Probability-2 Lecture-24

09 April 2024 11:18

$\underset{\sim}{a} = (a_1, \dots, a_k)^T \in \mathbb{R}^k$

↳ denotes a vector.

$\therefore \underset{\sim}{a}^T \cdot b = a_1 b_1 + \cdots + a_k b_k$. $\quad [\underset{\sim}{b} = (b_1, \dots, b_k)^T]$.

k-dimensional random vector is a vector $\underset{\sim}{X} = (X_1, \dots, X_k)^T$, where $X_1, \dots, X_k$ are real r.vs. on the probability space $(\Omega, \mathcal{A}, P)$.

<u>Fact</u>: for any Borel Set $B \subset \mathbb{R}^k$, $\{\omega : \underset{\sim}{X}(\omega) \in B\} \in \mathcal{A}$.

<u>Joint Distribution</u> of a random vector $\underset{\sim}{X}$ is a probability on $\mathbb{R}^k$ given by

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

for $\underset{\sim}{x}, \underset{\sim}{y} \in \mathbb{R}^k$, say $\underset{\sim}{x} \leq \underset{\sim}{y}$ (respectively, $\underset{\sim}{x} < \underset{\sim}{y}$). if $x_i \leq y_i \ \forall i$ (respectively, $x_i < y_i \ \forall i$).

Note that, these are only partial orders.
Also, $\underset{\sim}{x} \leq \underset{\sim}{y}$, $\underset{\sim}{x} \not< \underset{\sim}{y} \not\Rightarrow \underset{\sim}{x} = \underset{\sim}{y}$.

<u>Joint Distribution Function</u>:

For a random vector X, the joint distn fn is

$F_X : \mathbb{R}^k \longrightarrow \mathbb{R}$

$$F_X(\underset{\sim}{a}) = P(X \leq \underset{\sim}{a})$$
$$= P(X_i \leq a_i \ \forall i).$$

<u>Properties of F_X</u>:

Properties of F_X:

① $0 \leq F_{\underset{\sim}{X}} \leq 1$

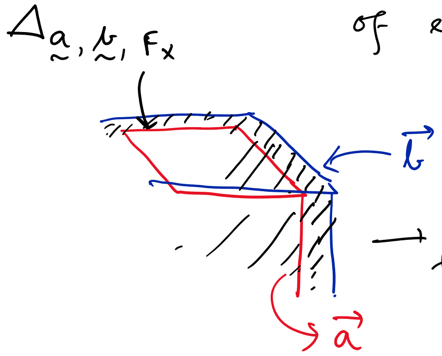
② Right continuous in each variable

($\Leftrightarrow$ if $\underset{\sim}{a_n} \to \underset{\sim}{a}$ in $\mathbb{R}^k$, $\underset{\sim}{a} \leq \underset{\sim}{a_n}$ $\forall n$, then $F_X(\underset{\sim}{a_n}) \longrightarrow F_X(\underset{\sim}{a})$.

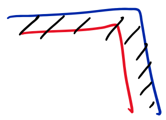
③ For $\underset{\sim}{a} < \underset{\sim}{b}$, $\Delta_{\underset{\sim}{a}, \underset{\sim}{b}, F_{\underset{\sim}{X}}} \geq 0$.

$$* \quad \Delta_{\underset{\sim}{a}, \underset{\sim}{b}, F_X} = \sum_{\underset{\sim}{\varepsilon} \in \{0,1\}^k} (-1)^{\sum \varepsilon_i} \cdot F_X\left(\underset{\sim}{b} - \left[(1-\varepsilon_1)(b_1 - a_1), \dots, (1-\varepsilon_k)(b_k - a_k)\right]^T\right)$$

$\longrightarrow$ this captures the probability "mass" of every k-dimensional rectangle.



$\longrightarrow$ for visualizing, imagine an cube



3 faces of red & blue same, 3 faces different (////// $\longrightarrow$ volume between the 3 different faces.)

④ $\lim_{\min a_i \nearrow \infty} F(\underset{\sim}{a}) = 1$

$\lim_{\min a_i \to -\infty} F(\underset{\sim}{a}) = 0$.

$\underset{\sim}{X}$ is a random vector, k-dimensional.

① A is a $m \times k$ real matrix. $\Rightarrow \underset{\sim}{Y} = A\underset{\sim}{X}$ is an <u>m-dimensional</u> random vector.

② for $\underset{\sim}{a} \in \mathbb{R}^k$, and a k-dimensional random vector X, ② for $\underset{\sim}{a} \in \mathbb{R}^k$, and a k-dimensional random vector X,

$\underset{\sim}{a}^T \underset{\sim}{X}$ is is a real random variable.

Fact: (to be seen later)

Joint distribution of $\underset{\sim}{X}$ is determined by joint distribution of $\underset{\sim}{a}^T \underset{\sim}{X}$, $a \in \mathbb{R}^k$.

Mean & Dispersion.

$\underset{\sim}{X}$ - k dimensional random vector.

The mean vector of $\underset{\sim}{X}$ is defined as

$$M(\underset{\sim}{X}) = E(\underset{\sim}{X}) = (EX_1, \ldots, EX_k)^T,$$

provided EX_i exists, & is finite $\forall i$

ie, $EX_i < \infty$

Consequence:

A – m×k matrix. $\underset{\sim}{\beta} \in \mathbb{R}^m$

$$E(A\underset{\sim}{X} + \underset{\sim}{\beta}) = AE(\underset{\sim}{X}) + \underset{\sim}{\beta}$$

Dispersion Matrix / Variance-Covariance Matrix :

$$D(\underset{\sim}{X}) = ((Cov(x_i, x_j)), \quad \text{provided } E(x_i^2) < \infty \ \forall i$$

Note: $D(\underset{\sim}{X}) = E\left((X - E(X))(X - E(X))^T\right)$. $\left[\text{check for } X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \text{ is, } D\text{-}2\times 2\right]$

A – m×k matrix, $\beta \in \mathbb{R}^m$.

$$D(A\underset{\sim}{X} + \underset{\sim}{\beta}) = A \cdot D(\underset{\sim}{X}) \cdot A^T$$

$$= E\left((A\underset{\sim}{X} + \cancel{\beta} - E(A\underset{\sim}{X} + \cancel{\beta}))(A\underset{\sim}{X} + \cancel{\beta} - E(A\underset{\sim}{X} + \cancel{\beta}))^T\right)$$

$\left[\underset{\sim}{\beta} \text{ just translates } A\underset{\sim}{X}. \text{ It has no impact on } D.\right]$

$$= A \; E\left((X - E(X))(X - E(X))^T\right) \cdot A^T$$

$$= A\, E\left((X - E(X))(X-E(X))^T\right) \cdot A^T$$

$$= A\, D(X)\, A^T$$

In particular,

for $\underset{\sim}{a} \in \mathbb{R}^k$,

$$V(\underset{\sim}{a}^T \underset{\sim}{X}) = \underset{\sim}{a}^T D(\underset{\sim}{X}) \cdot \underset{\sim}{a}.$$

$$\Rightarrow \underset{\sim}{a}^T D(\underset{\sim}{X})\, \underset{\sim}{a} \geq 0. \quad \forall\, \underset{\sim}{a} \in \mathbb{R}^k$$

Note that, $a^T X$ is 1×1.
$\therefore V(a^T X)$ is the only entry of 1×1 matrix $D(a^T X)$
$\therefore V(a^T X) = D(a^T X) = a\, D(X)\, a^T$ (using the last result.)

$a^T - 1\times n$, $X - n \times 1$

$\therefore a^T X - 1\times 1$

Hence, we can talk about Variance.

$\therefore D(\underset{\sim}{X})$ is a real, symmetric, non-negative definite (n.n.d) matrix.

Note: $\underset{\sim}{a}^T D(\underset{\sim}{X}) \cdot \underset{\sim}{a} = 0$

[* $V(X) = 0$. $E(X - E(X)) = 0$
$\Rightarrow P(X - E(X) = 0) = 1$. $\therefore P(X = E(X)) = 1$
i.e, almost surely.]

$\Rightarrow$ the random vector is degenerate.

i.e, $\underset{\sim}{a}^T \underset{\sim}{X} = \underbrace{\underset{\sim}{a}^T \underset{\sim}{\mu}}_{\text{Some constant.}}$ almost surely.

$\mu = (E(X_1), \ldots, E(X_n))^T$

(Set of $\omega \in \Omega$ where this equality does not hold is of Probability Measure 0.)

(**)

Proposition: for any k-dimensional random vector $\underset{\sim}{X}$,

$$P\left((X - E(X)) \in \mathcal{C}(D(X))\right) = 1.$$

Proof: Assume $D(\underset{\sim}{X})$ is singular.

$(\mathcal{C}(D(\underset{\sim}{X})))^\perp$ is non-trivial.

[if $D(\underset{\sim}{X})$ is non-singular $\mathcal{C}(D(X)) = \mathbb{R}^k$. So, nothing to prove.]

Choose a basis $\underset{\sim}{V}_1, \ldots, \underset{\sim}{V}_m$ for $(\mathcal{C}(D(\underset{\sim}{X})))^\perp$. $(m < k)$.

⊛ → [$\therefore V_i$ is orthogonal to every column in $D(X)$. i.e, $V_i^T \cdot (D(X))_{*j} = 0 \;\; \forall\, j = 1, \ldots, k$]

$$\therefore \underset{\sim}{V}_i^T (D(\underset{\sim}{X})) = \underset{\sim}{0} \Rightarrow \underset{\sim}{V}_i^T (D(\underset{\sim}{X}))\, V_i = \underset{\sim}{0}$$

⊛ —

$$\Rightarrow \underset{\sim}{V}_i^T \underset{\sim}{X} = \underset{\sim}{V}_i^T E(\underset{\sim}{X}) \text{ almost surely.}$$

[Using (**)]

By discarding the union of the null set, By discarding the union of the null set,

$$P\left(v_i^T(\underset{\sim}{X} - E(\underset{\sim}{X})) = 0 \text{ for ALL } i\right) = 1$$

these lie in ortho-complementary spaces

$$\Rightarrow P\left(\underset{\sim}{X} - E(\underset{\sim}{X}) \subset \mathcal{C}(D(X))\right) = 1.$$

<u>Exercise</u>:

$$\text{Rank}(D(\underset{\sim}{X})) = \min\left\{\dim(S) : S \subset \mathbb{R}^k \text{ satisfying } P(\underset{\sim}{X} - E(\underset{\sim}{X}) \in S) = 1\right\}.$$

<u>Corollary</u>: $D(\underset{\sim}{X})$ is singular.

$\Leftrightarrow \underset{\sim}{a}^T \underset{\sim}{X} \underset{\sim}{a}$ is degenerate for some $\underset{\sim}{a} \neq 0$.

<u>Fact</u>: $\sum_{k\times k}$ real, symmetric non-negative definite matrix

$$\Rightarrow \sum_{k\times k} = D(\underset{\sim}{X}) \text{ for some } \underset{\sim}{X} \in \mathbb{R}^k$$

<u>Proof</u>:

$\exists B$ s.t., $\sum = BB^T$.

Suppose $m = \#$ columns of B.

$Y_1, \ldots, Y_m \overset{iid}{\sim} N(0, 1)$

$\underset{\sim}{Y} = (Y_1, \ldots, Y_m)^T$.

then, $D(\underset{\sim}{Y}) = I_n$.

$\therefore$ Take $X = B\underset{\sim}{Y}$.

$$D(X) = D(B\underset{\sim}{Y}) = B \cdot \underbrace{D(\underset{\sim}{Y})}_{= I_n} \cdot B^T$$

$$D(X) = BB^T$$

▨

$\underset{\sim}{X}$ - k dimensional is <u>discrete</u> if $\exists$ countable $D \subset \mathbb{R}^k$ s.t. $P(\underset{\sim}{X} \in D) = 1$.

$\underset{\sim}{X}$ - discrete $\Leftrightarrow$ X_i - discrete $\forall i$

.. pmf.

$\underset{\sim}{X}$ - discrete $\Longleftrightarrow$ X_i [illegible]

.. pmf.

$$p_{\underset{\sim}{X}}(\underset{\sim}{x}) = P(\underset{\sim}{X} = \underset{\sim}{x}), \quad \underset{\sim}{x} \in D.$$

Marginals, conditionals etc are all analogous to the univariate discrete case.

Example: (Multinomial)

A trial can result in one of m types of outcomes with probabilities $p_1, \ldots, p_m$.

Consider n independent repetetions of the trial

$\underset{\sim}{X} = (X_1, \ldots, X_m)$, where

$X_i = \#$ times ith outcome appears in n trials.

clearly, $\underset{\sim}{X}$ is discrete.

X_i — non -ve integer, $\sum_{i=1}^{m} X_i = n$

$$\therefore \; p_{\underset{\sim}{X}}(x_1, \ldots, x_m) = \binom{n}{x_1 \; x_2 \cdots x_m} \cdot \prod_{i=1}^{m} p_i^{x_i} \quad \left[\begin{array}{l} \bullet\; x_i \geq 0, \; x_i \in \mathbb{Z}. \\ \bullet\; \sum_{i=1}^{m} x_i = n \end{array} \right]$$

Multinomial Coefficient

$$= \frac{\lfloor n}{\lfloor x_1 \cdot \lfloor x_2 \cdots \lfloor x_m} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdots p_m^{x_m}$$

$\underset{\sim}{X}$ is said to be absolutely continuous if $\exists$ non-negative function $f \in \mathbb{R}^k$ s.t. $\longrightarrow$ called the "joint density"

$$P(\underset{\sim}{X} \in B) = \int_B f(\underset{\sim}{x}) \cdot d\underset{\sim}{x}$$

$$= \int_B f(x_1, \ldots, x_k) \cdot dx_1 \ldots dx_k.$$

It is enough to have:

$$F_X(\underset{\sim}{a}) = \int_{x_1 \le a_1} \int_{x_2 \le a_2} \cdots \cdots \int_{x_k \le a_k} f(x_1, \dots, x_k) \, dx_1 \cdots dx_k .$$

$\underset{\sim}{X}$ has a density f, $f > 0$ on an open $U \subset \mathbb{R}^k$.

Suppose,

$$h : U \xrightarrow[\text{onto}]{1-1,} G \subset \mathbb{R}^k \quad \text{with} \quad g = h^{-1} : G \to U$$

G open.

Assume: all partial derivatives of g exist, & are continuous on G.

$\therefore$ Jacobian,

$$J_{g(\underset{\sim}{y})} = \det\left(\left(\frac{\partial x_i}{\partial y_i}\right)\right), \quad \underset{\sim}{y} \in G$$

$\neq 0$ for $\underset{\sim}{y} \in G_i$. with $\underset{\sim}{x} = g(\underset{\sim}{y}) \in D$.

if all these conditions are satisfied, then $h(\underset{\sim}{X}) = \underset{\sim}{Y}$ also has a density.

(Sufficient condition).

Note that: this is analogous to the change of density in univariate case.

Example:

Let $X_1, \dots, X_n$ be independent with

$$X_i \sim \text{Gamma}(\lambda, \alpha_i)$$

Density of $\underset{\sim}{X} = (X_1, \dots, X_n)^T$ is

$$f_{\underset{\sim}{X}}(\underset{\sim}{x}) = \frac{\lambda^{\sum_{i=1}^n \alpha_i} \cdot e^{-\lambda \sum_{i=1}^n x_i}}{\prod_{i=1}^n \Gamma(\alpha_i)} \cdot \prod_{i=1}^n x_i^{\alpha_i - 1}$$

$$\prod_{i=1}^{n} \Gamma(\alpha_i) \qquad i=1$$

Now, denote $Z_1 = \dfrac{X_1}{X_1+X_2}$

$$Z_2 = \frac{X_1+X_2}{X_1+X_2+X_3}$$

$$\vdots$$

$$Z_{n-1} = \frac{X_1+\cdots+X_{n-1}}{X_1+\cdots+X_n}.$$

$$Z_n = X_1+\cdots+X_n.$$

So, $\underset{\sim}{Z} = (Z_1, \ldots, Z_n)$

How to do the transformation? Doing directly seems complicated.

<u>Step-1:</u>

$(x_1, \ldots, x_n)^T \longrightarrow (y_1, \ldots, y_n)^T$, where

$y_1 = x_1$ $\qquad 0 < \underline{y_1 < y_2 < \cdots < y_n} < \infty$

$y_2 = x_1 + x_2$ $\qquad (*)$

$\vdots$

$y_n = x_1 + \cdots + x_n.$

(y_i's are just i^{th} partial sums.)

$$\therefore \quad (y_1, \ldots, y_n)^T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \cdot (x_1, \ldots, x_n)^T.$$

Lower Δ.

(this is a simple transformation by a triangular matrix.)

$$\therefore\ (x_1,\dots,x_n)^T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & & \vdots \\ 0 & & & \ddots & \vdots \\ 0 & \cdots & \cdots & -1 & 1 \end{bmatrix} \cdot (y_1,\dots,y_n)^T.$$

$$\left[\begin{aligned} \because\ & x_1 = 1 \\ & x_2 = y_2 - y_1 \\ & \ \vdots \\ & x_n = y_n - y_{n-1} \end{aligned}\right]$$

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

$$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \frac{\lambda^{\sum \alpha_i} \cdot e^{-\lambda y_n}}{\Gamma(\alpha_1)\cdot\dots\cdot\Gamma(\alpha_n)}\ y_1^{\alpha_1 - 1} \cdot (y_2 - y_1)^{\alpha_2 - 1} \cdot \dots \cdot (y_n - y_{n-1})^{\alpha_n - 1}$$

Now, we do the final transformation

$$\underset{\sim}{y} = (y_1,\dots,y_n) \longmapsto (z_1,\dots,z_n) = \underset{\sim}{z}$$

$$z_1 = \frac{y_1}{y_2}, \quad z_2 = \frac{y_2}{y_3}, \quad \dots\dots, \quad z_{n-1} = \frac{y_{n-1}}{y_n}, \quad z_n = y_n.$$

set over which inverse can be defined $\longrightarrow$ $G = \{ (z_1,\dots,z_n)^T : 0 < z_i < 1,\ i \in \{1,\dots,n-1\}$ and $0 < z_n < \infty \}$.

using $(*)$

$\therefore\ y_n = z_n$

$y_{n-1} = z_n \cdot z_{n-1}$

$\vdots$

$y_1 = z_n \cdot z_{n-1} \cdot \dots \cdot z_2 \cdot z_1$

Take a small eg. Say, $n = 4$.

$y_1 = z_1 z_2 z_3 z_4$

$y_2 = z_2 z_3 z_4$

$y_3 = z_3 z_4$

$y_4 = z_4$

$$\therefore J = \left(\left(\frac{\partial y_i}{\partial z_i}\right)\right)$$

$$= \begin{bmatrix} z_2 z_3 z_4 & z_1 z_3 z_4 & z_1 z_2 z_4 & z_1 z_2 z_3 \\ 0 & z_3 z_4 & z_2 z_4 & z_2 z_3 \\ 0 & 0 & z_4 & z_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$\therefore$ In this general case,

$$J = \begin{bmatrix} z_2 \cdots z_n & & & \\ & z_3 \cdots & & \\ & & \ddots & \\ & & & \end{bmatrix}$$

$$0 \quad 0 \quad 0 \quad 1$$

We have J, & have expressed y_i's in terms of z_i's.

$\therefore$ We can now compute

$$f_Z(z).$$