RP 2 Let $\underline{X}$ have covariance matrix $Cov[\underline{X}] = \underline{\Sigma}$ (pos definite)

and let $Y_i = \underline{e}_i^T \underline{X}$, $i = 1, 2, \ldots, p$ be the principal components.

Then $tr(\underline{\Sigma}) = \sum_{i=1}^{p} \sigma_{ii} = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} Var(Y_i)$

Proof. $tr(\underline{\Sigma}) = tr(\underline{P}\,\underline{\Lambda}\,\underline{P}^T) = tr(\underline{P}^T \underline{P}\,\underline{\Lambda}) = tr(\underline{\Lambda})$

$\sum_{i=1}^{p} \sigma_{ii}$ is the total population variance. Hence

$\dfrac{\lambda_k}{\sum_{i=1}^{p} \lambda_i}$ explains how much of the total population variance

that is explained by the k-th principal component.

RP 3. Let $\underline{Y}_i = \underline{e}_i^T \underline{X}$, $i = 1, 2, \ldots, p$ be the principal components

of $\underline{\Sigma}$. Then $\rho_{Y_i, X_k} = \dfrac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}$, $i, k = 1, 2, \ldots, p$.

Proof. Let $X_k = [0, \ldots, 0, 1, 0, \ldots, 1]\,\underline{X} = \underline{a}_k^T \underline{X}$

Then $Cov[X_k, Y_i] = E[\underline{a}_k^T [\underline{X} - \underline{\mu}][\underline{X} - \underline{\mu}]^T \underline{e}_i] = \underline{a}_k^T \underline{\Sigma}\, \underline{e}_i$

$= \underline{a}_k^T \lambda_i \underline{e}_i = \lambda_i e_{ik}$.

Therefore $\rho_{Y_i, X_k} = \dfrac{\lambda_i e_{ik}}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = \dfrac{\sqrt{\lambda_i}\, e_{ik}}{\sqrt{\sigma_{kk}}}$

The correlations between the principal components and the original variables are called principal component loadings. Sometimes only the weights (the elements in the principal components) are used as loads. (rotations)

Example.

$$\Sigma = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \begin{array}{l} \lambda_1 = 5.83 \\ \lambda_2 = 2.00 \\ \lambda_3 = 0.17 \end{array} \quad \begin{array}{l} e_1^T = [0.383, -0.924, 0] = [e_{11}, e_{12}, e_{13}] \\ e_2^T = [0, 0, 1] \\ e_3^T = [0.924, 0.383, 0] \end{array}$$

$$\rho_{y_1, x_1} = \frac{\sqrt{\lambda_1}\, e_{11}}{\sqrt{\sigma_{11}}} = \frac{\sqrt{5.83} \cdot 0.383}{\sqrt{1}} = 0.925$$

$$\rho_{y_1, x_2} = \frac{\sqrt{\lambda_1}\, e_{12}}{\sqrt{\sigma_{22}}} = \frac{\sqrt{5.83} \cdot -0.924}{\sqrt{5}} = -0.998$$

$e_{11} = 0.383$

$e_{12} = -0.924$

Let $Z_i = \dfrac{X_i - \mu_i}{\sqrt{\sigma_{ii}}}, \quad i = 1, 2, \ldots, p$ center and scale or

$$\underline{Z}_{p\times 1} = \begin{bmatrix} \sigma_{11}^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \sigma_{22}^{-\frac{1}{2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp}^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} = \underline{V}^{-\frac{1}{2}}(\underline{X} - \underline{\mu})$$

$\Rightarrow \quad Cov(\underline{Z}) = \underline{V}^{-\frac{1}{2}} \underline{\Sigma} \underline{V}^{-\frac{1}{2}} = \underline{\rho}$. i.e the correlation matrix

The principal components of $\underline{Z}$ are then found by means of the eigenvectors of $\underline{\rho}$

RP4. Let $\underline{Z}_{p\times 1}$ be standardized and let $Cov(\underline{Z}) = \underline{\rho}$

Let $Y_i = \underline{e}_i^T \underline{Z}$, $i = 1, 2, \cdots, p$ be the principal components of $\underline{Z}_{p\times 1}$. Then $\sum_{i=1}^{p} Var[Y_i] = \sum_{i=1}^{p} Var[Z_i] = p$

~~Proof.~~ and $\rho_{Y_i, Z_k} = e_{ik}\sqrt{\lambda_i}$, $i, k = 1, 2, \cdots, p$

where $(\lambda_1, \underline{e}_1), \cdots, (\lambda_p, \underline{e}_p)$ are the eigenvalue-eigenvector pairs for $\underline{\rho}$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

Proof. Follows from RP2.

Note. Principal component derived from $\underline{\Sigma}$ are different from those obtained from $\underline{\rho}$

Random variable with much higher variance than the others will dominate the first principal components.

Therefore it is normal procedure to standardize variables if they are measured on scales with different ranges or if their units are not commensurable.

(In questionnaires some question ~~s~~ may turn out to be less informative in the sense that all persons answers the same.)

Standardizing these may add too much values to these.

It is common to center the variables such that we work with $Y_i = \underline{e}_i^T(\underline{X} - \underline{\mu})$ or in practise $\hat{y}_i = \hat{e}_i^T(\underline{x} - \bar{\underline{x}})$

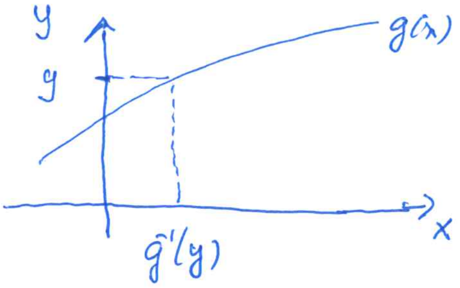
To determine how many principal components we should use, we can use a scree plot, or a plot of the variances of the principal components. Find the elbow and go one step to the left to decide the number of principal components.

values, (y_i) $i=1,2,\dots p$

We can also calculate the principal components for all the observations of the p vectors. These are called principal components scores and give n numbers for each principal component. n is the number of observation vectors.

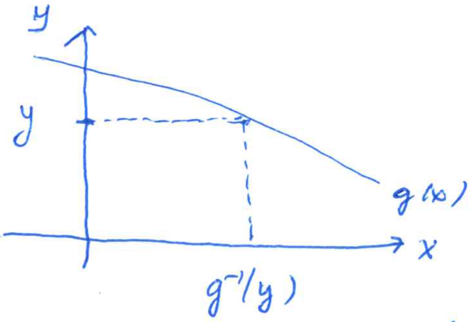
Transformations

Let X be a continuous random variable with pdf given by $f_X(x)$ and let $Y = g(X)$ where $g(x)$ is a one to one differentiable function. Then $g(x)$ is either monotonically increasing or decreasing



$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

$$= \begin{cases} P(X \le g^{-1}(y)), & g \text{ monoton increasing} \\ P(X \ge g^{-1}(y)), & g \text{ monoton decreasing} \end{cases}$$



$$= \begin{cases} F_X(g^{-1}(y)), & g \text{ monoton increasing} \\ 1 - F_X(g^{-1}(y)), & g \text{ monoton decreasing} \end{cases}$$

This implies

$$f_Y(y) = f_X(g^{-1}(y))\,\frac{dx}{dy}, \quad g \text{ monoton increasing}$$

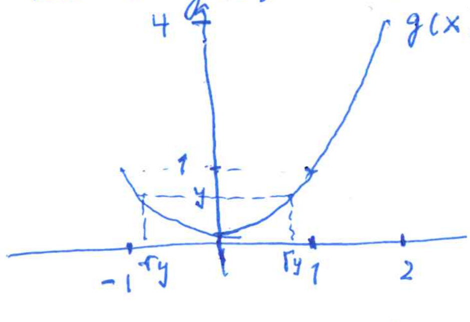
$$f_Y(y) = -f_X(g^{-1}(y))\,\frac{dx}{dy}, \quad g \text{ monoton decreasing}$$

But in the last case $\frac{dx}{dy}$ is negative $\Rightarrow$ we can write

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right| \quad \text{always.}$$

Example. $g(x)$ not monoton. Suppose X has support $-1 < x < 2$

Let $Y = g(X) = X^2$. Y has support $[0, 4]$



For $y \in (1, 4)$ we get $F_Y(y) = P(Y \le y) = P(X^2 \le y)$
$= P(X \le \sqrt{y}) = F_X(\sqrt{y})$ and $f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}}$

For $0 \le y \le 1$, $F_Y(y) = P(-\sqrt{y} \le X \le \sqrt{y})$
$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \Rightarrow f_Y(y) = f_X(\sqrt{y}) \cdot \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \cdot \frac{-1}{2\sqrt{y}}$

$$= \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}$$