A Review of Applied Probability Theory

Joint Random Variables

Dr. Bahman Honari

Autumn 2022

Joint Random Variables

- In this topic, we explain the theory for distributions of multiple random variables, called joint probability distributions.
- Foe better understanding, we'll start by an example focused on "Joint Frequency Table" before talking about "Joint probability Function".

Joint Random Variables (cont.)

- A survey reports the number of sons and daughters of 100 married couples. The results are summarized in the following table.

	B_0	B_1	B_2	Total
G_0	6	8	8	22
G_1	9	17	10	36
G_2	15	15	12	42
Total	30	40	30	100

- Here, G_0 = no daughters, G_1 = 1 daughter, G_2 = 2 or more daughters, B_0 = no sons, B_1 = 1 son, B_2 = 2 or more sons.

Joint Random Variables (cont.)

- Let X_1 be the number of daughters of a randomly chosen couple, and let X_2 be the number of sons.
- Then we can define the joint probability distribution for this pair of discrete random variables, namely:

$$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

- Therefore we can find

$$\begin{aligned} P(X_1 + X_2 \geq 2) &= P(G_2 \cup B_2 \cup (G_1 \cap B_1)) = P(G_2 \cup B_2) + P(G_1 \cap B_1) \\ &= P(G_2) + P(B_2) - P(G_2 \cap B_2) + P(G_1 \cap B_1) \\ &= \frac{42 + 30 - 12 + 17}{100} = \frac{77}{100} = 0.77 \end{aligned}$$

Joint Random Variables (cont.)

- Once again, the requirement that the sample space has probability 1 is equivalent to the

$$\sum_{\text{all } x_1} \sum_{\text{all } x_2} f(x_1, x_2) = 1$$

- In summary for discrete variables

Joint probability distribution	
$f(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$	$f(x_1, x_2) \geq 0$ for all x_1, x_2
$P(X_1, X_2 \in A) = \sum_{(x_1, x_2) \in A} \sum f(x_1, x_2)$	$\sum_{\text{all } x_1} \sum_{\text{all } x_2} f(x_1, x_2) = 1$

Joint Random Variables (cont.) – Marginal Distributions

- From the joint probability distribution, one can extract the probability distribution for any single random variable.
- Such a probability distribution is called a marginal distribution, and to keep the notation clear, we use a subscript to indicate which random variable is being retained.
- For example, the marginal distribution of X_1 is

$$f_1(x_1) = P(X_1 = x_1)$$

- The meaning of $f_1(x_1)$ is that it gives the probability of different values of X_1 regardless of the value of X_2 . For example,

$$f_1(1) = \mathrm{P}(X_1 = 1) = 0.36$$

Joint Random Variables (cont.) – Marginal Distributions

- The values of the marginal distributions $f_1(x_1)$ and $f_2(x_2)$ are obtained from the row and column totals, respectively, which are contained in the margins of the frequency table in slided 3.
- We obtain the marginal distribution $f_2(x_2)$, for example, by summing the joint distribution $f(x_1, x_2)$ over all possible values of x_1, and vice versa, to obtain $f_1(x_1)$, i.e.,

$$f_1(x_1) = \sum_{\text{all } x_2} f(x_1, x_2), \quad f_2(x_2) = \sum_{\text{all } x_1} f(x_1, x_2)$$

Joint Random Variables (cont.) – Marginal Distributions

- Thus, we have

$$f_1(x_1) = \begin{cases} 0.22 & x_1 = 0 \\ 0.36, & x_1 = 1 \\ 0.42, & x_1 = 2 \end{cases} \qquad f_2(x_2) = \begin{cases} 0.3 & x_2 = 0 \\ 0.4, & x_2 = 1 \\ 0.3, & x_2 = 2 \end{cases}$$

Joint Random Variables (cont.) – Marginal Distributions

- Now it is easy to generalize to the case of n discrete random variables $X_1, X_2, \dots, X_n$. The joint distribution function $f(x_1, x_2, \dots, x_n)$ is defined as follows:

$$f(x_1, x_2, \cdots, x_n) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

- The total probability is then

$$\sum_{\text{all } x_1} \cdots \sum_{\text{all } x_n} f(x_1, \cdots, x_n) = 1$$

Joint Random Variables (cont.) – Marginal Distributions

- The notation introduced for marginal distributions extends easily to arbitrary numbers of random variables; for example,

$$f_2(x_2) = \sum_{\text{all } x_1} \sum_{\text{all } x_3} \cdots \sum_{\text{all } x_n} f(x_1, x_2, \cdots, x_n)$$

- More complex marginal distributions are possible when there are more than 2 random variables. For example,

$$f_{2,4}(x_2, x_4) = \sum_{\text{all } x_1} \sum_{\text{all } x_3} f(x_1, x_2, x_3, x_4)$$

Joint Random Variables (cont.) – Continuous Distributions

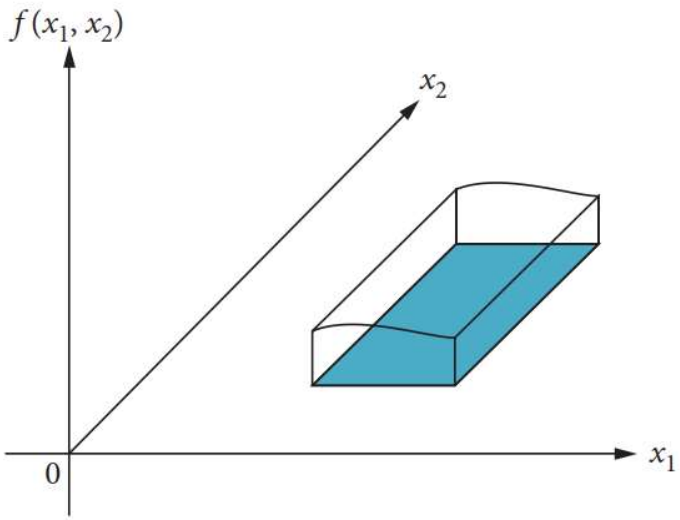
- For the case of multiple continuous random variables, the notation is the same as in the discrete case (just as it has been all along).
- However, the values of the function $f(x_1, x_2, \dots, x_n)$ now give probability densities.
- For n = 2, the meaning of the joint pdf $f(x_1, x_2)$ is as follows: For any real numbers a, b, c, d with $a \le b$ and $c \le d$

$$P(a \le X_1 \le b, c \le X_2 \le d) = \int_c^d \int_a^b f(x_1, x_2)\, dx_1 dx_2$$

Joint Random Variables (cont.) – Continuous Distributions

$f(x_1, x_2) \geq 0$ for all x_1, x_2 and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2)\,dx_1\,dx_2 = 1$$



Joint Random Variables (cont.) – Continuous Distributions

- For n continuous random variables $X_1, X_2, \ldots, X_n$ and constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$, the joint pdf is defined as

$$P(a_1 \le X_1 \le b_1, \cdots, a_n \le X_n \le b_n) = \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

- It is evident that the probability of any event is determined from an integral of the joint pdf over a region of the (x_1, x_2) plane.

$$P((X_1, X_2) \in A) = \iint_A f(x_1, x_2)\, dx_1\, dx_2$$

Joint Random Variables (cont.) – Continuous Distributions

- In direct analogy with the discrete case, we obtain the marginal pdf $f_1(x_1)$ by integrating the joint pdf over all possible values of x_2, and similarly for $f_1(x_1)$

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_2, \qquad f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx_1$$

- It is evident that the probability of any event is determined from an integral of the joint pdf over a region of the (x_1, x_2) plane.

Joint Random Variables (cont.) – Continuous Distributions

- Another useful function to generalize to the case of n random variables is the cumulative distribution function. The RVs $X_1, X_2, \dots, X_n$ have a joint cdf, defined as follows:

$$F(x_1, \cdots, x_n) = P(X_1 \le x_1, \cdots, X_n \le x_n)$$

- The marginal cdf for X_i is denoted as follows:

$$F_i(x_i) = P(X_i \le x_i), \qquad i = 1, \cdots, n$$

Independence

- Two random variables X_1 and X_2 are independent if for all possible values
- of x_1 and x_2

$$f(x_1, x_2) = f_1(x_1) f_2(x_2)$$

where $f(x_1, x_2)$ is the joint distribution function (or joint pdf) of X_1 and X_2 and $f_1(x_1)$ and $f_2(x_2)$ are the probability functions for X_1 and X_2 , respectively.

Independence – Example for Discrete RVs

Consider a joint distribution with random variables X_1 and X_2, given by the following table

$f(x_1, x_2)$		x_2: 0	10	20
x_1	5	0.22	0.10	0.10
	10	0.10	0.08	0.12
	20	0.05	0.05	0.18

We require, for their independence, that $f(x_1, x_2) = f_2(x_2)f_2(x_2)$ for all possible values of (x_1, x_2). However,

$$f_1(10)f_2(10) = (0.3) \cdot (0.23) = 0.069 \neq 0.08 = f(10, 10)$$

Therefore X_1 and X_2 are not independent.

Independence (Cont.)

An equivalent definition of independence is the following: X_1 and X_2 are independent if and only if, for all constants a, b, c, d with $a \leq b$ and $c \leq d$

$$P(a \leq X_1 \leq b, c \leq X_2 \leq d) = P(a \leq X_1 \leq b)P(c \leq X_2 \leq d)$$

We can also show that $X_1, X_2, \dots, X_n$ are independent if and only if

$$F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n)$$

Expectations

- The definition of expectation extends easily to joint distributions. Consider the case of n random variables $X_1, X_2, \ldots, X_n$ having a joint distribution $f(x_1, x_2, \ldots, x_n)$ and an arbitrary function $h(x_1, x_2, \ldots, x_n)$. Then

$$E(h(X_1, \cdots, X_n)) = \begin{cases} \sum_{x_1} \cdots \sum_{x_n} h(x_1, \cdots, x_n) f(x_1, \cdots, x_n) & \text{discrete} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \cdots, x_n) f(x_1 \cdots x_n) dx_1 \cdots dx_n & \text{continuous} \end{cases}$$

- In particular we have:

$$\text{discrete: } E[XY] = \sum_y \sum_x xy\, f(x,y) \qquad \text{continuous: } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy\, f(x,y) dx dy$$

Expectations (Cont.) – Moment Generating Function

- The joint moment generating function of RVs X and Y is defined as:

$$m_{X,Y}(t_1,t_2) = E[e^{t_1X+\,t_2Y}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+\,t_2y} f(x,y)dxdy$$

When X and Y are independent we have

$$m_{X,Y}(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x+\,t_2y} f(x)f(y)dxdy = \int_{-\infty}^{\infty} e^{t_1x} f(x)dx \int_{-\infty}^{\infty} e^{t_2y} f(y)dy = m_X(t_1)\,.m_Y\,(t_2)$$

Furthermore, if $Z = X + Y$, we have

$$m_Z(t) = E[e^{tZ}] = E\left[e^{t(X+Y)}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t(x+y)} f(x)f(y)dxdy = \int_{-\infty}^{\infty} e^{tx} f(x)dx \int_{-\infty}^{\infty} e^{ty} f(y)dy = m_X(t)\,.m_Y\,(t)$$

Expectations (Cont.) – Moment Generating Function

- Similar to the MGF for probability distributions in one dimension we can write:

$$E[X^m Y^n] = \left.\frac{\partial^{m+n} m_{X,Y}(t_1, t_2)}{\partial^m t_1 \partial^n t_2}\right|_{t_1 = t_2 = 0}$$

For instance

$$E[XY] = \left.\frac{\partial^2 m_{X,Y}(t_1, t_2)}{\partial t_1 \partial t_2}\right|_{t_1 = t_2 = 0}$$

Covariance and Correlation

- Covariance of random variables X and Y is defined as

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

and it has the properties listed below:

- $Cov(X,Y) = Cov(Y,X)$
- $Cov(X,X) = Var[X]$
- $Cov(X,c) = 0$
- $Cov(aX,Y) = aCov(X,Y)$
- $Cov(X_1 \pm X_2, Y) = Cov(X_1, Y) \pm Cov(X_2, Y)$
- $obviously\ combination\ of\ all\ above$
- $-\sqrt{Var[X]Var[Y]} \le Cov(X,Y) \le +\sqrt{Var[X]Var[Y]}$

Covariance and Correlation

- Correlation coefficient of random variables X and Y is defined as

$$Corr(X,Y) = \rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var[X]Var[Y]}}$$

that according to the last property of $Cov(X,Y)$ in last slide we have

$$-1 \leq Corr(X,Y) \leq 1$$

The sign and magnitude of the $Corr(X,Y)$ explains the direction and strength of relationship between random variables X and $Y.$

X and Y; Independent or Uncorrelated

If random variables X and Y are independent, then we have (we just show the proof for continuous RVs here, but the result is correct for the discrete RVs as well):

$$E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x)f(y)dxdy$$

$$= \int_{-\infty}^{\infty} xf(x)dx \times \int_{-\infty}^{\infty} yf(y)dy \; = E[X]E[Y]$$

Therefore if X and Y are independent, they are uncorrelated. However, if X and Y are uncorrelated, they are not necessarily independent, as if the double integrals above are equal, this does not mean that $f(x,y)$= $f(x)f(y)$.

Conditional Distributions

The conditional probability distribution (or conditional pdf) of X_2 given X_1 is defined as

$$f_2(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)}, \qquad f_1(x_1) > 0$$

If X_1 and X_2 are discrete, then $f_2(x_2|x_1)$ is the conditional probability given as

$$f_2(x_2|x_1) = P(X_2 = x_2|X_1 = x_1)$$

For continuous random variables, $f_2(x_2|x_1)$ should be read as the conditional pdf of $X_2 = x_2$ given $X_1 = x_1$.

Conditional Distributions (cont.)

Notice that both $f(x_1, x_2)$ and $f_2(x_2|x_1)$ may appear to be functions of the two variables X_1 and X_2 , yet the difference in the notation, and thus the meaning of the functions should be clear:

- $f(x_1, x_2)$ being the joint pdf for X_1 and X_2, is a function of two variables, whereas
- $f_2(x_2|x_1)$ is a function of only one variable X_2, with x_1 acting as a constant.

Conditional Distributions – Example 1

Consider the joint pdf

$$f(x, y) = \begin{cases} \frac{3}{8}(x + y^2), & 0 \le x \le 2, \quad 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

the marginal pdf of X is the given by

$$f_1(x) = \begin{cases} \frac{1}{8}(3x + 1), & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$$

Then we have

$$f_2(y|x) = \frac{f(x, y)}{f_1(x)} = \frac{\frac{3}{8}(x + y^2)}{\frac{1}{8}(3x + 1)} = \frac{3(x + y^2)}{3x + 1}, \qquad 0 \le y \le 1$$

Conditional Distributions – Example 1 (cont.)

Now assume we are looking for $P\left(Y < \frac{1}{2} \middle| X = \frac{4}{3}\right)$.

For $0 \leq y \leq 1$

$$f_2(y|4/3) = \frac{3\left(\frac{4}{3} + y^2\right)}{3\left(\frac{4}{3}\right) + 1} = \frac{3}{5}\left(\frac{4}{3} + y^2\right) = \frac{3}{5}y^2 + \frac{4}{5}$$

therefore

$$P(Y < 1/2|X = 4/3) = \int_{-\infty}^{1/2} f_2(y|4/3)dy = \int_0^{1/2}\left(\frac{3}{5}y^2 + \frac{4}{5}\right)dy = \left[\frac{y^3}{5} + \frac{4y}{5}\right]_0^{1/2} = \frac{1}{40} + \frac{2}{5}$$

$$= \frac{17}{40} = 0.425$$

Conditional Distributions – Example 2

Consider the following joint distribution for the discrete random variables X_1 and X_2

$f(x_1, x_2)$		x_2 = 0	1	2	Total
x_1	0	0.15	0.30	0.10	0.55
	1	0.25	0.15	0.05	0.45
Total		0.40	0.45	0.15	1

Then we have

$$f_1(0|0) = \frac{f(0,0)}{f_2(0)} = \frac{0.15}{0.40} = \frac{3}{8}, \qquad f_1(1|0) = \frac{f(1,0)}{f_2(0)} = \frac{0.25}{0.40} = \frac{5}{8},$$

$$f_2(0|0) = \frac{f(0,0)}{f_1(0)} = \frac{0.15}{0.55} = \frac{3}{11}, \qquad f_2(1|0) = \frac{f(0,1)}{f_1(0)} = \frac{0.30}{0.55} = \frac{6}{11}, \qquad f_2(2|0) = \frac{f(0,2)}{f_1(0)} = \frac{0.10}{0.55} = \frac{2}{11}$$

Bayes Rule in Joint Distributions and Expectations

Assume that $f_{X,Y}(x,y)$ is the joint pdf RVs X and Y. We can write

$$Cont.:\ f(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{\infty} f(y|x)f(x)dx \qquad Disc:\ f(y) = \sum_{x} f(x,y) = \sum_{x} f(y|x)f(x)$$

The above equations reminds us about the first Bayes rule:

$$P(A) = \sum_{i} P(A|E_i)P(E_i)$$

Furthermore, we can write

$$\mathrm{E[Y]} = \int_{-\infty}^{\infty} yf(y)dx = \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f(x,y)dx\right)dy = \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f(y|x)f(x)dx\right)dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y|x)f(x)dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y|x)f(x)dy\,dx = \int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} yf(y|x)dy\right)dx = \int_{-\infty}^{\infty} E[Y|x]f(x)\,dx$$

Bayes Rule in Joint Distributions and Expectations (cont.)

Furthermore, we can write

$$\mathrm{E[Y]} = \int_{-\infty}^{\infty} yf(y)dx = \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f(x,y)dx\right)dy = \int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f(y|x)f(x)dx\right)dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y|x)f(x)dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(y|x)f(x)dy\,dx = \int_{-\infty}^{\infty} f(x)\left(\int_{-\infty}^{\infty} yf(y|x)dy\right)dx = \int_{-\infty}^{\infty} E[Y|x]f(x)\,dx$$

In summary:

$$Cont.:\ E[Y] = \int_{-\infty}^{\infty} E[Y|x]f(x)\,dx \qquad Disc: E[Y] = \sum_{x} E[Y|x]f(x)$$

That again looks like the first Bayes rule.

pdf of a function of a continuous RV

- Assume X is a random variable with known pdf $f(x)$, and $Y = g(X)$ is a function of X. We want to find the pdf of Y.
- We first start to find the cdf of Y. We can then find the pdf of Y by differentiating CDF of Y.

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) =^{*} \begin{cases} P\left(X \le g^{-1}(y)\right) = \int_{-\infty}^{g^{-1}(y)} f_X(x)dx = F_X\left(g^{-1}(y)\right) - 0 \; (g(X) \; is \; increasing) \\ P\left(X \ge g^{-1}(y)\right) = \int_{g^{-1}(y)}^{\infty} f_X(x)dx = 1 - F_X\left(g^{-1}(y)\right) \; (g(X) \; is \; decreasing) \end{cases}$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \begin{cases} \frac{dF_X\left(g^{-1}(y)\right)}{dy} = \frac{dg^{-1}(y)}{dy} f_X\left(g^{-1}(y)\right) & (g(X) \; is \; increasing) \\ \frac{d\left(1 - F_X\left(g^{-1}(y)\right)\right)}{dy} = -\frac{dg^{-1}(y)}{dy} f_X\left(g^{-1}(y)\right) & (g(X) \; is \; decreasing) \end{cases} = \left|\frac{dg^{-1}(y)}{dy}\right| f_X\left(g^{-1}(y)\right)$$

* Assuming $\mathrm{Y} = g(X)$ is a one-to-one function

pdf of a function of a continuous RV - Example

- If X is an exponential random variable with $f_X(x) = \lambda e^{-\lambda x}$ $x \geq 0$, and $Y = \frac{1}{1+X}$, find pdf of Y.

$$Y = g(X) = \frac{1}{1+X} \quad \rightarrow \quad Y + XY = 1 \quad \rightarrow \quad X = \frac{1-Y}{Y} = \frac{1}{Y} - 1 = g^{-1}(Y)$$

$$f_Y(y) = \left|\frac{dg^{-1}(y)}{dy}\right| f_X\left(g^{-1}(y)\right) = \frac{1}{y^2} \lambda e^{-\lambda\left(\frac{1}{y} - 1\right)} \qquad 0 < y < 1$$

Joint pdf of functions of joint continuous RVs

- Assume that $f_{X,Y}(x, y)$ is the joint pdf RVs X and Y. In addition, $W = g(X, Y)$ and $Z = h(X, Y)$ are two functions of X and Y.
- We want to find the joint pdf of W and Z.
- We can extend the equation that we found for the similar question in one dimension to higher dimensions as below:

$$f_{W,Z}(w, z) = |J| f_{X,Y}(g^{-1}(w, z), h^{-1}(w, z))$$

where

$$J = \begin{vmatrix} \frac{\partial x}{\partial w} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial w} & \frac{\partial y}{\partial z} \end{vmatrix}$$

Joint pdf of functions of joint continuous RVs - Example

If $f_{X,Y}(x,y) = \frac{1}{2\pi} exp\left(-\frac{x^2+y^2}{2}\right) \; -\infty < both\ x\ and\ y < +\infty$, and $\Theta = g(X,Y) = tan^{-1}\frac{y}{x}$ and $Z = h(X,Y) = \sqrt{x^2+y^2}$.

Find $f_{\Theta,Z}(\theta,z)$.

We have

$$X = \ g^{-1}(W,Z) = Z\ cos\Theta \qquad , \qquad Y = h^{-1}(W,Z) = Z\ sin\Theta$$

therefore

$$J = \begin{vmatrix} \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \end{vmatrix} = \begin{vmatrix} -zsin\theta & cos\theta \\ zcos\theta & sin\theta \end{vmatrix} = -zsin^2\theta - zcos^2\theta = -z$$

and

$$f_{\Theta,Z}(\theta,z) = |\,J\,|\, f_{X,Y}(g^{-1}(\theta,z), h^{-1}(\theta,z)) = \frac{z}{2\pi} exp\left(-\frac{(z\ cos\theta)^2 + (z\ sin\theta)^2}{2}\right) = \frac{z}{2\pi} exp\left(-\frac{z^2}{2}\right)$$

$$0 \le z < \infty\ , 0 \le \theta \le 2\pi$$

Bivariate Normal Distribution

- If Z_1 and Z_2 have $N(0,1)$ distribution, the joint (bivariate) pdf of them equals:

$$f(z_1, z_2) = \frac{1}{2\pi} \exp\left(-\frac{z_1^2 + z_2^2}{2}\right)$$

We use the following transformations

$$X = \mu_X + \sigma_X Z_1$$

and

$$Y = \mu_Y + \sigma_Y \left(\rho Z_1 + \sqrt{1-\rho^2}\, Z_2\right)$$

We can then show (see sample questions):

Bivariate Normal Distribution (cont.)

$$X \sim N(\mu_X, \sigma_X^2) \quad , \quad Y \sim N(\mu_Y, \sigma_Y^2) \quad \text{and} \quad \rho_{X,Y} = \rho$$

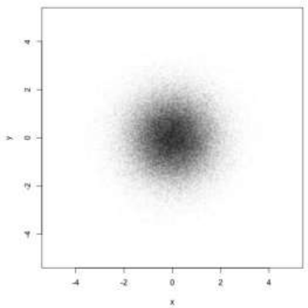
and

$$f_{X,Y}(x,y) = \frac{1}{2\pi\ \sigma_X \sigma_Y \sqrt{1-\rho^2}}\ e^{-\frac{1}{2}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{1}{1-\rho^2}\left(\frac{y-\mu_Y}{\sigma_Y} - \rho\ \frac{(x-\mu_X)}{\sigma_X}\right)^2\right]}$$

or

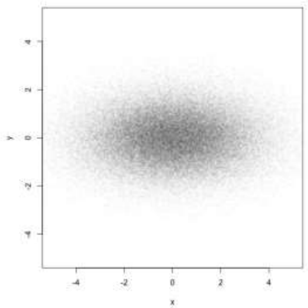
$$f_{X,Y}(x,y) = \frac{1}{2\pi\ \sigma_X \sigma_Y \sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\ \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X \sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right]$$

Visualisation of Bivariate Normal Sample



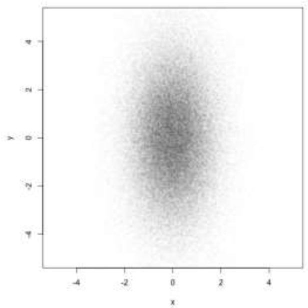
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0$



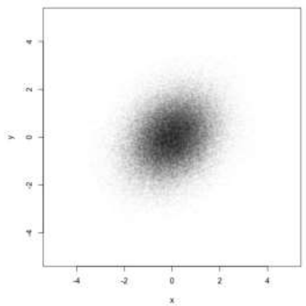
$X \sim \mathcal{N}(0,2),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0$



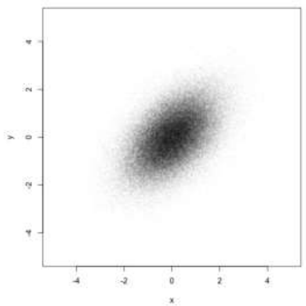
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,2)$

$\rho = 0$



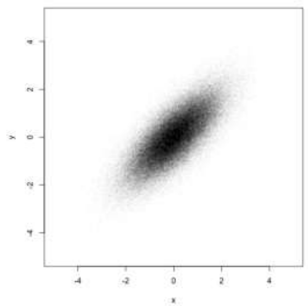
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0.25$



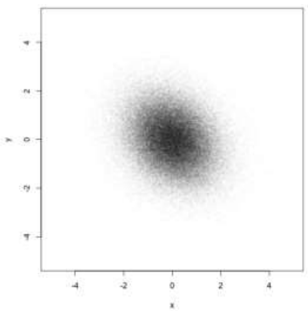
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0.5$



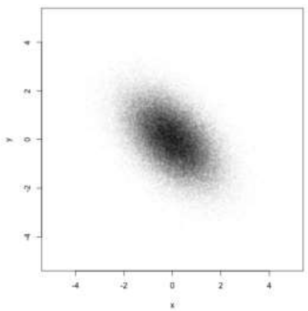
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = 0.75$



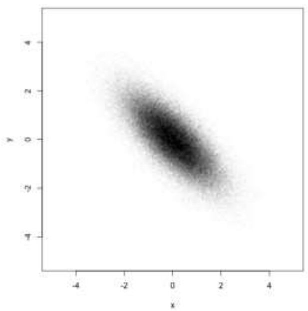
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.25$



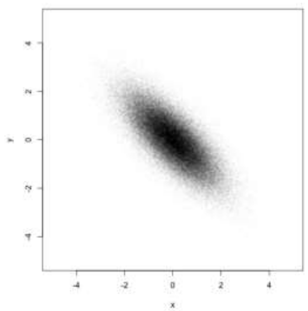
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.5$



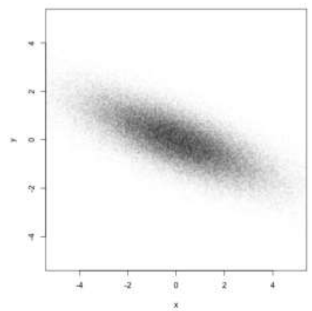
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.75$



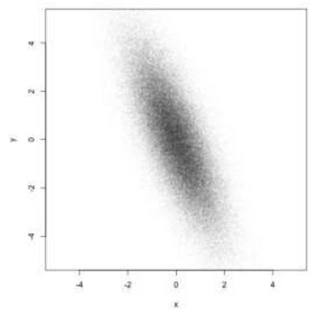
$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.75$



$X \sim \mathcal{N}(0,2),\ Y \sim \mathcal{N}(0,1)$

$\rho = -0.75$



$X \sim \mathcal{N}(0,1),\ Y \sim \mathcal{N}(0,2)$

$\rho = -0.75$

More Results and Properties

- We can also show (see sample questions)

$$E[Y|x] = \mu_Y + \rho_{X,Y} \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

and similarly

$$E[X|y] = \mu_X + \rho_{X,Y} \frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

which are estimation of Y knowing $X = x$, and estimation of X knowing $Y = y$, respectively.

More Results and Properties (Cont.)

The error of these estimate are given by

$$MSE[\,E[Y|x]\,] = \; E[(E[Y|x] - Y)^2] = Var[Y|x] = \sigma_Y^2\,(1 - \rho^2)$$

and

$$MSE[\,E[X|y]\,] = \; E[(E[X|y] - X)^2] = Var[X|y] = \sigma_X^2\,(1 - \rho^2)$$

Uncorrelated and/or Independent

For a bivariate Normal distribution, a very important result is observed by putting $\rho_{X,Y} = \rho = \mathbf{0}$ in

$$f_{X,Y}(x,y) = \frac{1}{2\pi\,\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\,\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right]$$

that give us

$$f_{X,Y}(x,y) = \frac{1}{2\pi\,\sigma_X\sigma_Y}\exp\left[-\left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right] = f_X(x)f_Y(y)$$

In other words, bivariate Normal variables X and Y are independent **iff** being uncorrelated.

$$being\ uncorrelated\ (\rho = 0) \begin{matrix}\rightarrow\\ \leftarrow\end{matrix} \quad independence$$

As you know this is not bi-directional for any other bivariate distribution, i.e. we only can say

$$being\ uncorrelated\ (\rho = 0) \leftarrow \quad independence$$

Further Notes on Conditional Expectation

- As we saw before, $E[Y|x]$ is the estimate of Y when $X = x$.

- $Var[Y|x]$ is defined as:

$$Var[Y|x] = E[Y^2|x] - E^2[Y|x]$$

This is the error of the estimate $E[Y|x]$ (proof in sample questions).

Further Notes on Conditional Expectation (cont.)

As we saw before, $E[Y]$ can be presented as expected value of conditional expectation of Y when $X = x$. Mathematically

$$E_Y[Y] = E_X\big[E_Y[Y|\, X]\big]$$

Example

Random variable X has a Poisson distribution with parameter θ. If the observed sample of X equals x, a Bernoulli experiment with $P(S) = p$ is run x times independently. If the number of successes is shown by Y, find the expected value Y.

Solution:

When $X = x$, random variable Y, the number of successes in x Bernoulli experiments, has a Binomial distribution with parameters ($n = x$ and $p = p$), and therefore $E_Y[Y|\, x] = xp$, or $E_Y[Y|\, X] = Xp$.

Therefore,

$$E_Y[Y] = E_X\big[E_Y[Y|\, X]\big] = E_X[Xp] = pE_X[X] = p\theta$$

Further Notes on Conditional Expectation cont.

$Var[Y]$ can be presented as in form of conditional variance and expectation as well. We can show:

$$Var[Y] = Var_X\big[E_Y[Y|\, X]\big] + E_X\big[Var_Y[Y|\, X]\big]$$

What does this mean?

This means:

In absence of any estimator like X to use to estimate Y, we can prove $E[Y]$ is the best estimate for Y. The error of this estimate is given by $Var[Y]$.

However, when a random variable X is available to use as Y estimator, the error of this estimate, $Var_X\big[E_Y[Y|\, X]\big]$, is smaller than the error of estimate with no estimator (as $Var_Y[Y|\, X]$ and therefore $E_X\big[Var_Y[Y|\, X]\big]$ is positive). This is obviously expected.

Min and Max Distributions

- For n independent/identical random variables X_i $(i = 1, \dots, n)$, with pdf $f_X(x)$, we define

$$U = min(X_1, \dots, X_n) \quad , \quad V = Max(X_1, \dots, X_n)$$

The pdf of U and V are then given by:

$$f_U(u) = nf_X(u)[1 - F_X(u)]^{n-1}$$

$$f_V(v) = nf_X(v)[F_X(v)]^{n-1}$$