Conditioning by a random variable

When an experiment produces a pair of random variables X and Y, observing a sample value of one of them provides partial Information about the other. The partial knowledge consists of the value of one of the random variables: either $B=\{X=x\}$ or $B=\{Y=y\}$. Learning $\{Y=y\}$ changes our knowledge of random variable X,Y. We now have complete knowledge of Y and modified knowledge of X. The new model is either a conditional PMF of X given Y or a conditional PDF of X given Y.

Conditional PMF

For any event $Y=y$ such that $P_Y(y)>0$, the conditional PMF of X given $Y=y$ is $P_{X|Y}(x|y)=P[X=x|Y=y]$

Theorem:
For discrete random variables X and Y with joint PMF $P_{X,Y}(x,y)$ and x and y such that $P_X(x)>0$ and $P_Y(y)>0$,

$$P_{X|Y}(x|y)=\frac{P_{X,Y}(x,y)}{P_Y(y)}, \quad P_{Y|X}(y|x)=\frac{P_{X,Y}(x,y)}{P_X(x)}$$

example
Random variables X and Y have the joint PMF $P_{X,Y}(x,y)$ as given

x \ y	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

Find the conditional PMF of Y given $X=x$ for each $x \in S_X$.

Sol $P_X(x)$ is

x	1	2	3	4
$P_X(x)$	$\frac{1}{4}$	$\frac{1}{8}+\frac{1}{8}$	$\frac{1}{12}+\frac{1}{12}+\frac{1}{12}$	$\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}$

x	1	2	3	4
$P_X(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} \cancel{= \frac{P_{X,Y}(x,y)}{\sum_{y\in S_Y} P_{X,Y}(x,y)}} \quad \text{where } x\in[1,2,3,4]$$

$$P_{Y|X}(y|1) = \begin{cases} \frac{1/4}{1/4} = 1, & y=1 \\ 0, & \text{otherwise.} \end{cases}$$

$$P_{Y|X}(y|2) = \begin{cases} \frac{1/8}{1/4} = \frac{1}{2}, & y=1 \\ \frac{1/8}{1/4} = \frac{1}{2}, & y=2 \\ 0, & \text{otherwise} \end{cases}$$

$$P_{Y|X}(y|3) = \begin{cases} \frac{1/12}{1/4} = \frac{1}{3}, & y=1 \\ \frac{1/12}{1/4} = \frac{1}{3}, & y=2 \\ \frac{1/12}{1/4} = \frac{1}{3}, & y=3 \\ 0, & \text{otherwise} \end{cases}$$

$$P_{Y|X}(y|4) = \begin{cases} \frac{1/16}{1/4} = \frac{1}{4}, & y=1 \\ \frac{1/16}{1/4} = \frac{1}{4}, & y=2 \\ \frac{1/16}{1/4} = \frac{1}{4}, & y=3 \\ \frac{1/16}{1/4} = \frac{1}{4}, & y=4 \\ 0, & \text{otherwise.} \end{cases}$$

Note:
If X and Y are independent,
$P_{X|Y}(x|y) = P_X(x)$,
$P_{Y|X}(y|x) = P_Y(y)$

Theorem:
For discrete random variables X and Y with joint PMF $P_{X,Y}(x,y)$ and x and y such that $P_X(x)>0$ and $P_Y(y)>0$

$$P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x) = P_{X|Y}(x|y)P_Y(y)$$

Conditional PDF

For y such that $f_Y(y) > 0$, the conditional PDF of X given $[Y = y]$ is $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$

example

Random Variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 2, & 0 \le y \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

For $0 \le x \le 1$, find the conditional PDF $f_{Y/X}(y/x)$.

For $0 \le y \le 1$, find the conditional PDF $f_{X|Y}(x|y)$.

sol^n

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_0^x 2\,dy = 2y\Big]_0^x = 2x, \text{ for } 0 \le x \le 1.$$

conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{2}{2x} = \frac{1}{x}, & \text{for } 0 \le y \le x \\ 0, & \text{otherwise}. \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_y^1 2\,dx = 2x\Big]_y^1 = 2(1-y), \text{ for } 0 \le y \le 1$$

conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{2}{2(1-y)} = \frac{1}{1-y}, & \text{for } y \le x \le 1 \\ 0, & \text{othorwise} \end{cases}$$

Theorem:

For continuous random variables X and Y with joint PDF $f_{X,Y}(x,y)$ and x and y such that $f_X(x) > 0$ and $f_Y(y) > 0$,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x) = f_{X|Y}(x|y) f_Y(y).$$

example;
Let R be the uniform (0,1) random variable. Given $R=r$, X is the uniform $(0,r)$ random variables. Find the conditional PDF of R given X.

sol^n Given $f_R(r)=\begin{cases} 1, & 0\le r<1, \\ 0, & \text{otherwise} \end{cases}$, $f_{X|R}(x|r)=\begin{cases} \frac{1}{r}, & 0\le x<r \\ 0, & \text{otherwise.} \end{cases}$

joint PDF of R and X is

$$f_{R,X}(r,x) = f_{X|R}(x|r)\,f_R(r) = \begin{cases} \frac{1}{r}, & 0\le x<r<1 \\ 0, & \text{otherwise} \end{cases}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{R,X}(r,x)\,dr = \int_x^1 \frac{1}{r}\,dr = +\ln r\Big]_x^1 = +\ln 1 \mp \ln x = -\ln x$$

for $0<x<1$

$$f_{R|X}(r|x) = \frac{f_{R,X}(r,x)}{f_X(x)} = \begin{cases} -\frac{1}{r\ln x}, & x\le r\le 1 \\ 0, & \text{otherwise} \end{cases}$$

Note
If X and Y are independent,

$$f_{X|Y}(x|y) = f_X(x), \quad f_{Y|X}(y|x) = f_Y(y)$$

Random vector

A random vector is a column vector $\mathbf{X}=[X_1 \ldots X_n]'$. Each X_i is a random variable.

vector sample value

A sample value of a random vector is a column vector $\mathbf{x}=[x_1 \ldots x_n]'$. The i^{th} component, x_i of the vector $\mathbf{x}$ is a sample value of a random variable, X_i.

Expected value vector

The expected value of a random vector X is a column vector $E[\mathbf{X}]=\mu_X=[E(X_1) \;\; E[X_2] \;\ldots\; E[X_n]]'$

Random matrix:

For a random vectors $\mathbf{X}$ with n components and $\mathbf{Y}$ with m components, the set of all products, $X_i Y_j$ is contained in the $n \times m$ random matrix $\mathbf{X}\mathbf{Y}'$. If $\mathbf{Y}=\mathbf{X}$, the random matrix $\mathbf{X}\mathbf{X}'$ contains all products, $X_i X_j$ of components of $\mathbf{X}$.

example:
If $\mathbf{X}=[X_1 \; X_2 \; X_3]'$, then the components of $\mathbf{X}\mathbf{X}'$ are

$$\mathbf{X}\mathbf{X}' = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} [X_1 \;\; X_2 \;\; X_3] = \begin{bmatrix} X_1^2 & X_1X_2 & X_1X_3 \\ X_2X_1 & X_2^2 & X_2X_3 \\ X_3X_1 & X_3X_2 & X_3^2 \end{bmatrix}$$

Expected value of a random matrix

For a random matrix $\mathbf{A}$ with the random variable A_{ij} as its i,jth element, $E[\mathbf{A}]$ is a matrix with i,jth element $E[A_{ij}]$.

Vector Correlation

The correlation of a random vector X is an $n \times n$ matrix R_X with i,jth element $R_X(i,j)=E[X_iX_j]$. In vector notation $R_X=E[\mathbf{X}\mathbf{X}']$.

example.

If $X = [X_1 \ X_2 \ X_3]'$, the correlation matrix of X is

$$R_X = \begin{bmatrix} E[X_1^2] & E[X_1X_2] & E[X_1X_3] \\ E[X_2X_1] & E[X_2^2] & E[X_2X_3] \\ E[X_3X_1] & E[X_3X_2] & E[X_3^2] \end{bmatrix} = \begin{bmatrix} E[X_1^2] & r_{X_1,X_2} & r_{X_1,X_3} \\ r_{X_2,X_1} & E[X_2^2] & r_{X_2,X_3} \\ r_{X_3,X_1} & r_{X_3,X_2} & E[X_3^2] \end{bmatrix}$$

The i,jth element of the correlation matrix is the expected value of the random variable X_iX_j.

Vector covariance

The covariance of a random vector X is an $n \times m$ matrix C_X with components $C_X(i,j) = Cov[X_i, X_j]$. In vector notation,

$$C_X = E[(X - \mu_X)(X - \mu_X)']$$

example:

If $X = [X_1 \ X_2 \ X_3]'$, the covariance matrix of X is

$$C_X = \begin{bmatrix} Var[X_1] & Cov[X_1,X_2] & Cov[X_1,X_3] \\ Cov[X_2,X_1] & Var[X_2] & Cov[X_2,X_3] \\ Cov[X_3,X_1] & Cov[X_3X_2] & Var[X_3] \end{bmatrix}$$

Theorem:

For a random vector X with correlation matrix R_X, covariance matrix C_X, and vector expected value μ_X.

$$C_X = R_X - \mu_X \mu_X'$$

example 1

Find the expected value $E[X]$, the correlation matrix R_X, and the covariance matrix C_X of the two dimensional random vector X with the PDF $f_{X,Y}(x,y) = \begin{cases} \frac{2}{9}, & 0 \le x_1 \le x_2 \le 3 \\ 0, & \text{otherwise.} \end{cases}$

soln

$$E[X_1] = \int_{x_2=-\infty}^{\infty}\int_{x_1=-\infty}^{\infty} x_1 f_{X,Y}(x,y)\,dx_1 dx_2 = \int_{x_2=0}^{3}\int_{x_1=0}^{x_2} x_1 \times \frac{2}{9}\,dx_1 dx_2 = \int_{x_2=0}^{3} \left.\frac{2x_1^2}{9\times 2}\right]_0^{x_2} dx_2$$

$$= \int_{x_2=0}^{3} \frac{x_2^2}{9}\,dx_2 = \left.\frac{x_2^3}{27}\right]_0^3 = 1$$

$$E[X_2] = \int_{x_2=0}^{3}\int_{x_1=0}^{x_2} x_2 \times \frac{2}{9}\,dx_1 dx_2 = \int_{x_2=0}^{3} \left.\frac{2}{9}x_2 x_1\right]_0^{x_2} dx_2 = \int_{x_2=0}^{3} \frac{2x_2^2}{9}\,dx_2 = \left.\frac{2x_2^3}{27}\right]_0^3 = 2.$$

$$E[X_1^2] = \int_{x_2=0}^{3}\int_{x_1=0}^{x_2} x_1^2 \times \frac{2}{9}\,dx_1 dx_2 = \int_{x_2=0}^{3} \left.\frac{2x_1^3}{27}\right]_0^{x_2} dx_2 = \int_{x_2=0}^{3} \frac{2x_2^3}{27}\,dx_2 = \left.\frac{2x_2^4}{27\times 4}\right]_0^3 = \frac{3}{2}$$

$$E[X_2^2] = \int_{x_2=0}^{3}\int_{x_1=0}^{x_2} x_2^2 \times \frac{2}{9}\,dx_1 dx_2 = \int_{x_2=0}^{3} \left.\frac{2x_2^2 x_1}{9}\right]_0^{x_2} dx_2 = \int_{x_2=0}^{3} \frac{2x_2^3}{9}\,dx_2 = \left.\frac{2x_2^4}{9\times 4}\right]_0^3 = \frac{9}{2}$$

$$E[X_1X_2] = \int_{x_2=0}^{3}\int_{x_1=0}^{x_2} \frac{2}{9}x_1 x_2\,dx_1 dx_2 = \int_{x_2=0}^{3} \left.\frac{2}{9}x_2 \frac{x_1^2}{2}\right]_0^{x_2} dx_2 = \int_{x_2=0}^{3} \frac{x_2^3}{9}\,dx_2 = \left.\frac{x_2^4}{9\times 4}\right]_0^3 = \frac{9}{4}.$$

$$R_X = \begin{bmatrix} E[X_1^2] & E[X_1X_2] \\ E[X_2X_1] & E[X_2^2] \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & \frac{9}{4} \\ \frac{9}{4} & \frac{9}{2} \end{bmatrix}$$

$$C_X = R_X - \mu_X \mu_X' = \begin{bmatrix} \frac{3}{2} & \frac{9}{4} \\ \frac{9}{4} & \frac{9}{2} \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{3}{2} & \frac{9}{4} \\ \frac{9}{4} & \frac{9}{2} \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

example

Find the covariance matrix for the two random variables X_1 and X_2 whose joint probability is represented as follows:

$x_1 \backslash x_2$	0	1
-1	0.24	0.06
0	0.16	0.14
1	0.40	0.00

sol^n

x_1	-1	0	1
$p(x_1)$	0.3	0.3	0.4

x_2	0	1
$P(x_2)$	0.8	0.2

$$E[X_1] = \sum x_1 p(x_1) = -1\times 0.3 + 0\times 0.3 + 1\times 0.4 = 0.1$$

$$E[X_2] = \sum x_2 p(x_2) = 0\times 0.8 + 1\times 0.2 = 0.2$$

$$\therefore \mu_X = \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix}$$

$$E[X_1^2] = \sum x_1^2 p(x_1) = (-1)^2\times 0.3 + 0^2\times 0.3 + 1^2\times 0.4 = 0.7$$

$$E[X_2^2] = \sum x_2^2 p(x_2) = 0^2\times 0.8 + 1^2\times 0.2 = 0.2$$

$$E[X_1X_2] = \sum x_1 x_2\, p(x_1, x_2)$$

$$= (-1)\times 0\times 0.24 + (-1)\times 1\times 0.06 + 0\times 0\times 0.16 + 0\times 1\times 0.14$$
$$+ 1\times 0\times 0.40 + 1\times 1\times 0.00$$

$$E[X_1X_2] = -0.06 = E[X_2X_1]$$

$$\therefore R_X = \begin{bmatrix} 0.7 & -0.06 \\ -0.06 & 0.2 \end{bmatrix}$$

$$C_X = R_X - \mu_X \mu_X'$$

$$= \begin{bmatrix} 0.7 & -0.06 \\ -0.06 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \begin{bmatrix} 0.1 & 0.2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.7 & -0.06 \\ -0.06 & 0.2 \end{bmatrix} - \begin{bmatrix} 0.01 & 0.02 \\ 0.02 & 0.04 \end{bmatrix} = \begin{bmatrix} 0.69 & -0.08 \\ -0.08 & 0.16 \end{bmatrix}$$

Transformation of random variables

Suppose that X_1 and X_2 are continuous random variables with joint p.d.f given by $f_{X_1,X_2}(x_1,x_2)$.
Let $(Y_1,Y_2)=T(X_1,X_2)$. We want to find the joint p.d.f of Y_1 and Y_2.

Suppose $T:(x_1,x_2)\to(y_1,y_2)$ is a one-to-one transformation is some region of $\mathbb{R}^2$ such that $x_1=H_1(y_1,y_2)$ and $x_2=H_2(y_1,y_2)$.
The Jacobian of $T^{-1}=(H_1,H_2)$ is defined by

$$J(y_1,y_2)=\begin{vmatrix}\frac{\partial H_1}{\partial y_1} & \frac{\partial H_1}{\partial y_2}\\ \frac{\partial H_2}{\partial y_1} & \frac{\partial H_2}{\partial y_1}\end{vmatrix}$$

Theorem:
Let $(Y_1,Y_2)=T(X_1,X_2)$ be some transformation of random variables. If T is a one-to-one function and the Jacobian of T^{-1} is non-zero in $T(A)$ where $A=\{(x_1,x_2): f_{X_1,X_2}(x_1,x_2)>0\}$, then the joint p.d.f of Y_1 and Y_2, $f_{Y_1,Y_2}(y_1,y_2)$, is given by,

$f_{X_1,X_2}(H_1(y_1,y_2),H_2(y_1,y_2))\,|J(y_1,y_2)|$ if $(y_1,y_2)\in T(A)$, and 0 otherwise

example:
let $X_1\sim U(0,1)$, $X_2\sim U(0,1)$ and suppose that X_1 and X_2 are independent. Let $Y_1=X_1+X_2$, $Y_2=X_1-X_2$. Find the joint p.d.f of Y_1 and Y_2.

solⁿ. joint p.d.f of X_1 and X_2 if $f_{X_1,X_2}(x_1,x_2)=f_{X_1}(x_1)f_{X_2}(x_2)$

$$=\begin{cases}1, & 0\le x_1\le 1\ \&\ 0\le x_2\le 1\\ 0, & \text{otherwise.}\end{cases}$$

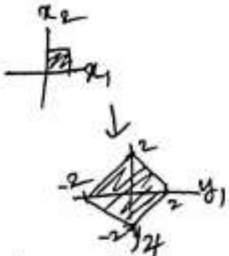
$T:(x_1,x_2)\to(y_1,y_2)$ is defined by $y_1=x_1+x_2$, $y_2=x_1-x_2$

$\therefore x_1=H_1(y_1,y_2)=\frac{y_1+y_2}{2}$, $x_2=H_2(y_2,y_1)=\frac{y_1-y_2}{2}$.

The Jacobian of $T^{-1}=J(y_1,y_2)=\begin{vmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & -\frac{1}{2}\end{vmatrix}=-\frac{1}{2}$

$\because A=\{(x_1,x_2)\mid 0\le x_1\le 1, 0\le x_2\le 1\}$ & since $x_1=0, x_1=1, x_2=0$ and $x_2=1$ map to $y_1+y_2=0$, $y_1+y_2=2$, $y_1-y_2=0$ and $y_1-y_2=2$

$T(A)=\{(y_1,y_2)\mid 0\le y_1+y_2\le 2, 0\le y_1-y_2\le 2\}$



$$\therefore f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{1}{2} f_{X_1,X_2}(H_1(y_1,y_2), H_2(y_1,y_2)), & \text{if } (y_1,y_2) \in T(A) \\ 0, & \text{otherwise.} \end{cases}$$

$$= \begin{cases} \frac{1}{2}, & \text{if } 0 \le y_1+y_2 \le 2 \text{ and } 0 \le y_1 - y_2 \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

<u>example</u>

Suppose that X_1 and X_2 are independent exponential random variables with parameter λ. Let $Y_1 = \frac{X_1}{X_2}$ and $Y_2 = X_1 + X_2$.
Find the joint p.d.f of Y_1 and Y_2.

<u>sol^n</u> Since X_1 and X_2 are independent exponential random variables with parameter λ, the joint p.d.f of X_1 and X_2 is given by

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$$
$$= \begin{cases} \lambda e^{-\lambda x_1} \lambda e^{-\lambda x_2}, & x_1, x_2 > 0 \\ 0, & \text{otherwise} \end{cases}$$
$$= \begin{cases} \lambda^2 e^{-\lambda(x_1+x_2)}, & x_1, x_2 > 0 \\ 0, & \text{otherwise.} \end{cases}$$

$$Y_1 = \frac{X_1}{X_2} \Rightarrow X_1 = Y_1 X_2 \Rightarrow Y_2 = Y_1 X_2 + X_2 = X_2(Y_1+1)$$

$$\Rightarrow X_2 = \frac{Y_2}{Y_1+1} = H_2(Y_1,Y_2) \text{ and } X_1 = Y_1 X_2 = \frac{Y_1 Y_2}{Y_1+1} = H_1(Y_1,Y_2)$$

$$\therefore \text{Jacobian of } T^{-1} = J(y_1,y_2) = \begin{vmatrix} \frac{y_2}{(y_1+1)^2} & \frac{y_1}{y_1+1} \\ -\frac{y_2}{(y_1+1)^2} & \frac{1}{y_1+1} \end{vmatrix} = \frac{y_2}{(y_1+1)^2}$$

$$A = \{(x_1,x_2) \mid f_{X_1,X_2}(x_1,x_2) > 0\} = \{(x_1,x_2) \mid x_1 > 0, x_2 > 0\}$$

Since $x_1 > 0$, $x_2 > 0$, $y_1 = \frac{x_1}{x_2} > 0$. Since $x_1 = \frac{y_1 y_2}{y_1+1} > 0 \Rightarrow y_1 y_2 > 0 \Rightarrow y_2 > 0$

$$\therefore T(A) = \{(y_1,y_2) \mid y_1 > 0, y_2 > 0\}$$

$$\text{p.d.f of } Y_1 \& Y_2 = f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \left|\frac{y_2}{(y_1+1)^2}\right| f_{X_1,X_2}(H_1(y_1,y_2), H_2(y_1,y_2)), & (y_1,y_2) \in T(A) \\ 0, & \text{otherwise} \end{cases}$$
$$= \begin{cases} \lambda^2 e^{-\lambda y_2} \cdot \left(\frac{y_2}{(1+y_1)^2}\right), & \text{if } y_1, y_2 > 0 \\ 0, & \text{otherwise.} \end{cases}$$

<u>example</u>

Let X and Y be independent random variables, each having probability density function,

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , x > 0 \\ 0 & , \text{otherwise.} \end{cases}$$

let $U = X+Y$ and $V = X-Y$.

Find the joint probability density function of U and V.

<u>Sol^n</u> Since X and Y are independent,

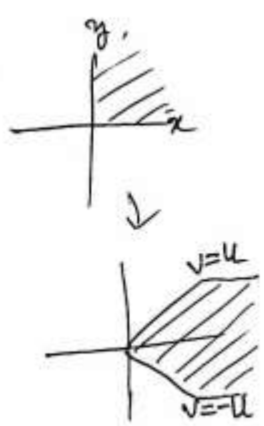
$$f_{X,Y}(x,y) = \begin{cases} \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda y} & , x, y > 0 \\ 0 & , \text{otherwise} \end{cases}$$

$$= \begin{cases} \lambda^2 e^{-\lambda(x+y)} & , x, y > 0 \\ 0 & , \text{otherwise.} \end{cases}$$

$U = X+Y$ and $V = X-Y \Rightarrow X = \frac{U+V}{2}, \; Y = \frac{U-V}{2}$.

$$\text{Jacobian of } T^{-1} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

pdf of U and V $= f_{U,V}(u,v) = \begin{cases} \frac{1}{2}\lambda^2 e^{-\lambda u} & , u+v > 0, \; u-v > 0 \\ 0 & , \text{otherwise} \end{cases}$



$$= \begin{cases} \frac{1}{2}\lambda^2 e^{-\lambda u} & , u > 0, \; -u < v < u \\ 0 & , \text{otherwise.} \end{cases}$$

Markov inequality

For any nonnegative random variable X with finite E[X], and any $k>0$, the following inequality holds:

$$P[X \geq k] \leq \frac{E[X]}{k}$$

example

A biased coin, with probability of tossing a head being $\frac{1}{5}$, is tossed 10 times. Estimate the probability of getting at least 8 heads in 10 tosses.

Sol^n $E[X] = np = 10 \times \frac{1}{5} = 2$.

By Markov inequality,

$$P[X \geq 8] \leq \frac{2}{8} = \frac{1}{4} = 0.25$$

The probability of atleast 8 heads is,

$n = 10 \quad p = \frac{1}{5}, \quad P[X \geq 8] = P[X=8] + P[X=9] + P[X=10]$

$${}^nC_x p^x q^{n-x} \qquad = {}^{10}C_8 \left(\frac{1}{5}\right)^8 \left(\frac{4}{5}\right)^2 + {}^{10}C_9 \left(\frac{1}{5}\right)^9 \left(\frac{4}{5}\right)^1 + {}^{10}C_{10} \left(\frac{1}{5}\right)^{10} \left(\frac{4}{5}\right)^0$$

$$= 0.0000779 < 0.25$$

example

A random variable X has the following PMF:

$$P_X(x) = \begin{cases} 1/25, & x = 5 \\ 24/25, & x = 0. \end{cases}$$

Estimate using Markov inequality a bound on the probability that X is at least 5.

Sol^n $E[X] = \sum x P_X(x) = 5 \times \frac{1}{25} + 0 \times \frac{24}{25} = \frac{1}{5}$.

$$P[X \geq 5] = \frac{1/5}{5} = \frac{1}{25}$$

and this is exactly the probability of $X = 5$.

Chebyshev's inequality

For any real-valued random variable X, with finite mean μ and finite variance $Var[X]$, and $k>0$, the following inequality holds: $P[|X-\mu| \geq k] \leq \frac{Var[X]}{k^2}$.

If we set $k = n\sigma$, where σ is the standard deviation of X, then we get $P[|X-\mu| \geq n\sigma] \leq \frac{Var[X]}{n^2\sigma^2} = \frac{1}{n^2}$.

example.
Let X be a random variable with mean 4 and variance 2. Use Chebyshev's inequality to obtain an upper bound on $P[|X-4| \geq 2]$.

sol^n $P[|X-4| \geq 2] \leq \frac{2}{2^2} = \frac{1}{2}$.

example
A random variable X is exponentially distributed with parameter λ. Find an upper bound on $P[|X-E[X]| \geq 1]$ using Chebyshev's inequality.

sol^n For $\lambda>0$, exponential distribution function is given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad E[X] = \frac{1}{\lambda} \text{ and } Var[X] = \frac{1}{\lambda^2}.$$

$$\therefore P[|X - E[X]| \geq 1] \leq \frac{1/\lambda^2}{1^2} = \frac{1}{\lambda^2}$$

Multivariate normal density and its properties

The multivariate normal density is a generalization of the univariate normal density to $p \geq 2$ dimensions.
The univariate normal distribution with mean μ and variance σ^2, has the probability density function,

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty \qquad -(1)$$

The term $\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$, $\qquad -(2)$

in the exponent of the univariate normal density function measures the square of the distance from x to μ in standard deviation units.
This can be generalized for a $p\times 1$ vector x of observations on several variables as,

$$(x-\mu)'\Sigma^{-1}(x-\mu) \qquad -(3)$$

The $p\times 1$ vector μ represents the expected value of the random vector X, and the $p\times p$ matrix Σ is the covariance matrix of X.
The expression in 3 is the square of the generalized distance from x to μ.

A p-dimensional normal density for the random vector $X = [X_1\ X_2 \dots X_p]'$ has the form

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \quad -\infty < x_i < \infty, \ i = 1, 2, \dots p \qquad -(4)$$

we shall denote this p-dimensional normal density by $N_p(\mu, \Sigma)$.

Bivariate Normal density

For two random variables X_1, X_2, the bivariate normal density function is given by:

$$f(x) = \frac{1}{2\pi|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}, \quad -\infty < x_1, x_2 < \infty$$

where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mu_1 = E[X_1]$, $\mu_2 = E[X_2]$, $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

$$x-\mu = \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix},$$

$\sigma_{11} = \mathrm{Var}(X_1)$, $\sigma_{22} = \mathrm{Var}(X_2)$, $\sigma_{12} = E[X_1 X_2] - E[X_1]E[X_2] = \sigma_{21}$.

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

Result:
If Σ is positive definite, so that Σ^{-1} exists, then $\Sigma e = \lambda e$ implies $\Sigma^{-1}e = \left(\frac{1}{\lambda}\right)e$, so (λ, e) is an eigenvalue-eigenvector pair for Σ corresponding to the pair $\left(\frac{1}{\lambda}, e\right)$ for Σ^{-1}. Also Σ^{-1} is positive definite.

Result,
From the expression in (4) for the density of a p-dimensional normal variable, the paths of x values yielding a constant height for the density are ellipsoids. That is, the multivariate normal density is constant on surfaces where the square of the distance $(x-\mu)'\Sigma^{-1}(x-\mu)$ is constant. These paths are called contours.

Constant probability density contour $= \{\text{all } x \text{ such that } (x-\mu)'\Sigma^{-1}(x-\mu) = c^2\}$
$=$ surface of an ellipsoid centered at μ.

The axes of each ellipsoid of constant density are in the direction of the eigenvectors of Σ^{-1}, and their lengths are proportional to the reciprocals of the square roots of the eigenvalues of Σ^{-1}.

Contours of constant density for the p-dimensional normal distribution are ellipsoids defined by x such that

$(x-\mu)'\Sigma^{-1}(x-\mu)=c^2$.

These ellipsoids are centered at μ and have axes $\pm c\sqrt{\lambda_i}\, e_i$ where $\Sigma e_i=\lambda_i e_i$ for $i=1,2,\ldots,p$.

~~A Con~~ example: Obtain the axes of constant probability density contours for a bivariate normal distribution when $\sigma_{11}=\sigma_{22}$

Solⁿ

$$\Sigma=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{bmatrix}=\begin{bmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{11}\end{bmatrix}$$

$$|\Sigma-\lambda I|=0 \Rightarrow (\sigma_{11}-\lambda)^2-\sigma_{12}^2=(\lambda-\sigma_{11})^2-\sigma_{12}^2=(\lambda-\sigma_{11}-\sigma_{12})(\lambda-\sigma_{11}+\sigma_{12})=0$$

$$\Rightarrow \lambda_1=\sigma_{11}+\sigma_{12} \text{ and } \lambda_2=\sigma_{11}-\sigma_{12}$$

for $\lambda_1=\sigma_{11}+\sigma_{12}$ $\quad (\Sigma-\lambda_1 I)e_1=0 \Rightarrow \begin{bmatrix}-\sigma_{12} & \sigma_{12}\\ \sigma_{12} & -\sigma_{12}\end{bmatrix}\begin{bmatrix}e_1\\ e_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \Rightarrow \frac{e_1}{-\sigma_{12}}=\frac{-e_2}{\sigma_{12}}$

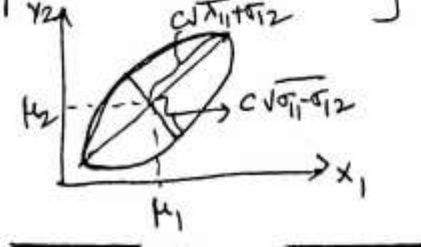
$$\Rightarrow e_1=\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$$

for $\lambda_2=\sigma_{11}-\sigma_{12}$ $\quad (\Sigma-\lambda_2 I)e_2=0 \Rightarrow \begin{bmatrix}\sigma_{12} & \sigma_{12}\\ \sigma_{12} & \sigma_{12}\end{bmatrix}\begin{bmatrix}e_1\\ e_2\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix} \Rightarrow \frac{e_1}{\sigma_{12}}=\frac{-e_2}{\sigma_{12}}$

$$\Rightarrow e_2=\begin{bmatrix}1\\-1\end{bmatrix}=\begin{bmatrix}1/\sqrt{2}\\-1/\sqrt{2}\end{bmatrix}$$

When the covariance (correlation 1,2) σ_{12} is positive, then $\lambda_1=\sigma_{11}+\sigma_{12}$ is the largest eigenvalue, and its associated eigenvector $e_1=\begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\end{bmatrix}$ lies along the 45° line through the point $\mu=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}$.

Since the axes of the constant-density ellipses are given by $\pm c\sqrt{\lambda_1}\, e_1$ and $\pm c\sqrt{\lambda_2}\, e_2$ and the eigenvectors each have length unity, the major axis will be associated with the largest eigenvalue. For positively correlated normal random variables, then, the major axis of the constant-density ellipses will be along the 45° line throug μ.



When the covariance (correlation) is negative, $\lambda_2 = \sigma_{11} - \sigma_{12}$ will be the largest eigenvalue, and the major axes of the constant-density ellipses will lie along a line at right angles to the 45° line through μ. (These results are true only for $\sigma_{11} = \sigma_{22}$)

To summarize, the axes of the ellipses of constant density for a bivariate normal distribution with $\sigma_{11} = \sigma_{22}$ are determined by

$$\pm c\sqrt{\sigma_{11}+\sigma_{12}} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \text{ and } \pm c\sqrt{\sigma_{11}-\sigma_{12}} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}.$$

Properties of multivariate Normal (Gaussian) Distribution.

~~4.~~ For a vector $X = [x_1 \dots x_n]'$, mean μ, covariance matrix Σ

1. The multivariate normal distribution has the joint density

$$f(x) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)} \qquad -\infty < x_i < \infty, \; i = 1, 2, \dots n.$$

2. The contours of the joint distribution are n-dimensional ellipsoids.

3. The joint distribution $N_n(\mu, \Sigma)$ is specified by μ & Σ only.

4. The distribution $N_n(\mu, \Sigma)$ has the moment generating function,

$$M(t) = e^{(\mu^T t + \frac{1}{2} t^T \Sigma t)},$$ where t is a real $n\times 1$ vector.

5. The distribution $N_n(\mu, \Sigma)$ has the characteristic function

$$\psi(t) = e^{(i\mu^T t - \frac{1}{2} t^T \Sigma t)},$$ where t is a real $n\times 1$ vector.

<u>example</u>:

Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be a multivariate normal random vector with mean $\mu = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and covariance matrix $\Sigma = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$. Find the mean and variance of the random variable $Y = X_1 + X_2$, which follows a normal distribution.

<u>Sol^n</u> $E[Y] = E(X_1 + X_2) = E[X_1] + E[X_2] = 1 + 2 = \underline{3}$

$$\begin{aligned}
Var[Y] &= Var[X_1 + X_2] = E[(X_1 + X_2)^2] - E[(X_1 + X_2)]^2 \\
&= E(X_1^2 + X_2^2 + 2X_1X_2) - [E[X_1] + E[X_2]]^2 \\
&= E(X_1^2) + E(X_2^2) + 2E[X_1X_2] - [E[X_1]]^2 - (E[X_2])^2 - 2E[X_1]E[X_2] \\
&= E[X_1^2] - (E[X_1])^2 + E(X_2^2) - (E[X_2])^2 + 2[E[X_1X_2] - E[X_1]E(X_2)] \\
&= Var[X_1] + Var[X_2] + 2Cov[X_1, X_2] \\
&= 3 + 2 + 2 \times 1 \\
&= \underline{7}
\end{aligned}$$

Also Y can be represented in vector form as $t = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \equiv \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Then mean $= \mu^T t = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3$

variance $= t^T \Sigma t = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} = 7$.

<u>example</u>

Measurements were taken on n heart-attack patients on their cholestrol levels. For each patient, measurements were taken 0, 2, 4 days following the attack. Treatment was given to reduce cholestrol levels. The sample mean vector and covariance matrix are as follows:

variable	mean
x_1 = 0 day	259.5
x_2 = 2 day	230.8
x_3 = 4 day	221.5

	0 day	2 day	4 day
0 day	2276	1508	813
2 day	1508	2206	1349
4 day	813	1349	1865

Suppose we are interested in the difference $x_1 - x_2$, the difference b/w 0-day and 2-day measurements.

Then the mean value of difference is $\mu^T t = \begin{bmatrix} 259.5 & 230.8 & 221.5 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 28.7$

Variance is $t^T \Sigma t = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} & & \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 1466$

example

Obtain the axes of the constant probability density contours for a bivariate normal distribution with the covariance matrix $\Sigma = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$.

$$\begin{array}{r|l} 2 & 48 \\ 2 & 24 \\ 2 & 12 \\ 2 & 6 \\ & 3 \end{array}$$

Sol^n: $\sigma_{11} = 0.5$, $\sigma_{12} = 0.25$, $\sigma_{22} = 0.5$

$$|\Sigma - \lambda I| = 0 \Rightarrow \lambda^2 - \lambda + \frac{3}{16} = 0 \Rightarrow 16\lambda^2 - 16\lambda + 3 = 0$$

$$\Rightarrow \lambda = \frac{3}{4}, \frac{1}{4}$$

for $\lambda = \frac{3}{4} = 0.75$

$$\Sigma - 0.75 I = \begin{bmatrix} -0.25 & 0.25 \\ 0.25 & -0.25 \end{bmatrix} \Rightarrow e = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

for $\lambda = \frac{1}{4} = 0.25$

$$\Sigma - 0.25 I = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix} \Rightarrow e = \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

The axes of the contours are $\pm c\sqrt{0.75} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ & $\pm c\sqrt{0.25} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$