Reminder of Random Variables II

Dr Ahmad Khonsari
ECE Dept.
The University of Tehran

5. System of RVs: jointly distributed RVs

Basic notes:

- sometimes it is required to investigate two or more RVs;

- we assume that RVs X and Y are defined on some probability space.

- Capital letters (i.e. X, Y) are random variables
and small letters (i.e. x, y are given constants)

5. System of RVs: jointly distributed RVs

Definition: joint probability distribution function (JPDF) of RVs X and Y is:

$$F_{XY}\,(x,y) = Pr\{X \le x, Y \le y\} \qquad (78)$$

For continuous RV., **Let us define**:

$$F_X\,(x) = Pr\{X \le x\} \quad F_Y\,(y) = Pr\{Y \le y\} \quad x, y \in R, \qquad (79)$$

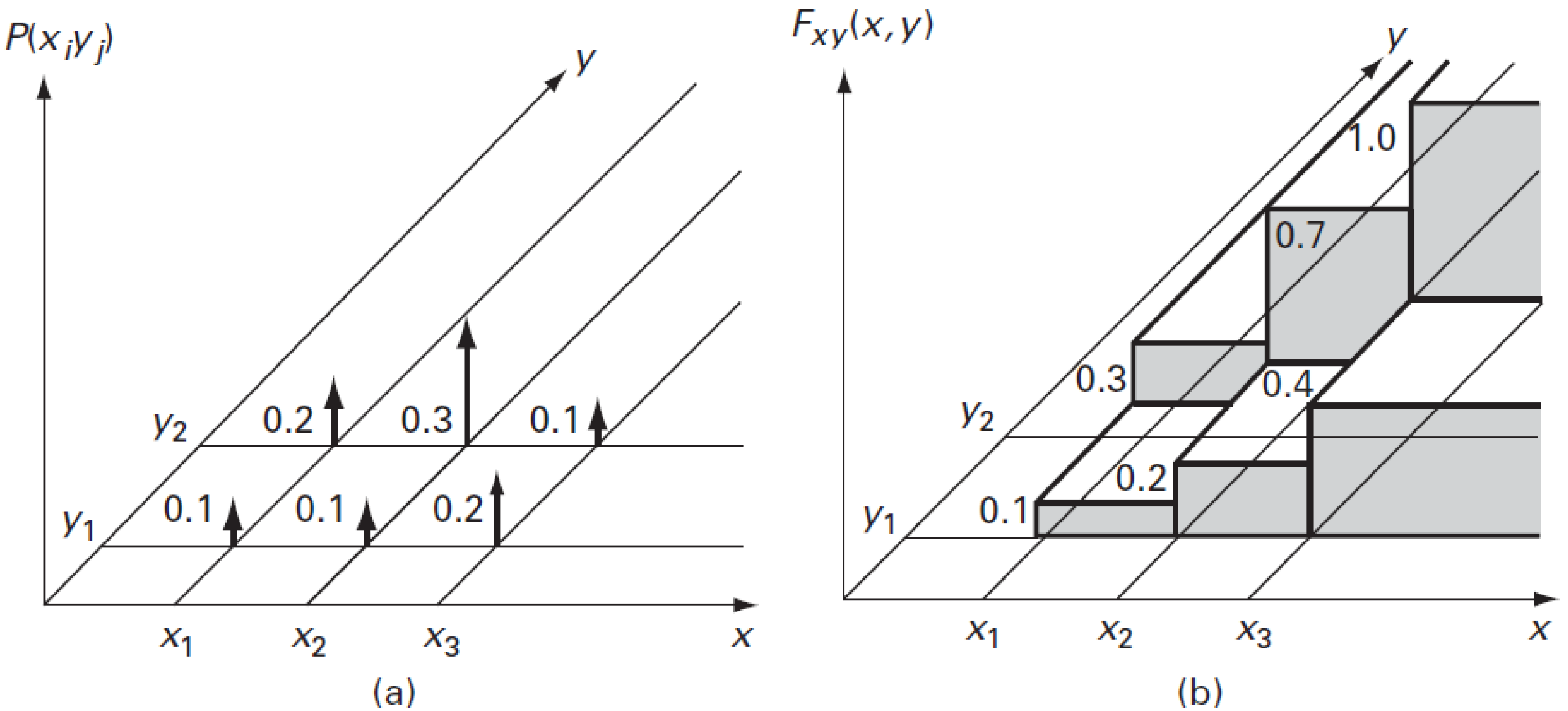
$F_X\,(x)$and $F_Y\,(y)$ are called marginal PDFs.

Marginal PDF can be derived form JPDF:

marginalize=neutralize=summing up to 1

$$F_X\,(x) = \lim_{y\to\infty} F_{XY}\,(x,y) = F_{XY}\,(x,\infty) \qquad (80)$$

$$F_Y\,(y) = \lim_{x\to\infty} F_{XY}\,(x,y) = F_{XY}\,(\infty,y)$$



(a) The joint probability distribution and
(b) the joint distribution function.

Definition: if $F_{XY}\ (x, y)$ is differentiable then the following function:

$$f_{XY}(x, y) = \frac{d^2}{dxdy} F_{XY}\ (x, y) \qquad (81)$$

$$= Pr\{x \leq X \leq x + dx, y \leq Y \leq y + dy\}$$

is called joint probability density function (jpdf).

Assume then that X and Y are discrete RVs.

Definition: joint probability mass function (Jpmf) of discrete RVs X and Y is:

$$f_{XY}(x,y)=\Pr\{X=x,Y=y\} \qquad (82)$$

Let us define:

$$f_X(x)=\Pr\{X=x\} \qquad f_Y(y)=\Pr\{Y=y\} \qquad (83)$$

- these functions are called marginal probability mass functions (Mpmf).

Marginal pmfs can be derived from Jpmf:

$$f_x(x)=\sum_{\forall y} f_{XY}(x,y), \qquad f_Y(y)=\sum_{\forall x} f_{XY}(x,y) \qquad (84)$$

با داشتن تابع توزیع توأم (یا تابع توزیع احتمال توأم) می توان جرم تک تک مولفه ها را بدست آورد، از جمله تابع توزیع حاشیه ای. ولی برعکس این موضوع درست نیست.

به عبارت دیگر با داشتن $P(X = x_i)$ و $P(Y = y_j)$ نمی توان $P(X = x_i, Y = y_j)$ را بدست آورد، ولی برعکس آن ممکن است.

$$P(X = x_i) = \sum_j P(x_i, y_j)$$

البته اگر پیشامدها مستقل باشند، به راحتی توزیع توأم از روی حاصلضرب ۲ توزیع کناری بدست می آید.

مثال: ۳ نوع باطری داریم: { نو=۳، کارکرده=۴ و خراب=۵}و میخواهیم سه باطری انتخاب کنیم.

پیشامدها $\begin{cases} X = \text{باطری برداشته شده نو باشد} \\ Y = \text{باطری برداشته شده کارکرده باشد} \end{cases}$

$$P(i,j) = P(X = i, Y = j) = ?$$

سه باطری برمی داریم احتمال آنکه صفر باطری سالم و صفر باطری کارکرده باشد. (احتمال اینکه هر سه باطری خراب باشد.)

$$P(0,0) = \frac{\binom{5}{3}}{\binom{12}{3}} = \frac{10}{220}$$

یک باطری کارکرده و دو تای دیگر خراب باشد. (صفر باطری سالم)

$$P(0,1) = \frac{\binom{4}{1}\binom{5}{2}}{\binom{12}{3}} = \frac{40}{220}$$

$i \backslash j$	$Y=0$	$Y=1$	$Y=2$	$Y=3$	$P(X=i)$
$X=0$	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
$X=1$	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
$X=2$	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
$X=3$	1	0	0	0	$\frac{1}{220}$
$P(Y=j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

pmf متغیر x با جمع سطری و pmf متغیر y با جمع ستونی بدست می آیدو چون این اطلاعات از روی حاشیه ها (کناره ها) جدول بدست می آید، به آن ها توزیع های حاشیه ای x وy می گویند.

نکته ۱: $P(X|Y=y)$ توزیع احتمال است.

مثالی از احتمال شرطی:

$$\sum_x P(X|Y=2) = \frac{P(0,2)}{P(Y=2)} + \frac{P(1,2)}{P(Y=2)} + \frac{P(2,2)}{P(Y=2)} + \frac{P(3,2)}{P(Y=2)} = 1$$

$$= \frac{\frac{30}{220}}{\frac{48}{220}} + \frac{\frac{18}{220}}{\frac{48}{220}} + \frac{0}{\frac{48}{220}} + \frac{0}{\frac{48}{220}} == \frac{30}{48} + \frac{18}{48} = 1$$

پس $P(X|Y=y)$ توزیع احتمال است.

نکته ۲: $P(Y=2)$ یک احتمال است و توزیع احتمال نیست، چون مقدار آن $\frac{48}{220}$ است.

نکته ۳: توزیع های حاشیه ای یک خلاصه ای از یک توزیع توأم است.

5.1. Conditional distributions and Mean (on Events / RV)

Discrete RV **Definition:** the following expression:

$$Pr_{X|Y}\{.,y\} = Pr_{X|Y}\{.|y\} = f_{X|Y}(.,y) = f_{X|Y}(.|y) = \frac{\text{Pr}\,\{X = \forall, Y = y\}}{\text{Pr}\,\{Y = y\}} \qquad (85)$$

• gives conditional PF of discrete RV X given that Y = y.

Conditional mean of RV X given Y = y can be obtained as:

$$\mathrm{E[X|Y\ =\ y]} = \sum_{\forall i} x_i\, Pr_{X|Y}\{\mathrm{x|y}\} \qquad (86)$$

Continous RV **Definition:** the following expression: $f_Y(y)$>0

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}\ , \qquad (87)$$

• gives conditional pdf of continuous RV X given that $Y = y$.

Conditional mean of RV X given Y = y from the following expression:

$$\mathrm{E[X|Y\ =\ y]} = \int_{-\infty}^{\infty} x\, f_{X|Y}\, dx \qquad (88)$$

5.1. Conditional distributions and Mean (conditioning by event / RV)

Conditional CDF:

$$F_{X|Y}(x|y) = Pr(X \leq x|Y \leq y) = \frac{\Pr\{X \leq x, Y \leq y\}}{\Pr\{Y \leq y\}} = \frac{F_{X,Y}(x,y)}{F_Y(y)}$$

Conditional pdf: $f_Y(y)>0$

$$f_{X|Y}(x|y) = \lim_{\Delta y \to 0} f_X(x|Y \approx y) = \lim_{\Delta y \to 0} \frac{\partial}{\partial x} F_X(x|Y \approx y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Note:

$$f_{X|Y}(x|y) \neq \frac{\partial}{\partial x} F_X(x|y)$$

Since the condition in pdf is Y=y and the condition in cdf is $Y \leq y$

Definition 2.19 ***Conditional PMF***

Given the event B, with $P[B] > 0$, the ***conditional probability mass function*** *of X is*

$$P_{X|B}(x) = P[X = x|B].$$

Theorem 2.16 *A random variable X resulting from an experiment with event space $B_1, \ldots, B_m$ has PMF*

yates page 82

$$P_X(x) = \sum_{i=1}^{m} P_{X|B_i}(x) P[B_i].$$

Proof The theorem follows directly from Theorem 1.10 with A denoting the event $\{X = x\}$.

Theorem 2.17

$$P_{X|B}(x) = \begin{cases} \dfrac{P_X(x)}{P[B]} & x \in B, \\ 0 & \text{otherwise.} \end{cases}$$

The theorem states that when we learn that an outcome $x \in B$, the probabilities of all $x \notin B$ are zero in our conditional model and the probabilities of all $x \in B$ are proportionally higher than they were before we learned $x \in B$.

Ref. book: Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers **2nd Edition**
by David J. Goodman (Author), Roy D. Yates (Author)

Theorem 4.6 *A joint PDF $f_{X,Y}(x,y)$ has the following properties corresponding to first and second axioms of probability (see Section 1.3):*

(a) $f_{X,Y}(x,y) \geq 0$ *for all* (x,y),

(b) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1.$

yates p161

Given an experiment that produces a pair of continuous random variables X and Y, an event A corresponds to a region of the X, Y plane. The probability of A is the double integral of $f_{X,Y}(x,y)$ over the region of the X, Y plane corresponding to A.

Theorem 4.7 *The probability that the continuous random variables (X, Y) are in A is*

$$P[A] = \iint_A f_{X,Y}(x,y)\,dx\,dy.$$

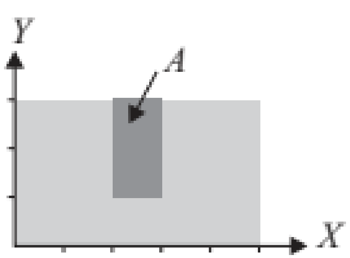
Example 4.4 Random variables X and Y have joint PDF

$$f_{X,Y}(x,y) = \begin{cases} c & 0 \leq x \leq 5, 0 \leq y \leq 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.22}$$

Find the constant c and $P[A] = P[2 \leq X < 3, 1 \leq Y < 3]$.

The large rectangle in the diagram is the area of nonzero probability. Theorem 4.6 states that the integral of the joint PDF over this rectangle is 1:

$$1 = \int_0^5 \int_0^3 c\,dy\,dx = 15c. \tag{4.23}$$



Therefore, $c = 1/15$. The small dark rectangle in the diagram is the event $A = \{2 \leq X < 3, 1 \leq Y < 3\}$. $P[A]$ is the integral of the PDF over this rectangle, which is

$$P[A] = \int_2^3 \int_1^3 \frac{1}{15}\,dv\,du = 2/15. \tag{4.24}$$

This probability model is an example of a pair of random variables uniformly distributed over a rectangle in the X, Y plane.

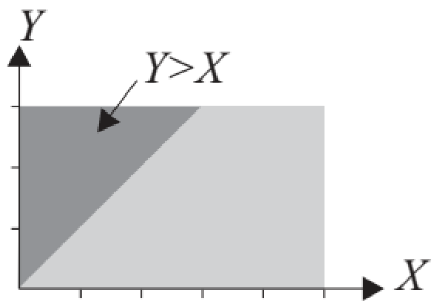
Example 4.6 As in Example 4.4, random variables X and Y have joint PDF

yates page 165

$$f_{X,Y}(x,y) = \begin{cases} 1/15 & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.30}$$

What is $P[A] = P[Y > X]$?

Applying Theorem 4.7, we integrate the density $f_{X,Y}(x,y)$ over the part of the X, Y plane satisfying $Y > X$. In this case,



$$P[A] = \int_0^3 \left(\int_x^3 \frac{1}{15} \right) dy\, dx \tag{4.31}$$

$$= \int_0^3 \frac{3-x}{15}\, dx = -\frac{(3-x)^2}{30} \bigg|_0^3 = \frac{3}{10}. \tag{4.32}$$

In this example, we note that it made little difference whether we integrate first over y and then over x or the other way around. In general, however, an initial effort to decide the simplest way to integrate over a region can avoid a lot of complicated mathematical maneuvering in performing the integration.

https://www.youtube.com/watch?v=hcBiYZuST7U

eg applying Fubini's theorem in calculating the expectation of a RV as tail prob yates P175

Definition 4.8 ***Correlation Coefficient***

The ***correlation coefficient*** *of two random variables X and Y is*

$$\rho_{X,Y} = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]}} = \frac{\mathrm{Cov}[X,Y]}{\sigma_X \sigma_Y}.$$

Note that the units of the covariance and the correlation are the product of the units of X and Y. Thus, if X has units of kilograms and Y has units of seconds, then $\mathrm{Cov}[X, Y]$ and $r_{X,Y}$ have units of kilogram-seconds. By contrast, $\rho_{X,Y}$ is a dimensionless quantity.

An important property of the correlation coefficient is that it is bounded by -1 and 1:

Theorem 4.17

$$-1 \le \rho_{X,Y} \le 1.$$

Proof Let σ_X^2 and σ_Y^2 denote the variances of X and Y and for a constant a, let $W = X - aY$. Then,

$$\mathrm{Var}[W] = E\left[(X - aY)^2\right] - (E[X - aY])^2. \tag{4.78}$$

Since $E[X - aY] = \mu_X - a\mu_Y$, expanding the squares yields

$$\mathrm{Var}[W] = E\left[X^2 - 2aXY + a^2Y^2\right] - \left(\mu_X^2 - 2a\mu_X\mu_Y + a^2\mu_Y^2\right) \tag{4.79}$$

$$= \mathrm{Var}[X] - 2a\,\mathrm{Cov}[X,Y] + a^2\,\mathrm{Var}[Y]. \tag{4.80}$$

Since $\mathrm{Var}[W] \ge 0$ for any a, we have $2a\,\mathrm{Cov}[X, Y] \le \mathrm{Var}[X] + a^2\,\mathrm{Var}[Y]$. Choosing $a = \sigma_X/\sigma_Y$ yields $\mathrm{Cov}[X, Y] \le \sigma_Y\sigma_X$, which implies $\rho_{X,Y} \le 1$. Choosing $a = -\sigma_X/\sigma_Y$ yields $\mathrm{Cov}[X, Y] \ge -\sigma_Y\sigma_X$, which implies $\rho_{X,Y} \ge -1$.

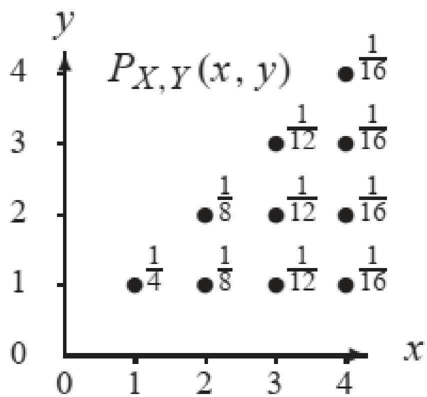
conditioning by event

Theorem 4.19 *For any event B, a region of the X, Y plane with $P[B] > 0$,*

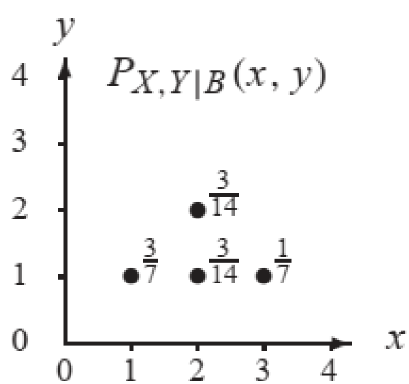
Yates p178

$$P_{X,Y|B}(x,y) = \begin{cases} \dfrac{P_{X,Y}(x,y)}{P[B]} & (x,y) \in B, \\ 0 & \textit{otherwise.} \end{cases}$$

Example 4.13



Random variables X and Y have the joint PMF $P_{X,Y}(x,y)$ as shown. Let B denote the event $X + Y \le 4$. Find the conditional PMF of X and Y given B.



Event $B = \{(1,1), (2,1), (2,2), (3,1)\}$ consists of all points (x,y) such that $x + y \le 4$. By adding up the probabilities of all outcomes in B, we find

$$P[B] = P_{X,Y}(1,1) + P_{X,Y}(2,1) + P_{X,Y}(2,2) + P_{X,Y}(3,1) = \frac{7}{12}.$$

The conditional PMF $P_{X,Y|B}(x,y)$ is shown on the left.

continuous RV

Definition 4.10 ***Conditional Joint PDF***

Given an event B with $P[B] > 0$, the conditional joint probability density function of X and Y is

Yates p178

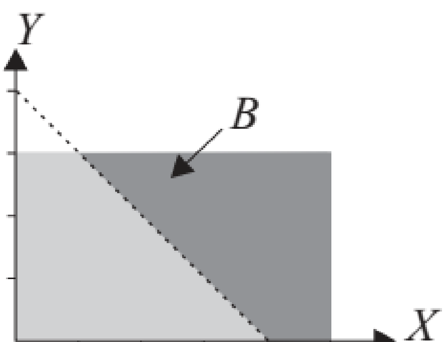
$$f_{X,Y|B}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P[B]} & (x,y) \in B, \\ 0 & \text{otherwise.} \end{cases}$$

Example 4.14 X and Y are random variables with joint PDF

$$f_{X,Y}(x,y) = \begin{cases} 1/15 & 0 \le x \le 5, 0 \le y \le 3, \\ 0 & \text{otherwise.} \end{cases} \tag{4.83}$$

Find the conditional PDF of X and Y given the event $B = \{X + Y \ge 4\}$.

We calculate $P[B]$ by integrating $f_{X,Y}(x,y)$ over the region B.



$$P[B] = \int_0^3 \int_{4-y}^5 \frac{1}{15}\, dx\, dy \tag{4.84}$$

$$= \frac{1}{15} \int_0^3 (1+y)\, dy \tag{4.85}$$

$$= 1/2. \tag{4.86}$$

Definition 4.10 leads to the conditional joint PDF

$$f_{X,Y|B}(x,y) = \begin{cases} 2/15 & 0 \le x \le 5, 0 \le y \le 3, x + y \ge 4, \\ 0 & \text{otherwise.} \end{cases} \tag{4.87}$$

Conditioning by a Random Variable

Yates p181

In Section 4.8, we use the partial knowledge that the outcome of an experiment $(x, y) \in B$ in order to derive a new probability model for the experiment. Now we turn our attention to the special case in which 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 variables X, Y. We now have complete knowledge of Y and modified knowledge of X. From this information, we derive a modified probability model for X. The new model is either a *conditional PMF of X given Y* or a *conditional PDF of X given Y*. When X and Y are discrete, the conditional PMF and associated expected values represent a specialized notation for their counterparts, $P_{X,Y|B}(x, y)$ and $E[g(X, Y)|B]$ in Section 4.8. By contrast, when X and Y are continuous, we cannot apply Section 4.8 directly because $P[B] = P[Y = y] = 0$ as discussed in Chapter 3. Instead, we define a conditional PDF as the ratio of the joint PDF to the marginal PDF.

Definition 4.12 ***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 4.22 *For random variables X and Y with joint PMF $P_{X,Y}(x, y)$, and x and y such that $P_X(x) > 0$ and $P_Y(y) > 0$,*

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

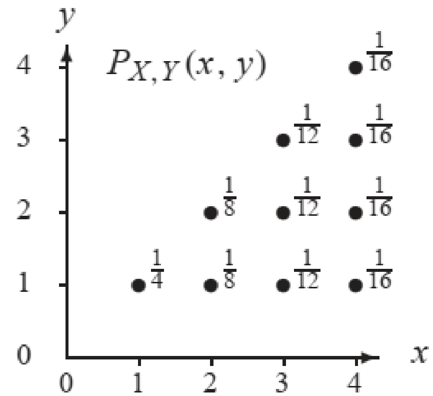
Proof Referring to Definition 4.12, Definition 1.6, and Theorem 4.3, we observe that

$$P_{X|Y}(x|y) = P[X = x|Y = y] = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{P_{X,Y}(x, y)}{P_Y(y)}. \tag{4.97}$$

Hence, $P_{X,Y}(x, y) = P_{X|Y}(x|y)P_Y(y)$. The proof of the second part is the same with X and Y reversed.

Example 4.17

Yates p182



Random variables X and Y have the joint PMF $P_{X,Y}(x,y)$, as given in Example 4.13 and repeated in the accompanying graph. Find the conditional PMF of Y given $X = x$ for each $x \in S_X$.

To apply Theorem 4.22, we first find the marginal PMF $P_X(x)$. By Theorem 4.3, $P_X(x) = \sum_{y \in S_Y} P_{X,Y}(x,y)$. For a given $X = x$, we sum the nonzero probablities along the vertical line $X = x$. That is,

$$P_X(x) = \begin{cases} 1/4 & x = 1, \\ 1/8 + 1/8 & x = 2, \\ 1/12 + 1/12 + 1/12 & x = 3, \\ 1/16 + 1/16 + 1/16 + 1/16 & x = 4, \\ 0 & \text{otherwise}, \end{cases} = \begin{cases} 1/4 & x = 1, \\ 1/4 & x = 2, \\ 1/4 & x = 3, \\ 1/4 & x = 4, \\ 0 & \text{otherwise}. \end{cases}$$

Theorem 4.22 implies that for $x \in \{1, 2, 3, 4\}$,

$$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = 4P_{X,Y}(x,y). \tag{4.98}$$

For each $x \in \{1, 2, 3, 4\}$, $P_{Y|X}(y|x)$ is a different PMF.

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

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

Given $X = x$, the conditional PMF of Y is the discrete uniform $(1, x)$ random variable.

Definition 4.13 ***Conditional PDF***

For y such that $f_Y(y) > 0$, the conditional PDF of X given $\{Y = y\}$ is

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

Definition 4.13 implies

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}. \tag{4.102}$$

Ref. book: Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers 2nd Edition
by David J. Goodman (Author), Roy D. Yates (Author)

Theorem 4.8 *If X and Y are random variables with joint PDF $f_{X,Y}(x,y)$,*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

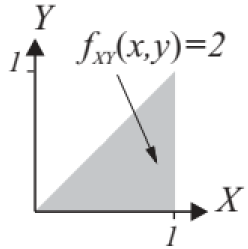
Proof From the definition of the joint PDF, we can write

$$F_X(x) = P[X \le x] = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{X,Y}(u,y)\,dy \right) du. \tag{4.34}$$

Taking the derivative of both sides with respect to x (which involves differentiating an integral with variable limits), we obtain $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$. A similar argument holds for $f_Y(y)$.

Example 4.19 Returning to Example 4.5, random variables X and Y have joint PDF

Yates p183



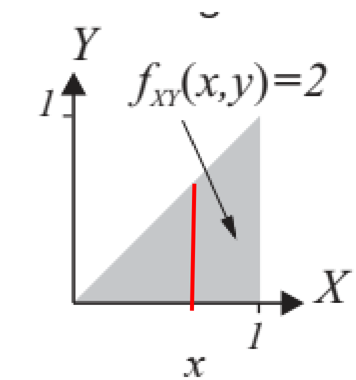
$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \le y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \qquad (4.103)$$

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 $f_{X|Y}(x|y)$.

$f_X(x) \neq Pr(X=x)$

For $0 \le x \le 1$, Theorem 4.8 implies

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy = \int_0^x 2\, dy = 2x. \qquad (4.104)$$

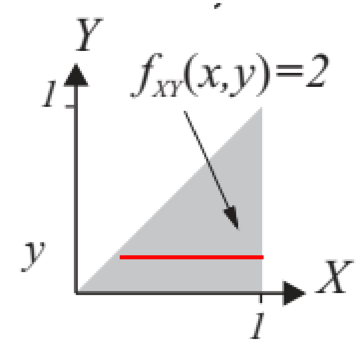


The conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} 1/x & 0 \le y \le x, \\ 0 & \text{otherwise.} \end{cases} \qquad (4.105)$$

Given $X = x$, we see that Y is the uniform $(0, x)$ random variable. For $0 \le y \le 1$, Theorem 4.8 implies

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx = \int_y^1 2\, dx = 2(1-y). \qquad (4.106)$$



Furthermore, Equation (4.102) implies

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} 1/(1-y) & y \le x \le 1, \\ 0 & \text{otherwise.} \end{cases} \qquad (4.107)$$

Conditioned on $Y = y$, we see that X is the uniform $(y, 1)$ random variable.

5.1. Conditional distributions and Mean (we saw Cond. Prob. Before)

Mixture Distribution:(page 239 Trivedi 1st ed.)

Conditoional density (pmf) can be extended to the case where X is discrete RV and Y is continuous RV (or vice versa)

5.2. Dependence and independence of RVs

Recall the definition of independent events E and F: P(EF)=P(E)P(F)

Definition: it is necessary and sufficient for two RVs X and Y to be independent:

$$F_{XY}(x,y) = F_X(x)F_Y(y) \text{ for all } x,y \tag{89}$$

- $F_{XY}(x,y)$ is the JPDF(=JCDF);
- $F_X(x)$ and $F_Y(y)$ are PDFs (CDFs) of RV X and Y .

Definition: it is necessary and sufficient for two continuous RVs X and Y to be independent:

$$f_{XY}(x,y) = f_X(x)f_Y(y) \text{ for all } x,y \tag{90}$$

- $f_{XY}(x,y)$ is the jpdf;
- $f_X(x)$ and $f_Y(y)$ are pdfs of RV X and Y .

Definition: it is necessary and sufficient for two discrete RVs X and Y to be independent:

$$p_{XY}(x,y) = p_{XY}(X = x, Y = \forall)p_Y(X = \forall, Y = y) \text{ for all } x,y \tag{91}$$

- $p_{XY}(x,y)$ is the Jpmf;
- $p_X(x)$and $p_Y(y)$ are pmfs (discrete RV) or pdfs (continuous RV)) of RV X and Y .

Let: D1, D2 be the outcomes of two rolls:
S=D1+D2. the sum of two rolls

Each roll of a 6-sided die is an independent trial,
D1,D2 are independent.
Are S ands D1 independent? No

1. p(D1=1,S=7)?
   = p(D1=1)p(s=7)

2. p(D1=1,S=5)?
   ≠ p(D1=1)p(s=5)

Let: D1, D2 be the outcomes of two rolls:
S=D1+D2. the sum of two rolls

- Each roll of a 6-sided die is an independent trial,
- D1,D2 are independent.

Are S ands D1 independent?

1. p(D1=1,S=7)?
Event (S=7) : {(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)}

p(D1=1)p(S=7)=(1/6)(1/6)
=1/36 =p(D1=1,S=7)

2. p(D1=1,S=5)?
Event (S=5) : {(1,4), (2,3),(3,2),(4,1)}
p(D1=1)p(S=5)=(1/6)(4/36)
≠ 1/36=p(D1=1,S=5)

Independent events (D1=1),(S=7)

Dependent events (D1=1),(S=5)

All events (X=x,Y=y) must be independent for X,Y to be independent variables.

5.3. Measure of dependence

Sometimes RVs are not independent:

- as a measure of dependence correlation moment (covariance) is used.

Definition: covariance of two RVs X and Y is defined as follows:

$$\sigma_{XY} = K_{XY} = cov(X, Y) = E[(X - E[X])(Y - E[Y])] \tag{92}$$

- where from definition , we find that $K_{XY} = K_{YX}$.

One can find the covariance using the following formulas:

- assume that RV X and Y are **discrete**:

$$K_{XY} = \sum_i \sum_j (x_i - E[X])(y_j - E[Y\ ])Pr\{X = x_i, Y = y_j\} \tag{93}$$

- assume that RV X and Y are **continuous**:

$$K_{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - E[X])(y_i - E[Y\ ])f_{XY}\ (x, y)dxdy \tag{94}$$

It is often easy to use the following expression :

$$\sigma_{XY} = K_{XY} = E[XY] - E[X]E[Y] \qquad (95)$$

Problem with covariance: can be arbitrary in $(-\infty, \infty)$:

- problem: hard to compare dependence between different pair of RVs;
- solution: use correlation coefficient to measure the dependence between RVs.

Definition: correlation coefficient of RVs X and Y is defined as follows:

$$\rho_{XY} = \frac{K_{XY}}{\sigma[X]\sigma[Y]} = \frac{\sigma_{XY}}{\sigma[X]\sigma[Y]} \qquad (96)$$

$$1 \le \rho_{XY} \le 1$$

- if $\rho_{XY} \neq 0$ then RVs X and Y are correlated and hence dependent;

- **Example:** assume we are given RVs X and Y such that Y = aX + b:

$$\rho_{XY} = +1 \qquad a>0$$

$$\rho_{XY} = -1 \qquad a<0 \qquad (97)$$

Very important note:

- ρ_{XY} is the measure telling how close the dependence to **linear.**

Question: what conclusions can be made when ρ_{XY} = 0? They are uncorrelated

•or RVs X and Y are not LINEARLY dependent;

- when $\rho_{XY} = 0$ is does not mean that they are independent.

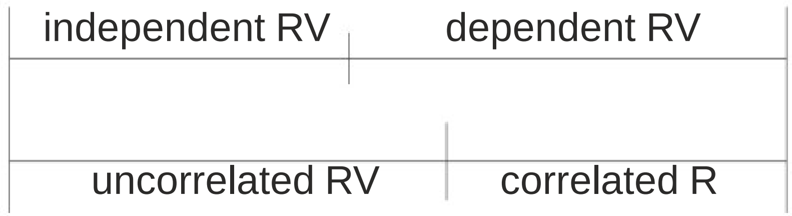


Fig: Independent and uncorrelated RVs.

What ρ_{XY} says to us:

•$\rho_{XY} \neq 0$: two RVs are correlated and also dependent;

- $\rho_{XY} = 0$: one can suggest that two RVs **MAY** BE independent;
- $\rho_{XY} = +1$ or $\rho_{XY} = -1$: RVs X and Y are linearly dependent.

5.4. (Expectations of product and Expectations of Sum) of correlated RVs

Mean:

- the mean of the product of two correlated RVs X,Y:

$$E[XY] = E[X]E[Y] + K_{XY} \tag{98}$$

- the mean of the product of two uncorrelated RVs X,Y:

$$E[XY] = E[X]E[Y] \tag{99}$$

Variance:

- the variance of the sum of two correlated RVs X,Y:

$$V[X+Y] = V[X] + V[Y] + 2K_{XY} \tag{100}$$

- the variance of the sum of two uncorrelated RVs X,Y:

$$V[X+Y] = V[X] + V[Y] \tag{101}$$

Now the Theory…

To capture this, define Covariance :

$$\boxed{\sigma_{XY} = E\{(X-\overline{X})(Y-\overline{Y})\}}$$

$$\boxed{\sigma_{XY} = \int\int (x-\overline{X})(y-\overline{Y})\, p_{XY}(x,y)\,dx\,dy}$$

If the RVs are both Zero-mean : $\boxed{\sigma_{XY} = \mathrm{E}\{XY\}}$

If X = Y: $\boxed{\sigma_{XY} = \sigma_X^2 = \sigma_Y^2}$

If X & Y are independent, then: $\boxed{\sigma_{XY} = 0}$

If $\sigma_{XY} = E\{(X - \overline{X})(Y - \overline{Y})\} = 0$

Say that X and Y are "uncorrelated"

If $\sigma_{XY} = E\{(X - \overline{X})(Y - \overline{Y})\} = 0$

Then $\underbrace{E\{XY\}}_{\text{Called "Correlation of X \&Y"}} = \overline{X}\,\overline{Y}$

So… RVs X and Y are said to be uncorrelated

if $E\{XY\} = E\{X\}E\{Y\}$

Independence vs. Uncorrelated

X & Y are Independent

$$f_{XY}(x,y) = f_X(x) f_Y(y)$$

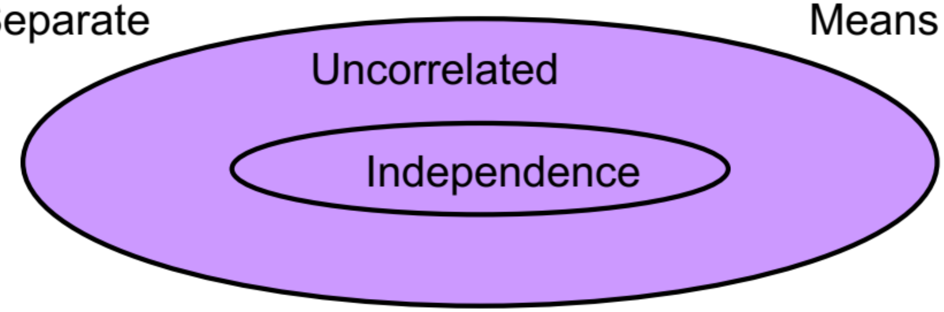
PDFs Separate

$\xrightarrow{\text{Implies}}$

X & Y are Uncorrelated

$$E\{XY\} = E\{X\}E\{Y\}$$

Means Separate



INDEPENDENCE IS A STRONGER CONDITION !!!!

Confusing Terminology…

Covariance : $\boxed{\sigma_{XY} = E\{(X - \bar{X})(Y - \bar{Y})\}}$

Correlation : $\boxed{E\{XY\}}$

Same if zero mean

Correlation Coefficient : $\boxed{\rho_{XY} = \dfrac{\sigma_{XY}}{\sigma_X \sigma_Y}}$

$\boxed{-1 \le \rho_{XY} \le 1}$

For Random Vectors…

$$\mathbf{x} = [X_1 \; X_1 \; \cdots \; X_N]^T$$

Correlation Matrix :

$$\mathbf{R}_{\mathbf{x}} = E\{\mathbf{x}\mathbf{x}^T\} = \begin{bmatrix} E\{X_1X_1\} & E\{X_1X_2\} & \cdots & E\{X_1X_N\} \\ E\{X_2X_1\} & E\{X_2X_2\} & \cdots & E\{X_2X_N\} \\ \vdots & \vdots & \ddots & \vdots \\ E\{X_NX_1\} & E\{X_NX_2\} & \cdots & E\{X_NX_N\} \end{bmatrix}$$

Covariance Matrix :

$$\mathbf{C}_{\mathbf{x}} = E\{(\mathbf{x} - \overline{\mathbf{x}})(\mathbf{x} - \overline{\mathbf{x}})^T\}$$

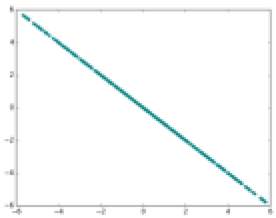
ارتباط شکل های 1 تا 4 را با روابطA,B,C,D را مشخص نمایید.

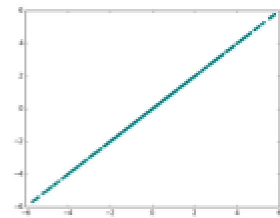
A. $\rho(X,Y) = 1$

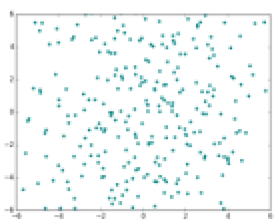
B. $\rho(X,Y) = -1$

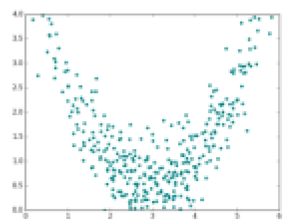
C. $\rho(X,Y) = 0$

D. Other

1.


2.


3.


4.


1-B $Y = -\frac{\sigma_Y}{\sigma_X} X + b$

خطی با ضریب زاویه منفی است . به مقدار ضریب زاویه هم توجه کنید و سعی کنید ان را متوجه شوید

2-A $Y = \frac{\sigma_Y}{\sigma_X} X + b$

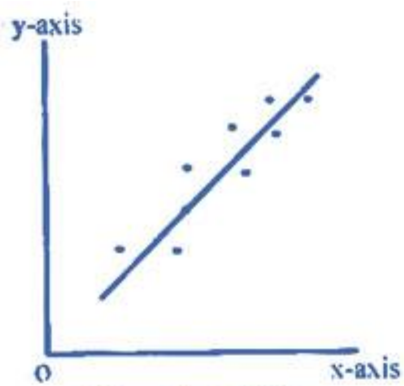
خطی با ضریب زاویه مثبت است . به مقدار ضریب زاویه هم توجه کنید و سعی کنید ان را متوجه شوید

3- C. $\rho(X,Y)$ =0 (ناهمبسته)

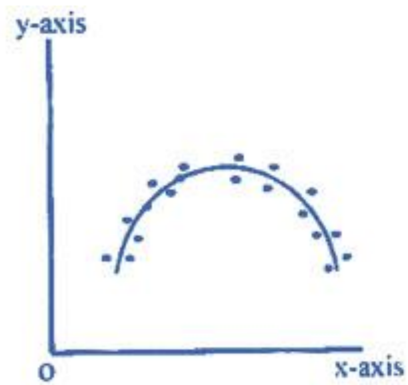
4- C. $\rho(X,Y)$ =0 $Y = X^2$

همانطور که تاکید شد ***همبستگی*** " خطی بودن " را اندازه می گیردو شکل 4 نشان میدهد با انکه کوواریانس XوY صفر می باشد . این دو متغیر به طور "غیر خطی " با هم رابطه دارند

- **<u>Linear Correlation</u>**
- Correlation is said to be linear if the ratio of change is constant. When the amount of output in a factory is doubled by doubling the number of workers, this is an example of linear correlation.
- In other words, when all the points on the scatter diagram tend to lie near a line which looks like a straight line, the correlation is said to be linear. This is shown in the figure on the left below.
- **<u>Non Linear (Curvilinear) Correlation</u>**
- Correlation is said to be non linear if the ratio of change is not constant. In other words, when all the points on the scatter diagram tend to lie near a smooth curve, the correlation is said to be non linear (curvilinear). This is shown in the figure on the right below.
-



Linear Correlation



Non Linear Correlation

6. Pdf of Sum of independent RVs

We consider independent RVs X and Y with probability functions:

$$P_X(x) = \Pr\{X = x\},\ P_Y(y) = \Pr\{Y = y\} \qquad (102)$$

PMF of RV Z, Z = X + Y is defined as follows (i.e. for independent RVs X and Y, convolution operation.)

$$\Pr\{Z = z\} = \sum_{k=-\infty} \Pr\{X = k\} \Pr\{Y = z - k\} \qquad (103)$$

- if $X = k$, then, Z take on z $(Z = z)$ if and only if $Y = z - k$.

If RVs X and Y are continuous:

$$f_Z(z) = f_X(x) \odot f_Y(y) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y) dy = \int_{-\infty}^{\infty} f_Y(z - x) f_X(x) dx \qquad (104)$$

Exercise: CDF of sum of 2 independent RVs : $\boldsymbol{F_z(z) = F_x(z) \odot f_y(z) = f_x(z) \odot F_y(z)}$

Q: what is pdf of the sum of two RVs generally

سؤال ① نشان دهید اگر $Z = X + Y$ را بدست آورید

(5-21) ~~$F_Z(z) = \iint I(x+y \le z)\, f_{XY}(x,y)\,dx\,dy$~~ $= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx\right]dy$

~~$f_Z(z) = \int_{-\infty}^{\infty}\left[\frac{\partial}{\partial z}\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx\right]dy$ مشتق برحسب z میگیریم~~

~~(مشتق لایب‌نیتز) مشتق انتگرال یعنی وقتی حدود انتگرال تابع~~ متغیر و توابع شکل هستند

(5-94) 133 $\frac{d}{dz}\int_{a(z)}^{b(z)} h(z,y)\,dy = h(z, b(z))\, b'(z) - h(z, a(z))\, a'(z) + \int_{a(z)}^{b(z)} \frac{\partial h(z,y)}{\partial z}\,dy$

پس $\frac{\partial}{\partial z}\int_{-\infty}^{z-y} f_{XY}(x,y)\,dx = f_{XY}(z-y, y) \times 1 - f_{XY}(-\infty, y)\times 0 + \int_{-\infty}^{z-y} \frac{\partial f_{XY}(x,y)}{\partial z}\,dx$

مشتق x,y نسبت به z صفر است

$= f_{XY}(z-y, y)$

ص 115 کوبایاشی

پس $f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z-y, y)\,dy$

سه ② صفحه ژرد و سطر آخر ک

در صورت استقلال x و y داریم: $f_{xy}(x,y)=f_x(x)f_y(y)$

(۴۷)

$$f_Z(z)=\int_{-\infty}^{\infty} f_x(z-y)f_y(y)\,dy=\int_{-\infty}^{\infty} f_x(x)f_y(z-x)\,dx=f_x(z)\circledast f_y(z)$$

برای حالت استقلال، CDF ، Z برحسب CDF های X و Y به این صورت است:

(5.21) را به کمک انتگرال استیلتیز بازنویسی کنیم:

$$F_Z(z)=\iint I(x+y\le z)\,dF_x(x)\,dF_y(y)=\int_{-\infty}^{\infty}\int_{-\infty}^{z-y} dF_x(x)\,dF_y(y)$$

پس

$$F_Z(z)=\int_{-\infty}^{\infty} F_X(z-y)\,dF_Y(y)=\int_{-\infty}^{\infty} F_y(z-x)\,dF_x(x)$$

تمام روابط فوق به کمک شامل x و y است.

با جایگزینی $dF_y(y)$ با $f_y(y)dy$ و $dF_x(x)$ با $f_x(x)dx$

شکل کانولوشن CDF را برای $F_Z(z)$ داریم:

$$F_Z(z)=F_X(z)\circledast f_y(z)=f_x(z)\circledast F_y(z)$$

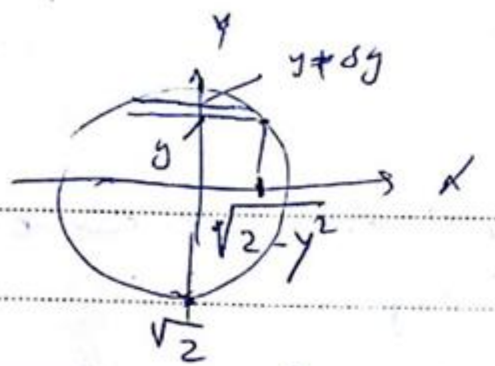
سوال (۳۲) $z = x^2 + y^2$ ، اگر چگالی توأم $f_{XY}(x,y)$ معلوم باشد، تابع احتمال Z

را بدست آورید.

$$F_Z(z) = P(X^2+Y^2 \leq z) = \iint I(x^2+y^2 \leq z)\, f_{XY}(x,y)\,dx\,dy$$

ناحیه $\{(x,y);\ x^2+y^2 \leq z\}$ دایره‌ای به شعاع $\sqrt{z}$ حول مبدأ است.

پس: $F_Z(z) = \int_{-\sqrt{z}}^{\sqrt{z}} \left[\int_{-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_{XY}(x,y)\,dx \right] dy$



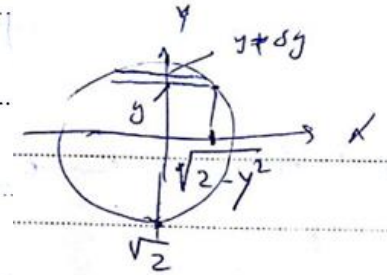
$$\left[\int_{-\sqrt{z-y^2}}^{\sqrt{z-y^2}} f_{XY}(x,y)\,dx\right]$$

$$\frac{\partial}{\partial z}F_Z = \frac{\partial}{\partial z}\int_{-\sqrt{z}}^{\sqrt{z}}[\;]\,dy = \int_{-\sqrt{z}}^{\sqrt{z}}\frac{\partial}{\partial z}\left[\int_{-\sqrt{z-y^2}}^{\sqrt{z-y^2}} dx\right]dy$$

از $F_Z(z)$ نسبت به z مشتق بگیریم و ترتیب انتگرال

در اینجا نمی‌خواد عوض کنیم. در ضمن تابع حد بالا و پایین تابع z هستند

مشتق حد بالا در [ ] داریم.



$$f_{xy}\left(\sqrt{z-y^2},y\right)\frac{1}{2\sqrt{z-y^2}} - f_{xy}\left(-\sqrt{z-y^2},y\right)\left(-\frac{1}{2\sqrt{z-y^2}}\right)+0$$

پس

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \int_{-\sqrt{z}}^{\sqrt{z}}\frac{1}{2\sqrt{z-y^2}}\left[f_{xy}\left(\sqrt{z-y^2},y\right)+f_{xy}\left(-\sqrt{z-y^2},y\right)\right]dy$$

An interesting case that often arises in signal detection problems, and for which we have a closed-form solution, is when X and Y are independent normal variables with zero mean and common variance (Problem 5.13). □

P 118 Kobayashi

5.13 Independent normal distribution and exponential distribution. Let X_1 and X_2 be independent normal variables with zero mean and common variance σ^2. Show that $Z = X_1^2 + X_2^2$ is exponentially distributed with mean $2\sigma^2$:

$$f_Z(z) = \frac{1}{2\sigma^2} e^{-z/2\sigma^2} u(z). \tag{5.104}$$

Example 5.5: $\boldsymbol{R = \sqrt{X^2 + Y^2}}$**.** Let us set $Z = R^2$ in the previous example. In the context of detecting a signal of the form $S(t) = X\cos(\omega t - \phi) + Y\sin(\omega t - \phi)$, the RV $R = \sqrt{X^2 + Y^2}$ represents the **envelope** of the signal, i.e., $S(t) = R\cos(\omega t - \theta)$, where $\theta - \phi = \tan^{-1}\frac{Y}{X}$.

The distribution function of R is given by

$$F_R(r) = \int_{-r}^{r}\left[\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} f_{XY}(x, y)\,dx\right]dy. \tag{5.35}$$

Differentiation of the expression inside the square brackets leads, using Leibniz's rule again, to the following expression:

$$f_{XY}(\sqrt{r^2 - y^2}, y)\frac{1}{2}\frac{2r}{\sqrt{r^2 - y^2}} - f_{XY}(-\sqrt{r^2 - y^2}, y)\left(-\frac{1}{2}\frac{2r}{\sqrt{r^2 - y^2}}\right) + 0. \tag{5.36}$$

Thus, we obtain

$$f_R(r) = \frac{dF_R(z)}{dr} = \int_{-r}^{r}\frac{r}{\sqrt{r^2 - y^2}}\left[f_{XY}(\sqrt{r^2 - y^2}, y) + f_{XY}(-\sqrt{r^2 - y^2}, y)\right]dy. \tag{5.37}$$

Again, an important and useful case is found when X and Y are independent normal variables with common variance (see Section 7.5.1). □

5.6* Leibniz's rule.[9] In deriving (5.23), we used a special case of Leibniz's rule for differentiation under the integral sign.

THEOREM 5.1 (Leibniz's rule). *The following rule holds for differentiation of a definite integral, when the integration limits are functions of the differential variable:*

$$\boxed{\frac{d}{dz}\int_{a(z)}^{b(z)} h(z,y)\,dy = h(z,b(z))b'(z) - h(z,a(z))a'(z) + \int_{a(z)}^{b(z)} \frac{\partial}{\partial z}h(z,y)\,dy.} \tag{5.94}$$

In particular, if h is a function of y only, the rule reduces to

$$\boxed{\frac{d}{dz}\int_{a(z)}^{b(z)} h(y)\,dy = h(b(z))b'(z) - h(a(z))a'(z).} \tag{5.95}$$

□

(a) Define

$$\int_{-\infty}^{y} h(x)\,dx \triangleq H(y).$$

Then prove (5.95).

(b) Define

$$\int_{-\infty}^{y} h(z,x)\,dx \triangleq H(z,y) \text{ and } \frac{\partial H(z,y)}{\partial y} \triangleq g(z,y).$$

Then prove (5.94).

(c) Alternative proof of (5.94). Consider a function $G(a, b, c)$, where a, b, and c stand for $a(z)$, $b(z)$, and $c(z)$ respectively. By applying the *chain rule* to the function G, we have

$$\frac{dG(a,b,c)}{dz} = \frac{\partial G}{\partial a}a'(z) + \frac{\partial G}{\partial b}b'(z) + \frac{\partial G}{\partial c}c'(z). \tag{5.96}$$

Consider a special case

$$c(z) = z \text{ and } G(a,b,c) \triangleq \int_a^b h(z,y)\,dy.$$

Then prove (5.94).

اگر از قاعده لایب نیتز، h فقط تابع y باشد داریم:

$$\frac{d}{dz}\int_{a(z)}^{b(z)} h(y)\,dy = h(b(z))\,b'(z) - h(a(z))\,a'(z) \qquad (5.95)$$

روش اثبات لایب نیتز:

الف) تعریف کنید $\int_{-\infty}^{y} h(x)\,dx \triangleq H(y)$

و از روی آن $(\underline{5.95})$ را بدست آورید.

ب) تعریف کنید: $\int_{-\infty}^{y} h(z,x)\,dx \triangleq H(z,y)$

and

$$\frac{\partial H(z,y)}{\partial y} \triangleq g(z,y)$$

و سپس 5.94 یا لایب نیتز کلی را ثابت کنید.

2) روش دوم برای اثبات 5.94: تابع $G(a, b, c)$ را در نظر بگیرید

که a، b، c به ترتیب برابر $a(z)$، $b(z)$، $c(z)$ هستند. با اعمال قاعده

زنجیری روی مشتق G داریم:

$$\frac{d\,G(a,b,c)}{dz} = \frac{\partial G}{\partial a} a'(z) + \frac{\partial G}{\partial b} b'(z) + \frac{\partial G}{\partial c} c'(z) \qquad (5\text{-}96)$$

از حالت خاص $c(z)=z$ و $G(a,b,c) \triangleq \int_a^b h(z,y)\,dy$ را در رابطه فوق قرار دهید، رابطه 5.94 را ثابت کنید

7. The distribution of max and min of independent random variables

Let $X1, \ldots, Xn$ be independent random variables

(distribution functions $F_i(x)$ and tail distributions $G_i(x)$, $i = 1, \ldots, n$)

Distribution of the maximum

$$\begin{aligned} P\{\max(X_1, \ldots, X_n) \le x\} &= P\{X_1 \le x, \ldots, X_n \le x\} \\ &= P\{X_1 \le x\} \cdots P\{X_n \le x\} \quad \text{(independence!)} \\ &= F_1(x) \cdots F_n(x) \end{aligned} \tag{105}$$

Distribution of the minimum

$$\begin{aligned} P\{\min(X_1, \ldots, X_n) > x\} = P\{X_1 > x, \ldots, X_n > x\} & \\ = P\{X_1 > x\} \cdots P\{X_n > x\} & \quad \text{(independence!)} \\ = G_1(x) \cdots G_n(x) & \end{aligned} \tag{106}$$

Appendix: General Case: Let $X_1, X_2, \ldots X_k$ be continuous random variables

i. Their joint **Cumulative Distribution Function**, $F(x_1, x_2, \ldots x_k)$ defines the probability that simultaneously X_1 is less than x_1, X_2 is less than x_2, and so on; that is

$$F(x_1, x_2, \ldots, x_k) = P(X_1 < x_1 \cap X_2 < x_2 \cap \cdots X_k < x_k)$$

i. The cumulative distribution functions $F_1(x_1), F_2(x_2), \ldots, F_k(x_k)$ of the individual random variables are called their **marginal distribution function**. For any i, $F_i(x_i)$ is the probability that the random variable X_i does not exceed the specific value x_i.

iii. The random variables are **independent** if and only if

$$F(x_1, x_2, \ldots, x_k) = F_1(x_1)F_2(x_2)\cdots F_k(x_k)$$

$$\textit{or}\ \text{equivalently}$$

$$f(x_1, x_2, \ldots, x_k) = f_1(x_1)\, f_2(x_2)\cdots f_k(x_k)$$