i^{(2)}],\ i = 1,\ldots, s.$$

When p ≠ q, there are $|p - q|$ additional unpaired canonical variables that not correlated with anything and have no significance.

You define *sample* canonical correlations and correlation canonical variables the same way using the sample eigenvalues $\hat{\theta}_i = \hat{\tau}_j^{\,2}$ and eigenvectors $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{v}}_i$ of

- $\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ relative to $\mathbf{S}_{11}$
- $\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$ relative to $\mathbf{S}_{22}$.


```
Cmd> eigs21 <- releigen(s12 %*% solve(s22) %*% s21, s11)

Cmd> eigs12 <- releigen(s21 %*% solve(s11) %*% s12, s22)

Cmd> uhat <- eigs21$vectors; vhat <- eigs12$vectors

Cmd> list(uhat,vhat)
uhat            REAL    3     3
vhat            REAL    4     4

Cmd> sqrt(eigs21$values) # canonical correlations
(1)    0.91156    0.17321   0.094556

Cmd> sqrt(eigs12$values[run(3)]) # canonical correlations
(1)    0.91156    0.17321   0.094556

Cmd> Z1 <- x1 %*% uhat; Z2 <- x2 %*% vhat
```

Z1 and Z2 contain canonical variables computed using relative eigenvectors.

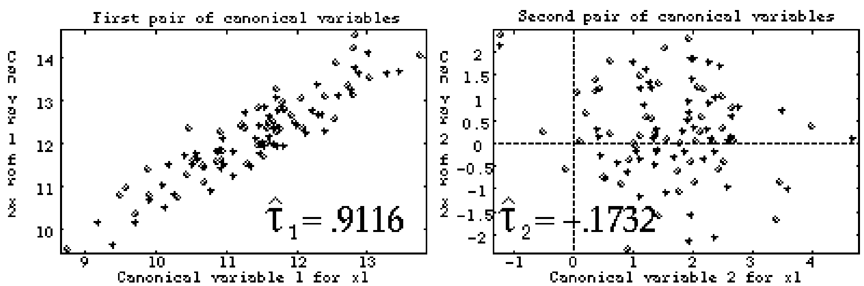
```
Cmd> print(format:"7.4f",cor(Z1,Z2))
MATRIX:
(1,1)  1.0000 -0.0000 -0.0000  0.9116  0.0000 -0.0000  0.0000
(2,1) -0.0000  1.0000 -0.0000 -0.0000 -0.1732 -0.0000 -0.0000
(3,1) -0.0000 -0.0000  1.0000  0.0000  0.0000  0.0946 -0.0000
(4,1)  0.9116 -0.0000  0.0000  1.0000  0.0000  0.0000 -0.0000
(5,1)  0.0000 -0.1732  0.0000  0.0000  1.0000 -0.0000 -0.0000
(6,1) -0.0000 -0.0000  0.0946  0.0000 -0.0000  1.0000  0.0000
(7,1)  0.0000 -0.0000 -0.0000 -0.0000 -0.0000  0.0000  1.0000
```

What do you do with canonical variables?

One thing to do is to make scatter plots of $\hat{z}_j^{(2)}$ vs $\hat{z}_j^{(1)}$.

```
Cmd> plot(Z1[,1],Z2[,1],xlab:"Canonical variable 1 for x1",\
        ylab:"Can var 1 for x2",\
        title:"First pair of canonical variables")

Cmd> plot(Z1[,2],Z2[,2],xlab:"Canonical variable 2 for x1",\
        ylab:"Can var 2 for x2",\
        title:"Second pair of canonical variables")
```



These are plots of $\hat{z}_1^{(2)}$ vs $\hat{z}_1^{(1)}$ (left) and $\hat{z}_2^{(2)}$ vs $\hat{z}_2^{(1)}$ (right).

And you can look at $\hat{\mathbf{u}}_j$ and $\hat{\mathbf{v}}_j$ to gain insight on what the canonical variables are made up from, much as you can do in MANOVA.

The $\hat{\theta}_i$ have the same information as the eigenvalues $\hat{\lambda}_1$, $\hat{\lambda}_2$, ... of $\mathbf{H}$ relative to $\mathbf{E}$ that appear in the multivariate regression tests of $\boldsymbol{\rho}_{12} = \mathbf{0}$.

$\hat{\theta}_i = \hat{\lambda}_i/(1 + \hat{\lambda}_i)$	$\hat{\lambda}_i = \hat{\theta}_i/(1 - \hat{\theta}_i)$

Only s = min(p,q) of these are non-zero.

The regression hypothesis and error matrices are

$\mathbf{H}_{1.2} = f_e\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$, $\mathbf{E}_{1.2} = f_e\mathbf{S}_{11} - \mathrm{H}_{1.2}$, $\mathbf{x}^{(1)}$ on $\mathbf{x}^{(2)}$

$\mathbf{H}_{2.1} = f_e\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$, $\mathbf{E}_{2.1} = f_e\mathbf{S}_{22} - \mathrm{H}_{2.1}$, $\mathbf{x}^{(2)}$ on $\mathbf{x}^{(1)}$

So $\hat{\lambda}_i$ is the i^{th} eigenvalue of $\mathbf{H}_{1.2}$ relative to $\mathbf{E}_{1.2}$ or of $\mathrm{H}_{2.1}$ relative to $\mathbf{E}_{2.1}$

When H_0: $\boldsymbol{\rho}_{12} = \mathbf{0}$ is true,

$$\{\hat{\lambda}_i\} = \{\hat{\theta}_i/(1 - \hat{\theta}_i)\}$$

You can use any of the MANOVA tests based on relative eigenvalues.

In terms of the canonical correlations and the matrix $\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{12}$

Hotelling's trace

$\sum\hat{\lambda}_i = \sum\hat{\theta}_i/(1 - \hat{\theta}_i)$

$= \mathrm{tr}(\mathbf{E}_{1.2}^{-1}\mathbf{H}_{1.2})$

$= \mathrm{tr}((\mathbf{S}_{11} - \mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})$

$= \mathrm{tr}((\mathbf{I}_p - \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})^{-1}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})$

LR test

$1/\Pi(1 + \hat{\lambda}_i) = \Pi(1 - \hat{\theta}_i)$

$= \det(\mathbf{E}_{1.2})/\det(\mathbf{H}_{1.2} + \mathbf{E}_{1.2})$

$= \det(\mathbf{I}_p - \mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})$

Pillai's trace

$\sum\hat{\lambda}_i/(1 + \hat{\lambda}_i) = \sum\hat{\theta}_i$

$= \mathrm{tr}\{(\mathbf{H}_{1.2} + \mathbf{E}_{1.2})^{-1}\mathbf{H}_{1.2}\}$

$= \mathrm{tr}(\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21})$

In these equations you can replace $\mathbf{S}_{11}$ by $\mathbf{S}_{22}$ and $\mathbf{S}_{11}^{-1}\mathbf{S}_{12}\mathbf{S}_{22}^{-1}\mathbf{S}_{21}$ by $\mathbf{S}_{22}^{-1}\mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12}$

Beyond Canonical Correlations

Here are two paths you might follow.

1. Use quadratic features instead of linear features. That is, try to find vectors $\mathbf{u}$ and $\mathbf{v}$ and symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ such that

$$\mathrm{Corr}[\mathbf{u}'\mathbf{x}^{(1)} + \mathbf{x}^{(1)\prime}\mathbf{A}\mathbf{x}^{(1)}, \mathbf{v}'\mathbf{x}^{(2)} + \mathbf{x}^{(2)\prime}\mathbf{B}\mathbf{x}^{(2)}]$$

is as large as possible

2. Describe the pattern of correlation among k > 2 sets of variables $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(k)}$. One possibility would be to find vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_k$ that minimize $\det(\mathbf{R}_U)$, where $\mathbf{R}_U$ is the correlation matrix of $\mathbf{u}_1'\mathbf{x}^{(1)}, \mathbf{u}_2'\mathbf{x}^{(2)}, \ldots, \mathbf{u}_k'\mathbf{x}^{(k)}$.

Since $\det(\mathbf{R}_U) = 1 - \rho^2[\mathbf{u}_1'\mathbf{x}^{(1)}, \mathbf{u}_2'\mathbf{x}^{(2)}]$ when k = 2, this leads to the ordinary canonical correlations when there are k = 2 groups of variables..

The classification problem

Situation: You have data **x** (1 or several variables) on an individual that is known to belong to one of g distinct populations π_1, π_2, ..., π_g.

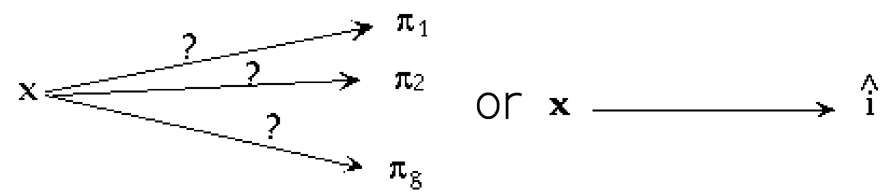
The ***classification problem***: Find a "rule" or formula which uses **x** to "guess" or "estimate" the population π_j the individual belongs to.

Example: When each population consists of patients with a particular disease and **x** contains an individual's medical history and test results, the classification problem would be to diagnose the correct disease from the information in **x**.

Statistics 5401 Lecture 33 November 23, 2005

More formally, suppose

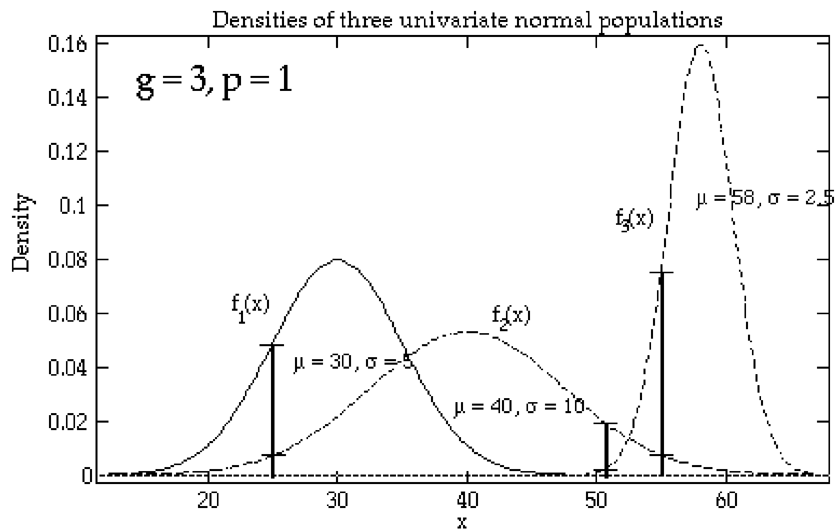
- You have a random vector **x** (the data) of p characteristics (variables).
- You know **x** has one of g densities $f_1(\mathbf{x})$, $f_2(\mathbf{x})$, ..., $f_g(\mathbf{x})$, where $f_j(\mathbf{x})$ defines the distribution of **x** in population π_j.
- You seek a procedure or formula (a "rule") that maps **x** to a population.



Here $\hat{i}$ is the guessed index of the population chosen.

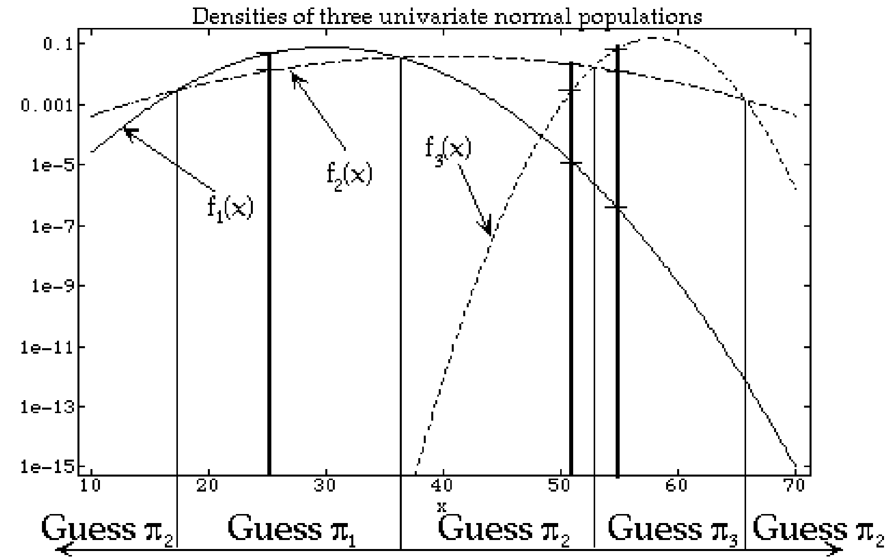
Suppose the observed **x** is much less likely to be observed in population π_1 (density $f_1(\mathbf{x})$) than in population π_2 (density $f_2(\mathbf{x})$). Then you might reasonably guess π_2 in preference to π_1.

Here are densities for three p=1 populations with normal distributions.



When x = 25, you would choose π_1 over π_2 or π_3; when x = 51, you would choose π_2; when x = 55, you would choose π_3.

It's often easier to compare densities when they are plotted in a log scale.

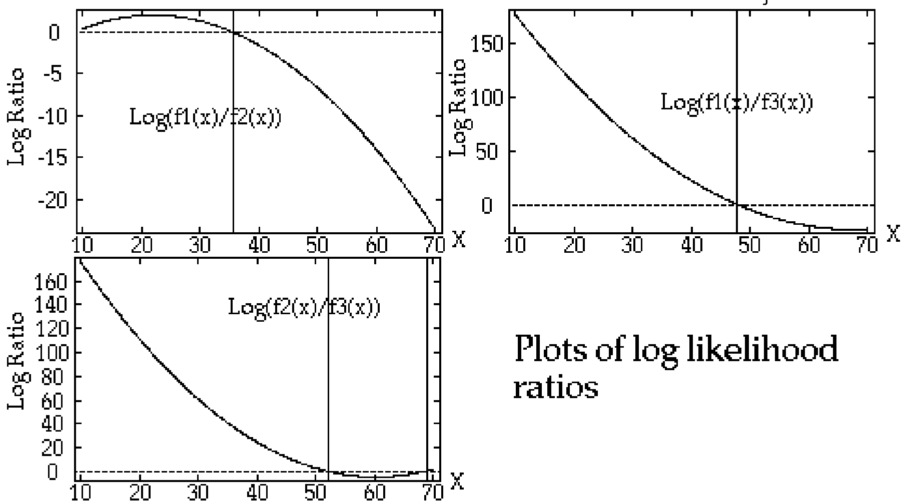


The extra vertical lines are where the densities intersect.

Under the graph is a sensible rule for picking one of these three populations - pick the population with largest density.

Near the boundary points you wouldn't be very sure about your decision based on this rule.

The logs of the ratios $f_j(x)/f_k(x)$ are informative for deciding between π_j and π_k.



Plots of log likelihood ratios

The 0 line is the line of equal likelihood. These let you choose between π_j and π_k

- When $10 < x < 35$, you would probably assign x to π_1 (above 0 in top 2 plots)
- x near 45 you would assign x to π_2
- $60 < x < 70$ you would assign x to π_3.

It looks like for $x < 10$ and $x > 70$, you should prefer π_2 to π_1 and π_3 even though x is nearer to μ_1 or μ_3 than to μ_2.

Effect of rarity

Suppose you knew, for example, that seeing *any* observation, regardless of value, from π_2 was extremely *rare* as compared to either π_1 or π_3. Then this "obvious" way to guess a population might change.

In that case, you might classify a value of x = 45 as coming from π_1, even though it would be an unlikely value to see from π_1, just because it is unlikely to see *any* individual from π_2.

In the extreme, if the chance of seeing any individual from π_2 was 1/1,000,000, for all practical purposes you can probably exclude π_2 from consideration and never pick π_2.