read Chapter 2

(II) Random Variables, Stochastic Processes

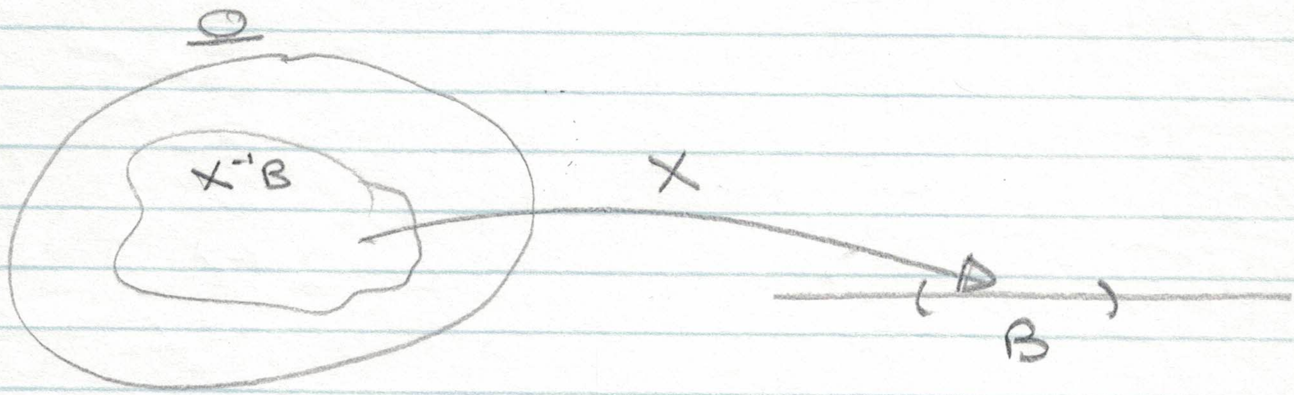
- suppose we have a prob. model $(\Omega, \mathcal{F}, P)$

Def A random variable is a function $X : \Omega \to \mathbb{R}^1$ satisfying $X^{-1}B \in \mathcal{F}$ for every $B \in \mathcal{B}^1$. In general we write $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^1, \mathcal{B}^1)$

inverse images properties

Def The marginal probability measure P_X on $(\mathbb{R}^1, \mathcal{B}^1)$ induced by the probability model $(\Omega, \mathcal{F}, P)$ and the r.v. X is given by

$$P_X(B) = P(X^{-1}B)$$



- note ① $\mathcal{F}_X = \{ X^{-1}B \mid B \in \mathcal{B}^1 \}$ is σ-field on Ω a sub-σ-field of $\mathcal{F}$ called the σ-field generated by X

Proof: (i) $\Omega = X^{-1}\mathbb{R}$

(ii) if $A \in \mathcal{F}_X$ then $\exists B \in \mathcal{B}^1$ s.t. $X^{-1}B = A$ and thus $X^{-1}B^c = A^c$

(iii) if $A_i \in \mathcal{F}_X$ then $\exists B_i \in \mathcal{B}^1$ s.t. $X^{-1}B_i = A_i$ and $\bigcup A_i = \bigcup X^{-1}B_i = X^{-1}\bigcup B_i \in \mathcal{F}$

- note ② $\mathfrak{F}(\{X^{-1}(a,b] \mid a,b \in \mathbb{R}\}) = \mathfrak{F}_X$

Proof: Let $\mathcal{A} = \{A \in \mathcal{B}^1 \mid X^{-1}A \in \mathfrak{F}(\ )\}$

Suppose $A \in \mathcal{A}$

Then $(X^{-1}A)^c = X^{-1}A^c \in \mathfrak{F}(\ )$ implying $A^c \in \mathcal{A}$

and $A_1, A_2, \ldots \in \mathcal{A}$ then $X^{-1}\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} X^{-1}A_i \in \mathfrak{F}(\ )$

$\mathcal{B}^+$

which implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Therefore $\mathcal{A}$ is a σ-algebra containing $\{(a,b] \mid a,b \in \mathbb{R}\}$ and this implies $\mathcal{B}^1 \subseteq \mathcal{A}$ and therefore $\mathfrak{F}_X \subseteq \mathfrak{F}(\ )$. Since it is obvious that $\mathfrak{F}(\ ) \subseteq \mathfrak{F}_X$ we have the result.

- note ③ P_X is a prob. measure on $\mathcal{B}^1$

Proof: E_X.

Def A random vector is a function
$\underset{\sim}{X} : (\Omega, \mathfrak{F}) \to (\mathbb{R}^k, \mathcal{B}^k)$

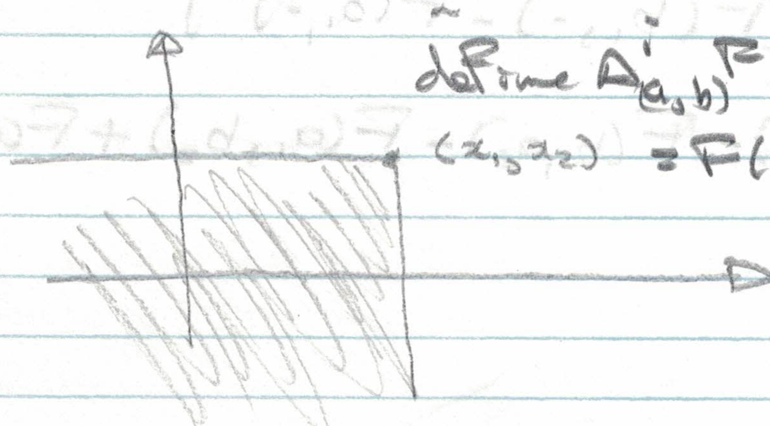
- similarly we get $\mathfrak{F}_{\underset{\sim}{X}}$ and $P_{\underset{\sim}{X}}$ and

$\mathfrak{F}_{\underset{\sim}{X}} = \mathfrak{F}(\{\underset{\sim}{X}^{-1}(\underset{\sim}{a}, \underset{\sim}{b}] \mid \underset{\sim}{a}, \underset{\sim}{b} \in \mathbb{R}^k\})$

where $(\underset{\sim}{a}, \underset{\sim}{b}] = \times_{i=1}^{k} (a_i, b_i]$

Def For random vector $\underset{\sim}{X}$ on $\mathbb{R}^k$ the distribution function $F_{\underset{\sim}{X}} : \mathbb{R}^k \to [0,1]$ is given by

$$F_{\underset{\sim}{X}}(x_1,\dots,x_k) = P_{\underset{\sim}{X}}\left((-\infty, x_1] \times \dots \times (-\infty, x_k] \right)$$



define $\Delta^{i}_{(a,b)}F$

(x_1, x_2) $= F(\cdot, \overset{i}{b}, \cdot) - F(\cdot, \overset{i}{a}, \cdot)$

Prop $F_{\underset{\sim}{X}}$ satisfies

(i) $P_{\underset{\sim}{X}}\left(\overset{k}{\underset{i=1}{\times}} (a_i, b_i] \right) = \Delta^{1}_{(a_1,b_1)} \cdots \Delta^{k}_{(a_k,b_k)} F \quad \forall\, \underset{\sim}{a}, \underset{\sim}{b} \in \mathbb{R}^k$ with $a_i \le b_i$, $i = 1, \dots, k$.

(ii) $\lim\limits_{x_i \to -\infty} F_{\underset{\sim}{X}}(\underset{\sim}{x}) = 0$, $\lim\limits_{\substack{x_i \to \infty \\ i=1,\dots,k}} F_{\underset{\sim}{X}}(\underset{\sim}{x}) = 1$

(iii) $F_{\underset{\sim}{X}}$ is right-continuous, i.e.
$$\lim_{\substack{\delta_i \downarrow 0 \\ i=1,\dots,k}} F_{\underset{\sim}{X}}(x_1+\delta_1, \dots, x_k+\delta_k) = F_{\underset{\sim}{X}}(x_1, \dots, x_k)$$

next page for (i)

Proof: (iii) $F_{\underset{\sim}{X}}(x_1+\delta_1, \dots, x_k+\delta_k) - F_{\underset{\sim}{X}}(x_1, \dots, x_k)$

$= P(X_1 \le x_1+\delta_1, \dots, X_k \le x_k+\delta_k) - P(X_1 \le x_1, \dots, X_k \le x_k)$

$\to 0$ as $\delta_i \downarrow 0$, $i = 1, \dots, k$ since

$\{X_1 \le x_1+\delta_1, \dots, X_k \le x_k+\delta_k\} \to \{X_1 \le x_1, \dots, X_k \le x_k\}$

and use cty of P

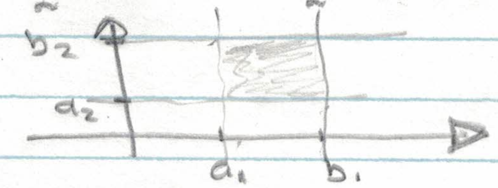
$$I_{(a_1,b_1]\times\mathbb{R}} = I_{(-\infty,b_1]\times\mathbb{R}} - I_{(-\infty,a_1]\times\mathbb{R}}.$$

$$I_{\mathbb{R}\times(a_2,b_2]} = I_{\mathbb{R}\times(-\infty,b_2]} - I_{\mathbb{R}\times(-\infty,a_2]}$$

$k=2$

(i) $P_{\underset{\sim}{X}}((\underset{\sim}{a},\underset{\sim}{b}]) = F_{\underset{\sim}{X}}(b_1,b_2) - F_{\underset{\sim}{X}}(a_1,b_2) - F_{\underset{\sim}{X}}(a_2,b_1) + F_{\underset{\sim}{X}}(a_1,a_2)$

Proof: Ex



note ① – in general we can evaluate $P_{\underset{\sim}{X}}((\underset{\sim}{a},\underset{\sim}{b}])$ using only the distribution function.

– this implies that we can evaluate $P_{\underset{\sim}{X}}(B)$ for all rectangles B using only the distribution function perhaps via limits

eg $k=1$

$$P_X(\{b\}) = F_X(b) - \lim_{a\uparrow b} F_X(a) = \lim_{a\uparrow b} P_X((a,b])$$

note ② – the Extension Theorem says the following: if you have a function $F_{\underset{\sim}{X}}: \mathbb{R}^k \to [0,1]$ satisfying (i),(ii),(iii) of Prop ① and if we define $P_{\underset{\sim}{X}}$ on rectangles using $F_{\underset{\sim}{X}}$ and the appropriate formula then there is a unique prob. measure on $\mathcal{B}^k$ that agrees with $P_{\underset{\sim}{X}}$ on rectangles.

note ③ – if $\underset{\sim}{Y} = (X_{i_1},\ldots,X_{i_j})$ for $\{i_1,\ldots,i_j\} \subseteq \{1,\ldots,k\}$ then $F_{\underset{\sim}{Y}}(y_1,\ldots,y_{i_j}) = P(Y_1 \le y_1,\ldots,Y_j \le y_j)$

$= P(X_{i_1} \le y_1,\ldots,X_{i_j} \le y_j)$

$= P((-\infty,\infty)\times\cdots\times(-\infty,\overset{i_1}{y_1}]\times\cdots\times(-\infty,\overset{i_j}{y_j}]\times\cdots\times(-\infty,\infty))$

$= F_{\underset{\sim}{X}}(\infty,\ldots$

- let S be a <u>state space</u> with σ-field $\mathcal{B}$ and T a <u>time domain</u>

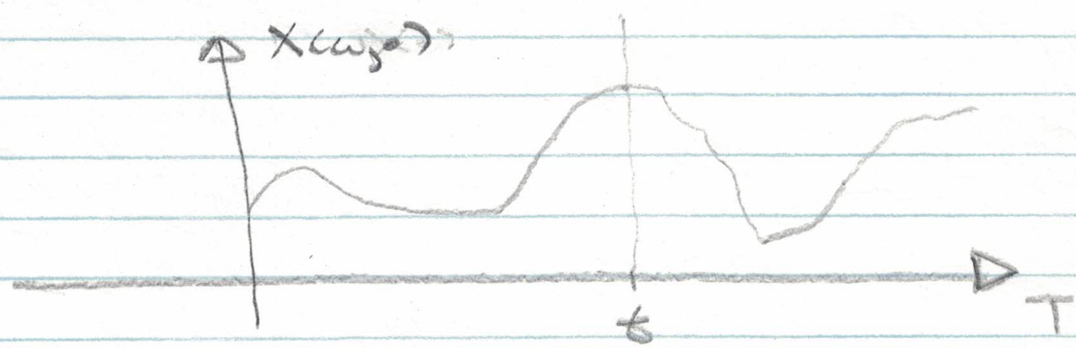
<u>Def</u> A function $X : (\Omega, \mathcal{F}) \to (S^T, \mathcal{B}^T)$ is called a <u>stochastic process</u>

↑ product space

- <u>note ①</u> – this gives rise to the probability model $(S^T, \mathcal{B}^T, P_X)$ where
$P_X(B) = P(X^{-1}B) \quad \forall B \in \mathcal{B}^T$

- <u>note ②</u> – a realization of the s.p. is a fn $X(\omega, \cdot) : T \to S$ sometimes called the <u>sample path</u>

<u>eg</u> $S = \mathbb{R}$, $T = [0, \infty)$



– we write $X(\cdot, t) = X_t$

- <u>note ③</u> – $S = \mathbb{R}$, $T = \{1\}$ gives random variables while $S = \mathbb{R}$, $T = \{1, \ldots, k\}$ gives random vectors

– typically we take $S = \mathbb{R}$, $\mathcal{B} = \mathcal{B}'$, $T = \mathbb{N}, \mathbb{Z}$, $[0, \infty)$ or $\mathbb{R}$

- how do you construct a s.p.?
- suppose $\mathcal{X} = \mathbb{R}$, $\mathcal{A} = \mathcal{B}^1$, $T \subseteq \mathbb{R}$
- for $i_1 < i_2 < \dots < i_k$ in T consider the joint prob. $P_{(i_1,\dots,i_k)}$ of $(X_{i_1},\dots,X_{i_k})$ with distribution function $F_{(i_1,\dots,i_k)}$
- we say this set of distribution functions is <u>consistent</u> if they marginalize appropriately; i.e. $F_{(i_1,i_2)}(x_1,x_2) = F_{(i_1,\dots,i_k)}(x_1,x_2,\infty,\dots,\infty)$ etc.

<u>Prop</u> ③ (<u>Kolmogorov Consistency Theorem</u>)
If we have a set of distribution fns indexed as above that are consistent then there is a unique prob. measure P on $(\mathbb{R}^T, \mathcal{B}^T)$ that satisfies

$$F_{(i_1,\dots,i_k)}(x_1,\dots,x_k) = P\left(\underset{t\in T}{\times} A_t \right)$$

where $A_t = \begin{cases} (-\infty, x_t] & \text{if } t \in \{i_1,\dots,i_k\} \\ \mathbb{R} & \text{otherwise.} \end{cases}$

Proof: accept

<u>note</u> - the stochastic process here is the identity, i.e. $X(\omega) = \omega$ for any $\omega \in \mathbb{R}^T$

- stopping times

- suppose $T \subseteq \mathbb{R}$ and $\mathcal{C}$ is a σ-field on T

- let $\mathcal{C}^* = \mathcal{F}(\{\mathcal{C}, \{\infty\}\})$

Def A function $T^* : (S^T, \mathcal{B}^T) \to (T \cup \{\infty\}, \mathcal{C}^*)$ is a stopping time if $\{T^* \le t^*\} \in \mathcal{F}(\{X_t : t \le t^*\})$ $\forall t^* \in T$.

- note - i.e. we stop at a time $\le t^*$ only based on process observed up to time t^* (we don't look into the future)

eg $T^*(x) \equiv t_0 \in T$ is a stopping time

$$T^*(x) = \begin{cases} t_0 & \text{when } X_{t_0} \in B \in \mathcal{A} \\ \infty & \text{otherwise} \end{cases}$$ is a stopping time

Def The probability model $(\mathbb{R}^k, \mathcal{B}^k, P)$ is <u>discrete</u> if there is a function $p : (\mathbb{R}^k, \mathcal{B}^k) \to (\mathbb{R}^1, \mathcal{B}^1)$ s.t.

$$P(A) = \sum_{\omega \in A} p(\omega)$$

for every $A \in \mathcal{B}^k$. In this case p is called the <u>probability function</u> of the model.

Prop ④ For discrete prob. model $(\mathbb{R}^k, \mathcal{B}^k, P)$ with prob. fn p we have

(i) $p(\underset{\sim}{x}) \geq 0$ | (ii) $\sum_{\underset{\sim}{x} \in \mathbb{R}^k} p(\underset{\sim}{x}) = 1$

(iii) $F(\underset{\sim}{x}) = \sum_{\underset{\sim}{y} \in (-\infty, \underset{\sim}{x}]} p(\underset{\sim}{y})$

(iv) $p(\underset{\sim}{x}) = P(\{\underset{\sim}{x}\}) = \lim_{\delta_1 \downarrow 0 \cdots \delta_k \downarrow 0} P((\underset{\sim}{x} - \underset{\sim}{\delta}, \underset{\sim}{x}])$

Proof: (i) $P(\{\underset{\sim}{x}\}) = p(\underset{\sim}{x}) \geq 0$, (ii) $1 = P(\mathbb{R}^k)$

$= \sum_{\underset{\sim}{x} \in \mathbb{R}^k} p(\underset{\sim}{x})$. (iii) $F(\underset{\sim}{x}) = P\left(\bigtimes_{i=1}^{k} (-\infty, x_i] \right)$.

(iv) $\lim_{\delta_1 \downarrow 0, \dots, \delta_k \downarrow 0} (\underset{\sim}{x} - \underset{\sim}{\delta}, \underset{\sim}{x}] = \{\underset{\sim}{x}\}$ and cty of P

- note ① $P \overset{\text{Extension Theorem}}{\Longleftrightarrow} F \overset{\text{discrete}}{\Longleftrightarrow} p$

- note ② if p sats (i), (ii) then $p(\underset{\sim}{x}) > 0$ for at most countably many $\underset{\sim}{x}$.

Proof: Let $A_n = p^{-1}\left(\frac{1}{n+1}, \frac{1}{n}\right]$. Then

$$1 = \sum_{\underset{\sim}{x} \in \mathbb{R}^k} p(\underset{\sim}{x}) = \sum_{n=1}^{\infty} \sum_{\underset{\sim}{x} \in A_n} p(\underset{\sim}{x}) \geq \sum_{n=1}^{\infty} \frac{1}{n+1} \#(A_n)$$

which implies $\#(A_n) < \infty$.

Prop ⑤ If $p : (\mathbb{R}^k, \mathcal{B}^k) \to (\mathbb{R}^1, \mathcal{B}^1)$ sats (i) and (ii) of Prop ④ then p is a prob. fn for the discrete prob. measure given by $P(B) = \sum_{\underset{\sim}{x} \in B} p(\underset{\sim}{x})$

Proof: Ex.

eg Multinomial $(n, p_1, \ldots, p_k)$

- where $n \in \mathbb{N}$, $p_i \geq 0$, $p_1 + \cdots + p_k = 1$

- $$p(x_1, \ldots, x_k) = \begin{cases} \binom{n}{x_1\, x_2 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} & x_i \in \{0, \ldots, n\},\ x_1 + \cdots + x_k = n \\ 0 & \text{otherwise} \end{cases}$$

- then p sats (i) and (ii) by the Multinomial Theorem.

eg Bernoulli (p) process (coin tossing)

- $\mathcal{X} = \mathbb{R}^1$, $\mathcal{A} = \mathcal{B}^1$, $T = \mathbb{N}$
- $$P_{(i_1, \dots, i_k)}(\underline{x}) = \begin{cases} p^{\sum_{j=1}^k x_{i_j}} (1-p)^{k - \sum_{j=1}^k x_{i_j}} & \underline{x} \in \{0,1\}^k \\ 0 & \text{otherwise} \end{cases}$$
- KCT then gives a s.p. $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, P)$

Def The probability model $(\mathbb{R}^k, \mathcal{B}^k, P)$ is absolutely continuous if there a function $f : (\mathbb{R}^k, \mathcal{B}^k) \to (\mathbb{R}^1, \mathcal{B}^1)$ s.t.

$$P(B) = \int_B f(\underline{x})\, d\underline{x}$$

for every $B \in \mathcal{B}^k$.

Prop ⑤ For abs. cont. model $(\mathbb{R}^k, \mathcal{B}^k, P)$ with density f we have

(i) $f(\underline{x}) \geq 0$ a.e. (ii) $\int_{\mathbb{R}^k} f(\underline{x})\, d\underline{x} = 1$

(iii) $F(\underline{x}) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f(z_1, \dots, z_k)\, dz_1 \dots dz_k$

(iv) $$f(\underline{x}) = \frac{\partial^k F(x_1, \dots, x_k)}{\partial x_1 \dots \partial x_k}$$

Proof: (i) $0 \leq P(f^{-1}(-\infty, 0)) = \int_{f^{-1}(-\infty,0)} f(\underline{x})\, d\underline{x} \leq 0$

(ii) $1 = P(\mathbb{R}^k) = \int_{\mathbb{R}^k} f(\underset{\sim}{x})\, d\underset{\sim}{x}$

(iii) $F(\underset{\sim}{x}) = P\left(\overset{k}{\underset{i=1}{\times}} (-\infty, x_i]\right) = \int_{\overset{k}{\underset{i=1}{\times}} (-\infty, x_i]} f(\underset{\sim}{z})\, d\underset{\sim}{z}$.

(iv) $\dfrac{\partial F(x_1, \dots, x_k)}{\partial x_k} = \dfrac{\partial}{\partial x_k} \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f(z_1, \dots, z_k)\, dz_1 \dots dz_k$

$\overset{FTC}{=} \int_{-\infty}^{x_{k-1}} \cdots \int_{-\infty}^{x_1} f(x_1, z_2, \dots, z_k)\, dz_1, \dots dz_k$.

-**note** - $P \Leftrightarrow F \Leftrightarrow f$ for abs. cont. models

Prop ⑥ If $f : (\mathbb{R}^k, \mathcal{B}^k) \to (\mathbb{R}^1, \mathcal{B}^1)$ sats (i) and (ii) of Prop ⑤ then f is a density fn for an abs. cont. model given by

$$P(B) = \int_B f(\underset{\sim}{x})\, d\underset{\sim}{x}$$

for $B \in \mathcal{B}^k$

Proof: (i) $P(\mathbb{R}) = 1$ (ii) $B_1, B_2, \dots \in \mathcal{B}^k$ mut. disj. then $P\left(\bigcup_{i=1}^{\infty} B_i\right) = \int_{\bigcup_{i=1}^{\infty} B_i} f(\underset{\sim}{x})\, dx \overset{fact}{\underset{④}{=}} \sum_{i=1}^{\infty} \int_{B_i} f(\underset{\sim}{x})\, d\underset{\sim}{x}$

$= \sum_{i=1}^{\infty} P(B_i)$.

MCT

eg <u>standard multivariate normal</u>

- $\underset{\sim}{X} \sim N_k(\underset{\sim}{0}, I)$

$f(\underset{\sim}{x}) = (2\pi)^{-\frac{k}{2}} \exp\{-\frac{1}{2}\underset{\sim}{x}'\underset{\sim}{x}\}$.

$f(\underset{\sim}{x}) \geq 0$

$\int_{\mathbb{R}^k} (2\pi)^{-k/2} \exp\{-\frac{1}{2}\underset{\sim}{x}'\underset{\sim}{x}\}\, d\underset{\sim}{x}$

$= \prod_{i=1}^{k} \int_{-\infty}^{\infty} (2\pi)^{-\frac{1}{2}} \exp\{-\frac{1}{2} x_i^2\}\, dx_i = 1$

eg <u>standard Gaussian process</u>

- $\mathcal{X} = \mathbb{R}^1$, $\mathcal{A} = \mathcal{B}^1$, $T = (-\infty, \infty)$

- $f_{(t_1, \dots, t_k)}(\underset{\sim}{x}) = N_k(\underset{\sim}{0}, I)$ density

- <u>change of variable</u>

- suppose $\underset{\sim}{X}$ is distributed on $\mathbb{R}^k$ with density $f_{\underset{\sim}{X}}$

- let $T: \mathbb{R}^k \to \mathbb{R}^k$ and put $\underset{\sim}{Y} = T(\underset{\sim}{X})$

- we want the distribution of $\underset{\sim}{Y}$

- $F_{\underset{\sim}{Y}}(y_1, \ldots, y_k) = P_{\underset{\sim}{Y}}(Y_1 \le y_1, \ldots, Y_k \le y_k)$

$$= P_{\underset{\sim}{X}}\left(T^{-1}\left((-\infty, y_1] \times (-\infty, y_2] \times \cdots \times (-\infty, y_k]\right)\right)$$

$$= \int_{T^{-1} \overset{\infty}{\underset{i=1}{X}} (-\infty, y_i]} f_{\underset{\sim}{X}}(\underset{\sim}{x})\, d\underset{\sim}{x}.$$

- now suppose T is 1-1 and smooth, namely, all derivatives of 1st order exist and are continuous.

- put $J_T(\underset{\sim}{x}) = \left| \det \begin{pmatrix} \frac{\partial T_1(\underset{\sim}{x})}{\partial x_1} & \cdots & \frac{\partial T_1(\underset{\sim}{x})}{\partial x_k} \\ \vdots & & \\ \frac{\partial T_k(\underset{\sim}{x})}{\partial x_1} & \cdots & \frac{\partial T_k(\underset{\sim}{x})}{\partial x_k} \end{pmatrix} \right|^{-1}$

- then change of variable formula gives for small nbd $B(\underset{\sim}{x}) \subseteq \mathbb{R}^k$ of $\underset{\sim}{x}$

$$\mathrm{vol}(T B(\underset{\sim}{x})) \simeq J_T^{-1}(\underset{\sim}{x})\, \mathrm{vol}(B)$$

- let $\underset{\sim}{y} = T(\underset{\sim}{x})$ so $\underset{\sim}{x} = T^{-1}(\underset{\sim}{y})$

- this leads to

$$P_{\underset{\sim}{Y}}(TB(T^{-1}(\underset{\sim}{y}))) \approx \mathrm{vol}(TB(T^{-1}(\underset{\sim}{y})))\, f_{\underset{\sim}{Y}}(\underset{\sim}{y})$$

$$\| $$

$$P_{\underset{\sim}{X}}(B(T^{-1}(\underset{\sim}{y})))$$

$$\approx \mathrm{vol}(B(T^{-1}(\underset{\sim}{y})))\, f_{\underset{\sim}{X}}(T^{-1}(\underset{\sim}{y}))$$

$$\therefore\ \mathrm{vol}(TB(T^{-1}(\underset{\sim}{y})))\, f_{\underset{\sim}{Y}}(\underset{\sim}{y})$$

$$\approx J_T(T^{-1}(\underset{\sim}{y}))\, \mathrm{vol}(B(T^{-1}(\underset{\sim}{y})))\, f_{\underset{\sim}{Y}}(\underset{\sim}{y})$$

$$\approx \mathrm{vol}(B(T^{-1}(\underset{\sim}{y})))\, f_{\underset{\sim}{X}}(T^{-1}(\underset{\sim}{y}))$$

$$\therefore\ f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = f_{\underset{\sim}{X}}(T^{-1}(\underset{\sim}{y}))\, J_T(T^{-1}(\underset{\sim}{y}))$$

eg $T : \mathbb{R}^k \to \mathbb{R}^k$ given by

$\underset{\sim}{y} = T(\underset{\sim}{x}) = \underset{\sim}{a} + B\underset{\sim}{x}$ where $B \in \mathbb{R}^{k \times k}$ is invertible

- then $J_T(\underset{\sim}{x}) = |\det B|^{-1}$ and density of $\underset{\sim}{y}$ is

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = f_{\underset{\sim}{X}}(B^{-1}\underset{\sim}{y})\, |\det B|^{-1}$$

eg **general multivariate normal**

- suppose $\underset{\sim}{Z} \sim N_k(\underset{\sim}{0}, I)$ so

$$f_{\underset{\sim}{Z}}(\underset{\sim}{z}) = (2\pi)^{-k/2} \exp\{-\underset{\sim}{z}'\underset{\sim}{z}/2\}$$

- put $\underset{\sim}{X} = \underset{\sim}{\mu} + B\underset{\sim}{Z}$ where $\underset{\sim}{\mu} \in \mathbb{R}^k$, $B \in \mathbb{R}^{k \times k}$ is invertible.

so $\underset{\sim}{z} = B^{-1}(\underset{\sim}{x} - \underset{\sim}{\mu})$. $(T^{-1}(\underset{\sim}{x}))$

- $\therefore f_{\underset{\sim}{X}}(\underset{\sim}{x}) = |\det B|^{-1} \exp\{-\frac{1}{2}(B^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}))'(B^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}))\}$

$$= |\det B|^{-1} \exp\{-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'(B^{-1})'B^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\}$$

$$(B^{-1})'B^{-1} = (B')^{-1}B^{-1} = (BB')^{-1}$$

$$= \Sigma^{-1} \text{ where } \Sigma = BB'$$

$$\det B = (\det B \det B')^{1/2} = (\det BB')^{1/2}$$

$$= (\det \Sigma)^{1/2}$$

$$= (\det \Sigma)^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\}$$

- we say $\underset{\sim}{X} \sim N_k(\underset{\sim}{\mu}, \Sigma)$

mean vector (for $\underset{\sim}{\mu}$) variance matrix (for Σ)

- note $\underset{\sim}{x}'\Sigma\underset{\sim}{x} = \underset{\sim}{x}'BB'\underset{\sim}{x} = \|B'\underset{\sim}{x}\|^2 \geq 0$ and > 0 when $\underset{\sim}{x} \neq \underset{\sim}{0}$ so Σ is positive definite.

- **note** - if $\underset{\sim}{x} \sim N_k(\mu, \Sigma)$ and let B be s.t. $\Sigma = BB'$ where $B \in \mathbb{R}^{k \times k}$ invertible
  - then $\underset{\sim}{z} = B^{-1}(\underset{\sim}{x} - \mu) \sim N_k(\underset{\sim}{0}, I)$ Ex
  - $\Sigma \overset{\text{spectral decomp}}{=} U \Lambda U'$ where
    $U \in \mathbb{R}^{k \times k}$ orthogonal
    $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ where $\lambda_i > 0$
  - put $\Sigma^{1/2} = U \Lambda^{1/2} U'$ where $\Lambda^{1/2} = \operatorname{diag}(\lambda_1^{1/2}, \ldots, \lambda_k^{1/2})$
  - then $(\Sigma^{1/2})' = \Sigma^{1/2}$ and $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})'$
  - $\Sigma^{1/2}$ is called the symmetric square root of Σ
  - apply Gram-Schmidt to columns of $\Sigma^{1/2}$ so $\Sigma^{1/2} = QR$ where $Q \in \mathbb{R}^{k \times k}$ is orthogonal and $R \in \mathbb{R}^{k \times k}$ is upper Δ with positive diagonal elements
  - $\Sigma = \Sigma^{1/2} \Sigma^{1/2} = (\Sigma^{1/2})' \Sigma^{1/2}$
    $= R'Q'QR = R'R$
  - R is called the Cholesky factor of Σ

- **note** $f_{\underset{\sim}{X}}(\underset{\sim}{x}) = (\det \Sigma)^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\}$

so level sets of $f_{\underset{\sim}{X}}$ are the sets

$$\{\underset{\sim}{x} : (\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) = c\} \qquad c>0$$

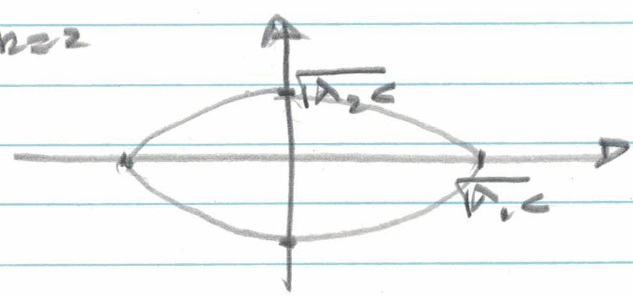
$$= \{\underset{\sim}{x} : (\underset{\sim}{x}-\underset{\sim}{\mu}) Q \Lambda^{-1} Q' (\underset{\sim}{x}-\underset{\sim}{\mu}) = c\}$$

$$= \underset{\sim}{\mu} + Q\{\underset{\sim}{z} : \underset{\sim}{z}'\Lambda^{-1}\underset{\sim}{z} = c\}$$

$$\underset{\sim}{z} = Q'(\underset{\sim}{x}-\underset{\sim}{\mu})$$

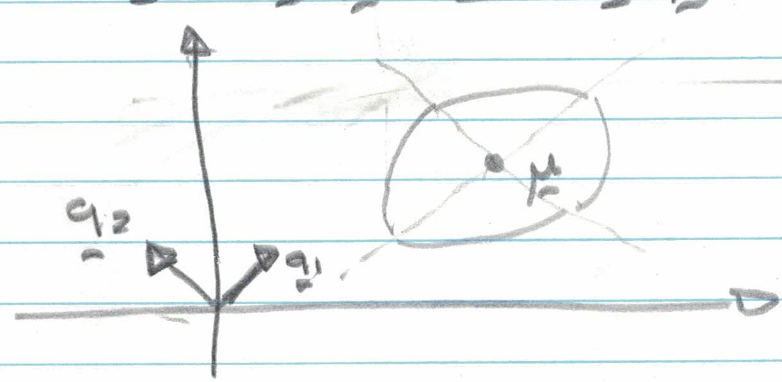
so $\{\underset{\sim}{z} : \underset{\sim}{z}'\Lambda^{-1}\underset{\sim}{z} = c\} = \{\underset{\sim}{z} : \sum_{i=1}^{n} z_i^2/\lambda_i = c\}$

$n=2$ $\qquad\qquad$ $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$



an ellipsoid in $\mathbb{R}^n$ centered at $\underset{\sim}{0}$ with semi-axes given by coordinate axes

$$\{\underset{\sim}{x} : (\underset{\sim}{x}-\underset{\sim}{\mu})'\Sigma^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu}) = c\}$$



- **note** – suppose $\underset{\sim}{X} \sim N_k(\underset{\sim}{\mu}, \Sigma)$ and $\underset{\sim}{Y} = \underset{\sim}{a} + C\underset{\sim}{X}$ where $\underset{\sim}{a} \in \mathbb{R}^{k \times k}$, $C \in \mathbb{R}^{k \times k}$ invertible

  - now $\Sigma = BB'$ for some $B \in \mathbb{R}^{k \times k}$ invertible and putting $\underset{\sim}{Z} = B^{-1}(\underset{\sim}{X} - \underset{\sim}{\mu})$ we have $\underset{\sim}{X} = \underset{\sim}{\mu} + B\underset{\sim}{Z}$ where $\underset{\sim}{Z} \sim N_k(\underset{\sim}{0}, I)$.

  - so $\underset{\sim}{Y} = \underset{\sim}{a} + C\underset{\sim}{X} = \underset{\sim}{a} + C\underset{\sim}{\mu} + CB\underset{\sim}{Z}$

$$\sim N_k(\underset{\sim}{a} + C\underset{\sim}{\mu}, (CB)(CB)')$$

$$= N_k(\underset{\sim}{a} + C\underset{\sim}{\mu}, CBB'C')$$

$$= N_k(\underset{\sim}{a} + C\underset{\sim}{\mu}, C\Sigma C')$$

eg **permutations**

- suppose $C \in \mathbb{R}^{k \times k}$ is a permutation matrix, namely each row and column contains one and only one 1 with the rest 0's

$C = I$ so $Cx = x$

$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$, $C\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_3 \\ x_2 \end{pmatrix}$ etc

$\underset{\sim}{X} \sim N_k(\underset{\sim}{\mu}, \Sigma)$ so $C\underset{\sim}{X} \sim N_k(C\underset{\sim}{\mu}, C\Sigma C')$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{13} & \sigma_{12} \\ \sigma_{13} & \sigma_{33} & \sigma_{23} \\ \sigma_{12} & \sigma_{23} & \sigma_{22} \end{pmatrix}$$

Def Random variables X and Y on $(\Omega, \mathcal{F}, P)$ are <u>statistically independent</u> if $\mathcal{F}_X$ and $\mathcal{F}_Y$ are statistically independent.

Prop 7 Random variables X and Y on $(\Omega, \mathcal{F}, P)$ are statistically independent iff
$$F_{(X,Y)}(x,y) = F_X(x)\,F_Y(y)$$
for every $x, y \in \mathbb{R}$.

Proof: $\Rightarrow]$ $F_{(X,Y)}(x,y) = P_{(X,Y)}\big((-\infty,x]\times(-\infty,y]\big)$

$= P\big((X,Y)^{-1}(-\infty,x]\times(-\infty,y]\big)$

$= P\big(X^{-1}(-\infty,x] \cap Y^{-1}(-\infty,y]\big) = P\big(X^{-1}(-\infty,x]\big)P\big(Y^{-1}(-\infty,y]\big)$

$= P_X\big((-\infty,x]\big)\,P_Y\big((-\infty,y]\big) = F_X(x)\,F_Y(y)$.

$\Leftarrow]$ $P_{(X,Y)}\big((\underline{a},\underline{b}]\big) = F_{(X,Y)}(b_1,b_2) - F_{(X,Y)}(a_1,b_2) - F_{(X,Y)}(a_2,b_1)$

$+ F_{(X,Y)}(a_1,a_2) = \big(F_X(b_1) - F_X(a_1)\big)\big(F_Y(b_2) - F_Y(a_2)\big)$

$= P_X\big((a_1,b_1]\big)\,P_Y\big((a_2,b_2]\big)$. The rest then follows from an argument involving the Extension Theorem.

Prop ⑧ If X, Y are r.v.'s on $(\Omega, \mathcal{F}, P)$ then X and Y are stat. ind iff $p_{(X,Y)}(x,y) = p_X(x)\,p_Y(y)$ in the discrete case and iff $f_{(X,Y)}(x,y) = f_X(x)\,f_Y(y)$ in the absolutely continuous case.

Proof: Ex.

Def The r.v.'s $\{X_t \mid t \in T\}$ are *mut. stat. ind.* if the σ-fields $\{\mathcal{F}_t \mid t \in T\}$ are mut. stat. ind.

- which means finite dimensional dist'n fns, prob fns or density functions factor as products.

Prop ⑨ If $\{X_t : t \in T\}$ are r.v.s on $(\Omega, \mathcal{F}, P)$ s.t. for each k and $t_1, \dots, t_k \in T$ $(X_{t_1}, \dots, X_{t_k})$ has a discrete (abs. cont) dist. then the r.v.s are mut. stat ind. iff $p_{(X_{t_1}, \dots, X_{t_k})}(x_1, \dots, x_k) = \prod_{i=1}^{k} p_{X_{t_i}}(x_i)$ $\left(f_{(X_{t_1}, \dots, X_{t_k})}(x_1, \dots, x_k) = \prod_{i=1}^{k} f_{X_{t_i}}(x_i)\right)$

- can generalize to $\{X_t : t \in T_1\}$, $\{X_t : t \in T_2\}$ are stat. ind. when $T_1 \cap T_2 = \phi$ and for $t_1, \dots, t_k \in T_1$, $t_{k+1}, \dots, t_{k+l} \in T_2$ then

$$f_{(X_{t_1}, \dots, X_{t_k}, X_{t_{k+1}}, \dots, X_{t_{k+l}})}(x_1, \dots, x_{k+l}) = f_{(t_1, \dots, t_k)}(x_1, \dots, x_k)\, f_{(t_{k+1}, \dots, t_{k+l})}(x_{k+1}, \dots, x_{k+l})$$

- <u>marginal distributions</u>

- suppose $X \in \mathbb{R}^k$ has density f and $\underset{\sim}{Y} = \begin{pmatrix} x_1 \\ \vdots \\ x_\ell \end{pmatrix}$ for $\ell < k$.

- so $\underset{\sim}{Y} = T(\underset{\sim}{X})$ where T is projection on the 1st ℓ coords

- so $P_{\underset{\sim}{Y}}(B) = P_{\underset{\sim}{X}}(T^{-1}B)$

$$T^{-1}B = \{\underset{\sim}{x} : (x_1, \ldots, x_\ell) \in B\}$$
$$= B \times \mathbb{R}^{k-\ell}$$

$$= \int_{B \times \mathbb{R}^{k-\ell}} f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k$$
$$= \int_B \left(\int_{\mathbb{R}^{k-\ell}} f(x_1, \ldots, x_k)\, dx_{\ell+1} \ldots dx_k \right) dx_1 \ldots dx_\ell$$
$$= \int_B f_{\underset{\sim}{Y}}(x_1, \ldots, x_\ell)\, dx_1 \ldots dx_\ell$$

$\therefore \underset{\sim}{Y}$ has density

$$f_{\underset{\sim}{Y}}(x_1, \ldots, x_\ell) = \int_{\mathbb{R}^{k-\ell}} f(x_1, \ldots, x_k)\, dx_{\ell+1} \ldots dx_k.$$

- take any ℓ coordinates say $(X_{i_1}, \ldots, X_{i_\ell}) = \underset{\sim}{Y}$ then marginal density of $\underset{\sim}{Y}$ is obtained by integrating out the remaining coordinates (let C be a permutation matrix so $(i_1, \ldots, i_\ell)$ are the first ℓ coords)

$$\underset{\sim}{X} = \begin{matrix} \ell \\ k-\ell \end{matrix}\begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \underset{\sim}{\mu}_1 \\ \underset{\sim}{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right)$$

Cholesky

$$\Sigma = R'R \qquad R \text{ upper } \Delta \text{ with pos. diag}$$

$$= \begin{pmatrix} R_1 & R_2 \\ 0 & R_3 \end{pmatrix}$$

$$= \begin{pmatrix} R_1' & 0 \\ R_2' & R_3' \end{pmatrix}\begin{pmatrix} R_1 & R_2 \\ 0 & R_3 \end{pmatrix} = \begin{pmatrix} R_1'R_1 & R_1'R_2 \\ R_2'R_1 & R_2'R_2 + R_3'R_3 \end{pmatrix}$$

$$\underset{\sim}{X} = \mu + R'\underset{\sim}{Z} \qquad \underset{\sim}{Z} = \begin{pmatrix} \underset{\sim}{Z}_1 \\ \underset{\sim}{Z}_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \underset{\sim}{0} \\ \underset{\sim}{0} \end{pmatrix}, \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \right)$$

$$= \begin{pmatrix} \underset{\sim}{\mu}_1 + R_1'\underset{\sim}{Z}_1 \\ \underset{\sim}{\mu}_2 + R_2'\underset{\sim}{Z}_1 + R_3'\underset{\sim}{Z}_2 \end{pmatrix}.$$

and noting that $\underset{\sim}{Z}_1 \sim N_\ell(\underset{\sim}{0}, I)$

$$\therefore \underset{\sim}{X}_1 \sim N_\ell(\underset{\sim}{\mu}_1, R_1'R_1) = N_\ell(\underset{\sim}{\mu}_1, \Sigma_{11})$$

- in general the marginal distribution of any ℓ coords of $\underset{\sim}{X} \sim N_k(\underset{\sim}{\mu}, \Sigma)$ is ℓ-dimensional normal with corresponding mean and variance entries

- conditional distributions

- recall for events when $P(C)>0$,

$$P(A|C) = P(A\cap C)\,/\,P(C)$$

- so when $\underset{\sim}{X} = \begin{pmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \end{pmatrix}$ and we want cond'l dist'n of $\underset{\sim}{X}_1$ given $\underset{\sim}{X}_2 = \underset{\sim}{x}_2$

- in discrete case when $P_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2) = P_{\underset{\sim}{X}_2}(\{\underset{\sim}{x}_2\})>0$

$$P_{\underset{\sim}{X}_1|\underset{\sim}{X}_2}(\underset{\sim}{x}_1|\underset{\sim}{x}_2) = \frac{P(\underset{\sim}{X}_1=\underset{\sim}{x}_1,\ \underset{\sim}{X}_2=\underset{\sim}{x}_2)}{P(\underset{\sim}{X}_2=\underset{\sim}{x}_2)}$$

$$= \frac{P_{(\underset{\sim}{X}_1,\underset{\sim}{X}_2)}(\underset{\sim}{x}_1,\underset{\sim}{x}_2)}{P_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2)}$$

- in the absolutely cont. case

$$f_{\underset{\sim}{X}_1|\underset{\sim}{X}_2}(\underset{\sim}{x}_1|\underset{\sim}{x}_2) = \lim_{\varepsilon\downarrow 0} \frac{P_{\underset{\sim}{X}_1|\underset{\sim}{X}_2}(B_\varepsilon(\underset{\sim}{x}_1)\,|\,B_\varepsilon(\underset{\sim}{x}_2))}{vol(B_\varepsilon(\underset{\sim}{x}_1))}$$

$$= \lim_{\varepsilon\downarrow 0} \frac{P_{(\underset{\sim}{X}_1,\underset{\sim}{X}_2)}(B_\varepsilon(\underset{\sim}{x}_1)\times B_\varepsilon(\underset{\sim}{x}_2))}{vol(B_\varepsilon(\underset{\sim}{x}_1))\,vol(B_\varepsilon(\underset{\sim}{x}_2))} \Bigg/ \frac{P_{\underset{\sim}{X}_2}(B_\varepsilon(\underset{\sim}{x}_2))}{vol(B_\varepsilon(\underset{\sim}{x}_2))}$$

$$= f_{(\underset{\sim}{X}_1,\underset{\sim}{X}_2)}(\underset{\sim}{x}_1,\underset{\sim}{x}_2)\,/\,f_{\underset{\sim}{X}_2}(\underset{\sim}{x}_2)$$

eg <u>multivariate normal</u>

- $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_k\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right)$

- consider $Y = X_1 - \Sigma_{12}\Sigma_{22}^{-1}X_2$.

- now $\begin{pmatrix} Y \\ X_2 \end{pmatrix} = \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

and so $\begin{pmatrix} Y \\ X_2 \end{pmatrix}$

$$\sim N_k\left(\begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \quad \end{pmatrix} \Sigma \begin{pmatrix} \quad \end{pmatrix}' \right)$$

$$= N_k\left(\begin{pmatrix} \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}' & 0 \\ 0 & \Sigma_{22} \end{pmatrix} \right)$$

- For any invertible matrix of the form

$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ we have $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{pmatrix}$

$$\det \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} = \det A \det B$$

Prop Y and X_2 are stat ind with
$Y \sim N_l(\mu_1 - \Sigma_{12}\Sigma_{22}^{-1}\mu_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}')$
$X_2 \sim N_{k-l}(\mu_2, \Sigma_{22})$.

Corollary $\underset{\sim}{X}_1 | \underset{\sim}{X}_2 = \underset{\sim}{x}_2$

$$\sim N_\ell\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right)$$

Proof: We have that $\underset{\sim}{X}_1 = \underset{\sim}{Y} + \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{X}_2$

and so $\underset{\sim}{X}_1 | \underset{\sim}{X}_2 = \underset{\sim}{x}_2 \sim \underset{\sim}{Y} + \Sigma_{12}\Sigma_{22}^{-1}\underset{\sim}{x}_2 \,|\, \underset{\sim}{X}_2$

$$\sim N_\ell\left(\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'\right)$$

- the vector $\underset{\sim}{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\underset{\sim}{x}_2 - \underset{\sim}{\mu}_2)$

 is called the <u>regression</u> of $\underset{\sim}{X}_1$ on $\underset{\sim}{X}_2$.