Kom 505E week 9

29.11.2016

13.5 Using conditioning simplify calculations of probability

<u>Goal:</u> X, Y jointly distributed r.v.s.

Let $Z = g(X, Y)$, find $p_Z(z)$

1. Fix $X = x$, let $Z|X=x = g(x, Y)$
$= g_x(Y)$
(fn. of Y)

2. Find $p_{Z|X}(z|x)$ using xform. from
Y to Z : $Z = g_x(Y)$

3. Uncondition to find $p_Z(z)$.

$$p_Z(z) = \int \underbrace{p_{Z|X}(z|x)\, p_X(x)}\, dx$$

<u>Ex:</u> $Z = \frac{Y}{X}$, find $p_Z(z)$

Let $X = x$; $(Z|X=x) = \frac{Y}{x}$

$$Z|x = g_x(Y) = \left(\frac{1}{x}\right) Y \Rightarrow Y = g^{-1}(z)$$
$$Y = xZ$$

$$Y \underset{g^{-1}}{\overset{g}{\rightleftarrows}} Z$$

$$\boxed{p_Z(z) = p_Y(\underbrace{g^{-1}(z)}_{xz}) \underbrace{\left|\frac{dg^{-1}(z)}{dz}\right|}_{|x|}}$$

$$p_{Z|X}(z|x) = p_{Y|X}(\underbrace{xz}_{g^{-1}(z)}) \underbrace{|x|}$$

$$p_Z(z) = \int_{-\infty}^{\infty} \underbrace{p_{Z|X}(z|x)} \, p_X(x)\, dx$$

$$= \int_{-\infty}^{\infty} p_{Y|X}(xz)\, p_X(x)\, |x|\, dx$$

- Assume X & Y independent:

$$\hookrightarrow p_Z(z) = \int_{-\infty}^{\infty} p_Y(xz)\, p_X(x)\, |x|\, dx$$

- Now, assume $X \sim \mathcal{N}(0, \sigma^2)$, $Y \sim \mathcal{N}(0, \sigma^2)$ } insert the $p_X(x)$ & $p_Y(\cdot)_{xz}$ in to the above eqn.

Exercise $p_Z(z) =$ Cauchy pdf.

Mean of Conditional pdfs (pmfs)

- Expected value of Y when $X = x$ is given

$\hookrightarrow$ mean of the pdf: $p_{Y|X}$
$\underline{\underline{\text{exp. value}}}$

$$\boxed{E_{Y|X}[y|x] = \int_{-\infty}^{\infty} y\, p_{Y|X}(y|x)\, dy}$$: Conditional mean

- Also, $E_X[\underbrace{E_{Y|X}[y|x]}_{\text{think of as a fn. of } x = g(x)}] = E_Y[Y]$

$$= \int_{-\infty}^{\infty} (g(x))\, p_X(x)\, dx$$

$E_{Y|X}[y|x]$

$E_X[g(x)]$

$$\Rightarrow = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \, \underbrace{p_{Y|X}(y|x)\, dy \, p_X(x)}_{p_{X,Y}(x,y)} \, dx$$

$$= \iint_{-\infty}^{\infty} y \, p_{X,Y}(x,y)\, dx\, dy = \int_{-\infty}^{\infty} y \underbrace{\left(\int_{-\infty}^{\infty} p_{X,Y}(x,y)\, dx \right)}_{p_Y(y)} dy$$

$$E_X[E_{Y|X}(y|x)] = \int_{-\infty}^{\infty} y \, p_Y(y)\, dy = E_Y[y]$$

Ex 8.4 : Experiments: 2 dice — 6 sides

1st: (fair) 1, 2, 3, 4, 5, 6

2nd (unfair) 2, 3, 2, 3, 2, 3

R.v.s. X: choice of die (1 or 2)

Y: # on the thrown die.

$$E_Y[Y] = ? \Rightarrow E_{Y|X}[Y|X=1] = ?$$
$$E_{Y|X}[Y|X=2] = ?$$

$$E_{Y|X}[Y|X=1] = \sum_{j=1}^{6} j \underbrace{p_{Y|X}[j|X=1]}_{\frac{1}{6}} = \frac{7}{2}$$

$$E_{Y|X}[Y|X=2] = \sum_{j=2,3} j \underbrace{p_{Y|X}[j|X=2]}_{1/2} = \frac{5}{2}$$

(we need $p_X(x)$)

↳ Assume $p_X(x=1) = p_X[x=2] = \frac{1}{2} \Rightarrow$ Find $E_Y[Y] =$

$$E_Y[Y] = E_X\left[E_{Y|X}[Y|X]\right] = \sum_{i=1,2} E_{Y|X}[Y|X_i]\,p_X(x_i)$$

$$= \frac{3.5+2.5}{2} = 3$$

* i.e the unconditional mean is the weighted average of the conditional means.

Conditional Variance: (pmf)

$$Var(Y|X_i) \triangleq \sum_j \left(y_j - E_{Y|X}[Y|X_i]\right)^2 . P_{Y|X}[y_j|x_i]$$

fix $X = x_i$

- Conditional of expectation of $g(Y)$

$$E_{Y|X}[g(Y)|X=x] = \int_{-\infty}^{\infty} g(y)\,p_{Y|X}(y|x)\,dy$$

Recall "optimal" linear predictor in the MMSE sense.

$$\hat{Y} = aX + b \; ; \; E[(Y-\hat{Y})^2]$$

X, Y jointly distributed ; predict Y based on X.

$$\Rightarrow \hat{Y} = \mu_y + \frac{Cov(X,Y)}{Var(X)}(x-\mu_x)$$

$Cov(X,Y) = \rho\,\sigma_x\sigma_y$

* Now allow nonlinear predictors; find the "best" predictor of Y given X; in MMSE sense:

MMSE = min MSE

Let $\hat{Y} = g(X)$, minimize:

$$\min E_{X,Y}\left[(Y-\underbrace{g(X)}_{\hat{Y}})^2\right] = \int_{-\infty}^{\infty}\int (y-g(x))^2 \, p_{X,Y}(x,y)\,dx\,dy$$

$$\min \int_{-\infty}^{\infty}\left[\int (y-g(x))^2 \, p_{Y|X}(y|x)\,dy\right] \underbrace{p_X(x)}_{\geq 0} dx$$

Let $\boxed{g(x)=c}$ unknown

We can equivalently minimize

$$\min \int_{-\infty}^{\infty} (y-c)^2 \, p_{Y|X}(y|x)\,dy = f(c)$$

$$\frac{\partial f}{\partial c} = \cancel{-2}\int_{-\infty}^{\infty} (y-c)\, p_{Y|X}(y|x)\,dy = 0$$

$$\int_{-\infty}^{\infty} y\, p_{Y|X}(y|x)\,dy = c\underbrace{\int_{-\infty}^{\infty} p_{Y|X}(y|x)\,dy}_{=1}$$

($p_{Y|X}$ is a pdf itself)

$$\boxed{\hat{Y}(X) = E_{Y|X}[y|x]} = c.$$

in the MSE sense.

This is MMSE estimator of Y given $X=x$

"optimal" over all predictors.

We can show for general bivariate Gaussian distributed X & Y:

MMSE = linear MMSE

can show: $$E_{Y|X}[Y|X] = \mu_Y + \frac{\rho\,\sigma_Y}{\sigma_X}(x-\mu_X).$$

General bivariate Gaussian:

$$p_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{(1-\rho^2)\sigma_x^2\sigma_y^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right]$$

$$E_{Y|X}[Y|x] = ? = \hat{y}$$

$$p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)} \rightarrow$$ exercise: derive the conditional pdf form: given here.

$$\sim \mathcal{N}\Big(\underbrace{\mu_y + \rho\sigma_y\left(\frac{x-\mu_x}{\sigma_x}\right)}_{\text{mean}}, \underbrace{\sigma_y^2(1-\rho^2)}_{\text{Var}}\Big)$$

Recall we derived for (X,Y) s.b.n: $\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right)$

$$Y|X \sim \mathcal{N}(\rho x, (1-\rho^2))$$

$$E_{Y|X}[Y|x] = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x)$$

☆ Let $Y = aX + b$, $a, b \in \mathbb{R}$,

$p_{X,Y} = ?$

$$E_Y[Y] = aE_X[X] + b = \mu_y$$

$$Var(Y) = a^2 Var(X)$$

$$\mathrm{Cov}(X,Y) = E_{X,Y}[(X - \underbrace{\mu_x}_{E_X[X]})(Y - \mu_y)]$$

$$(Y = aX + b)$$

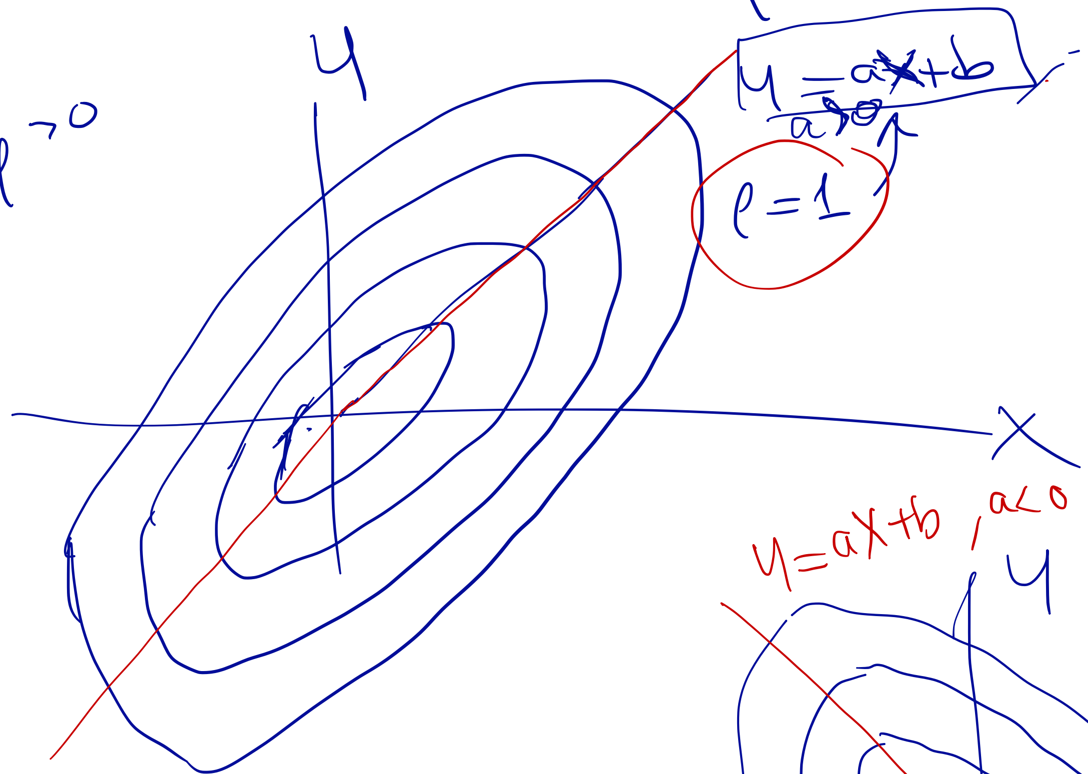
$$= E_{X,Y}[(X - \mu_x)(\underbrace{aX + \cancel{b} - (a\mu_x + \cancel{b})}_{a(X - \mu_x)})]$$

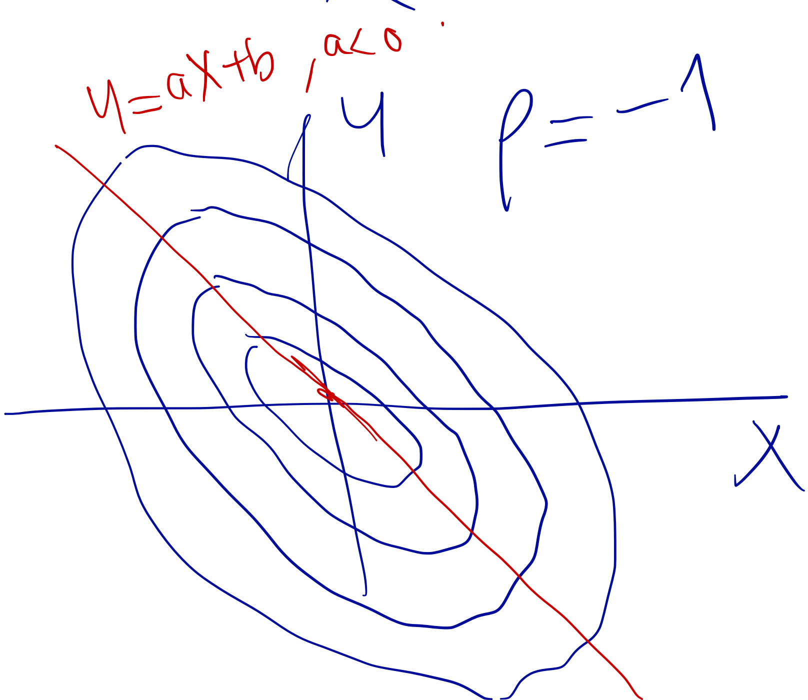
$$= a\,E[(X - \mu_x)^2] = a\,\mathrm{Var}(X)$$

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{a\,\mathrm{Var}(X)}{\sqrt{\mathrm{Var}X\, a^2 \mathrm{Var}X}}$$

$$= \begin{cases} +1, & a > 0 \\ -1, & a < 0 \end{cases}$$

$\rho > 0$





8.7 (13.7) Simulating Joint probabilities using conditional Prob.

Goal: generate joint r.v.s (X, Y) in (matlab or ---)

(1) First generate X using $p_X \Rightarrow X = i$

(2) Generate Y using $p_{Y|X}[Y|i] \Rightarrow Y = j$

(3) obtain $(X, Y) = (i, j)$

(eg. see Ex 8.6 at home).

12.11 Computer Simulation:

→ Given (X, Y) independent standard Gaussian r.v.s.

Generate a realizations of (W, Z) bivariate Gaussian

$$\sim \mathcal{N}\left(\begin{pmatrix} \mu_W \\ \mu_Z \end{pmatrix}, \underbrace{\begin{bmatrix} \sigma_W^2 & \rho\sigma_W\sigma_Z \\ \rho\sigma_W\sigma_Z & \sigma_Z^2 \end{bmatrix}}_{C_{W,Z}}\right).$$

How?

$$C_{X,Y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

$$\begin{pmatrix} W \\ Z \end{pmatrix} = G\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(G\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, G C_{X,Y} G^T\right)$$

$$C_{W,Z} = GG^T$$

You can go thru either evalue decomp / Cholesky decomp, or here an alternative simple approach → produce lower triangular G

$$G=\begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \Rightarrow GG^T=\begin{bmatrix} a^2 & ab \\ ab & b^2+c^2 \end{bmatrix}$$

$$\Downarrow$$

$$C_{w,z}=\begin{bmatrix} \sigma_w^2 & \rho\sigma_w\sigma_z \\ \rho\sigma_w\sigma_z & \sigma_z^2 \end{bmatrix}$$

$a=\sigma_w$

$b=\rho\sigma_z$

$c=\sigma_z\sqrt{1-\rho^2}$

$$\rightarrow G=\begin{bmatrix} \sigma_w & 0 \\ \rho\sigma_z & \sigma_z\sqrt{1-\rho^2} \end{bmatrix}$$

Summary: To generate $\mathcal{N}\left(\begin{bmatrix}\mu_w \\ \mu_z\end{bmatrix}, C_{w,z}\right)$ generate independent x, y std normals (randn)

$$\begin{bmatrix} w \\ z \end{bmatrix}=\begin{bmatrix} \sigma_w & 0 \\ \rho\sigma_z & \sigma_z\sqrt{1-\rho^2} \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}+\begin{bmatrix} \mu_w \\ \mu_z \end{bmatrix}$$

eg. $G=\begin{bmatrix} 1 & 0 \\ 0.9 & \sqrt{1-0.9^2} \end{bmatrix}$, $\begin{pmatrix}\mu_w \\ \mu_z\end{pmatrix}=\begin{bmatrix}1 \\ 1\end{bmatrix}$

```
M=2000;
for m=1:M
  x=randn(1,1); y=randn(1,1);
  wz = G*[x y]'+[1 1]';
end;
```

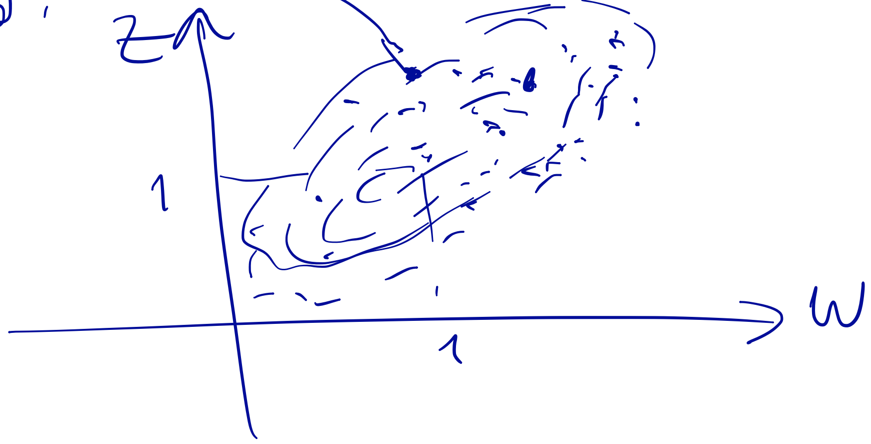
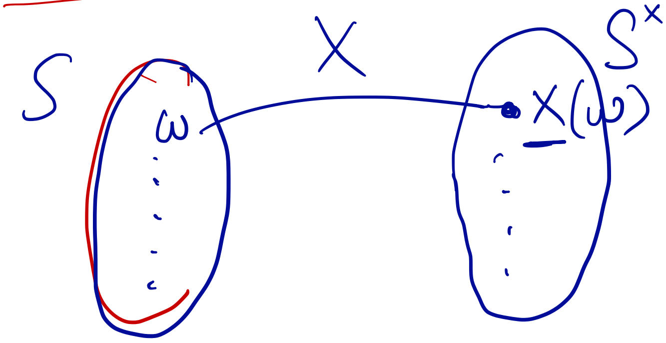


Fig. 12.21.

code is there.

N_dimensional Random Vectors (Ch 9 & 14)



$S^X = S_{X_1} \times S_{X_2} \times \dots S_{X_N}$

$X(\omega) = [X_1(\omega)\ X_2(\omega) \dots X_N(\omega)]_{N\times 1}$

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

Random vector

eg. An image $\underline{M \times M}$, $N = M^2$

$$\underline{X}(\omega) = \begin{bmatrix} I_1 \\ \vdots \\ I_N \end{bmatrix}$$

ω: An image

joint pmf of an N_dim r.v.

$$P_{X_1, X_2, \dots, X_N}[x_1, x_2, \dots, x_N] = P(X_1 = x_1, \dots X_N = x_N)$$

$$P_{\underline{X}}[\underline{x}] = P_{\underline{X}}(\underline{X} = \underline{x}) \quad \text{: vector notation.}$$

satisfies usual properties

$$0 \leq P_{\underline{X}}[\underline{x}] \leq 1, \quad \sum_{k_1, \dots, k_N} \sum \dots \sum P_{\underline{X}}[\underline{x}] = 1$$

marginals: $P_{X_1}[x_1] = \sum_{(k_2 \dots k_N)} \sum \dots \sum P_{X_1, X_2, \dots, X_N}[\dots]$ (with $X_2 \dots X_N$ under the sums)

$\boxed{\text{cont. joint pdfs:}}$

① $p_{\underline{X}}(\underline{x}) \geq 0$, ② $\int_{\underline{X}} p_{\underline{X}}(\underline{x})\, d\underline{x} = 1$

③ $P(A) = \int_{\underline{X} \in A} p_{\underline{X}}(\underline{x})\, d\underline{x}$

Marginals: $P_{X_1}(x_1) = \int_{\underline{X} \setminus X_1} P_{\underline{X}}(\underline{x})\, d\underline{x} \setminus x_1$

$P_{X_i X_j}(x_i, x_j) = \int_{X \setminus i,j} P_{\underline{X}}(\underline{x})\, d\underline{x} \setminus x_i, x_j$

integrate out other variables

for any subset

Ex: Multinomial pmf: M trials, N objects to choose from

X_1 = # times 1st object is seen → w/ p_1

⋮

X_N = # " N^{th} " " → w/ p_N.

$$\boxed{P_{X_1,\dots,X_N}(\;) = \binom{M}{k_1, k_2, \dots, k_N} p_1^{k_1} p_2^{k_2} \cdots p_N^{k_N}}$$

$$= \frac{M!}{k_1!\,k_2!\cdots k_N!}$$

$$\left(\sum_{i=1}^{N} k_i = M \qquad \sum_{i=1}^{k} p_i = 1\right)$$

e.g. An urn has 3 white, 5 black, 7 red balls. We pick 100 times (w/ replacement).

$N = 3$

$$P_{X_1,X_2,X_3}[k_1,k_2,k_3] = \frac{100!}{k_1!\,k_2!\,k_3!}\left(\frac{3}{15}\right)^{k_1}\left(\frac{5}{15}\right)^{k_2}\left(\frac{7}{15}\right)^{k_3}$$

Multivariate Gaussian $\Leftarrow \underline{X}$ a random vector; $X_1, \dots, X_N$ are jointly Gaussian distributed

$$\underline{p_X(x)} = \frac{1}{(2\pi)^{N/2} \det^{1/2}(\underset{N\times N}{\underline{\underline{C}}})} \exp\left[-\frac{1}{2} \underbrace{(\underline{x}-\underline{\mu})^T}_{1\times N} \underset{N\times N}{\underline{\underline{C}}^{-1}} \underset{N\times 1}{(\underline{x}-\underline{\mu})}\right]$$

$p_X(\underline{x}) : \mathbb{R}^N \to \underline{\mathbb{R}}$; $\underline{\underline{C}}_{ij} = \mathrm{Cov}[X_i, X_j]$

$$\underline{\underline{C}} = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \cdots & \mathrm{Cov}(X_1,X_N) \\ \mathrm{Cov}(X_2,X_1) & \mathrm{Var}(X_2) & & \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(X_N,X_1) & & \cdots & \mathrm{Var}(X_N) \end{bmatrix}_{N\times N}$$

Covariance matrix; symmetric & pos. definite $\Rightarrow$ $\underline{\underline{C}}$ is invertible.

$\underline{\underline{C}}$ positive definite $\Rightarrow \boxed{\forall\, \underline{x} : \underline{x}^T \underline{\underline{C}}\, \underline{x} > 0}$

Def: $$\underline{\underline{C}}_{N\times N} = E_{\underline{X}}\left[\underset{N\times 1}{(\underline{X}-\underline{\mu}_X)}\underset{1\times N}{(\underline{X}-\underline{\mu}_X)^T}\right]$$

* If $X_1, \dots, X_N$ are indep: $p_X(\underline{x}) = \prod_{i=1}^{N} p_{X_i}(x_i)$

* For a multivariate Gaussian if $X_1, X_2, \dots X_N$ are indep $\Rightarrow$ $\underline{\underline{C}} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots \ \sigma_N^2 \end{bmatrix}$ diagonal

$\mathrm{Cov}(X_i, X_j) = 0$; X_i's are uncorrelated.

$$p_{\underline{x}}(\underline{x}) = \frac{1}{(2\pi)^{N/2}\sqrt{\det \underline{C}}} \exp\left[-\frac{1}{2}(\underline{x}-\underline{\mu})^T \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{bmatrix} (\underline{x}-\underline{\mu})\right]$$

$$\hookrightarrow \sqrt{\prod_i \sigma_i^2}$$

$Cov(X_i, X_j) = 0 \quad \forall i,j \quad i \neq j.$

Assumption : uncorrelated.

$$p_{\underline{x}}(\underline{x}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right)$$

$\Downarrow$ $X_1, \ldots, X_N$ are independent.

For multivariate Gaussian : uncorrelated $\Rightarrow$ independence.

($\Leftarrow$ diagonal cov matrix $\underline{C}$)

Not true in general.

In General : Independence $\Rightarrow$ Uncorrelated

but uncorrelated $\not\Rightarrow$ independent

Joint Cdfs :

$$F_{\underline{X}}(\underline{x}) = P(X_i \leq x_i \;, i=1,\ldots,N)$$
$$= P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_N \leq x_N)$$

$$F_{\underline{X}}(-\infty, \ldots, -\infty) = 0$$
$$F_{\underline{X}}(+\infty, \ldots, +\infty) = 1$$

Check other properties from the book.

Transformations of Random Vectors: N-dim r. vectors

$$\underline{Y} = \underline{g}(\underline{X}) = \begin{bmatrix} g_1(\underline{X}) \\ g_2(\underline{X}) \\ \vdots \\ g_M(\underline{X}) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\left.\begin{aligned} Y_1 &= g_1(X_1, \dots, X_N) \\ &\vdots \\ Y_M &= g_M(X_1, \dots, X_N) \end{aligned}\right\}$$

If $M < N$; we add enough auxiliary r.v.s.

$$\left.\begin{aligned} Y_{M+1} &= X_{M+1} \\ &\vdots \\ Y_N &= X_N. \end{aligned}\right\}$$

Recall: in 2D case,

$W = X$

$Z = g(X, Y)$

to make it ($M = N$)

to make $\underline{g}$ 1-1

$$\Rightarrow \boxed{P_{\underline{Y}}(\underline{y}) = P_{\underline{X}}\left(\underline{g}^{-1}(\underline{y})\right) \left| \det\left(\frac{\partial \underline{X}}{\partial \underline{y}}\right) \right|}$$

$$\left(\frac{\partial \underline{X}}{\partial \underline{y}}\right)_{N \times N} = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_N} \\ \vdots & & & \\ \frac{\partial x_N}{\partial y_1} & \cdots & \cdots & \frac{\partial x_N}{\partial y_N} \end{bmatrix} = J^{-1}_{N \times N}$$

: Jacobian for $Y \to X$

Discrete case:

$$P_{\underline{Y}}(\underline{y}) = P_{\underline{X}}(\underline{x}) \Big|_{\underline{x} = \underline{g}^{-1}(\underline{y})}$$

Ex: $\underline{X} \sim \mathcal{N}(\underline{\mu}, \underline{\underline{C}})$

Let $\boxed{\underline{Y} = \underline{\underline{G}}\,\underline{X}}$ $\quad \underline{Y} \sim \mathcal{N}(\underline{\underline{G}}\,\underline{\mu}, \underline{\underline{G}}\,\underline{\underline{C}}\,\underline{\underline{G}}^T)$

You can show this using Jacobian inverse:

$$\underline{X} = \underline{\underline{G}}^{-1}\underline{Y} \Rightarrow \frac{\partial \underline{x}}{\partial \underline{y}} = \underline{\underline{G}}^{-1}$$

$$p_{\underline{Y}}(\underline{y}) = p_{\underline{X}}(\underline{\underline{G}}^{-1}\underline{y})\,|\det \underline{\underline{G}}^{-1}|$$

Exercise: derive the resulting $p_{\underline{Y}}(\underline{y})$, inserting into $p_{\underline{x}}(\cdot)$.