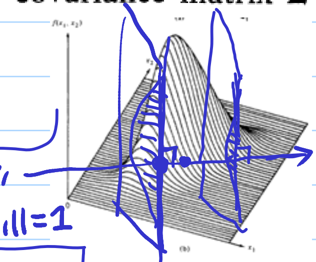
- Population Principal Components

- model: Let the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ have the covariance matrix $\boldsymbol{\Sigma}$ with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$.



- First principal component = linear combination

$$Y_1 = \mathbf{a}_1'\mathbf{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \qquad a_1=(a_{11},\cdots,a_{1p})$$

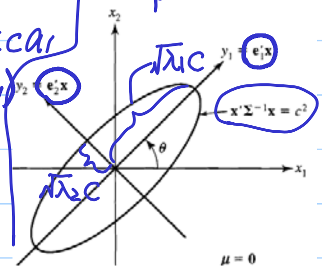
that maximizes $\mathrm{Var}(Y_1) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_1$

  - note: $\mathrm{Var}(Y_1)$ can be increased by multiplying any $\mathbf{a}_1$ by some constant $b_1=c\cdot a_1$, $b_1'X=c\cdot(a_1'X)$, $b_1'\Sigma b_1 = c a_1'\Sigma c a_1 = c^2(a_1'\Sigma a_1)$

    $\Rightarrow$ restrict attention to coefficient vectors of unit length $\Rightarrow \|a_1\|=1$

    $\Rightarrow$ $\mathbf{a}_1'\mathbf{X}$ is the projection of $\mathbf{X}$ on the direction of $\mathbf{a}_1$

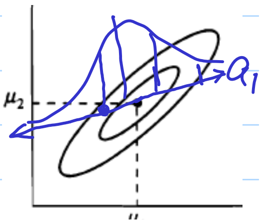
    $\Rightarrow$ alternative: $\max\limits_{\mathbf{a}\neq 0} \boxed{\dfrac{\mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}}{\mathbf{a}'\mathbf{a}}} = \dfrac{a'\Sigma a}{La^2} = \left(\dfrac{a}{La}\right)'\Sigma\left(\dfrac{a}{La}\right)$, $\left\|\dfrac{a}{La}\right\|=1$



- Second principal component = linear combination

$$Y_2 = \mathbf{a}_2'\mathbf{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p$$

that maximizes $\mathrm{Var}(\mathbf{a}_2'\mathbf{X}) = a_2'\Sigma a_2$ subject to $\mathbf{a}_2'\mathbf{a}_2 = 1$ and $\boxed{\mathrm{Cov}(\mathbf{a}_1'\mathbf{X}, \mathbf{a}_2'\mathbf{X}) = 0}$



  - note: $\mathrm{Cov}(Y_1, Y_2) = \mathbf{a}_1'\boldsymbol{\Sigma}\mathbf{a}_2$

- At the ith step, ith principal component = linear combination $\mathbf{a}_i'\mathbf{X}$ (Y_i) that maximizes $\mathrm{Var}(\mathbf{a}_i'\mathbf{X})$ subject to $\mathbf{a}_i'\mathbf{a}_i = 1$ and $\mathrm{Cov}(\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}) = 0$ for $k < i$
- The principal components are those *uncorrelated* linear combinations $Y_1, Y_2, \ldots, Y_p$ whose variances are as large as possible.

- Finding the Principal Components

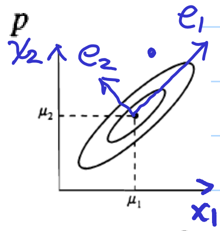
  - **Result 8.1.** Let $\boldsymbol{\Sigma}$ be the covariance matrix associated with the random vector $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$. Let $\boldsymbol{\Sigma}$ have the eigenvalue-eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$. Then the *ith principal component* is given by $\lambda_1 > \lambda_2 > \lambda_3 \cdots > \lambda_p > 0.$

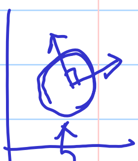
$$Y_i = \mathbf{e}_i'\mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \cdots + e_{ip}X_p, \qquad i = 1, 2, \ldots, p$$

With these choices, (any $c_1e_1+c_2e_2$ is an eigenvector; say $\lambda_1=\lambda_2$.)

$$\mathrm{Var}(Y_i) = \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_i = \lambda_i \qquad i = 1, 2, \ldots, p$$

$$\mathrm{Cov}(Y_i, Y_k) = \mathbf{e}_i'\boldsymbol{\Sigma}\mathbf{e}_k = 0 \qquad i \neq k$$



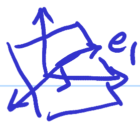


If some λ_i are equal, the choices of the corresponding coefficient vectors, $\mathbf{e}_i$, and hence Y_i, are not unique.

proof: 1st PC. $\max\limits_{a\neq 0} \dfrac{a'\Sigma a}{a'a} = \lambda_1$ (retained when $a=e_1$) LNp.2-17.

2nd PC $\mathrm{cov}(Y_1,Y_2) = e_1'\Sigma a = (\lambda_1 e_1)'a = \lambda_1(e_1'a) = 0 \Rightarrow e_1'a = 0.$ $a \perp e_1$

$\max\limits_{a\perp e_1} \dfrac{a'\Sigma a}{a'a} = \lambda_2$ (retained when $a=e_2$) $\cdots$



  - **Result 8.2.** Let $\mathbf{X}' = [X_1, X_2, \ldots, X_p]$ have covariance matrix $\boldsymbol{\Sigma}$, with eigenvalue–eigenvector pairs $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ where $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p \ge 0$. Let $Y_1 = \mathbf{e}_1'\mathbf{X}, Y_2 = \mathbf{e}_2'\mathbf{X}, \ldots, Y_p = \mathbf{e}_p'\mathbf{X}$ be the principal components. Then

$P=[e_1|\cdots|e_p]$

$\Lambda = \begin{bmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_p\end{bmatrix}$

$$\underbrace{\sigma_{11} + \sigma_{22} + \cdots + \sigma_{pp}}_{=tr(\Sigma)} = \sum_{i=1}^{p}\mathrm{Var}(X_i) = \underbrace{\lambda_1 + \lambda_2 + \cdots + \lambda_p}_{=tr(\Lambda)} = \sum_{i=1}^{p}\mathrm{Var}(Y_i)$$

$$tr(\Sigma) = tr(P\Lambda P') = tr(\Lambda P'P) = tr(\Lambda)$$

- $$\begin{pmatrix}\text{Proportion of total}\\ \text{population variance}\\ \text{due to } k\text{th principal}\\ \text{component}\end{pmatrix} = \frac{\lambda_k \; (= Var(Y_k))}{\underbrace{\lambda_1 + \lambda_2 + \cdots + \lambda_p}_{= \sum_{i=1}^{p} Var(X_i)}} \qquad k = 1, 2, \ldots, p$$

  If most (for instance, 80 to 90%) of the total population variance, for large p, can be attributed to the first one, two, or three components, then these components can "replace" the original p variables without much loss of information.

X_1=HW1
X_2=HW2
X_3=HW3

- Each component of the coefficient vector $\mathbf{e}_i' = [e_{i1}, \ldots, e_{ik}, \ldots, e_{ip}]$ also merits inspection. The magnitude of e_{ik} measures the importance of the kth variable (X_K) to the ith principal component, irrespective of the other variables. In particular, e_{ik} is proportional to the correlation coefficient between Y_i and X_k.

Y_1=0.63(HW1)
+0.57(HW2)
+0.52(HW3)
Y_2=0.67(HW1)
+0.08(HW2)
−0.75(HW3)
λ_1=0.9 λ_1=0.6
λ_2=0.01 λ_2=0.5

- **Result 8.3.** If $Y_1 = \mathbf{e}_1'\mathbf{X}$, $Y_2 = \mathbf{e}_2'\mathbf{X}, \ldots, Y_p = \mathbf{e}_p'\mathbf{X}$ are the principal components obtained from the covariance matrix $\mathbf{\Sigma}$, then $Y_i = e_i'X$, $X_k = [0, \cdots 0, 1, 0 \cdots 0]X$ ($= a'$)

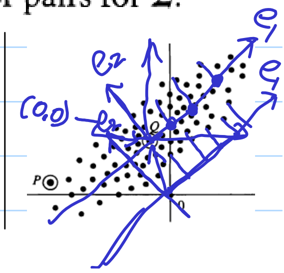
$$\frac{\lambda_i e_{ik}}{\sqrt{\lambda_i}\sqrt{\sigma_{kk}}} = \rho_{Y_i, X_k} = \frac{e_{ik}\sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}} \qquad i, k = 1, 2, \ldots, p$$

$Cov(X_k, Y_i) = a'\Sigma e_i = a'\lambda_i e_i = \lambda_i a' e_i = \lambda_i e_{ik}$

are the correlation coefficients between the components Y_i and the variables X_k. Here $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue–eigenvector pairs for $\mathbf{\Sigma}$.

  - $e_{ik}\sqrt{\lambda_i}$ is called *factor loading* $\Sigma = P\Lambda P' = P\Lambda^{1/2}\Lambda^{1/2}P' = (P\Lambda^{1/2})(P\Lambda^{1/2})'$

- the i^{th} principal component scores $P\Lambda^{1/2} = [e_1 | \cdots | e_p]\begin{bmatrix}\sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_p}\end{bmatrix}$
  - $Y_i = \mathbf{e}_i'\mathbf{X}$
  - $Y_i = \mathbf{e}_i'(\mathbf{X} - \boldsymbol{\mu})$

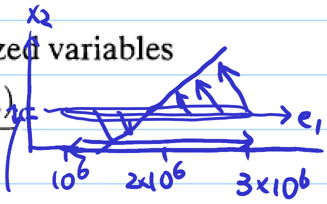


➢ Principal Components Obtained from Standardized Variables

- **Q**: What if one variables is measured in the millions whereas the others are measured in tens? or one variable has much larger scales than other variable?

  ⇒ The 1st PC will essentially be just that variable

- Principal components may also be obtained for the standardized variables



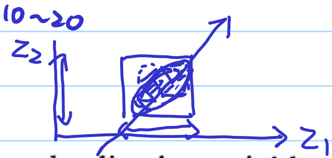
$$Z_1 = \frac{(X_1 - \mu_1)}{\sqrt{\sigma_{11}}}, \quad Z_2 = \frac{(X_2 - \mu_2)}{\sqrt{\sigma_{22}}}, \quad \ldots, \quad Z_p = \frac{(X_p - \mu_p)}{\sqrt{\sigma_{pp}}}$$

- $\mathbf{Z} = (\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu})$ and

  $\begin{bmatrix}Z_1\\ \vdots \\ Z_p\end{bmatrix}$ $\quad \mathbf{V}^{1/2} = \begin{bmatrix}\sigma_{11} & & 0\\ & \ddots & \\ 0 & & \sigma_{pp}\end{bmatrix}$

  $$\text{Cov}(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1}\mathbf{\Sigma}(\mathbf{V}^{1/2})^{-1} = \boldsymbol{\rho}$$

  where $\boldsymbol{\rho}$ is the *correlation* matrix of $\mathbf{X}$.



- **Result 8.4.** The ith principal component of the standardized variables $\mathbf{Z}' = [Z_1, Z_2, \ldots, Z_p]$ with $\text{Cov}(\mathbf{Z}) = \boldsymbol{\rho}$, is given by

$$Y_i = \mathbf{e}_i'\mathbf{Z} = \mathbf{e}_i'(\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu}), \qquad i = 1, 2, \ldots, p$$

PC scores

Moreover,

$$\sum_{i=1}^{p} \text{Var}(Y_i) = \sum_{i=1}^{p} \text{Var}(Z_i) = p \quad \Rightarrow \quad \boxed{\text{all variables equally important}}$$

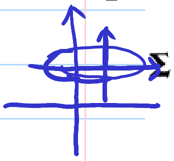
and

$$\rho_{Y_i, Z_k} = e_{ik}\sqrt{\lambda_i} \qquad i, k = 1, 2, \ldots, p$$

In this case, $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \ldots, (\lambda_p, \mathbf{e}_p)$ are the eigenvalue–eigenvector pairs for $\boldsymbol{\rho}$, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

- Note: *the* $(\lambda_i, \mathbf{e}_i)$ *derived from* $\mathbf{\Sigma}$ *are, in general, not the same as the ones derived from* $\boldsymbol{\rho}$.

- Principal Components for Covariance Matrices with Special Structures



- $\Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$

  $\mathbf{e}_i' = [0, \ldots, 0, 1, 0, \ldots, 0]$, with 1 in the ith position.
  $(\sigma_{ii}, \mathbf{e}_i)$ is the ith eigenvalue–eigenvector pair

- $\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}$ $\quad \boldsymbol{\rho} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}$

  $\lambda_1 = 1 + (p-1)\rho$

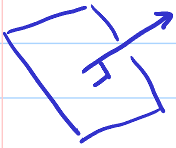
  $\mathbf{e}_1' = \left[\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \ldots, \frac{1}{\sqrt{p}}\right]$

  $\lambda_2 = \lambda_3 = \cdots = \lambda_p = 1 - \rho$

  $\mathbf{e}_2' = \left[\frac{1}{\sqrt{1 \times 2}}, \frac{-1}{\sqrt{1 \times 2}}, 0, \ldots, 0\right]$

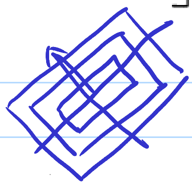
  $\mathbf{e}_3' = \left[\frac{1}{\sqrt{2 \times 3}}, \frac{1}{\sqrt{2 \times 3}}, \frac{-2}{\sqrt{2 \times 3}}, 0, \ldots, 0\right]$

  $\vdots \qquad \vdots$



$(1, \cdots, 1)$

$\boxed{\rho > 0}$

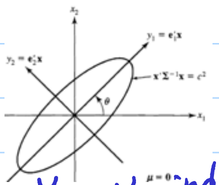


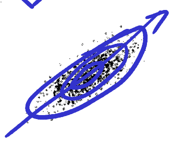
- Suppose $\mathbf{X}$ is distributed as $N_p(\boldsymbol{\mu}, \Sigma)$ — add normality assumption
  - contour of its pdf is the ellipsoid defined by

$P = [e_1 | \cdots | e_p]$ $\quad (\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2$

PC score: $Y = P'X \sim N_p(P'\mu, \Lambda)$

$Y = P'(X - \mu) \sim N_p(0, \Lambda)$





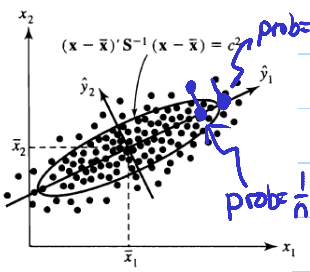
$Y_1, \ldots, Y_p$ indep.

❖ **Reading**: Textbook, 8.1, 8.2

- Sample Principal Components
  - Suppose the data $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ represent n independent drawings from some p-dimensional population with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ. These data yield the sample mean vector $\bar{\mathbf{x}}$, the sample covariance matrix $\mathbf{S}$, and the sample correlation matrix $\mathbf{R}$.

    parameters

    $\lambda_i, e_i \rightarrow$ parameters.
  - objective: construct uncorrelated linear combination of the measured characteristics that account for much of the variation in the sample
  - Finding principal components $\quad \boxed{P(x_i) = \frac{1}{n}, i = 1, \cdots, n}$
    - replace population distribution by empirical distribution
    - replace Σ by $\mathbf{S}$ (or $\mathbf{S}_n$)
    - replay $\boldsymbol{\rho}$ by $\mathbf{R}$
    - then, the rests the same as above
    - PCA based on $\mathbf{S}$ — random matrix.



If $\mathbf{S} = \{s_{ik}\}$ is the $p \times p$ sample covariance matrix with eigenvalue-eigenvector pairs $(\hat{\lambda}_1, \hat{\mathbf{e}}_1), (\hat{\lambda}_2, \hat{\mathbf{e}}_2), \ldots, (\hat{\lambda}_p, \hat{\mathbf{e}}_p)$, the ith sample principal component is given by $\rightarrow$ random.

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{x} = \hat{e}_{i1}x_1 + \hat{e}_{i2}x_2 + \cdots + \hat{e}_{ip}x_p, \qquad i = 1, 2, \ldots, p$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$ and $\mathbf{x}$ is any observation on the variables $X_1, X_2, \ldots, X_p$. Also,

$$\text{Sample variance}(\hat{y}_k) = \hat{\lambda}_k, \quad k = 1, 2, \ldots, p$$
$$\text{Sample covariance}(\hat{y}_i, \hat{y}_k) = 0, \quad i \neq k$$

In addition,

$$\text{Total sample variance} = \sum_{i=1}^{p} s_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

and

$$r_{\hat{y}_i, x_k} = \frac{\hat{e}_{ik}\sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}, \quad i, k = 1, 2, \ldots, p$$

- principal component scores

$$\hat{y}_i = \hat{\mathbf{e}}_i' \mathbf{x}, \quad i = 1, 2, \ldots, p$$
$$\hat{y}_i = \hat{\mathbf{e}}_i'(\mathbf{x} - \bar{\mathbf{x}}), \quad i = 1, 2, \ldots, p$$

- PCA based on R

If $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$ are standardized observations with covariance matrix $\mathbf{R}$, the ith sample principal component is

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{z} = \hat{e}_{i1}z_1 + \hat{e}_{i2}z_2 + \cdots + \hat{e}_{ip}z_p, \quad i = 1, 2, \ldots, p$$

where $(\hat{\lambda}_i, \hat{\mathbf{e}}_i)$ is the ith eigenvalue-eigenvector pair of $\mathbf{R}$ with $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$. Also,

$$\text{Sample variance } (\hat{y}_i) = \hat{\lambda}_i \quad i = 1, 2, \ldots, p$$
$$\text{Sample covariance } (\hat{y}_i, \hat{y}_k) = 0 \quad i \neq k$$

In addition,

$$\text{Total (standardized) sample variance} = \text{tr}(\mathbf{R}) = p = \hat{\lambda}_1 + \hat{\lambda}_2 + \cdots + \hat{\lambda}_p$$

fixed.

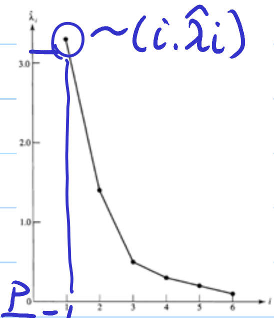
and

$$r_{\hat{y}_i, z_k} = \hat{e}_{ik}\sqrt{\hat{\lambda}_i}, \quad i, k = 1, 2, \ldots, p$$

factor loading.

- principal component scores

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{z}$$



- Number of Principal Components
  - **Q**: how many principal components to retain? → use $\hat{\lambda}_i$
  - scree plot
  - amount of total variance explained (say, 70~90%)
  - for PCA based on ρ, use the cutoff point 1 or 0.7 (i.e., keep PCs with $\hat{\lambda}_i \geq 1$ or 0.7) $\sum_{i=1}^{p} \hat{\lambda}_i = p$ $\hat{\lambda}_i \approx \frac{p}{p} = 1$
- Using Principal Components to display multivariate data
  - In addition to plotting $X_1, \ldots, X_p$, plot $Y_1, \ldots, Y_k$ to check normality assumption and detect suspect observations
  - The last few PCs can help pinpoint suspect observations

  → What's their difference?
- Large Sample Inference
  - **Q**: What are the distributions of the $(\hat{\lambda}_1, \ldots, \hat{\lambda}_p)$, $(\hat{e}_1, \ldots, \hat{e}_p)$, $(\hat{y}_1, \ldots, \hat{y}_p)$?
  - assumption $S \sim$ Wishart.
    - $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are a random sample from a normal population.
    - eigenvalues of $\boldsymbol{\Sigma}$ are distinct and positive, so that $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$.
  - asymptotic distribution
    1. Let $\boldsymbol{\Lambda}$ be the diagonal matrix of eigenvalues $\lambda_1, \ldots, \lambda_p$ of $\boldsymbol{\Sigma}$, then $\sqrt{n}\,(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is approximately $N_p(\mathbf{0}, 2\boldsymbol{\Lambda}^2)$. → $\begin{bmatrix} \lambda_1^2 & & 0 \\ & \lambda_2^2 & \\ 0 & & \lambda_p^2 \end{bmatrix}$

2. Let

$$\mathbf{E}_i = \lambda_i \sum_{\substack{k=1 \\ k \neq i}}^{p} \frac{\lambda_k}{(\lambda_k - \lambda_i)^2} \mathbf{e}_k \mathbf{e}_k'$$

then $\sqrt{n}\,(\hat{\mathbf{e}}_i - \mathbf{e}_i)$ is approximately $N_p(\mathbf{0}, \mathbf{E}_i)$.

3. Each $\hat{\lambda}_i$ is distributed independently of the elements of the associated $\hat{\mathbf{e}}_i$.

- A large sample $100(1-\alpha)\%$ confidence interval for λ_i is thus provided by

$$\frac{\hat{\lambda}_i}{(1 + z(\alpha/2)\sqrt{2/n})} \leq \lambda_i \leq \frac{\hat{\lambda}_i}{(1 - z(\alpha/2)\sqrt{2/n})}$$

$\hat{\lambda}_i - \lambda_i \overset{\mathcal{L}}{\approx} N(0, \frac{2}{n}\lambda_i^2)$

$\frac{\hat{\lambda}_i}{\lambda_i} - 1 \overset{\mathcal{L}}{\approx} N(0, \frac{2}{n})$

- Result 2 implies that the $\hat{\mathbf{e}}_i$'s are normally distributed about the corresponding $\mathbf{e}_i$'s for large samples. The elements of each $\hat{\mathbf{e}}_i$ are correlated, and the correlation depends to a large extent on the separation of the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$.

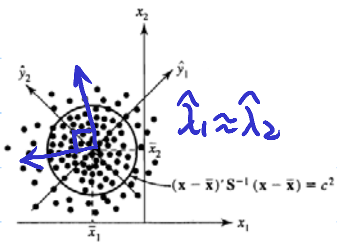
$\lambda_1 \approx p \quad \lambda_2 \cdots \lambda_p$ small.

- Some important issues
  - Analyses of principal components are more of a means to an end rather than an end in themselves, because they frequently serve as intermediate steps in much larger investigation. Example: regression, factor analysis, cluster analysis.
  - PCs with variance almost zero (i.e., $\hat{\lambda}_i \approx 0$)

$$\hat{y}_i = \hat{\mathbf{e}}_i'\mathbf{x} = \hat{e}_{i1}x_1 + \hat{e}_{i2}x_2 + \cdots + \hat{e}_{ip}x_p \approx c, \ c: \text{a constant}$$

$\Rightarrow$ an unusually small value for eigenvalue can indicate a linear dependency in the data set

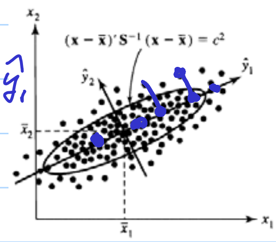
- equal eigenvalues (i.e., $\lambda_i = \lambda_{i+1} = \cdots = \lambda_j$)
  - population principal component
  - sample principal component



$\hat{\lambda}_1 \approx \hat{\lambda}_2$

- selecting a subset of variables $X_1, \ldots, X_p$
  - PCA does not quite reduce the data since PCs are linear combinations of all variables $X_1, \ldots, X_p$ so all variables are in some sense still needed. The technique, however, can help us discard some variables by keeping only those with the highest factor loading
- connection to the singular value decomposition (SVD)
  - Let us apply a SVD to the matrix $\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'$

0 $\hat{y}_1$



$$(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')_{n\times p} = U_{n\times p} L_{p\times p} V'_{p\times p}$$

where $U'U = I_m$, $V'V = I_m$, and L is a diagonal matrix. Then,

$$(n-1)\mathbf{S} = (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')'(\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}') = VLU'ULV' = V(LL)V' = VL^2V'$$

$$\mathbf{S} = V\left(\frac{1}{n-1}L^2\right)V' = V\left(\frac{1}{\sqrt{n-1}}L\right)^2 V' \equiv VL^{*2}V'$$

which means

- columns of V: eigenvectors of $\boldsymbol{S}$
- $\frac{l_i^2}{n-1}$: eigenvalues of $\boldsymbol{S}$, where l_i: singular value
- $\mathbf{Y} = (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}')V = UL$: PC scores