Lecture 6: Joint Distributions

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

April 18, 2024

Outline

Multivariate Distribution

$P(X=x, Y=y, Z=z)$

- Joint distribution provides complete information about how multiple r.v.s interact in high-dimensional space
- Marginal distribution is the individual distribution of each r.v.
- Conditional distribution is the updated distribution for some r.v.s after observing other r.v.s

Outline

Joint CDF

Definition

The *joint CDF* of r.v.s X and Y is the function $F_{X,Y}$ given by

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

The joint CDF of n r.v.s is defined analogously.

Joint PMF

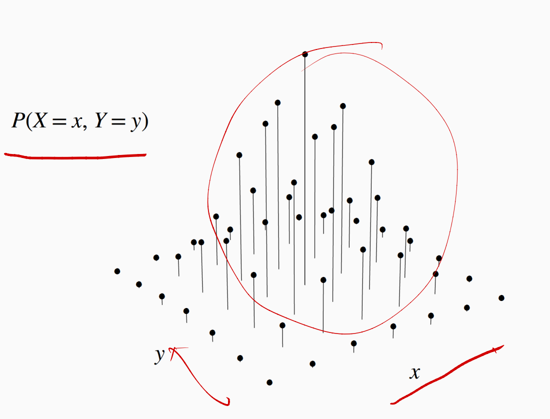
Definition

The joint PMF of discrete r.v.s X and Y is the function $p_{X,Y}$ given by

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

The joint PMF of n discrete r.v.s is defined analogously.

Joint PMF



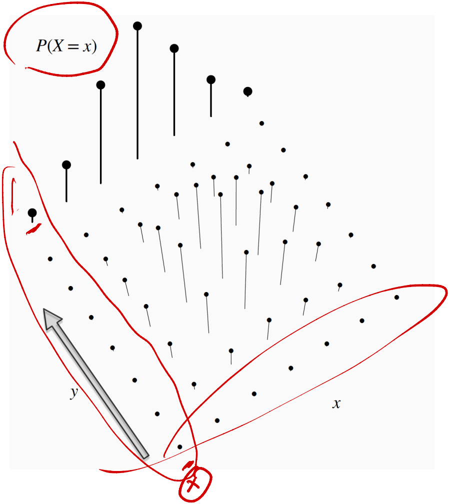
Marginal PMF

Definition

For discrete r.v.s X and Y, the *marginal PMF* of X is

$$P(X = x) = \sum_y P(X = x, Y = y).$$

Marginal PMF



Example

$$p_Y(2) = p(Y=2) = \sum_{i=1}^{4} p(X=i, Y=2) = \frac{1}{20} + \frac{2}{20} + \frac{3}{20} + \frac{1}{20} = \frac{7}{20}.$$

Joint PMF $p_{X,Y}(x,y)$ in tabular form

y \ x	1	2	3	4	Row sums: marginal PMF $p_Y(y)$
4	0	1/20	1/20	1/20	3/20
3	1/20	2/20	3/20	1/20	7/20
2	1/20	2/20	3/20	1/20	7/20
1	1/20	1/20	1/20	0	3/20
Column sums: marginal PMF $p_X(x)$	3/20	6/20	8/20	3/20	

$$p_X(1) = p(X=1) = p(X=1,Y=1) + p(X=1,Y=2) + p(X=1,Y=3) + p(X=1,Y=4) = 0 + \frac{1}{20} + \frac{1}{20} + \frac{1}{20} = \frac{3}{20}.$$

Conditional PMF

1°. Conditional PMF is also a valid PMF

2°. fixed y, $P_{X|Y}(\cdot|y)$ is a valid PMF.

$$\sum_x P_{X|Y}(x|y) \overset{\checkmark}{=} 1$$

$$\Leftrightarrow \quad \sum_x P(X=x|Y=y) \overset{\checkmark}{=} 1$$

Definition

For discrete r.v.s X and Y, the *conditional PMF* of X given $Y=y$ is

$$P_{X|Y}(x|y) = P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)}.$$

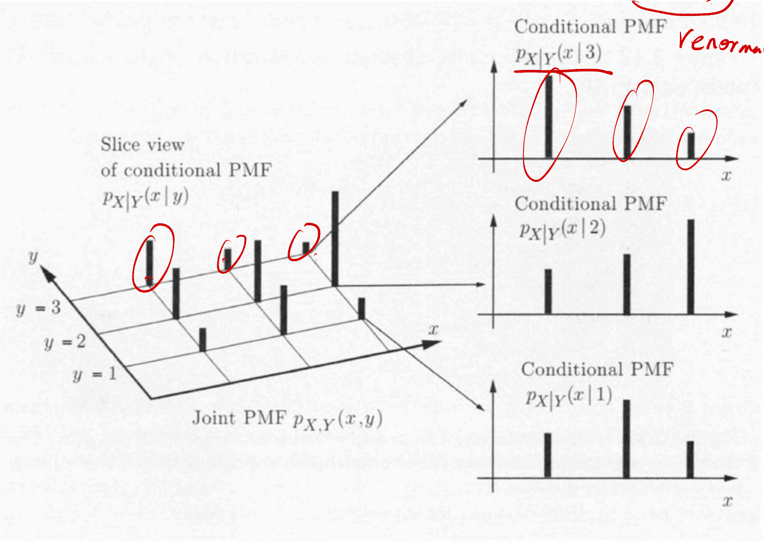
$$\Leftrightarrow \quad \sum_x \frac{P(X=x, Y=y)}{P(Y=y)} \overset{\checkmark}{=} 1$$

$$\Leftrightarrow \quad \sum_x P(X=x, Y=y) \overset{\checkmark}{=} P(Y=y)$$

Conditional PMF

$p_{X,Y}(x|3) = \frac{P(X=x, Y=3)}{P(Y=3)}$

renormalization



Independence of Discrete R.V.s

Definition

Random variables X and Y are *independent* if for all x and y,

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

If X and Y are discrete, this is equivalent to the condition

$$P(X=x, Y=y) = P(X=x)P(Y=y)$$

for all x and y, and it is also equivalent to the condition

$$P(Y=y|X=x) = P(Y=y)$$

for all y and all x such that $P(X=x) > 0$.

Example: Chicken-egg

① Joint PMF. $P(X=i, Y=j)$, $i \geq 0, j \geq 0$ integer.

② $X+Y=N$

$X+Y \mid N=n = n$

$X \mid N=n \sim \mathrm{Bin}(n,p)$

$Y \mid N=n \sim \mathrm{Bin}(n,q)$ $q=1-p$

Suppose a chicken lays a random number of eggs, N, where $N \sim \mathrm{Pois}(\lambda)$. Each egg independently hatches with probability p and fails to hatch with probability $q = 1 - p$. Let X be the number of eggs that hatch and Y the number that do not hatch, so $X + Y = N$. What is the joint PMF of X and Y?

③ $P(X=i, Y=j) \underset{\text{LOTP}}{=} \sum_{n=0}^{\infty} P(X=i, Y=j \mid N=n) \cdot P(N=n)$

$= P(X=i, Y=j \mid N=i+j) \cdot P(N=i+j)$

$= P(X=i \mid N=i+j)\, P(Y=j \mid X=i, N=i+j) \cdot P(N=i+j)$

Solution

$$P(X=i, Y=j) = P(X=i \mid N=i+j)\cdot P(N=i+j)$$

$X \mid N=i+j \sim Bin(i+j, p)$

$N \sim Pois(\lambda)$

$$= \binom{i+j}{i} p^i q^j \cdot \frac{e^{-\lambda}\cdot \lambda^{i+j}}{(i+j)!}$$

$p+q=1$

$$= \frac{\cancel{(i+j)!}}{i!\,j!}\cdot p^i q^j \cdot e^{-\lambda p}\cdot e^{-\lambda q}\cdot \frac{\lambda^{i+j}}{\cancel{(i+j)!}}$$

$$= e^{-\lambda p}\cdot \frac{p^i\cdot\lambda^i}{i!}\cdot e^{-\lambda q}\cdot\frac{q^j\cdot\lambda^j}{j!}$$

$$= \underbrace{e^{-\lambda p}\cdot\frac{(\lambda p)^i}{i!}}_{\sim Pois(\lambda p)} \cdot \underbrace{e^{-\lambda q}\cdot\frac{(\lambda q)^j}{j!}}_{\sim Pois(\lambda q)}$$

$X+Y \mid N=n$

$= n$

X and Y are NOT conditional independent

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

$= P(X=i)\cdot P(Y=j)$

$\forall i, j \geq 0.$

$\Rightarrow X$ and Y are independent!

④ $$P(X=i) = \sum_{j=0}^{\infty} P(X=i, Y=j) = \frac{e^{-\lambda p}\cdot(\lambda p)^i}{i!}$$

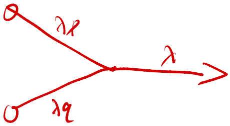
$X \sim Pois(\lambda p)$

$$P(Y=j) = \sum_{i=0}^{\infty} P(X=i, Y=j) = \frac{e^{-\lambda q}\cdot(\lambda q)^j}{j!}$$

$Y \sim Pois(\lambda q)$

Solution

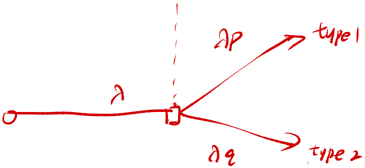
Related Theorem



Theorem

If $X \sim \mathrm{Pois}\,(\lambda p)$, $Y \sim \mathrm{Pois}\,(\lambda q)$, *and* X *and* Y *are independent, then* $N = X + Y \sim \mathrm{Pois}(\lambda)$ *and* $X|N = n \sim \mathrm{Bin}(n, p)$.

Related Theorem



Theorem

If $N \sim \mathrm{Pois}(\lambda)$ *and* $X|N = n \sim \mathrm{Bin}(n, p)$, *then* $X \sim \mathrm{Pois}(\lambda p)$, $Y = N - X \sim \mathrm{Pois}(\lambda q)$, *and* X *and* Y *are independent.*

Outline

Conditional PDF Given an Event

3°. $P(X \le x) \overset{\text{LOTP}}{=} \sum_{i=1}^{n} P(A_i) P(X \le x | A_i)$

$\int_{-\infty}^{x} f_X(t)dt = \sum_{i=1}^{n} P(A_i) \int_{-\infty}^{x} f_{X|A_i}(t)dt$

Taking the derivate of both sides, w.r.t. x.

Conditional PDF Given an Event

1° • The conditional PDF $f_{X|A}$ of a continuous random variable X, given an event A with $\mathbf{P}(A) > 0$, satisfies

$$\mathbf{P}(X \in B \,|\, A) = \int_B f_{X|A}(x)\, dx.$$

$\Rightarrow f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)$

2° • If A is a subset of the real line with $\mathbf{P}(X \in A) > 0$, then

$$f_{X|\{X\in A\}}(x) = \begin{cases} \dfrac{f_X(x)}{\mathbf{P}(X \in A)}, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

3° • Let $A_1, A_2, \ldots, A_n$ be disjoint events that form a partition of the sample space, and assume that $\mathbf{P}(A_i) > 0$ for all i. Then,

$$f_X(x) = \sum_{i=1}^{n} \mathbf{P}(A_i) f_{X|A_i}(x)$$

(a version of the total probability theorem).

Proof

2° $f_{X|\{X\in A\}}(x) = \lim\limits_{\delta\to 0} \dfrac{P(x\le X\le x+\delta \mid X\in A)}{\delta} = \lim\limits_{\delta\to 0} \dfrac{P(x\le X\le x+\delta,\ X\in A)}{\delta\cdot P(X\in A)}$

$= \lim\limits_{\delta\to 0} \dfrac{P(X\in A \mid x\le X\le x+\delta)\cdot P(x\le X\le x+\delta)}{\delta\cdot P(X\in A)}$

$= \dfrac{1}{P(X\in A)} \cdot \underline{\lim\limits_{\delta\to 0} \dfrac{P(x\le X\le x+\delta)}{\delta}} \cdot \underline{P(X\in A \mid x\le X\le x+\delta)}$

$= \dfrac{1}{P(X\in A)} \cdot f_X(x) \cdot P(X\in A \mid X=x)$

$= \dfrac{f_X(x)}{P(X\in A)} \cdot 1_{\{x\in A\}}$

$P(X\in A \mid X=x) = \dfrac{P(X\in A, X=x)}{P(X=x)} = 1_{\{x\in A\}}$

if $x\in A$, 1

$x\notin A$, 0

Joint PDF

① Valid joint PDF. $\begin{cases} f_{X,Y}(x,y) \geq 0 \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1 \end{cases}$

② Example. $P(X<3, 1<Y<4)$

$= \int_{-\infty}^{3}\int_{1}^{4} f_{X,Y}(x,y)\,dx\,dy$

Definition

If X and Y are continuous with joint CDF $F_{X,Y}$, their joint PDF is the derivative of the *joint CDF* with respect to x and y:

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

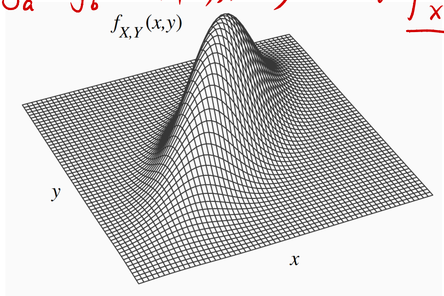
③ $B \subseteq R^2$. $P((X,Y) \in B) = \iint_B f_{X,Y}(x,y)\,dx\,dy$.

Joint PDF

$\delta \approx 0 \quad P(a \le X \le a+\delta) = \int_a^{a+\delta} f_X(x)dx \approx f_X(a)\cdot\delta$

$P(a \le X \le a+\delta,\ b \le Y \le b+\delta)$

$= \int_a^{a+\delta}\int_b^{b+\delta} f_{X,Y}(x,y)\,dx\,dy \approx \underline{f_{X,Y}(a,b)}\cdot\delta^2$



Marginal PDF

Definition

For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the *marginal PDF* of X is

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

This is the PDF of X, viewing X individually rather than jointly with Y.

Conditional PDF is a valid PDF. given fixed y or x.

$f_{Y|X}(\cdot|x)$ is a valid PDF. 1° ≥ 0 ✓

2°. $\int_{-\infty}^{\infty} f_{Y|X}(y|x)dy \overset{?}{=} 1 \Leftarrow \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)} dy \overset{\checkmark}{=} 1$

Definition

For continuous r.v.s X and Y with joint PDF $f_{X,Y}$, the *conditional PDF* of Y given $X = x$ is

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

$\Leftarrow \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \overset{\checkmark}{=} f_X(x)$

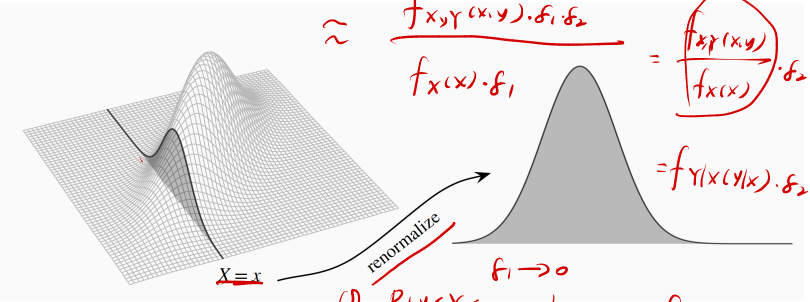
Conditional PDF

① $P(y \le Y \le y+\delta_2 \mid x \le X \le x+\delta_1)$

$$= \frac{P(y \le Y \le y+\delta_2,\ x \le X \le x+\delta_1)}{P(x \le X \le x+\delta_1)}$$

$$\approx \frac{f_{X,Y}(x,y)\cdot\delta_1\cdot\delta_2}{f_X(x)\cdot\delta_1} = \frac{f_{X,Y}(x,y)}{f_X(x)}\cdot\delta_2$$

$$= f_{Y|X}(y|x)\cdot\delta_2$$



$\delta_1 \to 0$

② $P(y \le Y \le y+\delta_2 \mid X=x) \approx f_{Y|X}(y|x)\cdot\delta_2$

Technique Issue

- What is the meaning of conditioning on zero-probability event $X = x$ for a continuous r.v. X.
- We are actually conditioning on the event that X falls within a small interval of x: $X \in (x - \epsilon, x + \epsilon)$ and then taking a limit as $\epsilon \to 0$.

Example

① $0<x<1,\ 0<y<1,$

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

$$= \frac{\frac{12x(2-x-y)}{5}}{\int_0^1 \frac{12x(2-x-y)}{5}\,dx} = \frac{x(2-x-y)}{\int_0^1 x(2-x-y)\,dx} = \frac{6x(2-x-y)}{4-3y}.$$

The joint PDF of X and Y is given by

$$f(x,y) = \begin{cases} \frac{12x(2-x-y)}{5} & \text{if } 0<x<1, 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the conditional PDF of X given that $Y = y$, where $0 < y < 1$.

Example

① Conditioning PDF. $0<x<\infty$, $0<y<\infty$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_0^\infty f(x,y)dx} = \frac{\frac{e^{-x/y-y}}{y}}{\int_0^\infty \frac{e^{-x-y}}{y}dx}$$

$$= \frac{1}{y}e^{-x/y}$$

Suppose that the joint PDF of X and Y is given by

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

Find $P\{X > 1 | Y = y\}$.

② $P(X>1|Y=y) = \int_1^\infty f_{X|Y}(x|y)dx$

$$= \int_1^\infty \frac{1}{y}e^{-x/y}dx = e^{-\frac{1}{y}}$$

Continuous form of Bayes' Rule and LOTP

① $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$ $\Rightarrow$ $f_{Y|X}(y|x) \cdot f_X(x)$

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

Theorem

For continuous r.v.s X and Y,

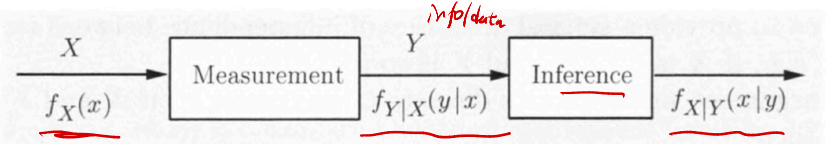
$$① \quad f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)\, f_Y(y)}{f_X(x)}$$

$$② \quad f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy$$

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

Proof

Bayes' Rule: Inference Perspective



prior PDF

Posterior PDF

Example

① r.v. Λ, $f_\Lambda(\lambda) = 2$, $1 \le \lambda \le \frac{3}{2}$

prior PDF.

② Conditional PDF. $f_{\Lambda|Y}(\lambda|y) = \dfrac{f_\Lambda(\lambda)\, f_{Y|\Lambda}(y|\lambda)}{f_Y(y)}$ $\sim expo(\lambda)$

$Y=y$ posterior PDF of Λ under $Y=y$

A light bulb produced by the GE company is known to have an exponential distributed lifetime Y. However, the company has been experiencing quality control problems. On any given day, the parameter λ of the PDF of Y is actually a random variable. uniformly distributed in the interval$[1, 3/2]$. We test a light bulb and record its lifetime. What we can say about the underlying parameter λ?

y

$$= \frac{2 \cdot \lambda e^{-\lambda y}}{f_Y(y)}$$

$$f_Y(y) \overset{\text{LoTP}}{=} \int_{-\infty}^{\infty} f_\Lambda(t) f_{Y|\Lambda}(y|t)\,dt = \int_1^{\frac{3}{2}} 2te^{-ty}\,dt$$

General Bayes' Rule

① X discrete, Y continuous.

$$P(Y \in (y-\varepsilon, y+\varepsilon) \mid X=x) = \frac{P(X=x \mid Y \in (y-\varepsilon, y+\varepsilon)) \cdot P(Y \in (y-\varepsilon, y+\varepsilon))}{P(X=x)}$$

	Y **discrete**	Y **continuous**
X **discrete**	$P(Y=y \mid X=x) = \frac{P(X=x \mid Y=y)P(Y=y)}{P(X=x)}$	$f_Y(y \mid X=x) = \frac{P(X=x \mid Y=y)f_Y(y)}{P(X=x)}$ ①
X **continuous**	$P(Y=y \mid X=x) = \frac{f_X(x \mid Y=y)P(Y=y)}{f_X(x)}$ ②	$f_{Y \mid X}(y \mid x) = \frac{f_{X \mid Y}(x \mid y)f_Y(y)}{f_X(x)}$

$$\Rightarrow \lim_{\varepsilon \to 0} \frac{P(Y \in (y-\varepsilon, y+\varepsilon) \mid X=x)}{2\varepsilon} = \lim_{\varepsilon \to 0} \frac{1}{P(X=x)} P(X=x \mid Y \in (y-\varepsilon, y+\varepsilon)) \cdot \frac{P(Y \in (y-\varepsilon, y+\varepsilon))}{2\varepsilon}$$

$$f_Y(y \mid X=x) = \frac{P(X=x \mid Y=y)}{P(X=x)} \cdot f_Y(y)$$

✓

Proof

② $P(Y=y \mid X=x) = \dfrac{f_X(x \mid Y=y)}{f_X(x)} \cdot P(Y=y)$

Y discrete

X continuous.

$$\lim_{\varepsilon \to 0} P(Y=y \mid X \in (x-\varepsilon, x+\varepsilon)) = \frac{P(X \in (x-\varepsilon, x+\varepsilon) \mid Y=y) \cdot P(Y=y)}{P(X \in (x-\varepsilon, x+\varepsilon))}$$

$$= \lim_{\varepsilon \to 0} \frac{\dfrac{P(X \in (x-\varepsilon, x+\varepsilon) \mid Y=y)}{2\varepsilon} \cdot P(Y=y)}{\dfrac{P(X \in (x-\varepsilon, x+\varepsilon))}{2\varepsilon}}$$

$$P(Y=y \mid X=x) = \frac{f_X(x \mid Y=y)}{f_X(x)} \cdot P(Y=y)$$

Proof

General LOTP

② $\lim_{\varepsilon \to 0} \frac{P(X \in (x-\varepsilon, x+\varepsilon))}{2\varepsilon} \overset{\text{LOTP}}{=} \lim_{\varepsilon \to 0} \sum_y \frac{P(X \in (x-\varepsilon, x+\varepsilon) | Y=y)}{2\varepsilon} \cdot p(Y=y)$

$$f_X(x) = \sum_y f_X(x | Y=y) \cdot p(Y=y)$$

	Y **discrete**	Y **continuous**
X **discrete**	$P(X = x) = \sum_y P(X = x \mid Y = y) P(Y = y)$	$P(X = x) = \int_{-\infty}^{\infty} P(X = x \mid Y = y) f_Y(y) dy$ ①
X **continuous**	$f_X(x) = \sum_y f_X(x \mid Y = y) P(Y = y)$ ②	$f_X(x) = \int_{-\infty}^{\infty} f_{X\mid Y}(x \mid y) f_Y(y) dy$

① $P(X=x | Y=y) = \frac{f_Y(y | X=x)}{f_Y(y)} \cdot p(X=x)$

$\Rightarrow P(X=x | Y=y) \cdot f_Y(y) = f_Y(y | X=x) \cdot p(X=x)$

$\Rightarrow \int_{-\infty}^{\infty} P(X=x | Y=y) \cdot f_Y(y) dy = \int_{-\infty}^{\infty} f_Y(y | X=x) p(X=x) dy$

$= P(X=x) \cdot \int_{-\infty}^{\infty} f_Y(y | X=x) dy = P(X=x)$

Proof

Example

① $Y = N + S$; $Y \mid S=s = N + s \sim \mathcal{N}(s,1)$ $\quad N \sim \mathcal{N}(0,1)$.

② $P(S=1 \mid Y=y) = \dfrac{f_{Y|S}(y|1) \cdot P(S=1)}{f_Y(y)} = \dfrac{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-1)^2} \cdot p}{f_Y(y)}$ $\quad$ $S=1$. PDF. $\mathcal{N}(1,1)$

$f_Y(y) = f_{Y|S}(y|1)P(S=1) + f_{Y|S}(y|-1)P(S=-1)$

A binary signal S is transmitted, and we are given that $P(S=1) = p$ and $P(S=-1) = 1-p$. The received signal is $Y = N + S$, where N is normal noise, with zero mean and unit variance, independent of S. What is the probability that $S = 1$, as a function of the observed value y of Y?

$P(S=1 \mid Y=y)$ posterior prob

$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-1)^2} \cdot p + \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y+1)^2} \cdot (1-p)$

③ $P(S=1 \mid Y=y) = \dfrac{pe^{y}}{p \cdot e^{y} + (1-p)e^{-y}} = \dfrac{p}{p + (1-p)e^{-2y}} = \begin{cases} >p & y>0 \\ <p & y<0 \\ p & y=0 \end{cases}$

Example: Comparing Exponentials of Different Rates

$$P(T_1<T_2) \overset{LOTP}{=} \int_0^\infty P(T_1<T_2 \mid T_2=t)\, f_{T_2}(t)\, dt$$

$$= \int_0^\infty P(T_1<t \mid T_2=t)\cdot \lambda_2 e^{-\lambda_2 t}\, dt$$

$$= \int_0^\infty P(T_1<t)\cdot \lambda_2 e^{-\lambda_2 t}\, dt$$

$$= \int_0^\infty (1-e^{-\lambda_1 t})\cdot \lambda_2 e^{-\lambda_2 t}\, dt$$

$$= \frac{\lambda_1}{\lambda_1+\lambda_2}$$

$$P(T_2<T_1) = \frac{\lambda_2}{\lambda_1+\lambda_2}$$

$$P(T_2=T_1) = 0$$

$$P(T_1=\min(T_1,T_2)) = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

Let $T_1 \sim Expo(\lambda_1)$, $T_2 \sim Expo(\lambda_2)$, T_1 and T_2 are independent. Find $P(T_1 < T_2)$.

$T_1, \dots T_n$: $Expo(\lambda_1) \dots Expo(\lambda_n)$

independent

$$\Rightarrow P(T_i = \min(T_1, \dots T_n)) = \frac{\lambda_i}{\lambda_1+\dots+\lambda_n}$$

$$\min(T_1, \dots T_n) \sim Expo(\lambda_1+\dots+\lambda_n)$$

Independence of Continuous R.V.s

Definition

Random variables X and Y are *independent* if for all x and y,

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

If X and Y are continuous with joint PDF $f_{X,Y}$, this is equivalent to the condition

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

for all x and y, and it is also equivalent to the condition

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

for all y and all x such that $f_X(x) > 0$.

Proposition

$f_{X,Y}(x,y) = \begin{cases} 8xy, & 0 \le x \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

$f_X(x) = 4x(1-x^2), 0 \le x \le 1$; $f_Y(y) = 4y^3, 0 \le y \le 1$

Theorem

$f_{X,Y}(x,y) \neq f_X(x) \cdot f_Y(y)$

Suppose that the joint PDF $f_{X,Y}$ of X and Y factors as

$$f_{X,Y}(x,y) = g(x)\,h(y)$$

for all x and y, where g and h are nonnegative functions. Then X and Y are independent. Also, if either g or h is a valid PDF, then the other one is a valid PDF too and g and h are the marginal PDFs of X and Y, respectively. (The analogous result in the discrete case also holds.)

Proof

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

① $f_{X,Y}(x,y) = g(x)\cdot h(y) = c\cdot g(x)\cdot \frac{h(y)}{c}$.

$c = \int_{-\infty}^{\infty} h(y)dy \Rightarrow 1 = \int_{-\infty}^{\infty} \frac{1}{c} h(y)dy$

$\Rightarrow f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{\infty} c\cdot g(x)\cdot \frac{h(y)}{c} dy$

$= c\cdot g(x) \int_{-\infty}^{\infty} \frac{h(y)}{c} dy = c\cdot g(x)$

② $\Rightarrow \int_{-\infty}^{\infty} f_X(x)dx = 1 \Rightarrow \int_{-\infty}^{\infty} c\cdot g(x)dx = 1$

$\Rightarrow f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-\infty}^{\infty} c\, g(x)\cdot \frac{h(y)}{c} dx$

$= \frac{h(y)}{c} \cdot \underbrace{\int_{-\infty}^{\infty} c\cdot g(x)dx}_{1} = \frac{h(y)}{c}$.

③ $f_{X,Y}(x,y) = c\cdot g(x)\cdot \frac{h(y)}{c} = f_X(x)\cdot f_Y(y)$. $\Rightarrow X, Y$ are independent.

if g is a valid PDF. $\Rightarrow c = 1$ $\Rightarrow h$ is a valid PDF.

2D LOTUS

Theorem

Let g be a function from $\mathbb{R}^2$ to $\mathbb{R}$. If X and Y are discrete, then

$$E(g(X,Y)) = \sum_x \sum_y g(x,y)\, P(X=x, Y=y)\,.$$

If X and Y are continuous with joint PDF $f_{X,Y}$, then

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f_{X,Y}(x,y)\, dxdy.$$

Expected Distance between Two Uniforms

① $E(|X-Y|) = \int_0^1 \int_0^1 |x-y| f_X(x) f_Y(y)\, dx\, dy = \int_0^1 \int_0^1 |x-y|\, dx\, dy$

$= \int_0^1 \int_y^1 (x-y)\, dx\, dy + \int_0^1 \int_0^y (y-x)\, dx\, dy = \frac{1}{3}$

② $M = \max(X,Y)$, $L = \min(X,Y)$, $M + L = X + Y$.

For $X, Y \overset{\text{i.i.d.}}{\sim} \text{Unif}(0,1)$, find $E(|X-Y|)$, $E(\max(X,Y))$, and $E(\min(X,Y))$.

$\Rightarrow E(M+L) = E(X+Y) = E(X) + E(Y) = \frac{1}{2} + \frac{1}{2} = 1$

$\Rightarrow E(M) + E(L) = 1$

③ $M - L = \max(X,Y) - \min(X,Y) = \begin{cases} X-Y & \text{if } X>Y \\ Y-X & \text{if } X<Y \end{cases} = |X-Y|$

$\Rightarrow E(M-L) = E(|X-Y|) = \frac{1}{3} \Rightarrow E(M) - E(L) = \frac{1}{3}$

④ $E(M) = \frac{2}{3}$; $E(L) = \frac{1}{3}$.

Expected Distance between Two Normals

X, Y indep

$Var(X-Y)$

$= Var(X)$

$+ Var(-Y)$

$= Var(X)$

$+ Var(Y)$

① Method 1 : $E(|X-Y|) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y| \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \, dx\, dy$

② Method 2 : $Y \sim N(0,1)$; $-Y \sim N(0,1)$

$X-Y \sim N(0,2)$; $X-Y = \sqrt{2} Z$, $Z \sim N(0,1)$

For $X, Y \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, find $E(|X-Y|)$.

$\Rightarrow E(|X-Y|) = E(\sqrt{2}|Z|) = \sqrt{2}\, E(|Z|)$

$E(|Z|) = \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = 2\int_0^{\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \sqrt{\frac{2}{\pi}}$.

$\Rightarrow E(|X-Y|) = \frac{2}{\sqrt{\pi}}$.

Outline

Covariance

Definition

The *covariance* between r.v.s X and Y is

$$\operatorname{Cov}(X, Y) = E((X - EX)(Y - EY)).$$

Multiplying this out and using linearity, we have an equivalent expression:

$$\operatorname{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Key Properties of Covariance

$Cov(X,Y)$
$= E[(X-EX)(Y-EY)]$
$= E[XY] - E[X] \cdot E[Y]$

- $Cov(X, X) = Var(X)$.
- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X, c) = 0$ for any constant c.
- $Cov(a \cdot X, Y) = a \cdot Cov(X, Y)$ for any constant a.
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.
- $Cov(X + Y, Z + W) =$
  $Cov(X, Z) + Cov(X, W) + Cov(Y, Z) + Cov(Y, W)$.
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.
- For n r.v.s $X_1, \ldots, X_n$,

$$Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n) + 2\sum_{i<j} Cov(X_i, Y_j).$$

Proof

Correlation

Definition

The correlation between r.v.s X and Y is

$$\operatorname{Corr}(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}.$$

(This is undefined in the degenerate cases $\operatorname{Var}(X) = 0$ or $\operatorname{Var}(Y) = 0$.)

Definition

Given r.v.s X and Y, if $\operatorname{Cov}(X, Y) = 0$ or $\operatorname{Corr}(X, Y) = 0$, X and Y are uncorrelated.

Uncorrelated

$$\mathrm{Cov}(X,Y) = E[(X-EX)(Y-EY)]$$

$$= \underline{E[(X-EX)]} \cdot E[Y-EY]$$

$$= (EX-EX)\cdot(EY-EY)$$

$$= 0$$

Theorem

If X and Y are independent, then they are uncorrelated.

Uncorrelated $\not\Rightarrow$ Independent

$\underbrace{Cov(X,Y)}_{0} = \underbrace{E[XY]}_{0} - \underbrace{E[X]}_{0} \cdot E[Y]$

Example :

1°. $X \sim N(0,1)$. $Y = X^2$

$E(X) = 0$; $E(XY) = E(X^3) = 0$

$\Rightarrow Cov(X,Y) = 0$

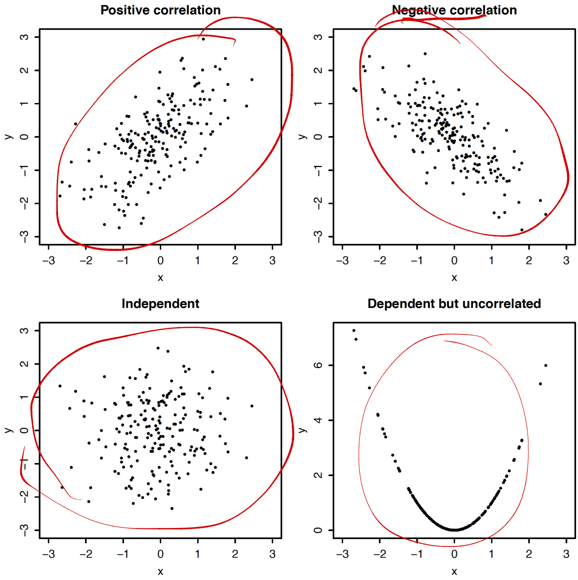
2°. X, Y are uncorrelated

However. X and Y are NOT Independent.

Covariance & Correlation

- Measure a tendency of two r.v.s X & Y to go up or down together
- Positive covariance (Correlation): when X goes up, Y also tends to go up
- Negative covariance (Correlation): when X goes up, Y tends to go down

Correlation



Cauchy - Schwarz Inequality.

Correlation Bounds

$$\rightarrow E^2[X \cdot Y] \le E[X^2] \cdot E[Y^2]$$

$$f(t) = E[(X - tY)^2] \ge 0$$

$$= E[X^2 - 2tXY + t^2Y^2]$$

$$= t^2 E[Y^2] - 2t E[XY] + E[X^2]$$

$$\Delta = (2E[XY])^2 - 4 \cdot E[X^2] \cdot E[Y^2] \le 0$$

$$ax^2 + bx + c \ge 0$$

$$\Delta = b^2 - 4ac \le 0$$

Theorem

For any r.v.s X and Y,

$$-1 \le \mathrm{Corr}(X, Y) \le 1.$$

$$\left(\sum_{i=1}^n a_i b_i\right)^2 \le \left(\sum_{i=1}^n a_i^2\right) \cdot \left(\sum_{i=1}^n b_i^2\right)$$

$$|\langle a, b\rangle|^2 \le \langle a, a\rangle \langle b, b\rangle$$

$$\cos\theta_{a,b} = \frac{\langle a, b\rangle}{|a| \cdot |b|}$$

$$|a|^2 = \langle a, a\rangle$$

$$\left(\int_a^b f(x) g(x) dx\right)^2 \le \int_a^b f^2(x) dx \cdot \int_a^b g^2(x) dx$$

$$E^2[(X - EX)(Y - EY)] \le E[(X - EX)^2] \cdot E[(Y - EY)^2]$$

$$\mathrm{Cov}^2(X, Y) \le \mathrm{Var}(X) \cdot \mathrm{Var}(Y)$$

$$\Rightarrow \mathrm{Corr}^2(X, Y) \le 1$$

$$\Rightarrow -1 \le \mathrm{Corr}(X, Y) \le 1$$

Outline

Story

$k=2$; $p_1+p_2=1$

$p_1=p$, $p_2=1-p$

Each of n objects is independently placed into one of k categories. An object is placed into category j with probability p_j, where the p_j are nonnegative and $\sum_{j=1}^{k} p_j = 1$. Let X_1 be the number of objects in category 1, X_2 the number of objects in category 2, etc., so that $X_1 + \ldots + X_k = n$. Then $X = (X_1, \ldots, X_k)$ is said to have the Multinomial distribution with parameters n and $\mathbf{p} = (p_1, \ldots, p_k)$. We write this as $\mathbf{X} \sim Mult_k(n, \mathbf{p})$.

Multinomial Joint PMF

n_1 objects $\longrightarrow$ category 1 : $p_1^{n_1}$

$\vdots$ 2 $\vdots$

n_k k $p_k^{n_k}$

Theorem

If $\mathbf{X} \sim \mathrm{Mult}_k(n, \mathbf{p})$, *then the joint PMF of* $\mathbf{X}$ *is*

$$P(\underline{X_1 = n_1}, \ldots, \underline{X_k = n_k}) = \frac{n!}{n_1! n_2! \ldots n_k!} \underline{p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}}$$

for $n_1, \ldots, n_k$ *satisfying* $n_1 + \cdots + n_k = n$.

$$\underline{\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdots \binom{n_k}{n_k}} = \frac{n!}{n_1! n_2! \ldots n_k!}$$

Proof

Multinomial Marginals

Successful events:

objects landing on jth category.

prob( ) = p_j

Theorem

If $\mathbf{X} \sim \mathrm{Mult}_k(n, \mathbf{p})$*, then* $X_j \sim \mathrm{Bin}(n, p_j)$*.*

Multinomial Lumping

A_i , A_j $\quad A_i \cap A_j = \phi$.

$P(A_i \cap A_j) = P(A_i) + P(A_j)$

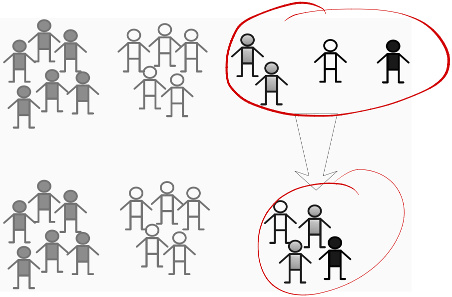
$= p_i + p_j$

Theorem

If $\mathbf{X} \sim \mathrm{Mult}_k(n, \mathbf{p})$*, then for any distinct i and j,* $X_i + X_j \sim \mathrm{Bin}(n, p_i + p_j)$*. The random vector of counts obtained from merging categories i and j is still Multinomial. For example, merging categories 1 and 2 gives*

$$(X_1 + X_2, X_3, \ldots, X_k) \sim \mathrm{Mult}_{k-1}(n, (p_1 + p_2, p_3, \ldots, p_n)).$$

Multinomial Lumpling



Multinomial Conditioning

1°. Given n_1 objects in Category 1, the remaining $n-n_1$ objects landing into Categories $2, \dots k$, independent of each other

2°. $p_j' \overset{j=2,\dots,k,\ j\neq 1}{=} \text{Prob}(\text{landing in Category } j \mid \text{not landing in category } 1)$

$= \dfrac{\text{Prob}(\text{landing in Category } j)}{\text{Prob}(\text{not landing in Category } 1)} = \dfrac{p_j}{1-p_1} = \dfrac{p_j}{p_2+\dots+p_k}$

$p_1+p_2+\dots+p_k=1$

Theorem

If $\mathbf{X} \sim \text{Mult}_k(n, p)$, *then*

$$(X_2, \dots, X_k) \mid X_1 = n_1 \sim \text{Mult}_{k-1}\left(n - n_1, \left(p_2', \dots, p_k'\right)\right),$$

where $p_j' = p_j / (p_2 + \cdots + p_k)$.

Covariance in A Multinomial

$Z \sim Bin(n,p)$

$Var(Z) = np(1-p)$

1°. w.l.o.g. $i=1, j=2$

$X_1 \sim Bin(n, p_1)$

$X_2 \sim Bin(n, p_2)$

$X_1 + X_2 \sim Bin(n, p_1+p_2)$

Theorem

Let $(X_1, \dots, X_k) \sim \mathrm{Mult}_k(n, \mathbf{p})$, *where* $\mathbf{p} = (p_1, \dots, p_k)$. *For* $i \neq j$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$.

2°. $Var(X_1+X_2) = Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

$n(p_1+p_2)(1-p_1-p_2) = np_1(1-p_1) + np_2(1-p_2) + 2Cov(X_1,X_2)$

$\Rightarrow \quad Cov(X_1, X_2) = -np_1p_2$

Proof

Outline

Multivariate Normal Distribution

Definition

A random vector $\mathbf{X} = (X_1, ..., X_k)$ is said to have a *Multivariate Normal* (MVN) distribution if every linear combination of the X_j has a Normal distribution. That is, we require

$$t_1X_1 + \cdots + t_kX_k$$

to have a Normal distribution for any choice of constants $t_1, ..., t_k$. If $t_1X_1 + \cdots + t_kX_k$ is a constant (such as when all $t_i = 0$), we consider it to have a Normal distribution, albeit a degenerate Normal with variance 0. An important special case is $k = 2$; this distribution is called the *Bivariate Normal* (BVN).

Non-example of MVN

① $X \sim N(0,1)$

$$S = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2}. \end{cases} \quad \text{random sign}$$

S independent of X.

② $Y = S \cdot X \sim N(0,1)$

$$P(Y \le y) = P(SX \le y)$$

LOTP

$$= P(SX \le y \mid S=1) P(S=1) + P(SX \le y \mid S=-1) P(S=-1)$$

$$= P(X \le y \mid S=1) \cdot \tfrac{1}{2} + P(-X \le y \mid S=-1) \cdot \tfrac{1}{2}$$

$$= P(X \le y) \cdot \tfrac{1}{2} + P(X \ge -y) \cdot \tfrac{1}{2}$$

$$= P(X \le y) \cdot \tfrac{1}{2} + P(X \le y) \cdot \tfrac{1}{2} = P(X \le y), \quad \forall y \in \mathbb{R}.$$

③ $(X,Y) \ne$ MVN $\quad P(X+Y=0) = P(S=-1) = \frac{1}{2}.$

$X+Y$ NOT Continuous. $\not\Rightarrow$ NOT Normal.

Actual MVN

① $Z, W \overset{iid}{\sim} N(0,1)$

$(Z, W) \sim$ Bivariate Normal

$\underline{t_1 Z + t_2 W}$

$\sim$ Normal

MGF
Sum of Independent
Normal Variable
is still Normal

② $(Z+2W, 3Z+5W)$ is Bivariate Normal

$t_1(Z+2W) + t_2(3Z+5W)$

$= (t_1+3t_2)Z + (2t_1+5t_2)W$

Theorem

$t_1X_1 + t_2X_2 + t_3X_3 \sim \text{Normal}$

$\forall t_1, t_2$, $t_3 = 0$

Theorem

If (X_1, X_2, X_3) is Multivariate Normal, then so is the subvector (X_1, X_2).

Theorem

Theorem

If $\mathbf{X} = (X_1, ..., X_n)$ *and* $\mathbf{Y} = (Y_1, ..., Y_m)$ *are MVN vectors with* $\mathbf{X}$ *independent of* $\mathbf{Y}$, *then the concatenated random vector* $\mathbf{W} = (X_1, ..., X_n, Y_1, ..., Y_m)$ *is Multivariate Normal.*

Parameters of MVN

① Bivariate Normal (X, Y)

mean $E(X), E(Y)$

variance $Var(X), Var(Y)$

ρ = Correlation = $\dfrac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

$$\begin{bmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{bmatrix}$$ Symmetric

Parameters of an MVN random vector $(X_1, \ldots, X_k)$ are:

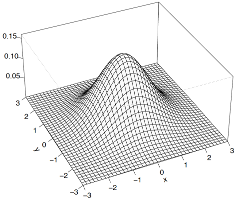
- the mean vector $(\mu_1, \ldots, \mu_k)$, where $E(X_j) = \mu_j$.
- the covariance matrix, which is the $k \times k$ matrix of covariance between components, arranged so that the row i, column j entry is $Cov(X_i, X_j)$.

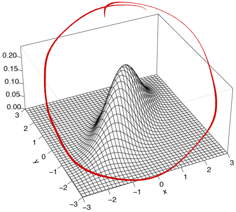
① $X, Y \sim N(0,1)$, ρ.

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

Joint PDF

Joint PDF of Bivariate Normal Distributions



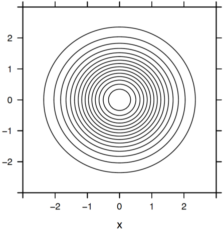


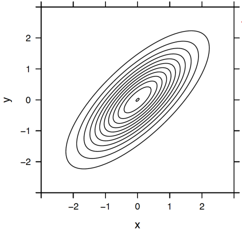
$x^2+y^2-2\rho xy = c$

Contour:

$\rho = 0.75$

$x^2+y^2=c$





$\rho=0$

Joint MGF

Definition

The *joint MGF* of a random vector $\mathbf{X} = (X_1, \ldots, X_k)$ is the function which takes a vector of constants $\mathbf{t} = (t_1, \ldots, t_k)$ and returns

$$M(\mathbf{t}) = E\left(e^{\mathbf{t}'\mathbf{X}}\right) = E\left(e^{t_1X_1+\cdots+t_kX_k}\right).$$

We require this expectation to be finite in a box around the origin in $\mathbb{R}^k$; otherwise we say the joint MGF does not exist.

Theorem

① For any Normal r.v. W, $E[e^{tW}] = e^{E[W] \cdot t + \frac{1}{2}Var(W)t^2}$

② $(X_1 \ldots X_k)$ MVN $(t_1X_1 + \cdots + t_kX_k \sim$ Normal$)$

Joint MGF $E[e^{t_1X_1 + \cdots + t_kX_k}] = \exp\{t_1E[X_1] + \cdots + t_kE[X_k] + \frac{1}{2}Var(t_1X_1 + \cdots + t_kX_k)\}$

Theorem

Within an MVN random vector, uncorrelated implies independent. That is, if $\mathbf{X} \sim \mathrm{MVN}$ *can be written as* $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2})$, *where* $\mathbf{X_1}$ *and* $\mathbf{X_2}$ *are subvectors, and every component of* $\mathbf{X_1}$ *is uncorrelated with every component of* $\mathbf{X_2}$, *then* $\mathbf{X_1}$ *and* $\mathbf{X_2}$ *are independent.*
In particular, if (X, Y) *is Bivariate Normal and* $\mathrm{Corr}(X, Y) = 0$, *then* X *and* Y *are independent.*

Proof

① Bivariate Normal (X,Y)

$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, $\mathrm{Corr}(X,Y)=\rho$.

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)$$

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\rho\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

joint MGF

② $$M_{X,Y}(s,t) = E[e^{sX+tY}] = \exp\{s\mu_1 + t\mu_2 + \tfrac{1}{2}\mathrm{Var}(sX+tY)\}$$

$$= \exp\{s\mu_1 + t\mu_2 + \tfrac{1}{2}(s^2\sigma_1^2 + t^2\sigma_2^2 + 2\rho s t\sigma_1\sigma_2)\}$$

③ if $\rho=0$ $\Rightarrow$ $M_{X,Y}(s,t) = \exp\{s\mu_1 + t\mu_2 + \tfrac{1}{2}(s^2\sigma_1^2 + t^2\sigma_2^2)\}$

$$= \exp\{s\mu_1 + \tfrac{1}{2}s^2\sigma_1^2\} \cdot \exp\{t\mu_2 + \tfrac{1}{2}t^2\sigma_2^2\}$$

$$= M_X(s)\cdot M_Y(t)$$ $\Rightarrow$ X and Y independent

Bivariate Normal Generation

[$t_1 Z + t_2 W \sim$ Normal]

1°. $Z = aX + bY$ $\quad a, b, c, d \Rightarrow (Z, W)$ BVN ✓

$W = cX + dY$ $\quad \mathrm{corr}(Z, W) = \rho$. $\quad Z, W \sim N(0,1)$

2°. $E(Z) = E(W) = 0$ $\quad \forall a, b, c, d$ $\quad [E(X) = E(Y) = 0]$

Suppose that we have access to i.i.d. r.v.s $X, Y \sim \mathcal{N}(0, 1)$, but want to generate a Bivariate Normal (Z, W) with $\mathrm{Corr}(Z, W) = \rho$ and Z, W marginally $\mathcal{N}(0, 1)$, for the purpose of running a simulation. How can we construct Z and W from linear combinations of X and Y?

$$\mathrm{Var}(Z) = a^2 \mathrm{Var}(X) + b^2 \mathrm{Var}(Y) = a^2 + b^2 = 1$$

$$\mathrm{Var}(W) = c^2 \mathrm{Var}(X) + d^2 \mathrm{Var}(Y) = c^2 + d^2 = 1$$

$$\mathrm{Corr}(Z, W) = \rho \Rightarrow \mathrm{Cov}(Z, W) = \rho \Rightarrow \mathrm{Cov}(aX + bY, cX + dY) = \rho$$

$$\Rightarrow a c \cdot \overset{1}{\mathrm{Cov}(X, X)} + b \cdot d \cdot \overset{1}{\mathrm{Cov}(Y, Y)} + 0 = \rho \Rightarrow ac + bd = \rho.$$

Solution

3°. $\begin{cases} a^2+b^2=1 \\ c^2+d^2=1 \\ ac+\underline{bd}=\rho \end{cases}$ Find one solution is enough.

$b=0 \Rightarrow a^2=1$ pick $a=1$

$\Rightarrow c=\rho \Rightarrow d^2=1-\rho^2$ pick $d=\sqrt{1-\rho^2}$

4°. $Z = aX+bY = X$.

$W = cX+dY = \rho X + \sqrt{1-\rho^2}\,Y$.

$\underline{(Z,W)}$ BVN.

$(X,Y) \longrightarrow (Z,W)$

Outline

Change of Variables in One Dimension

Theorem

Let X be a continuous r.v. with PDF f_X, and let $Y = g(X)$, where g is differentiable and strictly increasing (or strictly decreasing). Then the PDF of Y is given by

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|,$$

where $x = g^{-1}(y)$. The support of Y is all $g(x)$ with x in the support of X.

Proof

① W.L.O.G. Let g be strictly increasing.

② We consider the CDF of Y.

$Y = g(X)$

$y = g(x)$

$x = \underline{g^{-1}(y)}$

$\underline{F_Y(y)} = P(Y \le y) = P(g(X) \le y)$

$= P(X \le g^{-1}(y)) = F_X(g^{-1}(y)) = \underline{F_X(x)}$

Then by the chain rule, PDF of Y is

$f_Y(y) = F_Y'(y) = \dfrac{d F_X(x)}{\underline{dx}} \cdot \dfrac{dx}{dy}$

$= f_X(x) \cdot \dfrac{dx}{dy}$

③ g is decreasing

$f_Y(y) = f_X(x) \left(-\dfrac{dx}{dy}\right)$ $\Rightarrow$ $f_Y(y) = f_X(x) \left| \dfrac{dx}{dy} \right|$

Example: Log-Normal PDF

$\log Y \sim N(0,1)$

PDF of Y ?

① $X = \log Y$, $X \sim N(0,1)$

$Y = g(X) = e^X$ [$y = e^x > 0 \Rightarrow x = \log(y)$ $\Rightarrow \frac{dx}{dy} = \frac{1}{y}$]

② $f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right| = f_X(x) \cdot \frac{1}{y}$

$= f_X(\log y) \cdot \frac{1}{y}$ $y > 0$

③ $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

$\Rightarrow f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\log^2 y} \cdot \frac{1}{y}$, $y > 0$

Change of Variables

$(X, Y, Z) \to (\theta, \phi)$?

Theorem

Let $\mathbf{X} = (X_1, \ldots, X_n)$ *be a continuous random vector with joint PDF* $f_{\mathbf{X}}(x)$*, and let* $\mathbf{Y} = g(\mathbf{X})$ *where* g *is an invertible function from* $\mathbb{R}^n$ *to* $\mathbb{R}^n$*. Let* $y = g(\mathbf{x})$ *and suppose that all the partial derivatives* $\frac{\partial x_i}{\partial y_j}$ *exists and are continuous, so we can form the* ***Jacobian matrix***

$f_Y(y)$

$= f_X(x) \left| \frac{\partial y}{\partial x} \right|^{-1}$

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

$n \times n$

Also assume that the determinant of the Jacobian matrix is never 0. Then the joint PDF of $\mathbf{Y}$ *is*

$|\cdot|$ absolute value of determinant of Jacobian matrix

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(x) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|$$

Jacobian or not

① Discrete r.v. $X, Y > 0$. $Y = X^3$

NO Jacobian. $P(Y=y) = P(X = y^{\frac{1}{3}})$

② Continuous r.v. $X, Y > 0$, $Y = X^3$

$X = Y^{\frac{1}{3}}$ $\left(x = y^{\frac{1}{3}}, \frac{dx}{dy} = \frac{1}{3} y^{-\frac{2}{3}} \right)$

$y > 0$, $f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$

$= f_X(y^{\frac{1}{3}}) \cdot \frac{1}{3} y^{-\frac{2}{3}}$

Box-Muller

① $(X, Y) \to g(U, T)$. $\frac{\partial(u,t)}{\partial(x,y)}$

$\frac{\partial(x,y)}{\partial(u,t)}$ ✓

$x = \sqrt{2t}\cos u$
$y = \sqrt{2t}\sin u$
$x^2+y^2 = 2t \Rightarrow t = \frac{1}{2}(x^2+y^2)$
u ?

$$f_{X,Y}(x,y) = f_{U,T}(u,t) \frac{1}{\left|\frac{\partial(x,y)}{\partial(u,t)}\right|}$$

($|\det(\cdot)|$)

Let $U \sim \mathrm{Unif}(0, 2\pi)$, and let $T \sim \mathrm{Expo}(1)$ be independent of U. Define $X = \sqrt{2T}\cos U$ and $Y = \sqrt{2T}\sin U$. Find the joint PDF of (X, Y). Are they independent? What are their marginal distributions?

② $f_{U,T}(u,t) = f_T(t) \cdot f_U(u) = e^{-t} \cdot \frac{1}{2\pi}$ $\quad u \in (0, 2\pi)$, $t > 0$

③ Jacobian $\frac{\partial(x,y)}{\partial(u,t)} = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial t} \end{pmatrix} = \begin{pmatrix} -\sqrt{2t}\sin u & \frac{1}{\sqrt{2t}}\cos u \\ \sqrt{2t}\cos u & \frac{1}{\sqrt{2t}}\sin u \end{pmatrix}$

$\det\left(\frac{\partial(x,y)}{\partial(u,t)}\right) = -\sin^2 u - \cos^2 u = -1$

④ $f_{X,Y}(x,y) = e^{-t} \cdot \frac{1}{2\pi} \cdot \frac{1}{1} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$

Solution

① $f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} = \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}}_{\substack{N(0,1) \\ h(x)}} \cdot \underbrace{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}}_{\substack{N(0,1) \\ g(y)}}$

$\Rightarrow$ X and Y are independent

$X, Y \sim$ i.i.d. $N(0,1)$

Bivariate Normal Joint PDF

1°. $X, Y \sim$ i.i.d $N(0,1)$

$$\begin{cases} Z = X \\ W = \rho X + \sqrt{1-\rho^2}\, Y \end{cases} \qquad -1 < \rho < 1$$

$f_{Z,W}(z,w)$?

Bivariate Normal.

$\mathrm{Corr}(Z, W) = \rho$.

$Z, W \quad N(0,1)$.

2°. $(Z, W) = g(X, Y)$

$$f_{Z,W}(z,w) = f_{X,Y}(x,y) \cdot \left| \frac{\partial(x,y)}{\partial(z,w)} \right|$$

3°. Jacobian. $\begin{cases} z = x \\ w = \rho x + \sqrt{1-\rho^2}\, y \end{cases} \Rightarrow \begin{cases} x = z \\ y = \frac{1}{\sqrt{1-\rho^2}} w - \frac{\rho}{\sqrt{1-\rho^2}} z \end{cases}$

$$\frac{\partial(x,y)}{\partial(z,w)} = \begin{pmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix}$$

$$\Rightarrow \det\left(\frac{\partial(x,y)}{\partial(z,w)} \right) = \frac{1}{\sqrt{1-\rho^2}} \qquad \Rightarrow f_{Z,W}(z,w) = f_{X,Y}(x,y) \frac{1}{\sqrt{1-\rho^2}}.$$

Bivariate Normal Joint PDF

$X, Y \sim$ iid $N(0,1)$

$$\begin{cases} x = z \\ y = \frac{1}{\sqrt{1-\rho^2}} w - \frac{\rho}{\sqrt{1-\rho^2}} z \end{cases}$$

$4°$. $f_{Z,W}(z,w) = f_{X,Y}(x,y) \cdot \frac{1}{\sqrt{1-\rho^2}} = f_X(x) \cdot f_Y(y) \cdot \frac{1}{\sqrt{1-\rho^2}}$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \cdot \frac{1}{\sqrt{1-\rho^2}}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}(x^2+y^2)}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}\left[z^2 + \left(\frac{1}{\sqrt{1-\rho^2}} w - \frac{\rho}{\sqrt{1-\rho^2}} z\right)^2\right]}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[z^2 + w^2 - 2\rho z w\right]}$$

, $z, w \in \mathbb{R}$.

---

if $\rho = 0$ $\Rightarrow f_{Z,W}(z,w) = \frac{1}{2\pi} e^{-\frac{1}{2}(z^2+w^2)}$

Z, W independent

Outline

Convolution Sums and Integrals

$T' = X - Y$

$T'' = \frac{X}{Y}$

$T''' = X \cdot Y$

Theorem

If X and Y are independent discrete r.v.s, then the PMF of their sum $T = X + Y$ is

$$P(X+Y=t) \overset{\text{LOTP}}{=} \sum_x P(X+Y=t \mid X=x)\, P(X=x)$$

$$\begin{aligned} P(T=t) &= \sum_x P(Y=t-x)\, P(X=x) = \sum_x P(Y=t-x \mid X=x) \cdot P(X=x) \\ &= \sum_y P(X=t-y)\, P(Y=y) = \sum_x P(Y=t-x) \cdot P(X=x) \end{aligned}$$

If X and Y are independent continuous r.v.s, then the PDF of their sum $T = X + Y$ is

$$\begin{aligned} f_T(t) &= \int_{-\infty}^{\infty} f_Y(t-x)\, f_X(x)\, dx \\ &= \int_{-\infty}^{\infty} f_X(t-y)\, f_Y(y)\, dy \end{aligned}$$

Proof

① For continuous r.v.s.

$$\underline{F_T(t)} = P(X+Y\le t) \overset{\text{LOTP}}{=} \underline{\int_{-\infty}^{\infty} P(X+Y\le t \mid X=x)\cdot f_X(x)\,dx}$$

$$= \int_{-\infty}^{\infty} P(Y\le t-x \mid X=x)\cdot f_X(x)\,dx = \int_{-\infty}^{\infty} P(Y\le t-x)\, f_X(x)\,dx$$

$$= \underline{\int_{-\infty}^{\infty} F_Y(t-x)\, f_X(x)\,dx} \Rightarrow \text{Diff w.r.t. } t$$

$$f_T(t) = \underline{\int_{-\infty}^{\infty} f_Y(t-x) f_X(x)\,dx}$$

Continuous r.v.

$T = X+Y$; $V = X$; $(T, V) = g(X, Y)$

$$\begin{cases} t = x+y \\ v = x \end{cases} \Rightarrow \begin{cases} x = v \\ y = t-v \end{cases} \Rightarrow J = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

$$\Rightarrow f_{T,V}(t,v) = f_{X,Y}(x,y)\cdot |J| = f_X(x)\cdot f_Y(y)\cdot 1 = f_X(x) f_Y(y)$$

$$\Rightarrow f_T(t) = \int_{-\infty}^{\infty} f_{T,V}(t,v)\,dv = \int_{-\infty}^{\infty} f_X(\underline{x})\, \underline{f_Y(y)}\,dv$$

$$= \underline{\int_{-\infty}^{\infty} f_X(v)\cdot f_Y(t-v)\,dv} = \int_{-\infty}^{\infty} f_X(x)\, f_Y(t-x)\,dx$$

Exponential Convolution

$t-x \geq 0$

$x \geq 0$

$\Rightarrow 0 \leq x \leq t$

$X \geq 0$, $Y \geq 0$, $T = X+Y \geq 0$

$\forall t \geq 0 \quad f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_X(x)\,dx$

$\lambda e^{-\lambda t}$

$$= \int_0^t \lambda e^{-\lambda(t-x)} \cdot \lambda e^{-\lambda x}\,dx$$

Let $X, Y \overset{\text{i.i.d.}}{\sim} \mathrm{Expo}(\lambda)$. Find the distribution of $T = X + Y$.

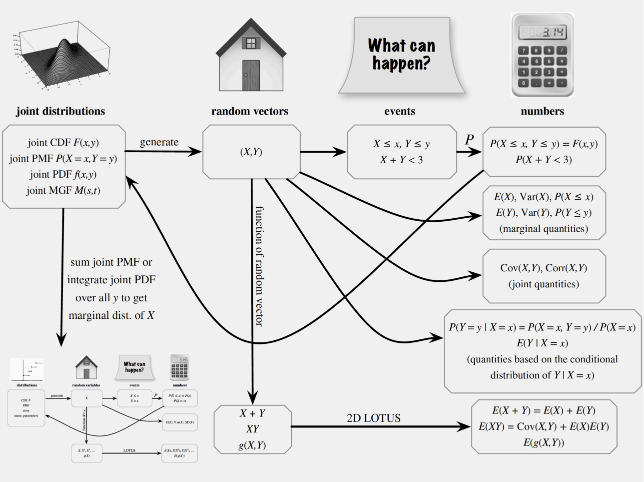
$$= \int_0^t \lambda^2 e^{-\lambda t}\,dx$$

$$= \lambda^2 e^{-\lambda t} \int_0^t dx = \lambda^2 t e^{-\lambda t}$$

Summary 1: Discrete & Continuous

	Two discrete r.v.s	Two continuous r.v.s
Joint CDF	$F_{X,Y}(x,y) = P(X \le x, Y \le y)$	$F_{X,Y}(x,y) = P(X \le x, Y \le y)$
Joint PMF/PDF	$P(X = x, Y = y)$ • Joint PMF is nonnegative and sums to 1: $\sum_x \sum_y P(X = x, Y = y) = 1$.	$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$ • Joint PDF is nonnegative and integrates to 1: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$. • To get probability, integrate joint PDF over region of interest.
Marginal PMF/PDF	$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y P(X = x \mid Y = y)P(Y = y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{\infty} f_{X\mid Y}(x\mid y) f_Y(y)dy$
Conditional PMF/PDF	$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)}$	$f_{Y\mid X}(y\mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X\mid Y}(x\mid y) f_Y(y)}{f_X(x)}$
Independence	$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$ $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all x and y. $P(Y = y \mid X = x) = P(Y = y)$ for all x and y, $P(X = x) > 0$.	$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$ $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all x and y. $f_{Y\mid X}(y\mid x) = f_Y(y)$ for all x and y, $f_X(x) > 0$.
LOTUS	$E(g(X,Y)) = \sum_x \sum_y g(x,y) P(X = x, Y = y)$	$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dxdy$

Summary 2: Multivariate Distribution



References

- Chapters 7 & 8 of **BH**
- Chapters 2 & 3 & 4 of **BT**