Week 8

[See Zheng, Jiang, Bai, He (2014).]

Understanding correlation between various 'features' of your data or model is important. We are now going to look at measures of population correlation & sample correlation in the $\underbrace{\text{two-variate (review)}}_{\text{"ordinary correlation"}}$ and $\underbrace{\text{multivariate}}_{\text{"multiple correlation"}}$ setting.

Ordinary correlation coefficients

Population correlation

The correlation between two variables X_1 and X_2 is defined by

$$\rho = \rho(X_1, X_2) = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}$$

If we write $\mathbb{X} := (X_1, X_2)'$ then the mean, covariance matrix and correlation matrix of $\mathbb{X}$ are

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \qquad R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Writing $\sigma_{ii} = \sigma_i^2$, the correlation between X_1 & X_2 is

$$\rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$$

Correlations are useful as they are invariant under scaling and shifts. This can be easily seen; consider

$$Z_1 = aX_1 + b \qquad Z_2 = cX_2 + d.$$

a, b, c, d constants. $a > 0$, $c > 0$.

$$Var(Z_1) = a^2 Var(X_1) \qquad Var(Z_2) = c^2 Var(X_2)$$

$$Cov(Z_1, Z_2) = ac\, cov(X_1, X_2).$$

$$\Rightarrow \rho(aX_1 + b, cX_2 + d) = \rho(X_1, X_2).$$

Recall that $-1 \leq \rho \leq 1$.

<u>Simple model</u>: Predict X_1 by a linear function of X_2. ie., $\alpha X_2 + \beta$ and choose optimal α, β in least-squares sense.

<u>Thm</u>: $\min\limits_{\alpha, \beta} \mathbb{E}[(X_1 - \alpha - \beta X_2)^2] = \sigma_1^2(1 - \rho^2)$.

The best linear predictor $\hat{X}_1 = \mu_1 + b(X_2 - \mu_2)$, $b = \rho \frac{\sigma_2}{\sigma_1}$.

$$\underbrace{\mathbb{E}(X_1 - \hat{X}_1)^2}_{\geq 0} = \sigma_1^2(1 - \rho^2).$$

notice this implies $|\rho| \leq 1$.

Sample correlation

Let $\mathbb{X}_1 = (X_{11}, X_{12})'$, $\mathbb{X}_2 = (X_{21}, X_{22})'$, $\ldots$ $\mathbb{X}_N = (X_{N1}, X_{N2})'$ be N random samples drawn from a population with mean μ and covariance Σ.

The sample mean and covariance are

$$\overline{\mathbb{X}} = \begin{pmatrix} \overline{X}_1 \\ \overline{X}_2 \end{pmatrix} \qquad \mathbb{S} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$$

and sample correlation

$$R = \frac{S_{12}}{\sqrt{S_{11} S_{22}}},$$

Thm: Let R be the sample correlation coefficient of a sample of size $N = n+1$ drawn from a bivariate normal distribution with correlation ρ. If $\rho = 0$, then

$$\sqrt{n-1}\, \frac{R}{\sqrt{1-R^2}}$$

has a t-distribution with $n-1$ degrees of freedom.

Proof: Recall that if $\mathbb{X}=(X_1,X_2,\dots X_N)'$ & $\mathbb{Y}=(Y_1,Y_2,\dots,Y_N)'$ then without loss of generality we can assume

$$\begin{pmatrix}X_1\\Y_1\end{pmatrix},\begin{pmatrix}X_2\\Y_2\end{pmatrix},\dots,\begin{pmatrix}X_N\\Y_N\end{pmatrix} \text{ iid } \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}\right)$$

$$S=\frac{1}{N-1}\sum_{i=1}^{N}(Z_i-\bar{Z})'(Z_i-\bar{Z})' \qquad Z_i:=\begin{pmatrix}X_i\\Y_i\end{pmatrix}$$

So since $N=n+1$,

$$nS\sim W(n,\Sigma) \qquad \Sigma=\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}$$

Wishart.

It follows that (since zero mean)

$$R=\frac{\sum_{i=1}^{N}X_iY_i}{\sqrt{\sum X_i^2\sum Y_i^2}}=\frac{\mathbb{X}'\mathbb{Y}}{\sqrt{\|\mathbb{X}\|^2\|\mathbb{Y}\|^2}}=\frac{A'\mathbb{Y}}{\sqrt{\|\mathbb{Y}\|^2}}$$

& $A:=\left(\frac{\mathbb{X}}{\|\mathbb{X}\|}\right.$.

Since $\rho=0$, $\mathbb{Y}$ independent from $\mathbb{X}$. and

$$\mathbb{Y}\mid\mathbb{X}\sim N(0,I_N)$$

conditional

If H is an orthogonal matrix (has real entries & $H'H'=I$)

then $\mathbb{Y}$ and $H'\mathbb{Y}$ have same dist. (since $\mathbb{X}\sim N(\mu,\Sigma)$, $B\mathbb{X}+b\sim N(B\mu+b, B\Sigma B')$)

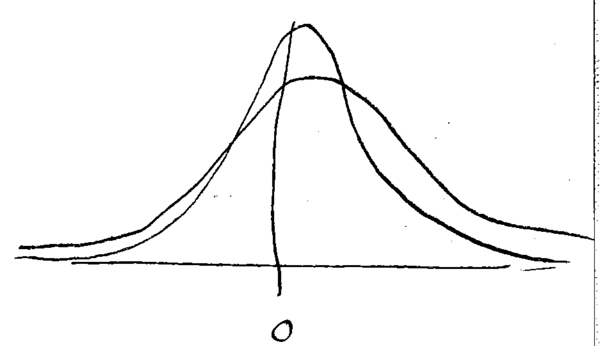
5

Take H as its first column A. Then

$$\frac{R}{\sqrt{1-R^2}} = \frac{Y_1}{\sum_{i=1}^{2} Y_i^2}.$$

$\blacksquare$

Recall that t-distribution ("Student's t") comes from an estimate of a mean of a normal distributed population when the sample size is small and population std. dev is unknown.



Density given by

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

ν : degrees of freedom.

$X_1, X_2 \ldots X_n \sim N(\mu, \sigma^2)$ iid.

$\overline{X} = \frac{1}{n}\sum X_i \quad S^2 = \frac{1}{n-1}\sum^{n} (X_i - \overline{X})^2$

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

Multiple correlation coefficient.

Population multiple correlation coefficient.

We consider relationship between X_1 and $\mathbb{X}_2 = (X_2, \ldots, X_p)'$

$$\mathbb{X} := (X_1, \mathbb{X}_2')'$$

with mean and covar. given by

Σ_{22} $(p-1)\times(p-1)$ cov matrix

$$\bar{\mu} = \begin{pmatrix} \mu_1 \\ \bar{\mu}_2 \end{pmatrix} \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \bar{\sigma}_{21}' \\ \bar{\sigma}_{21} & \Sigma_{22} \end{pmatrix}$$

The multiple correlation coefficient (MCC) can be characterised in different ways.

Consider predicting X_1 by the linear predictor of $\mathbb{X}_2$ given

$$\alpha + \bar{\beta}'\mathbb{X}_2 \qquad \bar{\beta} := (\beta_2, \ldots, \beta_p)' \qquad \alpha \in \mathbb{R}, \ \beta_i \in \mathbb{R}$$

then MCC is the maximum correlation between X_1 and any linear function $\alpha + \bar{\beta}'\mathbb{X}_2$. It is explicitly given by

$$\rho_{1(2\cdots p)} := \sqrt{\frac{\bar{\sigma}_{21}' \Sigma_{22}^{-1} \sigma_{21}}{\sigma_{11}}} \qquad (*)$$

Thm: For linear predictor $\alpha + \bar{\beta}' \mathbb{X}_2$ it holds that:

(1) $$\min_{\alpha, \bar{\beta}} \mathbb{E}\left[\left(X_1 - \alpha - \bar{\beta}' \mathbb{X}_2\right)^2\right] = \sigma_1^2 \left(1 - \rho^2_{1(2\cdots p)}\right).$$

(2) $$\max_{\alpha, \bar{\beta}} \rho\left(X_1, \alpha + \bar{\beta}' \mathbb{X}_2\right) = \max_{\bar{\beta}} \rho\left(X_1, \bar{\beta}' \mathbb{X}_2\right) = \rho_{1(2\cdots p)}.$$

Proof: see classic multivariate stats books. $\blacksquare$

Sample MCC

Let $$\mathbb{X}_1 = \begin{pmatrix} X_{11} \\ \mathbb{X}_{12} \end{pmatrix}, \ldots, \mathbb{X}_N = \begin{pmatrix} X_{N1} \\ \vdots \\ \mathbb{X}_{N2} \end{pmatrix}$$

be a random sample drawn from a population with mean vector $\bar{\mu}$ and covariance Σ. The sample covar. & corr. matrices are:

$$\bar{\mathbb{X}} = \begin{pmatrix} \bar{X}_1 \\ \bar{\mathbb{X}}_2 \end{pmatrix} \qquad \mathbb{S} = \begin{pmatrix} S_{11} & \mathbb{S}_{21}' \\ \mathbb{S}_{21} & \mathbb{S}_{22} \end{pmatrix} \qquad \mathbb{R} = \begin{pmatrix} 1 & \mathbb{R}_{21}' \\ \mathbb{R}_{21} & \mathbb{R}_{22} \end{pmatrix}$$

The sample MCC between X_1 and $\mathbb{X}_2$ is defined by

$$R_{1(2\cdots p)} = \sqrt{\frac{\mathbb{S}_{21}' \mathbb{S}_{22}^{-1} \mathbb{S}_{21}}{S_{11}}} = \sqrt{\mathbb{R}_{21}' \mathbb{R}_{22}^{-1} \mathbb{R}_{21}}$$

(which is equiv. to (*) but substituting $\mathbb{S}$ for Σ.)

When $\bar{\rho} := \rho_{1(2\cdots p)} = 0$ and writing $\bar{R} = R_{1(2\cdots p)}$. We can determine the distribution of $\bar{R}$ under the assumption that the population is a p-variate normal distribution, ie.

$\mathbf{X}_i \sim N(\mu, \Sigma)$, $i = 1, \ldots, N$.

Write $\mathbf{V} = n\mathbf{S}$ then $\mathbf{V} \sim W(n, \Sigma)$. Partition $\mathbf{V}$ similar to $\mathbf{S}$ (or Σ) as

$$\mathbf{V} = \begin{pmatrix} V_{11} & \mathbf{V}_{21}' \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$$

then sample MCC is given by

$$\bar{R}^2 = \frac{\mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}}{V_{11}}$$

$$= \frac{\mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}}{V_{11.2} + \mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}}$$

where $V_{11.2} = V_{11} - \mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}$. and note $\bar{\rho} = 0 \iff \bar{\sigma}_{21} = \underset{\sim}{0}$.

Since we know that $V_{11.2} \sim \chi^2(n-(p-1))$ and

$$\frac{\mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}}{\sigma_{11}} \sim \chi^2(p-1)$$

are they are independent this gives $\bar{R} \sim \dfrac{\chi^2_{p-1}}{\chi^2_{p-1} + \chi^2_{n-(p-1)}}$

Theorem: Let $\bar{R}$ be the sample MCC between X_1 and $\mathbb{X}_2 = (X_2, \dots, X_p)'$ based on a sample of size $N = n+1$ of $\mathbb{X} = (X_1, \mathbb{X}_2')'$ whose distribution is $N_p(\mu, \Sigma)$.
If population MCC $\bar{\rho} = 0$, then

$$\frac{n-(p-1)}{p-1} \cdot \frac{\bar{R}^2}{1-\bar{R}^2}$$

has F-distribution with $(p-1, n-(p-1))$ degrees of freedom.

Proof: See Anderson (2004), chap 4.

$\bar{R}^2$ is always nonnegative, this means that as an estimator of the population MCC $(\bar{\rho}^2 = 0)$ it has positive bias. This means that sometimes people prefer the adjusted MCC

$$(\bar{R}^*)^2 := \bar{R}^2 - \frac{p-1}{n-p}(1-\bar{R}^2)$$

which attempts to correct this bias.

Notice that $\bar{R}^{*2}$ is always smaller than $\bar{R}^2$ (unless $p=1$ or $\bar{R}^2 = 1$) and it has a smaller bias than $\bar{R}^2$. Unfortunately it can also become negative (with positive probability) and this contradicts the original interpretation of MCC.

Suppose p is fixed, and $n \to \infty$ (classic setting)

then:

$$\bar{R}^2 \xrightarrow{P} \bar{\rho}^2 \quad \text{and} \quad \bar{R}^{*2} \xrightarrow{P} \bar{\rho}^2$$

The case of $\bar{\rho} = 0$ can be seen [if you know that if $X \sim F(d_1, d_2)$ then $Y = \lim_{d_2 \to \infty} d_1 X \sim \chi^2_{d_1}$; see Wikipedia] as $F_{p-1, n-(p-1)} \xrightarrow{\mathcal{D}} \chi^2_{p-1}/(p-1)$

$\left(X \sim \chi^2_{p-1} \quad \mathbb{E}[X] = p-1 \right)$

Hence,

$$\frac{\bar{R}^2}{1-\bar{R}^2} \xrightarrow{P} 0$$

and $\bar{R}^2 \to 0$.

We are interested in the large dimensional case when $p, n \to \infty$ $p/n \to y \in [0, 1)$.

For simplicity we assume observations drawn from normal population.

Given a sample $\mathbb{X}_1, \mathbb{X}_2, \cdots \mathbb{X}_n$ from $N_p(\mu, \Sigma)$, then instead of $\mathbb{S} = \frac{1}{n-1}\sum(\mathbb{X}_i - \bar{\mathbb{X}})(\mathbb{X}_i - \bar{\mathbb{X}})'$ we equivalently consider $\mathbb{A}$ given by

$$\mathbb{A} := \sum_{i=1}^{N} \mathbb{Z}_i \mathbb{Z}_i' \qquad \mathbb{Z}_i \sim N_p(0, \Sigma) \text{ iid.}$$

then $\mathbb{A} \sim W(N, \Sigma)$ with $N = (n-1)$ degrees of freedom.

We can also write

$$\mathbb{A} = \underbrace{(\mathbb{Z}_1, \mathbb{Z}_2, \cdots \mathbb{Z}_N)}_{p \times N}(\mathbb{Z}_1, \cdots, \mathbb{Z}_N)^*$$
$$= (\mathbb{Y}_1, \cdots, \mathbb{Y}_p)^* \underbrace{(\mathbb{Y}_1, \cdots, \mathbb{Y}_p)}_{N \times p}$$

where $\mathbb{Y}_k$ are n-dimensional vectors.

We define matrices $\mathbb{Y}_2$ and $\mathbb{Y}_3$.

$$(\mathbb{Y}_1, \cdots \mathbb{Y}_p) = (\mathbb{Y}_1, \mathbb{Y}_2) = (\mathbb{Y}_1, \mathbb{Y}_2, \mathbb{Y}_3)$$

We can write $\bar{R} = \sqrt{\dfrac{S_{21} S_{22}^{-1} S_{21}}{S_{11}}} = \sqrt{\dfrac{a_1' A_{22}^{-1} a_1}{a_{11}}}$ in terms of $\mathbb{A}$.

Hence $\bar{R}^2 = \frac{a_1' \mathbb{A}_{22}^{-1} a_1}{a_{11}}$ where we have

$$a_{11} = Y_1' Y_1' \quad \in \mathbb{R}$$

$$a_{11} = \mathbb{Y}_2' Y_1 = \begin{pmatrix} Y_2' Y_2 \\ \mathbb{Y}_3' Y_1 \end{pmatrix}$$

$$\mathbb{A}_{22} = \mathbb{Y}_2' \mathbb{Y}_2 \qquad \mathbb{Y}_2 = (Y_2, \mathbb{Y}_3) = (Y_2, Y_3, \dots Y_p)$$

The MCC $\bar{R}^2$ is invariant under linear transformations of Y_1 or $\mathbb{Y}_2$ so we can assume

$$E[Y_k] = 0 \qquad \operatorname{cov}(Y_k) = I_N$$

$$\operatorname{cov}(Y_1, Y_2) = \bar{\rho} I_N \qquad \operatorname{cov}(Y_i, Y_j) = 0 \quad i<j \quad (i,j) \neq (1,2)$$

Since $\mathbb{A}_{22} = \begin{pmatrix} Y_2' Y_2 & Y_2' \mathbb{Y}_3 \\ \mathbb{Y}_3' Y_2 & \mathbb{Y}_3' \mathbb{Y}_3 \end{pmatrix}$ by the inversion formula for block matrices:

$$\mathbb{A}_{22}^{-1} = a_{22\cdot 3}^{-1} \begin{bmatrix} 1 & -Y_2 \mathbb{Y}_3 (\mathbb{Y}_3' \mathbb{Y}_3)^{-1} \\ -(\mathbb{Y}_3' \mathbb{Y}_3)^{-1} \mathbb{Y}_3' Y_2 & (\mathbb{Y}_3' \mathbb{Y}_3)^{-1} + (\mathbb{Y}_3' \mathbb{Y}_3)^{-1} \mathbb{Y}_3' Y_2 Y_2' \mathbb{Y}_3 (\mathbb{Y}_3' \mathbb{Y}_3)^{-1} \end{bmatrix}$$

where $a_{22\cdot 3} = \mathbb{Y}_2'(I_N - \mathbb{Y}_3(\mathbb{Y}_3'\mathbb{Y}_3)^{-1}\mathbb{Y}_3')\mathbb{Y}_2$.

We get

$$\overline{R}^2 = a_{11}^{-1}\left[\frac{(\mathbb{Y}_1'\mathbb{Y}_2 - \mathbb{Y}_2'\mathbb{Y}_3(\mathbb{Y}_3'\mathbb{Y}_3)^{-1}\mathbb{Y}_3'\mathbb{Y}_1)^2}{a_{22\cdot 3}} + \mathbb{Y}_1'\mathbb{Y}_3(\mathbb{Y}_3'\mathbb{Y}_3)^{-1}\mathbb{Y}_3'\mathbb{Y}_1)\right]$$

and by direct calculation and SLLN: $\left(y = \frac{p}{n}.\right)$

$$\frac{a_{11}}{n} \to 1, \quad \frac{a_{22\cdot 3}}{n} \to 1 - y,$$

$$\frac{\mathbb{Y}_1'\mathbb{Y}_2}{n} \to \overline{\rho}, \quad \frac{1}{n}\mathbb{Y}_2'\mathbb{Y}_3(\mathbb{Y}_3'\mathbb{Y}_3)^{-1}\mathbb{Y}_3'\mathbb{Y}_1 \to y\overline{\rho},$$

$$\frac{1}{n}\mathbb{Y}_2'\mathbb{Y}_3(\mathbb{Y}_3'\mathbb{Y}_3)^{-1}\mathbb{Y}_3'\mathbb{Y}_1 \to y\overline{\rho}$$

$$\frac{1}{n}\mathbb{Y}_1'\mathbb{Y}_3(\mathbb{Y}_3'\mathbb{Y}_3)^{-1}\mathbb{Y}_3'\mathbb{Y}_1 \to y$$

This implies:

<u>Thm:</u> Under Gaussian assumption, $p/n \to y \in [0,1)$

$$\overline{R}^2 \longrightarrow (1-y)\overline{\rho}^2 + y.$$

#

The previous theorem implies that, for $p \approx n$ case, the sample MCC $\bar{R}$ will nearly always overestimate the population MCC $\bar{\rho}$.

However, doing the same analysis for $\overline{R^*}^2$ one observes that the adjusted MCC remains consistent in the large-dimensional case.

CLT for sample MCC.

There exist some CLT results for $\bar{R}^2$ in the large-dimensional setting: $p, n \to \infty$, $p/n \to y$

Thm: $$\sqrt{n}\left(\frac{\bar{R}^2}{1-\bar{R}^2} - \frac{y_n + (1+y_n)\bar{\rho}^2}{(1-y_n)(1-\bar{\rho}^2)}\right) \to N\left(0, \frac{\sigma^2\bar{\rho}^2}{(1-y)^4(1-\bar{\rho}^2)^4}\right)$$

Thm: $$\sqrt{n}\left(\bar{R} - \sqrt{y_n + (1-y_n)\bar{\rho}^2}\right) \to N\left(0, \frac{\sigma^2\bar{\rho}^2}{4[y+(1-y)\bar{\rho}^2]}\right)$$