CL205: AI & DS

MB

Mani Bhushan,
Department of Chemical Engineering,
Indian Institute of Technology Bombay
Mumbai, India- 400076

mbhushan@iitb.ac.in
Ack: Prof. Sachin Patwardhan

Autumn 2024

This handout

Cumulative or probability distribution function: CDF
Probability density function: PDF

- Multiple random variables
- Joint, marginal, conditional distribution and density functions
- Independence

Parallel Notions of Events

- Multiple random variables $\equiv$ Multiple events
- Joint distribution/density functions $\equiv$ Probability of intersection of events
- Conditional distribution/density functions $\equiv$ Conditional probability
- Independent random variables $\equiv$ independent events

Slight Change of Notation

- Random variable denoted by upper-case letter (earlier it was bold-face letter): X (earlier was $\boldsymbol{x}$)
- Lower case english letter (earlier it was greek letter while writing density or distribution functions: x (earlier was ξ).
- For example: $F_X(x) = P(\{\omega : X(\omega) \leq x\})$
- ξ, ρ, etc. will be used as variables of integration.

Extension of Ideas:

- Multiple (Multivariate) Random Variables: Jointly distributed random variables
- Event ω occurs in sample space Ω. Associate many, $X_1, X_2, ..., X_n$, random variables with ω.
- Each random variable is a valid mapping from Ω to $\mathbb{R}$.

Bivariate Random Variables

- For simplicity of notation consider two random variables: X, Y.
- Special case of multiple random variables.
- Examples:
  - Average number of cigarettes smoked daily and the age at which an individual gets cancer,
  - Height and weight of an individual,
  - Height and IQ of an individual.
  - Flow-rate and pressure drop of a liquid flowing through a pipe.
  - Number of heads and number of tails in an experiment involving toss of several coins.

Jointly distributed random variables

Often interested in answering questions on X, Y taking values in a specified region D in $\mathbb{R}^2$ (xy plane).

- The distribution functions $F_X(x)$ and $F_Y(y)$ of X and Y determine their individual probabilities but not their joint probabilities. The probability of event

$$
\begin{aligned}
&\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\} \\
&= \{\omega : X(\omega) \leq x, Y(\omega) \leq y\}
\end{aligned}
$$

  cannot be expressed in terms of $F_X(x)$ and $F_Y(y)$.

- Joint probabilities of X, Y completely determined if probability of above event known for every x and y.

Joint Probability Distribution Function or Joint Cumulative Distribution Function

For random variables (discrete or continuous) X, Y, the joint (bivariate) probability distribution function is:

$$F_{X,Y}(x, y) = P(\{X \leq x, Y \leq y\})$$

RHS to be interpreted as: $P(\{\omega : \; X(\omega) \leq x, Y(\omega) \leq y\})$,
where x, y are two arbitrary real numbers.
Often, the subscript X, Y omitted.

Properties of Joint Probability Distribution Function (Papoulis and Pillai, 2002)

1. $F(-\infty, y) = F(x, -\infty) = 0, \quad F(\infty, \infty) = 1.$
2. $$P(x_1 < X \le x_2, Y \le y) = F(x_2, y) - F(x_1, y)$$
   $$P(X \le x, y_1 < Y \le y_2) = F(x, y_2) - F(x, y_1)$$
3. $$P(x_1 < X \le x_2, y_1 < Y \le y_2) =$$
   $$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$$

Joint Density Function: I

- The joint density of X and Y is the function (defn.)

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

  [When both X, Y are continuous random variables]

- It follows that,

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(\xi, \rho) d\xi d\rho$$

$$P((X, Y) \in D) = \iint_{D} f(x, y) dx dy$$

Joint Density Function: II

- In particular, as $\Delta x \to 0$ and $\Delta y \to 0$,

$$P(x < X \leq x + \Delta x, y < Y \leq y + \Delta y) \approx f(x, y)\Delta x \Delta y$$

- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$; $f(x, y) \geq 0 \ \forall x, y \in \mathbb{R}$.

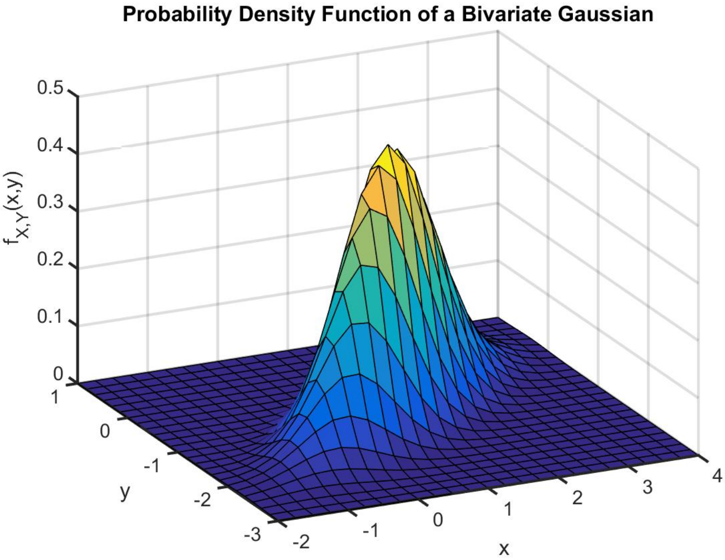
Joint Density Example: Bivariate Gaussian Random Variable

$$f(x, y) = \alpha \exp(-0.5(\xi - \mu)^T P^{-1}(\xi - \mu))$$

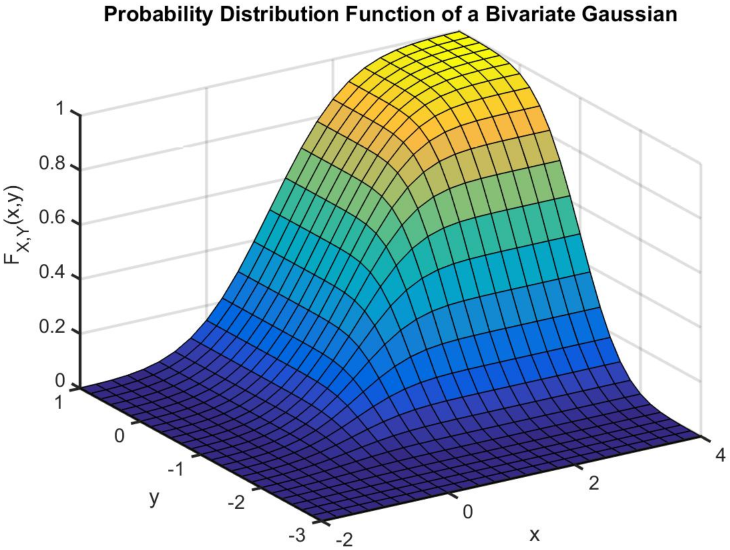
with

$$\xi = \begin{bmatrix} x \\ y \end{bmatrix}, \ \mu = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \ P = \begin{bmatrix} 0.9 & 0.4 \\ 0.4 & 0.3 \end{bmatrix},$$
$$\alpha = \frac{1}{2\pi\sqrt{|P|}}$$

Joint Density Visualization



Joint Distribution Visualization



Marginal Distribution or Density Functions of Individual Random Variables: I

- Marginal Probability Distribution Functions: $F_X(x), F_Y(y)$:
  - Extract $F_X(x)$ from $F(x,y)$ as:

$$\begin{aligned} F_X(x) &= P(X \leq x) = \\ &= P(X \leq x, Y < \infty) = F(x, \infty) \end{aligned}$$

  - Similarly, extract $F_Y(y)$ as:

$$F_Y(y) = P(Y \leq y) = P(X < \infty, Y \leq y) = F(\infty, y)$$

Marginal Distribution or Density Functions of Individual Random Variables: II

- Marginal Probability Density Functions: $f_X(x), f_Y(y)$:
  - Extract these from $f(x, y)$ as:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Marginal Probability Density

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

Makes sense, since

$$\begin{aligned} P(X \in A) &= P(X \in A, Y \in (-\infty, \infty)) \\ &= \int_A \int_{-\infty}^{\infty} f(x,y)dydx = \int_A f_X(x)dx \end{aligned}$$

where $f_X(x)$ is as defined above.
Similarly,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

Example 4.3c from Ross: I

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & 0 < x < \infty, \quad 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

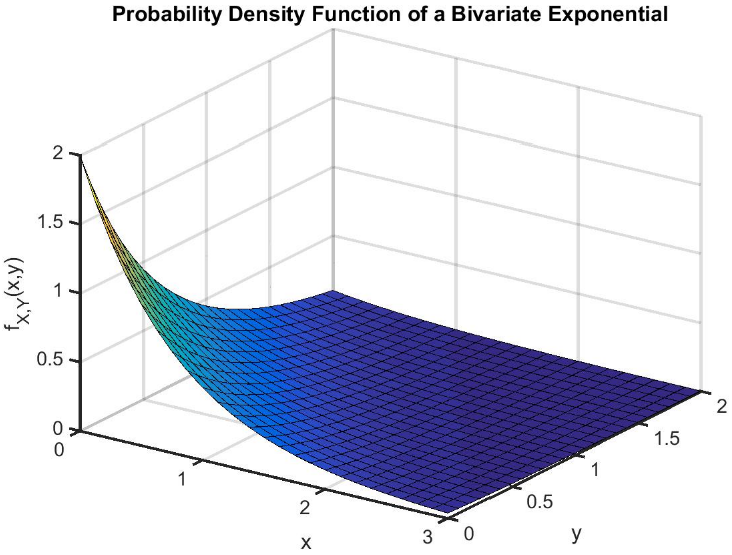
Compute: (a) $P(X > 1, Y < 1)$, (b) $P(X < Y)$, (c) $P(X < a)$

$$\begin{aligned} P(X > 1, Y < 1) &= \int_0^1 \int_1^\infty 2e^{-x}e^{-2y}\,dxdy \\ &= \int_0^1 2e^{-2y}(-e^{-x}\ |_1^\infty)dy \\ &= e^{-1}\int_0^1 2e^{-2y}dy = e^{-1}(1-e^{-2}) \end{aligned}$$

Example 4.3c from Ross: II

$$
\begin{aligned}
P(X < Y) &= \int_0^\infty \int_0^y 2e^{-x} e^{-2y} dxdy = 1/3 \\
P(X < a) &= \int_0^a \int_0^\infty 2e^{-x} e^{-2y} dydx = 1 - e^{-a}
\end{aligned}
$$

Joint Density Visualization: Exponential

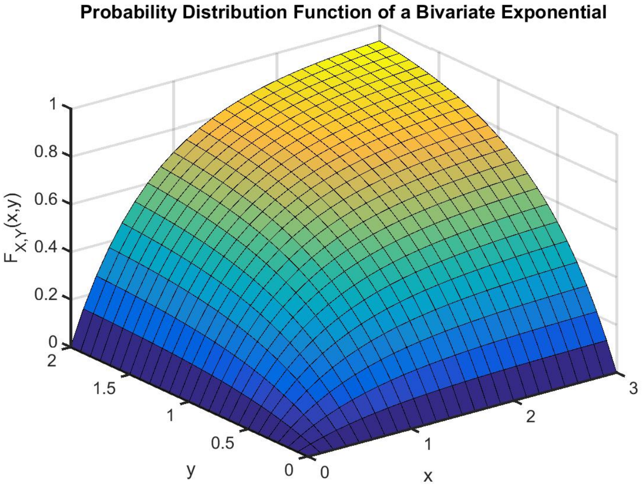


Example: Joint Distribution Function

For this example:

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(\xi,\rho)\,d\xi\,d\rho$$
$$= \begin{cases} (1-e^{-x})(1-e^{-2y}), & x>0, y>0 \\ 0 & \text{otherwise} \end{cases}$$

Joint Distribution Visualization: Exponential



Joint Probability Mass Function (PMF)

- Given two discrete random variables X and Y in the same experiment, the joint PMF of X and Y is

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

  for all pairs of (x_i, y_j) values that X and Y can take. $p(x_i, y_j)$ also denoted as $p_{X,Y}(x_i, y_j)$.
- The marginal probability mass functions for X and Y are

$$\begin{aligned} p_X(x) = P(X = x) &= \sum_{y} p_{X,Y}(x, y) \\ p_Y(y) = P(Y = y) &= \sum_{x} p_{X,Y}(x, y) \end{aligned}$$

Computation of Marginal PMF from Joint PMF: I

Formally:

$$\{X = x_i\} = \bigcup_j \{X = x_i, Y = y_j\}$$

All events on RHS are mutually exclusive. Thus,

$$\begin{aligned} p_X(x_i) = P(X = x_i) &= \sum_j P(X = x_i, Y = y_j) \\ &= \sum_j p(x_i, y_j) \end{aligned}$$

Computation of Marginal PMF from Joint PMF: II

Similarly,
$p_Y(y_j) = P(Y = y_j) = \sum_i p(x_i, y_j)$.

Note: $P(X = x_i, Y = y_j)$ cannot be constructed from knowledge of $P(X = x_i)$ and $P(Y = y_j)$.

Example: 4.3a, Ross

3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. Let X, Y denote the number of new, and used but working batteries that are chosen, respectively. Find $p(x_i, y_j) = P(X = x_i, Y = y_j)$.
Solution: Let $T =^{12} C_3$
$p(0,0) = (^5C_3)/T$
$p(0,1) = (^4C_1)(^5C_2)/T$
$p(0,2) = (^4C_2)(^5C_1)/T$
$p(0,3) = (^4C_3)/T$
$p(1,0) = (^3C_1)(^5C_2)/T$
$p(1,1) = (^3C_1)(^4C_1)(^5C_1)/T$
$p(1,2) = ...$, $p(2,0) = ...$, $p(2,1) = ...$, $p(3,0) = ...$

Tabular Form

	0	1	2	3	Row Sum ($P(X = i)$)
0	10/220	40/220	30/220	4/220	84/220
1	30/220	60/220	18/220	0	108/220
2	15/220	12/220	0	0	27/220
3	1/220	0	0	0	1/220
Col sum ($P(Y = j)$)	56/220	112/220	48/220	4/220	

i represents row and j represents column:

Both row and column sums add upto 1.

Marginal probabilities in the margins of the table.

n Random Variables: I

- Joint cumulative probability distribution function $F(x_1, x_2, ..., x_n)$ of n random variables $X_1, X_2, ..., X_n$ is defined as:

$$F(x_1, x_2, ..., x_n) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n)$$

- If random vars. discrete: joint probability mass function

$$p(x_1, x_2, ..., x_n) = P(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$$

n Random Variables: II

- If random vars. continuous: joint probability density function $f(x_1, x_2, ..., x_n)$ such that for any set C in n-dimensional space

$$P((X_1, X_2, ..., X_n) \in C) =$$
$$\underset{(x_1,...,x_n)\in C}{\int\int \dots \int} f(x_1, x_2, ..., x_n) dx_1 dx_2 ... dx_n$$

where,

$$f(x_1, x_2, ..., x_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

Obtaining Marginals

$$
\begin{aligned}
F_{X_1}(x_1) &= F(x_1, \infty, \infty, \ldots, \infty) \\
f_{X_1}(x_1) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)dx_2 dx_3 ... dx_n \\
p_{X_1}(x_1) &= \sum_{x_2}\sum_{x_3}\cdots\sum_{x_n} p(x_1, x_2, \ldots, x_n)
\end{aligned}
$$

Independence of Random Variables: I

- Random variables X and Y are independent if for any two sets of real numbers A and B:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

  i.e. events $E_A = \{X \in A\}$ and $E_B = \{Y \in B\}$ are independent.
- Height and IQ
- In particular: $P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b)$, or
- In terms of joint cumulative distribution function F of X and Y:

$$F(a, b) = F_X(a)F_Y(b); \quad \forall a, b \in \mathbb{R}$$

Independence of Random Variables: II

Random variables that are not independent are called dependent

Independence: Probability Mass and Density Functions

Random variables X, Y independent if:

- Discrete random variables: Probability mass function

$$p(x_i, y_j) = p_X(x_i)p_Y(y_j) \;\; \text{for all} \;\; x_i, y_j$$

- Continuous random variables: Probability density function

$$f(x, y) = f_X(x)f_Y(y) \;\; \text{for all} \;\; x, y$$

Independence: Equivalent Statements

1) $P(X \in A, Y \in B) = P(X \in A)P(Y \in B);$
   $\forall A, B$ sets in $\mathbb{R}$
2) $F(x, y) = F_X(x)F_Y(y); \;\; \forall x, y$
3) $f(x, y) = f_X(x)f_Y(y); \;\; \forall x, y;$ continuous RVs
3) $p(x_i, y_j) = p_X(x_i)p_Y(y_j); \;\; \forall x_i, y_j;$ discrete RVs

Example 5.2 (Ogunnaike, 2009): I

The reliability of the temperature control system for a commercial, highly exothermic polymer reactor is known to depend on the lifetimes (in years) of the control hardware electronics, X_1, and of the control valve on the cooling water line, X_2. If one component fails, the entire control system fails. The random phenomenon in question is characterized by the two-dimensional random variable (X_1, X_2) whose joint probability distribution is given as:

$$f(x_1, x_2) = \begin{cases} \frac{1}{50} e^{-(0.2x_1 + 0.1x_2)}, & 0 < x_1 < \infty,\ 0 < x_2 < \infty \\ 0, & \text{otherwise} \end{cases}$$

Example 5.2 (Ogunnaike, 2009): II

1. Establish that above is a legitimate joint probability density function, To show: $\int_0^\infty \int_0^\infty f(x_1, x_2) dx_1 dx_2 = 1$.

$$\int_0^\infty \int_0^\infty \frac{1}{50} e^{-(0.2x_1 + 0.1x_2)} dx_1 dx_2$$
$$= \frac{1}{50}(-5e^{-0.2x_1}|_0^\infty)(-10e^{-0.1x_2}|_0^\infty) = 1$$

Example 5.2 (Ogunnaike, 2009): III

1. What's the probability of the system lasting more than 2 years. To find: $P(X_1 > 2, X_2 > 2) = \int_2^\infty \int_2^\infty \frac{1}{50} e^{-(0.2x_1+0.1x_2)} dx_1 dx_2 = 0.549.$
2. Find marginal density function of X_1.

$$f_{X_1}(x_1) = \int_0^\infty \frac{1}{50} e^{-(0.2x_1+0.1x_2)} dx_2 = \frac{1}{5} e^{-(0.2x_1)}$$

3. Find marginal density function of X_2?

$$f_{X_2}(x_2) = \int_0^\infty \frac{1}{50} e^{-(0.2x_1+0.1x_2)} dx_1 = \frac{1}{10} e^{-(0.1x_2)}$$

Example 5.2 (Ogunnaike, 2009): IV

4 Are X_1, X_2 independent? Yes, since $f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2)$.

Independence of n Random Variables: I

Random variables $X_1, X_2, ..., X_n$ are said to be independent if

- For all sets of real numbers $A_1, A_2, ..., A_n$:

$$P(X_1 \in A_1, X_2 \in A_2, ..., X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$

- In particular: $\forall a_1, a_2, ..., a_n \in \mathbb{R}$

$$\begin{aligned} P(X_1 \le a_1, X_2 \le a_2, ..., X_n \le a_n) &= \prod_{i=1}^{n} P(X_i \le a_i), \\ \text{or,} \quad F(a_1, a_2, ..., a_n) &= \prod_{i=1}^{n} F_{X_i}(a_i) \end{aligned}$$

Independence of n Random Variables: II

- For discrete random variables: probability mass function factorizes:

$$p(x_1, x_2, ..., x_n) = p_{X1}(x_1)p_{X2}(x_2)...p_{X_n}(x_n)$$

- For continuous random variables: probability density function factorizes:

$$f(x_1, x_2, ..., x_n) = f_{X_1}(x_1)f_{X_2}(x_2)...f_{X_n}(x_n)$$

Independent, Repeated Trials

- In statistics, one usually does not consider just a single experiment, but that the same experiment is performed several times.
- Associate a separate random variable with each of those experimental outcomes.
- If the experiments are independent of each other, then we get a set of independent random variables.
- Example: Tossing a coin n times. Random variable X_i is the outcome (0 or 1) in the i^{th} toss.

Independent and Identically Distributed (IID) Variables: I

A collection of random variables is said to be IID if

- The variables are independent
- The variables have the same probability distribution
- Example 1: Tossing a coin n times. The probability of obtaining a head in a single toss does not vary and all the tosses are independent.
  - Each toss leads to a random variable with the same probability distribution function. The random variables are also independent. Thus, IID.

Independent and Identically Distributed (IID) Variables: II

- Example 2: Measuring temperature of a beaker at n time instances in the day. The true water temperature changes throughout the day. The sensor is noisy.
  - Each sensor reading leads to a random variable.
  - Variables are independent but not identically distributed.

Conditional Distributions

Remember for two events A and B: conditional probability of A given B is:

$$P(A \mid B) = \frac{P(A, B)}{P(B)}$$

for $P(B) > 0$.

Conditional Probability Mass Function

- For X, Y discrete random variables, define the conditional probability mass function of X given $Y = y$ by

$$\begin{aligned} p_{X|Y}(x|y) &= P(X = x \mid Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for $p_Y(y) > 0$.

Examples 4.3b,f from Ross: I

Question: In a community, 15% families have no children, 20% have 1, 35% have 2 and 30% have 3 children. Each child is equally likely to be a boy or girl. We choose a family at random. Given that the chosen family has one girl, compute the probability mass function of the number of boys in the family?

G: number of girls, B: number of boys, C: number of children

To find: $P(B = i|G = 1), \; i = 0, 1, 2, 3.$

$$P(B = i|G = 1) = \frac{P(B = i, G = 1)}{P(G = i)}, \; i = 0, 1, 2, 3$$

Examples 4.3b,f from Ross: II

First find $P(G = 1)$

$$\begin{aligned}\{G = 1\} &= \{G = 1\} \cap (\{C = 0\} \cup \{C = 1\} \cup \{C = 2\} \\ &\quad \cup \{C = 3\}) \\ P(G = 1) &= P(G = 1, C = 0) + P(G = 1, C = 1) \\ &\quad + P(G = 1, C = 2) + P(G = 1, C = 3)\end{aligned}$$

since $C = 0, C = 1, C = 2, C = 3$ are mutually exclusive events with union as Ω.
Then,

$$\begin{aligned}P(G = 1) &= P(G = 1 \mid C = 0)P(C = 0) \\ &\quad + P(G = 1 \mid C = 1)P(C = 1) + ... \\ &= 0 + (1/2) \times 0.2 + ... = 0.3875\end{aligned}$$

Examples 4.3b,f from Ross: III

Then,

$$P(B = 0 \mid G = 1) = \frac{P(B = 0, G = 1)}{P(G = 1)}$$

Numerator
$= P(G = 1 and C = 1) = P(G = 1 \mid C = 1)P(C = 1) = (1/2)0.2 = 0.1.$
Then,

$$P(B = 0 \mid G = 1) = 0.1/0.3875 = 8/31$$

Similarly: $P(B = 1 \mid G = 1) = 14/31, P(B = 2 \mid G = 1) = 9/31, P(B = 3 \mid G = 1) = 0$. Check: Sum of conditional probabilities is 1.

Conditional Probability Density Function

For Random Variables X, Y, conditional probability density of X given that $Y = y$ is defined as:

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

for $f_Y(y) > 0$.
Hence, can make statements on probabilities of X taking values in some set A given the value obtained by Y as:

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y) dx$$

Independence and Conditional Probabilities

If X, Y are independent, then

$$p_{X|Y}(x|y) = p_X(x)$$
$$f_{X|Y}(x|y) = f_X(x)$$

Temperature Control Example (Continued), Example 5.2 (Ogunnaike, 2009) Earlier

1. Find Conditional density function: $f_{X_1|X_2}(x_1|x_2)$.

$$f(x_1, x_2)/f_{X_2}(x_2) = \frac{1}{5}e^{-0.2x_1}$$

   which is same as $f_{X_1}(x_1)$ in this example.

2. Similarly, $f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2)$ in this example.
   Generic Question: If $f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1)$, then is $f_{X_2|X_1}(x_2|x_1) = f_{X_2}(x_2)$?
   Answer: Yes

Example 5.5 (Ogunnaike, 2009): I

$$f_{X_1,X_2} = \begin{cases} x_1 - x_2, & 1 < x_1 < 2, 0 < x_2 < 1 \\ 0, & \text{otherwise} \end{cases}$$

Find: Conditional probability densities.
Answer: Compute marginals

$$f_{X_1}(x_1) = \begin{cases} (x_1 - 0.5), & 1 < x_1 < 2 \\ 0, & \text{otherwise} \end{cases}$$

$$f_{X_2}(x_2) = \begin{cases} (1.5 - x_2), & 0 < x_2 < 1 \\ 0, & \text{otherwise} \end{cases}$$

Example 5.5 (Ogunnaike, 2009): II

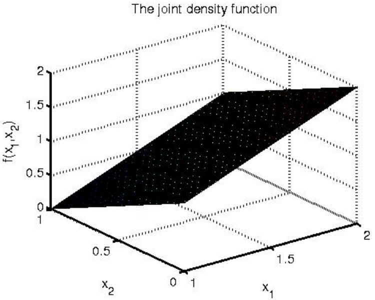
Then compute conditionals

$$f_{X_1|X_2}(x_1|x_2) = \frac{(x_1 - x_2)}{(1.5 - x_2)}, \quad 1 < x_1 < 2$$

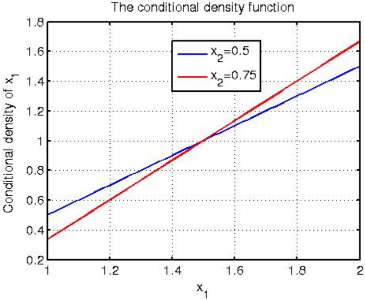
$$f_{X_2|X_1}(x_2|x_1) = \frac{(x_1 - x_2)}{(x_1 - 0.5)}, \quad 0 < x_2 < 1$$

The random variables X_1, X_2 are not independent.

Plots



The joint density function
2
1.5
1
0.5
0
f(x1,x2)
1
0.5
0
x2
1
1.5
2
x1



The conditional density function
x2=0.5
x2=0.75
Conditional density of x1
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
1
1.2
1.4
1.6
1.8
2
x1

Independence of Transformations

If random variables X, Y are independent, then the random variables

$$Z = g(X), U = h(Y)$$

are also independent.
Proof: Let A_z denote the set of points on the x-axis such that $g(x) \leq z$ and B_u denote the set of points on the y-axis such that $h(y) \leq u$. Then,

$$\{Z \leq z\} = \{X \in A_z\}; \;\; \{U \leq u\} = \{Y \in B_u\}$$

Thus, the events $\{Z \leq z\}$ and $\{U \leq u\}$ are independent because events $\{X \in A_z\}$ and $\{Y \in B_u\}$ are independent.

THANK YOU