EC381/MN308
Probability and Some Statistics

Yannis Paschalidis

yannisp@bu.edu, http://ionia.bu.edu/

BOSTON UNIVERSITY   CISE

Dept. of Manufacturing Engineering

Dept. of Electrical and Computer Engineering

Center for Information and Systems Engineering

Lecture 11 - Outline

1. **Pairs of Random Variables**
   - **PDFs: joint, marginal, conditional, independence**

4.3 Joint PDF (Continuous Bivariate RVs)

Definition: Two RVs (X,Y) are said to be jointly continuous only if (X,Y) can take on all possible values in a region of nonzero area.

The probability mass is spread (usually with varying density) over this area.

Another definition: The CDF $F_{X,Y}(x,y)$ is continuous in x and y and is differentiable almost everywhere.

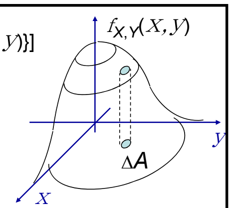
Joint PDF

$f_{X,Y}(x,y)$ = density of the probability mass ≥ 0
Units are probability mass per unit area

$P[(X,Y) \in$ {small area ΔA containing (x,y)}]

≈ $f_{X,Y}(x,y) \times \Delta A$

= *volume* of a column of area ΔA above the $x-y$ plane and below the $f_{X,Y}$ surface = height x base area

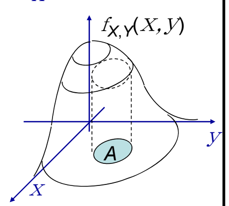


Probability from Joint PDF

The probability that the random point (X,Y) lies in a given region A is the integral of the joint pdf over the region A

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

= volume between A & $f_{X,Y}$ surface

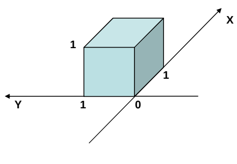


Total volume between $f_{X,Y}$ surface and the $x-y$ plane is 1

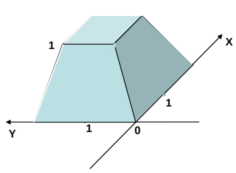
Example

X and Y are independent and each is uniform in [0,1].

PDF



CDF



Example

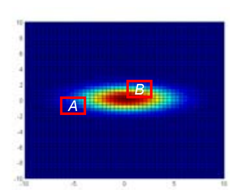
X, Y independent, $X \sim N(0,\sigma_1)$, $Y \sim N(0,\sigma_2)$.

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{y^2}{2\sigma_2^2}\right)$$

Range is entire plane!

$\sigma_2 = 3\,\sigma_1$

$P[B] > P[A]$



Example

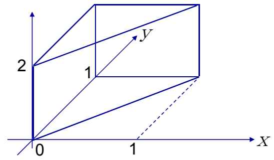
$f_{X,Y}(x,y) = 2$, for $0 < x < y < 1$

$= 0$, otherwise.

Determine:

a) $P[X + Y > 1]$

b) $P[(X - 0.5)^2 + (Y - 0.5)^2 \le 0.5^2]$

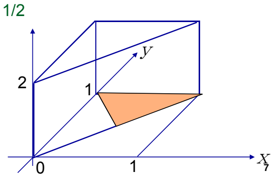


a) $P[X + Y > 1]$

Base Area is triangle of dimensions 1 x 1/2

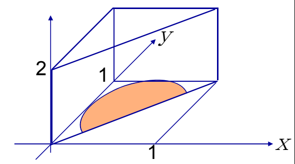
Height = 2

Volume = 2 x 1/4 = 1/2



b) $P[(X - 0.5)^2 + (Y - 0.5)^2 \le 0.5^2]$

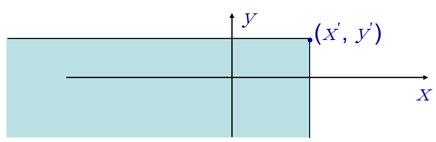
- Base area is half-circle
- Radius ½
- Base area = $\pi/8$
- Height = 2
- Volume = $\pi/4$



Joint CDF from Joint PDF

$F_{X,Y}(x',y') = P[\{X \le x', Y \le y'\}]$ is the probability that (X,Y) is in the shaded area

The volume under an infinite rectangle with northeast corner at (x',y')



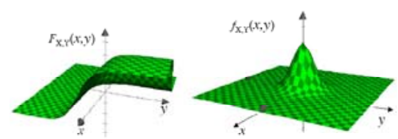
$$F_{X,Y}(x',y') = \int_{-\infty}^{x'}\int_{-\infty}^{y'} f_{X,Y}(x,y)\,dx\,dy$$

Joint PDF from Joint CDF

The joint PDF of jointly continuous RVs is the derivative of its joint CDF

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

assuming that $F_{X,Y}(x,y)$ is differentiable



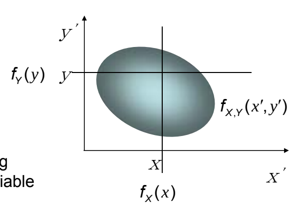
Marginal PDFs of X and Y

If X and Y are jointly continuous, then X and Y are continuous RVs with PDFs $f_X(x)$ and $f_Y(y)$ called the marginal PDFs. These can be obtained from the joint CDF $F_{X,Y}(x, y)$ by differentiation

$$F_X(x) = P(X \le x) = F_{X,Y}(x,\infty) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x',y')dy'dx'$$

$$f_X(x) = \frac{d}{dx}F_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y')dy'$$

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



Get marginal PDFs by integrating joint PDF over the unwanted variable

Example A

The joint PDF of X and Y is given by

$$f_{X,Y}(x,y) = c(1 - x - y), \quad 0 \le x \le 1 - y \le 1$$
$$= 0, \quad \text{otherwise}$$

a) What is the value of c?

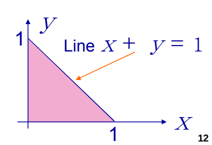
b) Find the marginal PDF of X

c) Determine $P[X < Y]$

a) PDF region $0 \le x \le 1 - y \le 1$

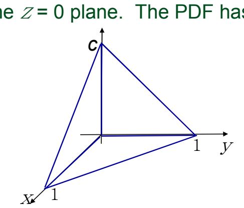
Both x and y must be between 0 and 1

$x \le 1 - y \quad \Rightarrow \quad x + y \le 1$



PDF Shape

$z = c(1 - x - y)$ is the equation of a plane in 3D space. The plane passes through the point (0, 0, c) and through the line $x + y = 1$ in the $z = 0$ plane. The PDF has the shape of a pyramid.



Finding c

Total volume under the PDF surface = 1

Volume of pyramid = base area × height / 3

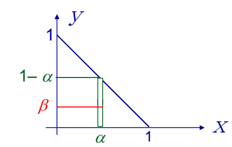
= (1/2×1×1) × c/3 = c/6 $\Rightarrow c = 6$

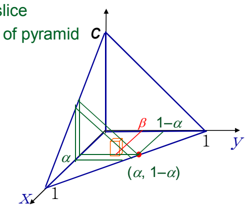
b) Find the marginal PDF of X

For any fixed α between 0 and 1, take an elemental volume at (α,β) as shown

– Vary β from 0 to $1-\alpha$ to get volume of slice

– Vary α from 0 to 1 to get entire volume of pyramid





$$F_{X,Y}(\infty,\infty) = 1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(\alpha,\beta)\,d\alpha\, d\beta$$

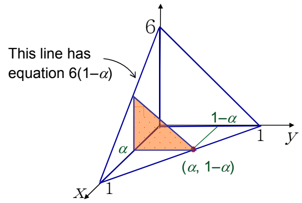
$$= \int_{\alpha=0}^{1}\int_{\beta=0}^{1-\alpha} c(1-\alpha-\beta)\,d\beta\, d\alpha$$

$$= \int_{\alpha=0}^{1}\frac{c(1-\alpha)^2}{2}\,d\alpha = \frac{c}{6}$$

c = 6, as obtained earlier

Marginal PDF

$$f_X(\alpha) = \int_{y=-\infty}^{\infty} f_{X,Y}(\alpha,y)\,dy = \text{area of cross section}$$



$f_X(\alpha)$ = area of triangle = base × altitude / 2

= $(1-\alpha)[6(1-\alpha)]/2 = 3(1-\alpha)^2$

Crosscheck

- $f_X(x) = 3(1-x)^2$ for $0 \le x \le 1$, and $f_X(x) = 0$ otherwise
- An important check on the correctness of the work done is the test that the marginal PDF is indeed a valid PDF
- Test failed? Your calculations are wrong
- Test passed? Your calculations may still be wrong, but they are not obviously incorrect

Marginal PDF of Y

•By symmetry,

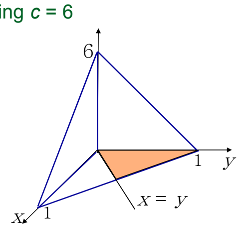
$f_Y(y) = 3(1-y)^2$ for $0 \le y \le 1$,

and $f_Y(y) = 0$, otherwise

c) Determining $P[X < Y]$

- $P\{X < Y\}$ = $P\{(X,Y) \in$ orange triangle$\}$

  = 1/2 by symmetry
- We could have written this down even if he had made an error in computing c = 6



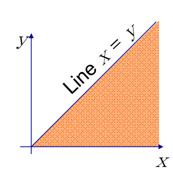
Example B

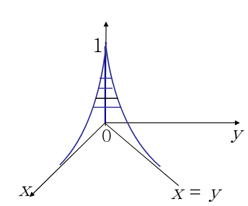
$f_{X,Y}(x,y) = \exp(-x), \quad 0 \le y \le x \le \infty$

$= 0, \quad$ otherwise

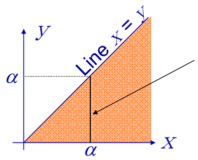
a) Find the marginal PDFs of X and Y

b) Find $P\{X + Y \le \alpha\}$ where $\alpha \ge 0$





a) Marginal PDF of *X*



joint pdf has constant value $\exp(-\alpha)$ on this line

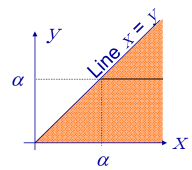
$$f_X(\alpha) = \int_{y=-\infty}^{\infty} f_{X,Y}(\alpha, y)dy = \text{area of cross section}$$

- Cross-section is a rectangle of base α and height $e^{-\alpha}$

$$f_X(\alpha) = \alpha e^{-\alpha}, \alpha \geq 0$$

➔ *X* = Erlang RV with parameters (n = 2, λ = 1)

Marginal PDF of *Y*



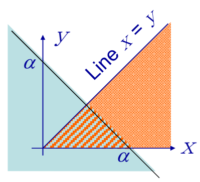
$$f_Y(a) = \int_{-\infty}^{\infty} f_{X,a}(x,y)dx \ (area\ of\ cross-section)$$
$$= \int_{a}^{\infty} e^{-x}dx = e^{-a}, a \geq 0$$

➔ *Y* = exponential RV with parameter λ = 1

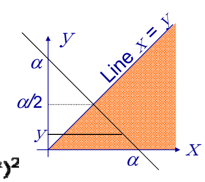
b) $P[X + Y \leq \alpha]$

$P\{X + Y \leq \alpha\}$ = integral of $f_{X,Y}(x,y)$ over shaded region

Easier if we integrate with respect to x first and then with respect to y



For any fixed value of y, $0 \leq y \leq \alpha/2$, x varies from y to $\alpha - y$

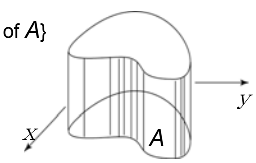


$$P(X+Y \leq \alpha) = \int_{y=0}^{\alpha/2}\int_{x=y}^{\alpha-y} e^{-x}dxdy$$
$$= 1 - 2e^{-\alpha/2} + e^{-\alpha} = (1-e^{-\alpha})^2$$

Uniformly-distributed RVs

The random point (*X*,*Y*) is said to be uniformly distributed over a region *A* if the joint PDF has constant value over *A* (and is zero elsewhere)

$f_{X,Y}(x,y)$ = constant = 1/{area of *A*}

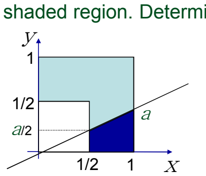


For any given region *B* in the plane

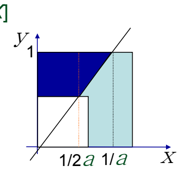
$P[(X,Y) \in B]$ = {area of $A \cap B$}/{area of *A*}

Example C

The random point (*X*,*Y*) is uniformly distributed over the shaded region. Determine $P[Y \leq aX]$



$a \leq 1$



$a \geq 1$

$f_{X,Y}(u,v)$ = 4/3 in shaded region

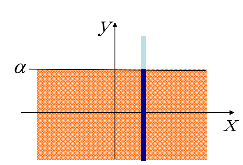
$P\{Y \leq aX\}$= (4/3)• blue area = (4/3) (3a/8) = a/2.

$P\{Y \leq aX\}$ = 1 – (4/3)•blue area = 1 – (4/3)•(3/8a) = 1 – 1/2a

Conditional PDF

X and *Y* are jointly continuous RVs, and $A = \{Y \leq \alpha\}$. Given that *A* occurred, the conditional PDF $f_{X|A}(x|A)$ is given by:

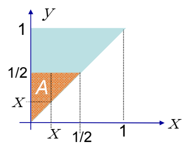
$$f_{X|A}(x|A) = \frac{P[\{x \leq X \leq x+\delta x\} \cap A]}{P[A]} = \frac{\int_{-\infty}^{\alpha} f_{X,Y}(x,y)dy\delta x}{\int_{-\infty}^{\alpha}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy}$$
$$= \frac{\text{volume above dark blue strip}}{\text{volume above orange area}}$$



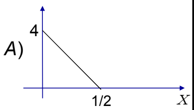
Example D

$f_{X,Y}(x,y) = 2$, for $0 < x < y < 1$
$= 0$, otherwise

$A = \{Y \le 1/2\}$. $P[A] = 1/4$; orange region



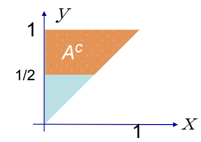
$$f_{X|A}(x|A) = \frac{\int_{-\infty}^{0.5} f_{X,Y}(x,y)dy}{P(A)} = \frac{\int_x^{0.5} 2dy}{P(A)} = \frac{2(0.5-x)}{1/4}, x \in (0, 0.5) = \begin{cases} 4(1-2x) & x \in (0, 0.5) \\ 0 & \text{otherwise} \end{cases}$$

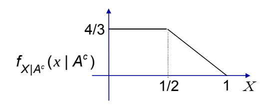


Conditional PDF given A^c

$A^c = \{Y > 1/2\}$. $P[A^c] = 3/4$; orange region

$$f_{X|A^c}(x|A^c) = \frac{\int_{0.5}^{1} f_{X,Y}(x,y)dy}{P(A^c)} = \begin{cases} \frac{2(0.5)}{0.75} & x \in (0, 0.5) \\ \frac{2(1-x)}{0.75} & x \in (0.5, 1) \\ 0 & \text{otherwise} \end{cases}$$

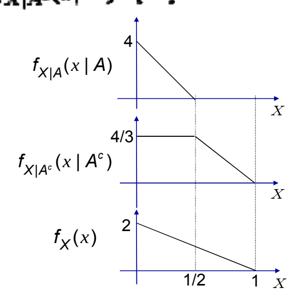




Total Probability Theorem

A and A^c are disjoint, collectively exhaustive

$$f_X(x) = f_{X|A}(x|A)P[A] + f_{X|A^c}(x|A^c)P[A^c]$$



Conditional PDF (cont.)

X and Y are jointly continuous RVs, and $A = \{Y = \alpha\}$. Given that A occurred, determine the conditional PDF $f_{X|A}(x|A)$.

Assume first that $A = \{\alpha \le Y \le \alpha + \Delta\alpha\}$

$$f_{X|A}(x|A)\Delta x = \frac{P[\{x \le X \le x + \Delta x\} \cap A]}{P(A)} \approx \frac{f_{X,Y}(x,\alpha)\Delta x \Delta\alpha}{f_Y(\alpha)\Delta\alpha} = \frac{f_{X,Y}(x,\alpha)}{f_Y(\alpha)}\Delta x$$

In the limit as $\Delta\alpha \to 0$, the event $A \to \{Y = \alpha\}$

$$f_{X|Y}(x|Y=\alpha) = \frac{f_{X,Y}(x,\alpha)}{f_Y(\alpha)} \text{ provided } f_Y(\alpha) > 0$$

It is a PDF over X with a dependence on the parameter α

Convention: if $f_Y(\alpha) = 0$, $f_{X|Y}(x|\alpha) = 0$

Total probability Theorem

Since $f_{X|Y}(x|\alpha) = \frac{f_{X,Y}(x,\alpha)}{f_Y(\alpha)}$ provided $f_Y(\alpha) > 0$

$$\Rightarrow f_{X,Y}(x,\alpha) = f_{X|Y}(x|\alpha) f_Y(\alpha)$$

and $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,\alpha)d\alpha$

we have $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|\alpha) f_Y(\alpha)d\alpha$ Total probability theorem!

Bayes' Rule

Combine conditional probability & total probability theorem

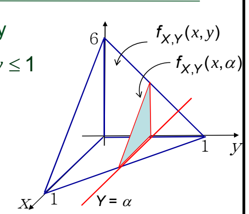
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx}$$

Example E

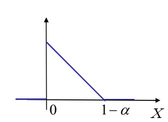
The joint PDF of X and Y is given by

$f_{X,Y}(x,y) = 6(1-x-y)$, $0 \le x \le 1-y \le 1$
$= 0$, otherwise

Determine $f_{X|Y}(x \mid Y = \alpha)$



$$f_{X|Y}(x \mid Y = \alpha) = \frac{f_{X,Y}(x,\alpha)}{f_Y(\alpha)}, \quad f_Y(\alpha) > 0$$
$$= \frac{6(1-x-\alpha)}{0.5[6(1-\alpha)][1-\alpha]}, \quad x < 1-\alpha$$



Scaling factor of $1/f_Y(\alpha)$ guarantees that area under function is 1

Independence

If X, Y are independent, then their joint CDF factors as a product of the marginals:

$$F_{X,Y}(x,y) = F_X(x)\, F_Y(y).$$

Differentiating,

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{X,Y}(x,y) \\ &= \frac{\partial}{\partial x}F_X(x)\frac{\partial}{\partial y}F_Y(y) \\ &= f_X(x) f_Y(y) \end{aligned}$$

so that the joint PDF also factors as a product of marginals

This holds if and only if $f_{Y|X}(y|x) = f_Y(y)$ is independent of x, i.e., knowing the value x of X tells us nothing more than we knew *a priori* about Y

Proof of necessity:

If $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$,
then $f_{Y|X}(y|x) = f_X(x)\, f_Y(y)/f_X(x) = f_Y(y)$ = independent of x

Proof of sufficiency:

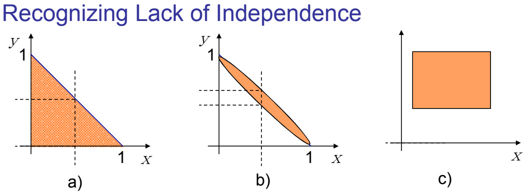
If $f_{Y|X}(y|x)$ is independent of x, then

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

$$= f_{Y|X}(y|x)\int_{-\infty}^{\infty} f_X(x)\,dx = f_{Y|X}(y|x) = f_{X,Y}(x,y)/f_X(x)$$

so that $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$,

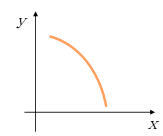
Recognizing Lack of Independence



- If a joint PDF is nonzero on the nonrectangular regions shown in a) or b), we can assert that X and Y are dependent RVs without any calculations
- Y can take on values between 0 and 1. But, if $X = 1/2$, Y can take on values over a smaller range only! DEPENDENT!!
- If that region is rectangular, we still have to check the factorization test

Fully dependent RVs

$Y = g(X)$



- Probability mass is spread along the curve $y = g(x)$
- X and Y are not jointly continuous RVs
- All questions involving the probabilistic behavior of the random point (X,Y) can be translated into questions involving X alone (there is only one degree of randomness)