Lecture 6: Joint Distributions

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

November 14, 2024

Outline

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal
6. Change of Variables
7. Convolutions

Multivariate Distribution

- 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

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal
6. Change of Variables
7. Convolutions

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

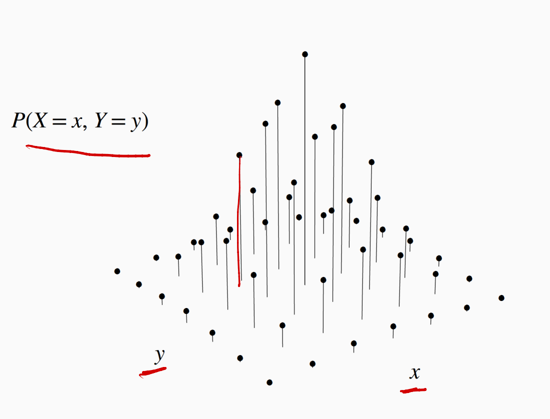
Definition

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

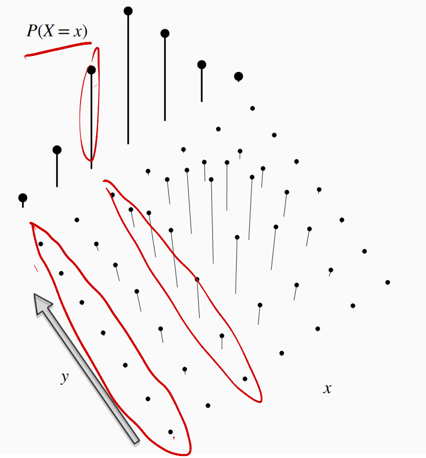
Definition

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

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

LOTP

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

$y \backslash x$	1	2	3	4	Row sums
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	3/20	6/20	8/20	3/20	

Row sums: marginal PMF $p_Y(y)$

Column sums: marginal PMF $p_X(x)$

$p_X(1)$

$= P(X=1)$

$= P(X=1, Y=1)$

$+ P(X=1, Y=2)$

$+ P(X=1, Y=3)$

$+ P(X=1, Y=4)$

$= \frac{1}{20} + \frac{1}{20} + \frac{1}{20}$

$+ \frac{\cancel{1}}{20} = \frac{3}{20}$

Conditional PMF

1°. Conditional PMF is also a valid PMF.

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

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

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

Definition

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

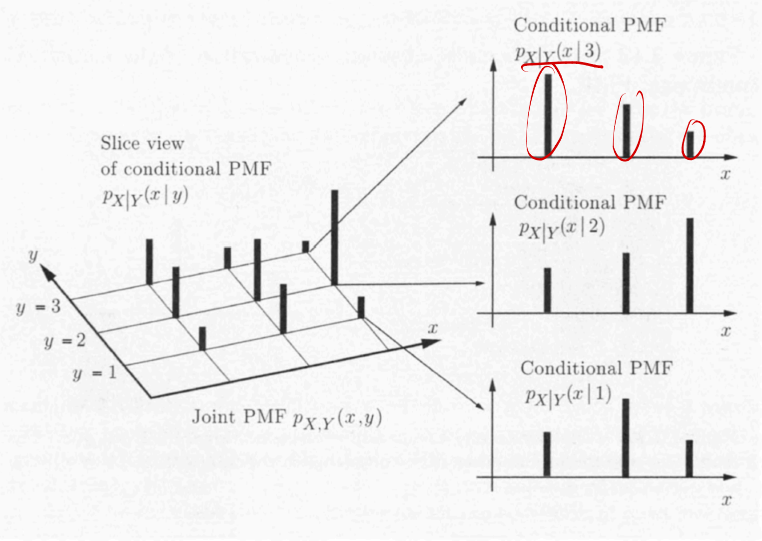
$$P_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

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

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

Conditional PMF

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



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

$i, j \geq 0$, integer.

② $X+Y=N$.

$X+Y|_{N=n} = n$.

$X|N=n \sim \text{Bin}(n,p)$

$Y|N=n \sim \text{Bin}(n,q)$ $q=1-p$.

Suppose a chicken lays a random number of eggs, N, where $N \sim \text{Pois}(\lambda)$. 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 $P(X=i, Y=j) \overset{\text{LOTP}}{=} \sum_{n=0}^{\infty} P(X=i, Y=j \mid N=n) \cdot P(N=n)$

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

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

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

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

$X \mid \sim Pois(\lambda)$

$p+q=1$

$\lambda p + \lambda q = \lambda$

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

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

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

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

$\sim Pois(\lambda p) \cdot Pois(\lambda q)$

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

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

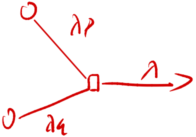
$$P(Y=j) = \sum_{i=0}^{\infty} P(X=i, Y=j) = \frac{e^{-\lambda q} \cdot (\lambda q)^j}{j!} \sim Pois(\lambda q)$$

X, Y are independent.

Are NOT conditional independent.

Solution

Related Theorem

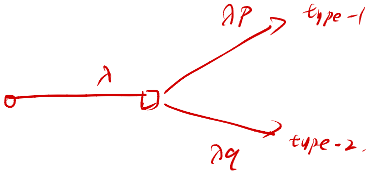


Theorem

$p+q=1$

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

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal
6. Change of Variables
7. Convolutions

Conditional PDF Given an Event

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

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

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

(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).

$\int_{-\infty}^{x} f_X(t)\,dt$

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

Taking the derivative of both sides w.r.t. x.

Proof

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

$$= \lim_{\delta\to 0} \frac{P(x\le X\le x+\delta,\ X\in A)}{\delta\cdot P(X\in A)}$$

$$= \lim_{\delta\to 0} \frac{P(X\in A \mid x\le X\le x+\delta)\cdot P(x\le X\le x+\delta)}{\delta\cdot P(X\in A)}$$

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

$$= \frac{1}{P(X\in A)}\cdot f_X(x)\cdot \underbrace{\boxed{P(X\in A \mid X=x)}}_{1_{\{x\in A\}}} = \begin{cases} 1 & \text{if } x\in A \\ 0 & \text{otherwise} \end{cases}$$

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

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

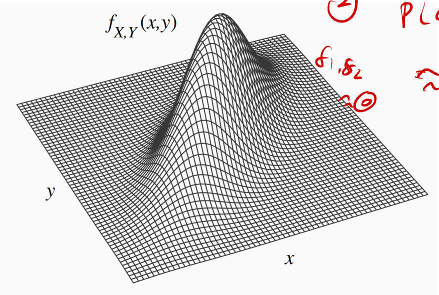
① $\delta \approx 0$ $\quad P(a \le X \le a+\delta) = \int_a^{a+\delta} f_X(x)\,dx$

$\approx f_X(a)\cdot\delta.$

② $P(a \le X \le a+\delta_1,\ b \le Y \le b+\delta_2)$

$\delta_1, \delta_2 \approx 0$

$\approx f_{X,Y}(a,b)\,\delta_1\delta_2$



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

Definition

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

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

$$\Leftrightarrow \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)}{f_X(x)}\,dy \overset{?}{=} 1$$ ✓

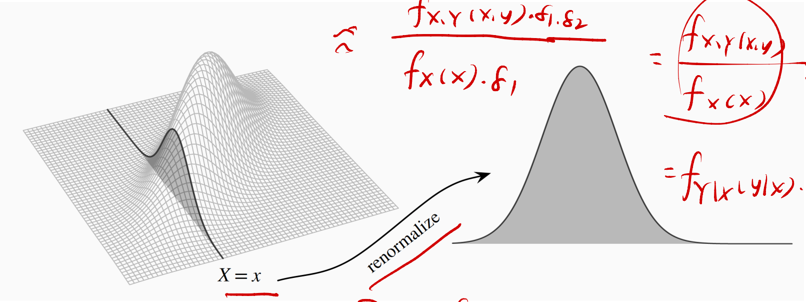
$$\Leftrightarrow \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \overset{\checkmark}{=} f_X(x)$$

Conditional PDF

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

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

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



② $\delta_1 \to 0$

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

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 $0<x<1$, $0<y<1$,

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

The joint PDF of X and Y is given by

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

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

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

Example

① Conditional PDF $\quad 0<x<\infty,\ 0<y<\infty.$

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_0^\infty f(x,y)dx} = \frac{\frac{e^{-x/y-y}}{y}}{\int_0^\infty \frac{e^{-x/y-y}}{y}dx} = \frac{1}{y}e^{-x/y}$$

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

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

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

Theorem

For continuous r.v.s X and Y,

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

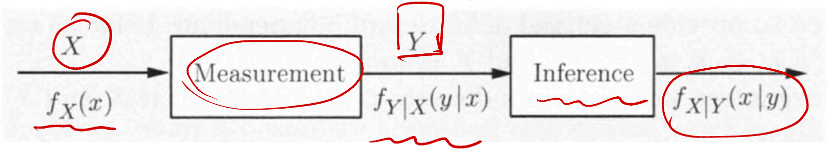
② $$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)\, f_Y(y)\, dy$$

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

Proof

Bayes' Rule: Inference Perspective

info/data



Prior PDF

Likelihood

posterior PDF

Example

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

Prior PDF $\sim \text{unif}(1, \frac{3}{2})$

② posterior PDF $Y = y$

$$f_{\Lambda|Y}(\lambda|y) = \frac{f_\Lambda(\lambda) \cdot f_{Y|\Lambda}(y|\lambda)}{f_Y(y)}$$

$\sim \text{Expo}(\lambda)$

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?

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

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

General Bayes' Rule

① X discrete ; Y continuous.

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

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

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

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

✓

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.

$\downarrow$

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

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

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

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

Proof

General LOTP

② X continuous, Y discrete.

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

$$f_X(x) = \sum_y f_X(x \mid Y=y) \cdot P(Y=y)$$

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

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

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

$$= P(X=x)$$

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

1

Proof

Example

① $Y = N + S$; $Y|_{S=s} = N + s \sim \mathcal{N}(s, 1)$ PDF

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

$f_{Y|S}(y|1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-1)^2}$

$f_{Y|S}(y|-1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y+1)^2}$

② $P(S=1 | Y=y) = \dfrac{f_{Y|S}(y|1)}{f_Y(y)} \cdot P(S=1)$

posterior prob.

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

Prior

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?

③ $$P(S=1 | Y=y) = \frac{\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(y-1)^2} \cdot p}{\cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(y-1)^2} \cdot p + \cancel{\frac{1}{\sqrt{2\pi}}} e^{-\frac{1}{2}(y+1)^2} (1-p)}$$

$$= \frac{pe^{y}}{p \cdot e^{y} + (1-p)e^{-y}} = \frac{p}{p + (1-p)e^{-2y}} = \begin{cases} > p & y > 0 \\ < p & y < 0 \\ p & y = 0 \end{cases}$$

Example: Comparing Exponentials of Different Rates

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

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

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

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

$P(T_1=T_2) = 0$

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

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

$$= 1 - \frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{\lambda_1}{\lambda_1+\lambda_2}$$

$P(T_1 = \min(T_1,T_2))$

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

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

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

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

$f_{X,Y}(x,y) = 8xy$, $0 \le x \le y \le 1$

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

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

decouple of joint support zone

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

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

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

① $f_{X,Y}(x,y) = g(x)\cdot h(y) = \underline{c\cdot g(x)\cdot \frac{h(y)}{c}}$, $c = \int_{-\infty}^{\infty} h(y)dy$

$\Rightarrow 1 = \int_{-\infty}^{\infty} \underline{\frac{1}{c}h(y)}dy \quad \Rightarrow f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$

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

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

$$\Rightarrow f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{h(y)}{c}\cdot \underbrace{\int_{-\infty}^{\infty} c\cdot g(x)dx}_{1} = \underline{\frac{1}{c}h(y)}$$

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

X, Y are independent.

if g is a valid PDF, $\Rightarrow c=1 \Rightarrow h(y)$ 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

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 $E(|X-Y|)$.

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

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

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

Outline

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal
6. Change of Variables
7. Convolutions

Covariance

Definition

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

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

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

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

Key Properties of Covariance

- $Cov(X, X) = Var(X)$.
- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X, c) = 0$ for any constant c.
- $Cov(a \cdot X, Y) = a \cdot Cov(X, Y)$ for any constant a.
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.
- $Cov(X + Y, Z + W) =$
  $Cov(X, Z) + Cov(X, W) + Cov(Y, Z) + Cov(Y, W)$.
- $\boxed{Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).}$
- 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

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

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

Definition

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

Uncorrelated

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

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

$$= (EX-EX)(EY-EY)$$

$$= 0$$

Theorem

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

Uncorrelated $\not\Rightarrow$ Independent

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

0

Example:

$X \sim N(0,1)$; $Y = X^2$.

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

$\Rightarrow \mathrm{Cov}(X,Y) = 0$

X, Y are uncorrelated;

NOT Independent.

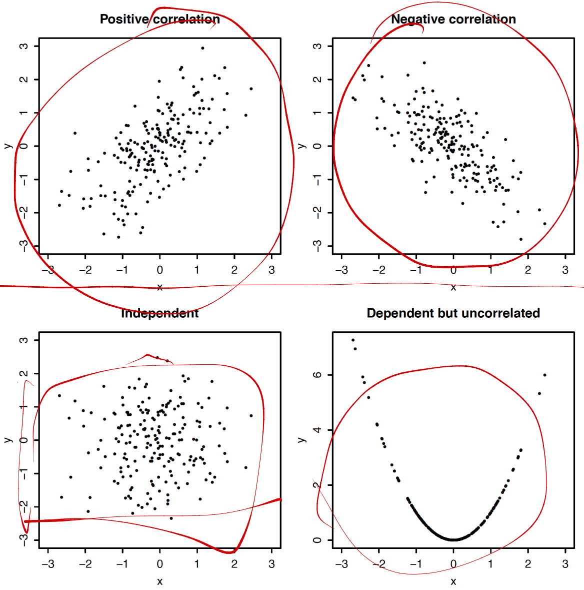
Correlation (linear correlation)

NOT Capture Nonlinear correlation

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



Correlation Bounds

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

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

$1^\circ.$ $E^2[X \cdot Y] \leq E[X^2] \cdot E[Y^2]$ $\quad (E[X^2] < \infty)$

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

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

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

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

$\Rightarrow E^2[X \cdot Y] \leq E[X^2] \cdot E[Y^2]$

Theorem

For any r.v.s X and Y,

$$= \frac{E[(X-EX)(Y-E(Y))]}{\sqrt{Var(X) \cdot Var(Y)}}$$

$$-1 \leq \operatorname{Corr}(X, Y) \leq 1.$$

$2^\circ.$ $X \leftarrow X - EX$, $Y \leftarrow Y - EY$.

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

$Cov^2(X,Y) \leq Var(X) \cdot Var(Y)$

$\Rightarrow Corr^2(X,Y) \leq 1 \Rightarrow -1 \leq Corr(X,Y) \leq 1$

Outline

Cauchy-Schwarz Inequality

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

$$|\langle \vec{a}, \vec{b} \rangle|^2 \le \langle a, a \rangle \cdot \langle b, b \rangle$$

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

$$|\vec{a}|^2 = \langle \vec{a}, \vec{a} \rangle$$

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal
6. Change of Variables
7. Convolutions

Story

$k=2$; $\rightarrow$ Binomial

(p_1, p_2) $p_1+p_2=1$

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 $\mathbf{X} \sim Mult_k(n, \mathbf{p})$.

$X = (X_1, \ldots, X_k)$

$X = (X_1, \ldots, X_k)^T \left(\begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \right)$

Multinomial Joint PMF

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

n_k - - - $\longrightarrow$ k : $p_k^{n_k}$

Theorem

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

$p_1^{n_1} \cdot p_2^{n_2} \cdots p_k^{n_k}$

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

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

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

Proof

In practice:

Categorical Distribution

$$Y_1, \ldots, Y_n \overset{iid}{\sim} \mathrm{Cat}(k, \mathbf{p}), \quad \mathbf{p} = (p_1, \ldots, p_k)$$

$$\begin{cases} p_1 + \cdots + p_k = 1; \\ p_i > 0 \end{cases}$$

$$p_i = P(\underbrace{Y_1 = i}_{\text{the } i^{th} \text{ category is selected}}), \quad i = 1, \ldots, k.$$

$X_1, \ldots, X_k$

$$X_j = \underbrace{\sum_{m=1}^{n} I_{\{Y_m = j\}}}_{\#\text{ of objects landing into category } j}, \quad j = 1, \ldots, k.$$

$$(X_1, \ldots, X_k) \sim \mathrm{Mult}_k(n, \mathbf{p}), \quad \mathbf{p} = (p_1, \ldots, p_k)$$

Multinomial Marginals

Model & Story

of Binomial Distribution

Theorem

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

Define "success"!

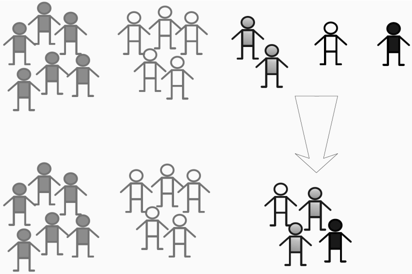
Multinomial Lumping

Theorem

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

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

Multinomial Lumpling



Multinomial Conditioning

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

Theorem

2°. $p_j' = \text{Prob}(\underbrace{\text{landing into Category } j}_{\text{event } A} \mid \underbrace{\text{not landing into Category 1}}_{\text{event } B})$ $(j=2,\ldots k, (j \neq 1))$

$= \text{Prob}(A \mid B)$, $A \subseteq B$;

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

$$(X_2, \ldots, X_k) \,|X_1 = n_1 \sim \mathrm{Mult}_{k-1}(n - n_1, (p_2', \ldots, p_k')),$$

where $p_j' = p_j / (p_2 + \cdots + p_k)$. $= \dfrac{\text{Prob}(A)}{\text{Prob}(B)} = \dfrac{p_j}{1-p_1} = \dfrac{p_j}{p_2 + \cdots + p_k}$

$p_1 + \cdots + p_k = 1$

Covariance in A Multinomial

$Z \sim Bin(n,p)$

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

1°. W.L.O.G. let $i=1$, $j=2$

$X_1 \sim Bin(n, p_1)$

$X_2 \sim Bin(n, p_2)$

Theorem

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

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

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

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

$\Rightarrow Cov(X_1, X_2) = -np_1p_2 < 0$

Proof

Outline

1 Discrete Multivariate R.V.s

2 Continuous Multivariate R.V.s

3 Covariance and Correlation

4 Multinomial Distribution

5 Multivariate Normal

6 Change of Variables

7 Convolutions

Multivariate Normal Distribution

Definition

A random vector $\mathbf{X} = (X_1, \dots, X_k)$ 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

$$t_1 X_1 + \cdots + t_k X_k$$

to have a Normal distribution for any choice of constants $t_1, \dots, t_k$. If $t_1 X_1 + \cdots + t_k X_k$ 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 $X \sim N(0,1)$ $\quad \underline{X} \sim \underline{-X}$

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

S is independent of X

② $Y = SX \sim N(0,1)$

CDF of Y

$P(Y \le y) = P(S \cdot X \le y)$

$\overset{\text{LOTP}}{=} P(S \cdot X \le y \mid S=1) \cdot P(S=1) + P(S \cdot X \le y \mid S=-1) \cdot P(S=-1)$

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

$\overset{X \perp\!\!\!\perp S}{=} P(X \le y) \cdot \frac{1}{2} + \underbrace{P(-X \le y)}_{=P(X \le y)} \cdot \frac{1}{2}$

$\overset{X \sim -X}{=} P(X \le y), \ \forall y \in \mathbb{R}$

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

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

$= \frac{1}{2}$.

$X+Y$ NOT Continuous r.v $\Rightarrow$ Not Normal.

Actual MVN

① $Z, W \overset{iid}{\sim} N(0,1)$

$(Z, W) \sim$ Bivariate Normal.

$t_1 Z + t_2 W \sim$ Normal $\quad \forall t_1, t_2 \in \mathbb{R}$.

② $(Z+2W,\ 3Z+5W)$ is a Bivariate Normal R.V.

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

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

Theorem

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

$\forall t_1, t_2, t_3$

Theorem

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

$t_3 = 0$

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

$\forall t_1, t_2$

Theorem

$\mathbf{X}$ independent of $\mathbf{Y}$

iff X_i independent of Y_j

$i=1,\dots,n$; $j=1,\dots,m$

Theorem

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

Parameters of MVN

$k=2$; $\det\left(\begin{bmatrix} Var(X_1) & Cov(X_1,X_2) \\ Cov(X_2,X_1) & Var(X_2) \end{bmatrix}\right)$; ≥ 0

mean $E[X_1], E[X_2]$

Variance $Var[X_1], Var[X_2]$; $\rho = \dfrac{Cov(X_1,X_2)}{\sqrt{Var(X_1)\cdot Var(X_2)}}$;

$Cov(X_1,X_2) = Cov(X_2,X_1)$

Joint PDF

$[Var(X_1)\cdot Var(X_2) = Cov^2(X_1,X_2)] \Rightarrow \rho^2=1$

Positive Semidefinite

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 $Cov(X_i, X_j)$.

$i=j$ $Cov(X_i,X_j) = Var(X_i)$

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x_1^2+x_2^2-2\rho x_1x_2)\right\}$$

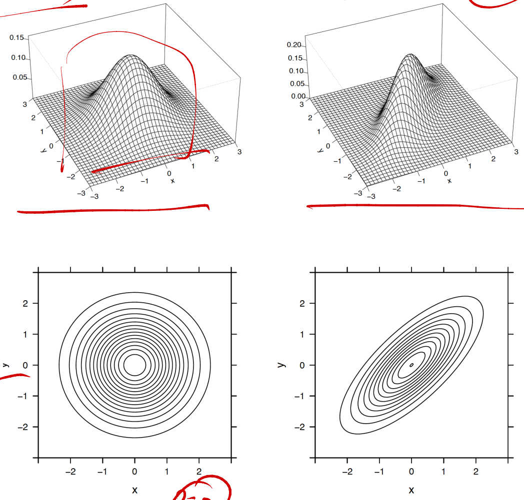
Joint PDF $f(X_1, X_2)$ $\rho^2 \neq 1$

Covariance Matrix Positive Define

Joint PDF of Bivariate Normal Distributions

$x_1^2+x_2^2-2\rho x_1x_2=c \Rightarrow$ joint PDF $f_{X_1,X_2}(x_1,x_2)=c'$

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



$x^2+y^2=c$

$\rho=0.75$

$\rho=0$

Joint MGF

2° if $t_1X_1 + \cdots + t_kX_k \sim$ Normal

$\Rightarrow M(t) = ?$

3°. MVN: $t_1X_1 + \cdots + t_kX_k \sim$ Normal

Definition

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

1°. $X \sim$ MVN $\Rightarrow M(t)$?

Theorem

$\forall\ W \sim \text{Normal}$

$$E[e^{tW}] = e^{E[W]\cdot t + \frac{1}{2}\text{Var}(W)\cdot t^2} = e^{E[tW] + \frac{1}{2}\text{Var}(tW)}$$

$(X_1, \ldots, X_k)$ $\quad (t_1X_1 + \cdots + t_kX_k)$

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

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)

$X \sim N(\mu_1, \sigma_1^2)$; $Y \sim N(\mu_2, \sigma_2^2)$.

$\mathrm{Corr}(X, Y) = \rho$;

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

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

$t_1 X + t_2 Y \sim$ Normal $\forall t_1, t_2$

$t_1 = 0$, $t_2 = 1$.

$t_1 = 1$, $t_2 = 0$

$X, Y \sim$ Normal.

② Joint MGF. $sX + tY \sim$ Normal (W)

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

$= e^{s\mu_1 + t\mu_2 + \frac{1}{2}(s^2\sigma_1^2 + t^2\sigma_2^2 + 2\rho s \cdot t \cdot \sigma_1 \cdot \sigma_2)}$

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

$= e^{s\mu_1 + \frac{1}{2}s^2\sigma_1^2} \cdot e^{t\mu_2 + \frac{1}{2}t^2\sigma_2^2} = M_X(s) \cdot M_Y(t)$ $\Rightarrow$ X and Y are independent.

Bivariate Normal Generation

Find $a, b, c, d \Rightarrow (Z, W) \sim BVN$, $Corr(Z,W) = \rho$, $Z, W \sim N(0,1)$

1°. $Z = aX + bY$

$W = cX + dY$

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

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 $\operatorname{Corr}(Z, W) = \rho$ 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 ?

$Var(Z) = a^2 Var(X) + b^2 Var(Y) = a^2 + b^2 = 1$

$Var(W) = c^2 Var(X) + d^2 Var(Y) = c^2 + d^2 = 1$

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

$Var(Z) = Var(W) = 1$

$\Rightarrow Cov(aX, cX) + Cov(bY, dY) = \rho$

$\Rightarrow ac\,Var(X) + bd\,Var(Y) = \rho$

$\Rightarrow ac + bd = \rho$

Solution

3°. $\begin{cases} a^2+b^2=1 \\ c^2+d^2=1 \\ ac+bd=\rho \end{cases}$

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°. $Z = aX+bY = X$

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

$(X, Y) \overset{i.i.d}{\sim} N(0,1) \xrightarrow{a,b,c,d} (Z, W)$ BVN

Check the joint PDF of (Z, W)

Outline

$X \quad g \quad g(X)$

PDF
PMF

PDF
PMF ?

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal
6. Change of Variables
7. Convolutions

$(X_1, X_2) \quad g=(g_1, g_2) \quad g(X_1, X_2)$

$\begin{pmatrix} g_1(X_1) \\ g_2(X_2) \end{pmatrix}$

joint PDF
joint PMF

joint PDF
joint PMF ?

Change of Variables in One Dimension

$y = g(x)$

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

$dx = h'(y)\,dy$

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

x

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

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.

$Y = g(X)$
$y = g(x)$
$x = g^{-1}(y)$

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

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

Then by the chain rule, PDF of Y is

$f_Y(y) = F_Y'(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dy} = \frac{dF_X(x)}{dx} \cdot \frac{dx}{dy}$

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

③ g is strictly decreasing.

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

Example: Log-Normal PDF

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

Y ?

① $X = \log Y$, $X \sim N(0,1)$

$Y = g(X) = e^X$

$[\; y = e^x > 0$, $x = \log y$

$\Rightarrow \frac{dx}{dy} = \frac{1}{y} \;]$

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

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

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

Change of Variables

$f_Y(y)\, dV_y = f_X(x) \cdot dV_x$

Theorem

$f_Y(y) = f_X(x) \left|\frac{dV_x}{dV_y}\right|$

Let $\mathbf{X} = (X_1, ..., X_n)^T$ *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**

$f_X(x) = f_Y(y) \frac{dV_y}{dV_x}$

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

Local Linear Transformation

$n \times n$ square matrix

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.

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

Jacobian or not

① Discrete r.v. / r.v

No Jacobian.

$X, Y > 0$

$Y = X^3$

$P(Y=y)$

$= P(X = y^{\frac{1}{3}})$

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

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

$y > 0$, $f_Y(y) = f_X(x) \cdot \left|\frac{dx}{dy}\right| = f_X(y^{\frac{1}{3}}) \cdot \frac{1}{3} y^{-\frac{2}{3}}$

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

Box-Muller

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

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

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

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

Let $U \sim \text{Unif}(0, 2\pi)$, and let $T \sim \text{Expo}(1)$ 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?

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

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

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

Solution

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

$\Rightarrow$ X and Y are independent.

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

Bivariate Normal Joint PDF

1°. $X, Y \overset{i.i.d}{\sim} N(0,1)$

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

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

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

Bivariate Normal Joint PDF

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

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

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

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

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

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

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[z^2 + w^2 - 2\rho z w\right]}, \quad z, w \in \mathbb{R}.$$

BVN, $\underline{\text{Corr}} = \rho$, Marginal $N(0,1)$.

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

Outline

1 Discrete Multivariate R.V.s

2 Continuous Multivariate R.V.s

3 Covariance and Correlation

4 Multinomial Distribution

5 Multivariate Normal

6 Change of Variables

7 Convolutions

Convolution Sums and Integrals

$T = X \cdot Y$; $T = X/Y$

$T = X - Y$; ...

Theorem

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

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

$$P(T=t) = \sum_x P(Y=t-x) P(X=x)$$

$= \sum_x P(Y=t-x \mid X=x) \cdot P(X=x)$

$$= \sum_y P(X=t-y) P(Y=y)$$

$= \sum_x P(Y=t-x) \cdot P(X=x)$

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

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

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

Proof

① continuous r.v.s.

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

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

$$= \int_{-\infty}^{\infty} F_Y(t-x)\, f_X(x) \cdot dx$$

Diff w.r.t. t

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

② $T = X+Y$; $V = X$ $\Rightarrow (T, V) = g(X, Y)$

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

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

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

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

Exponential Convolution

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

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

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

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

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

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

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

$= \lambda^2 t \cdot e^{-\lambda t}$

Summary 1: Discrete & Continuous

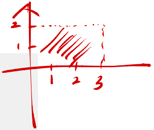
	Two discrete r.v.s	Two continuous r.v.s
Joint CDF	$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$	$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$
Joint PMF/PDF	$P(X = x, Y = y)$ • Joint PMF is nonnegative and sums to 1: $\sum_x \sum_y P(X = x, Y = y) = 1$.	$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$ • Joint PDF is nonnegative and integrates to 1: $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx dy = 1$. • 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	$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$ $P(X = x, Y = y) = P(X = x) P(Y = y)$ for all x and y.	$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$ $f_{X,Y}(x,y) = f_X(x) f_Y(y)$ for all x and y.
	$P(Y = y \mid X = x) = P(Y = y)$ for all x and y, $P(X = x) > 0$.	$f_{Y\mid X}(y\mid x) = f_Y(y)$ for all x and y, $f_X(x) > 0$.
LOTUS	$E(g(X,Y)) = \sum_x \sum_y g(x,y) P(X = x, Y = y)$	$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dx dy$

Summary 2: Multivariate Distribution

$A = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ $\det(A) = 6$

Area $= 6$

①



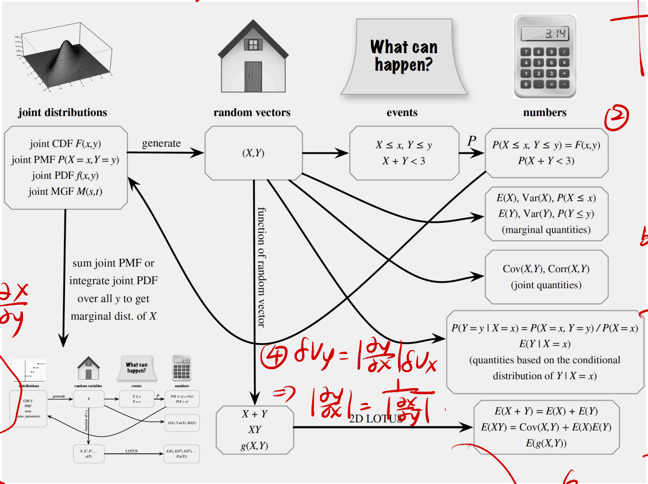
③ $dx_1 = \frac{\partial x_1}{\partial y_1} dy_1 + \cdots + \frac{\partial x_1}{\partial y_n} dy_n$

$\vdots$

$dx_n = \frac{\partial x_n}{\partial y_1} dy_1 + \cdots + \frac{\partial x_n}{\partial y_n} dy_n$

$\Rightarrow [dx_1, \dots dx_n] = [dy_1, \dots dy_n] \frac{\partial x}{\partial y}$

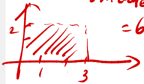
Local Linear Transformation.



② A is a linear transformation

before unit-area $= 1$

after unit-area $= 6$



④ $\delta V_y = \left|\frac{\partial y}{\partial x}\right| \delta V_x$

$\Rightarrow \left|\frac{\partial y}{\partial x}\right| = \frac{1}{\left|\frac{\partial x}{\partial y}\right|}$

$\frac{6}{1} = 6 = |\det(A)|$

$\delta V_x = \prod_{i=1}^{n} dx_i$

$\delta V_y = \prod_{i=1}^{n} dy_i$

$\left.\begin{array}{l} \delta V_x = \left|\frac{\partial x}{\partial y}\right| \delta V_y \\ f_X(x)\,\delta V_x = f_Y(y)\,\delta V_y \end{array}\right\} \Rightarrow f_Y(y) = f_X(x) \left|\frac{\partial x}{\partial y}\right|$

References

Geometric view of Jacobian Determinant :

① Static view of determinant of $n\times n$ matrix A: the volume ($n\geq 3$) (area, $n=2$) of parallelohedron ($n\geq 3$) 平行多面体 (parallelogram ($n=2$) 平行四边形) spanned by all row vectors (column vectors) of A.

② dynamic view of determinant of $n\times n$ matrix A: A is a linear transformation, the ratio of unit volume ($n\geq 3$) (area, $n=2$) after/before transformation.

③ $n\times n$ Jacobian matrix J : Local Linear transformation (化曲为直, 分段线性)

determinant of $n\times n$ Jacobian matrix J is the ratio of δ-volume ($n=3$) (area, $n=2$) after/before Local Linear Transformation.

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

Unit volume 单位体积

δ-volume 微元体积