Continue with Multivariate Normal Distribution:

$$V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix} \sim N_p\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \underbrace{\begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}}_{\Gamma}\right]$$

Γ_{11}: $p_1 \times p_1$, Γ_{12}: $p_1 \times p_2$, Γ_{21}: $p_2 \times p_1$, Γ_{22}: $p_2 \times p_2$

V_1 as $p_1 \times 1$

V_2 as $p_2 \times 1$

V_1 and V_2 are independent if and only if $\Gamma_{12} = 0$ i.e V_1 and V_2 are uncorrelated.

$$\underset{p_1 \times p_2}{\Gamma_{12}} = \mathbb{E}\left[\underset{p_1 \times 1}{(V_1 - \mu_1)}\ \underset{1 \times p_2}{(V_2 - \mu_2)^T}\right]$$

We showed that marginals of normals are normals..

What about conditional distributions?

Let $\underline{X}_{p\times 1} = \begin{pmatrix} X_{1\,p_1\times 1} \\ X_{2\,p_2\times 1} \end{pmatrix} \sim N\left[\underset{\sim}{0}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21}, & \Sigma_{22} \end{pmatrix}\right]$

$p_1 + p_2 = p$

Interested in finding the best predictor of X_1 in terms of X_2.

$X_{1,j}$: jth comp. of X_1

It's best predictor in terms of X_2 is

$$\mathbb{E}[X_{1,j} \mid X_2]$$

$$\mathbb{E}[X_1 \mid X_2] = \begin{bmatrix} \mathbb{E}[X_{1,1} \mid X_2] \\ \vdots \\ \mathbb{E}[X_{1,p_1} \mid X_2] \end{bmatrix}$$

I'm guessing this form is linear taking the cue from BVN!

Consider $\underline{BLP[X_{1,j} \mid X_2]}$

BLP: best linear predictor

So, looking at:

Minimizing $\mathbb{E}[X_{1,j} - \underline{\gamma^T X_2}]^2$ ⊛

over all $\gamma_{p_2 \times 1}$.

(Where is the intercept? I'm dropping it.)

Let γ_j minimize (⊛).

Then $BLP[X_{1,j} \mid X_2] = \gamma_j^T X_2$.

Minimize $\mathbb{E}[X_{1,j} - \gamma^T X_2]^2$

Set $\nabla_\gamma \left[\mathbb{E}\left(X_{1,j} - \gamma^T X_2\right)^2\right] = 0$

$\Rightarrow -\mathbb{E}\left[\left(X_{1,j} - \gamma^T X_2\right) X_2\right] = 0$

i.e. $\mathbb{E}[X_{1j} X_2] - \mathbb{E}[X_2 X_2^T \gamma] = 0$

i.e. $\mathbb{E}[X_{1j} X_2] = \underbrace{\mathbb{E}[X_2 X_2^T]}\gamma = \Sigma_{22}\gamma$

so $\gamma_j = \Sigma_{22}^{-1}\, \mathbb{E}[x_{1j}\, x_2]$

$$\mathbb{E}[x_{1j}\, x_2] = \begin{pmatrix} \mathbb{E}[x_{1j}\, x_{21}] \\ \vdots \\ \mathbb{E}[x_{1j}\, x_{2p_2}] \end{pmatrix} = \tilde{\sigma}_{j,2}$$

What is Σ_{12}?

$$\Sigma_{12} = \mathbb{E}[x_1 x_2^T]$$

$$= \begin{pmatrix} \tilde{\sigma}_{1,2}^T \\ \tilde{\sigma}_{2,2}^T \\ \vdots \\ \tilde{\sigma}_{p_1,2}^T \end{pmatrix} \leftarrow \text{(check)}$$

$$\gamma_j = \Sigma_{22}^{-1}\, \tilde{\sigma}_{j2}$$

and so $\gamma_j^T = \tilde{\sigma}_{j2}^T\, \Sigma_{22}^{-1}$

$$BLP[x_{1j} \mid x_2] = \tilde{\sigma}_{j2}^T\, \Sigma_{22}^{-1}\, x_2$$

$\downarrow$

now stack these up for $1 \le j \le p_1$

$\underline{BLP[x_1 \mid x_2]}$

$$= \begin{pmatrix} \underline{\tilde{\sigma}_{1,2}^T \Sigma_{22}^{-1}} \\ \underline{\tilde{\sigma}_{2,2}^T \Sigma_{22}^{-1}} \\ \vdots \\ \tilde{\sigma}_{p,1,2}^T \Sigma_{22}^{-1} \end{pmatrix} \cdot \underline{x_2}$$

X_1

$$= \boxed{\Sigma_{12} \Sigma_{22}^{-1} x_2} .$$

Linear Residual of X_1 on X_2:

$$LR \equiv X_1 - BLP(X_1 \mid X_2)$$

$$= \underline{x_1 - \Sigma_{12} \Sigma_{22}^{-1} x_2} .$$

I'm going to consider ~~the~~ transformation.

$$(X_1, \underline{X_2}) \longmapsto \begin{pmatrix} \underline{LR} \\ \underline{X_2} \end{pmatrix}$$

$$\begin{pmatrix} LR \\ X_2 \end{pmatrix} = \underbrace{\begin{bmatrix} I_{p_1 \times p_1} & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_{p_2 \times p_2} \end{bmatrix}}_{= M} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

M is a non-singular matrix.

$$\begin{pmatrix} \underline{LR} \\ \underline{X_2} \end{pmatrix} \sim N\left[\underset{\sim}{0}, \; M \Sigma M^T\right]$$

We're going to now show that LR and X_2 are independent.

Suffices to show ~~that~~ LR and X_2 are uncorrelated.

Consider:

$$\mathbb{E}[LR \cdot X_2] = \mathbb{E}\left[(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2)X_2^T\right]$$

$$= \mathbb{E}[\underline{X_1 X_2^T}] - \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}[\underline{X_2 X_2^T}]$$

$$= \Sigma_{12} - \Sigma_{12}\cancel{\Sigma_{22}^{-1}}\cancel{\Sigma_{22}} = 0$$

so LR and x_2 are independent.

Now: $x_1 = \underline{\Sigma_{12}\Sigma_{22}^{-1}x_2}$
$=$
$+ \underline{(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2)}$.

consider

$$\underline{\mathbb{E}[x_1|x_2] = \Sigma_{12}\Sigma_{22}^{-1}x_2} + \underline{\mathbb{E}\left[(x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2)\,|\,x_2\right]}$$

$\|$ (by indep.)

$$\mathbb{E}[x_1 - \Sigma_{12}\Sigma_{22}^{-1}x_2] = 0$$

so $\mathbb{E}[x_1|x_2]$ is precisely the BLP!

and residual: LR is independent

of BLP!

<u>This is precisely a linear model.</u>

with a vector response x_1.

or p_1 multiple linear regression models with scalar responses which are the p_1 co-ordinates of X_1.

LR: $\underline{X_1 - \Sigma_{12}\Sigma_{22}^{-1} X_2}$

$\sim N[0, \underline{H}]$

$$H = \mathbb{E}\left[(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2)(X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2)^T\right]$$

$$= \mathbb{E}\left[(\underline{X_1} - \Sigma_{12}\Sigma_{22}^{-1}X_2)(X_1^T - \underline{X_2^T\Sigma_{22}^{-1}}\Sigma_{21})\right]$$

$$= \mathbb{E}[X_1X_1^T] - \mathbb{E}[X_1X_2^T]\Sigma_{22}^{-1}\Sigma_{21}$$

$$- \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}[X_2X_1^T]$$

$$+ \Sigma_{12}\Sigma_{22}^{-1}\mathbb{E}[X_2X_2^T]\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$- \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} + \Sigma_{12}\cancel{\Sigma_{22}^{-1}\Sigma_{22}}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$= \underline{\Sigma_{11\cdot 2}} \quad (\text{Schur complement})$$

$$X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2 \sim N[0, \Sigma_{11\cdot 2}]$$

and is independent of X_2

$$\underline{X_1 \mid X_2} \stackrel{d}{=} (\underline{X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2} + \underline{\Sigma_{12}\Sigma_{22}^{-1}X_2}) \mid X_2$$

$$= \Sigma_{12}\Sigma_{22}^{-1}X_2 + (X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2) \mid X_2$$

$$= \Sigma_{12}\Sigma_{22}^{-1}X_2 + N[0, \Sigma_{11\cdot 2}]$$

$$= N[\Sigma_{12}\Sigma_{22}^{-1}X_2, \Sigma_{11\cdot 2}]$$

conditional mean is linear

conditional variance free of X_2

Exercise:

Now suppose $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma\right]$

as partitioned before.

Find the conditional distributions of $x_1 | x_2$ and $x_2 | x_1$.

Fact: We can have for a pair (X, Y) where X and Y are random variables that:

X is normally distributed

$Y|X$ is normally distributed but

(X, Y) is not JOINTLY normal.

All I need is $Y|X \sim N[\mu(X), \sigma^2(X)]$

where $\mu(X)$ is non-linear in X or $\sigma^2(X)$ depends nontrivially on X.

Long term behavior of sequences of random variables: Notions of Convergence.

Convergence in distribution already defined..

$\{X_n\}$. $X_n \xrightarrow{d} X$ if

$F_{X_n}(x) \longrightarrow F_X(x)$ for all x

such that F_X is continuous at x.

$X_n \xrightarrow{d} X$ and $X_n \xrightarrow{d} -X$ are both simultaneously possible..

yes, when X is symmetric.

We are not considering here how close the actual values of X_n come close to the value of X.

$\{X_n\}$ of random variables defined on a given probability space $(\Omega, \mathcal{A}, P)$.

We say ~~that~~ X_n converges a.s (almost surely) to X if $\exists\, \Omega_0 \subseteq \Omega$ with $P(\Omega_0) = 1$, such that for any $\omega \in \Omega_0$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$

---

weaker version:

convergence in probability

We say ~~that~~ $X_n \to_p X$ if for any $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon)$$

$$= P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) \longrightarrow 0 \text{ as } n \to \infty.$$

convergence in probability $\Rightarrow$ convergence in distribution.

X can be a constant

SRSWR

N individuals

N_D Democrats

N_R Republicans

$$p = \frac{N_D}{N}.$$

Sample n individuals WITH replacement.

$\{X_1, \ldots, X_n\}$ $X_i = 1$ if i'th person D

$= 0$ o.w

looked at $\frac{1}{n}\sum_{i=1}^{n} X_i = \hat{p}_n$

surrogate for population proportion p.

What happens to $\hat{p}_n$ as $n \to \infty$?

Statistican wants $\hat{p}_n$ to be close to p when n is large . . .

Question : Does $\hat{p}_n$ converge to p in probability ?

i.e $P\left(|\underline{\hat{p}_n} - \underline{p}| > \varepsilon\right) \longrightarrow 0$ as $n \to \infty$?

Yes, by WLLN (weak law of large numbers)

Statisticians want their estimates ($\underline{\underline{X_n}}$) based on sample (of size n) to get close to truth with more info. (i.e larger n)

Simple example illustrating convergence in probability:

$\{T_n\}$: sequence of random variables

$$\left.\begin{aligned} P(T_n = 0) &= 1 - \frac{1}{\sqrt{n}} \\ P(T_n = n^5) &= \frac{1}{\sqrt{n}}. \end{aligned}\right\}$$

$E\,T_n \to \infty$!

$\neq E(0) = 0.$

Does T_n converge in probability at all?

$T_n \xrightarrow{P} 0.$

$P(|T_n - 0| > \varepsilon) = \underline{P(|T_n| > \varepsilon)}$

ε fixed

$\underline{P(|T_n| \le \varepsilon)}$

$= P(T_n = 0)$ whenever $\varepsilon < n^5$.

$= 1 - \frac{1}{\sqrt{n}} \to 1$!