联合分布

Lecture 6: Joint Distributions

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

April 11, 2023

Outline

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. 协方差 相关性
   Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal

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

Joint CDF

Definition

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

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

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

$$F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n).$$

Joint PMF

联合分布

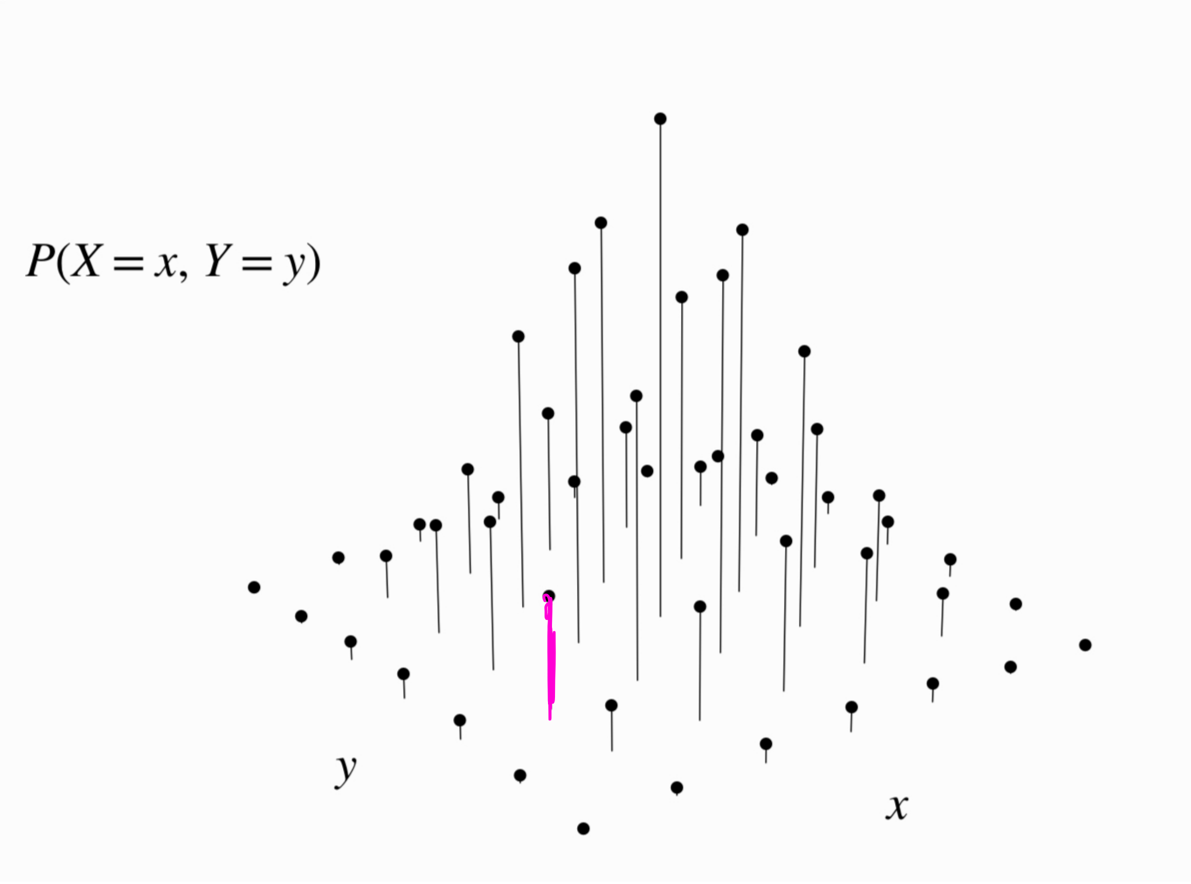
Definition

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

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

边际分布

Definition

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

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

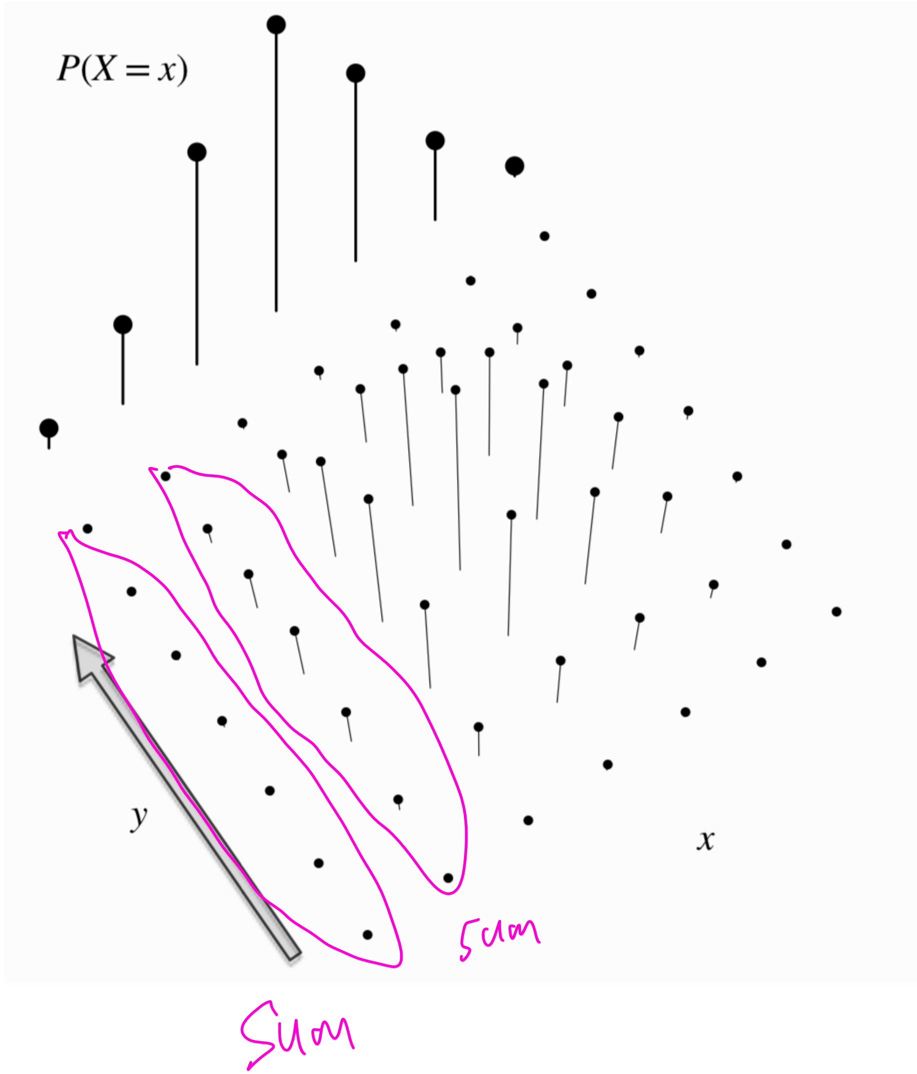
$$P(X=x) = \sum_y P(X=x|Y=y)P(Y=y)$$

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

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

Marginal PMF

$P(X=x)$: total height of the bars in the corresponding column of the joint PMF.



Example

$$P_X(1) = P(X=1) = \sum_{y=1}^{4} P(X=1, Y=y) = \frac{1}{20} + \frac{1}{20} + \frac{1}{20} + 0 = \frac{3}{20}$$

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	

$$P_Y(3) = P(Y=3) = \sum_{x=1}^{4} P(X=x, Y=3)$$

$$= \frac{1}{20} + \frac{2}{20} + \frac{3}{20} + \frac{1}{20}$$

$$= \frac{7}{20}$$

Conditional PMF

1° conditional PMF is also a valid PMF. $\begin{cases} \sum_x P_{X|Y}(x|y) = 1. \\ P_{X|Y}(x|y) \geq 0 \end{cases}$

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

Definition

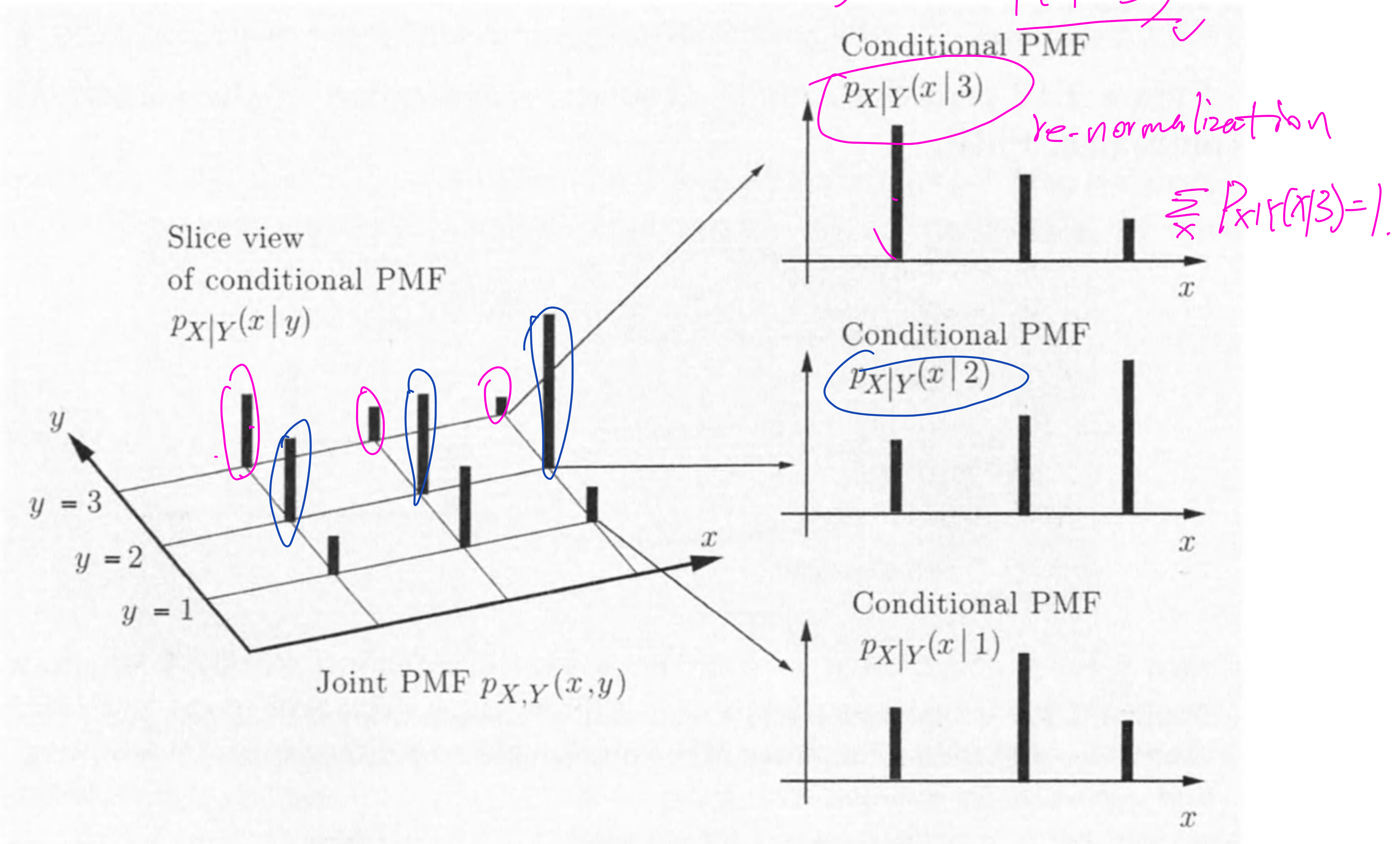
For discrete r.v.s X and Y, the *conditional PMF* of X given $Y = y$ is

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

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

Conditional PMF

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



Independence of Discrete R.V.s

Definition

$\boxed{\text{Random variables}}$ X and Y are *independent* if for all x and y,

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

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

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 $\boxed{P(X = x) > 0.}$

Example: Chicken-egg

① Joint PMF $P(X=i, Y=j)$, $i \in N$, $j \in N$.

② Conditioning on $N=n$. $X|N=n \sim Bin(n,p)$, $Y|N=n \sim Bin(n,q)$

$X+Y|N=n = n$.

$P(N=k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}$

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 LOTP. $P(X=i, Y=j) = \sum_{n=0}^{\infty} P(X=i, Y=j, N=n) = P(X=i, Y=j, N=i+j)$

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

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

$P(A,B|C)$

$= P(A|C)\, P(B|A,C)$

Solution

$e^{-\lambda} = e^{-\lambda(p+q)} = e^{-\lambda p} \cdot e^{-\lambda q}$

$X|_{N=i+j} \sim Bin(i+j, p)$

$N \sim Pois(\lambda)$

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

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

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

④ $$P(X=i) = \sum_{j=0}^{\infty} P(X=i, Y=j) = \frac{e^{-\lambda p} (\lambda p)^i}{i!} \cdot \underbrace{\left(\sum_{j=0}^{\infty} \frac{e^{-\lambda q} \cdot (\lambda q)^j}{j!} \right)}_{1}$$

$X \sim Pois(\lambda p)$ $\quad$ $Y \sim Pois(\lambda q$

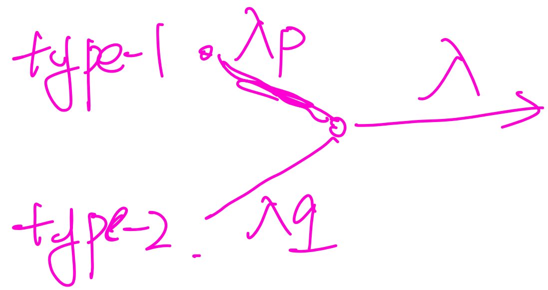
Solution

(5) $P(X=i, Y=j) = P(X=i)P(Y=j)$. $\forall i,j \in \mathbb{N}^2$.

X, Y are independent. $(X+Y=N,\ N \sim Pois(\lambda))$

Conditioning on $N=n$, $X+Y=n$, X, Y not independent.

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* $X|N = n \sim \mathrm{Bin}(n, p)$. $p+q=1$.

$\mathrm{Pois}(\lambda p + \lambda q) = \mathrm{Pois}(\lambda)$

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

Conditional PDF Given an Event

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

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

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

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

- 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

LOTP: $P(X\le x) = \sum_{i=1}^{n} P(X\le x | A_i)\cdot P(A_i)$

$$P(X\le x|A_i) = \int_{-\infty}^{x} f_{X|A_i}(t)\,dt.$$

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

Taking derivatives of both sides

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

Joint PDF

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

① Valid PDF $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy = 1$. $f_{X,Y}(x,y) \ge 0$.

② Example: $P(X<3, 1 \le Y < 4) = \int_{-\infty}^{3}\int_{1}^{4} f_{X,Y}(x,y)\,dx\,dy$.

Definition

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

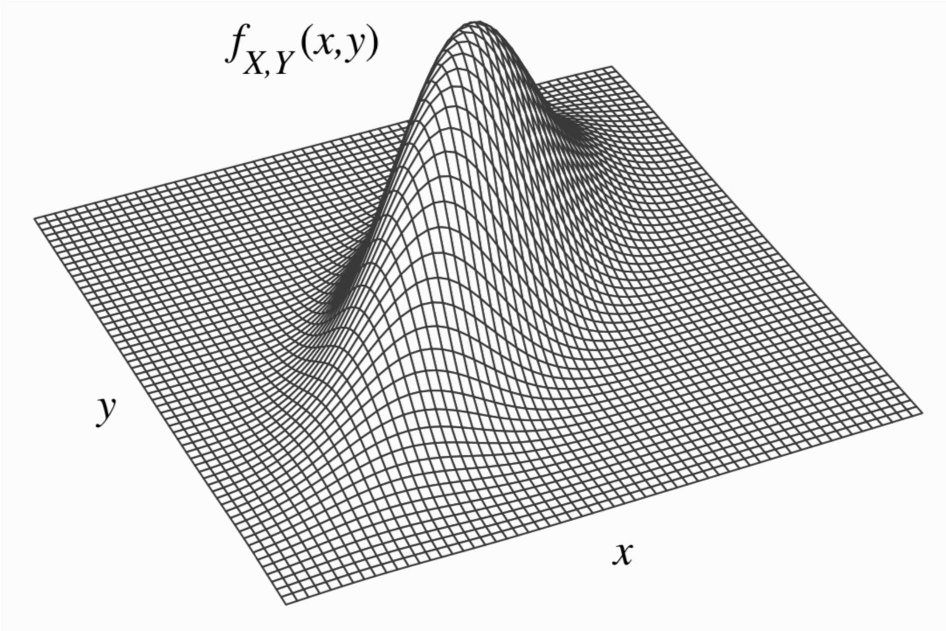
$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

③ Generally. $(x,y) \in B$. $B \subseteq \mathbb{R}^2$.

$$P((x,y)\in B) = \iint_B f_{X,Y}(x,y)\,dx\,dy$$

Joint PDF

1-D: $\delta \approx 0$. $P(a \leq X \leq a+\delta) = \int_a^{a+\delta} f(x)dx \approx f_X(a)\cdot\delta$.



2-D: $\delta_1 \approx 0$
$\delta_2 \approx 0$.

$P(a \leq X \leq a+\delta_1, b \leq Y \leq b+\delta_2) = \int_a^{a+\delta_1}\int_b^{b+\delta_2} f_{X,Y}(x,y)\,dx\,dy$

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

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

① Conditional PDF is also a valid PDF, given fixed x.

② $f_{Y|X}(\cdot|x)$ is a valid PDF: 1° $f_{Y|X}(y|x) \geq 0$, $\forall y$.

Definition

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

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \qquad f_{X,Y}(x,y) \geq 0, \quad f_X(x) > 0$$

$X = x$ 处 PDF > 0

$$\int_{-\infty}^{+\infty} f_{Y|X}(y|x)\,dy = \frac{\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dy}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1.$$

renormalization

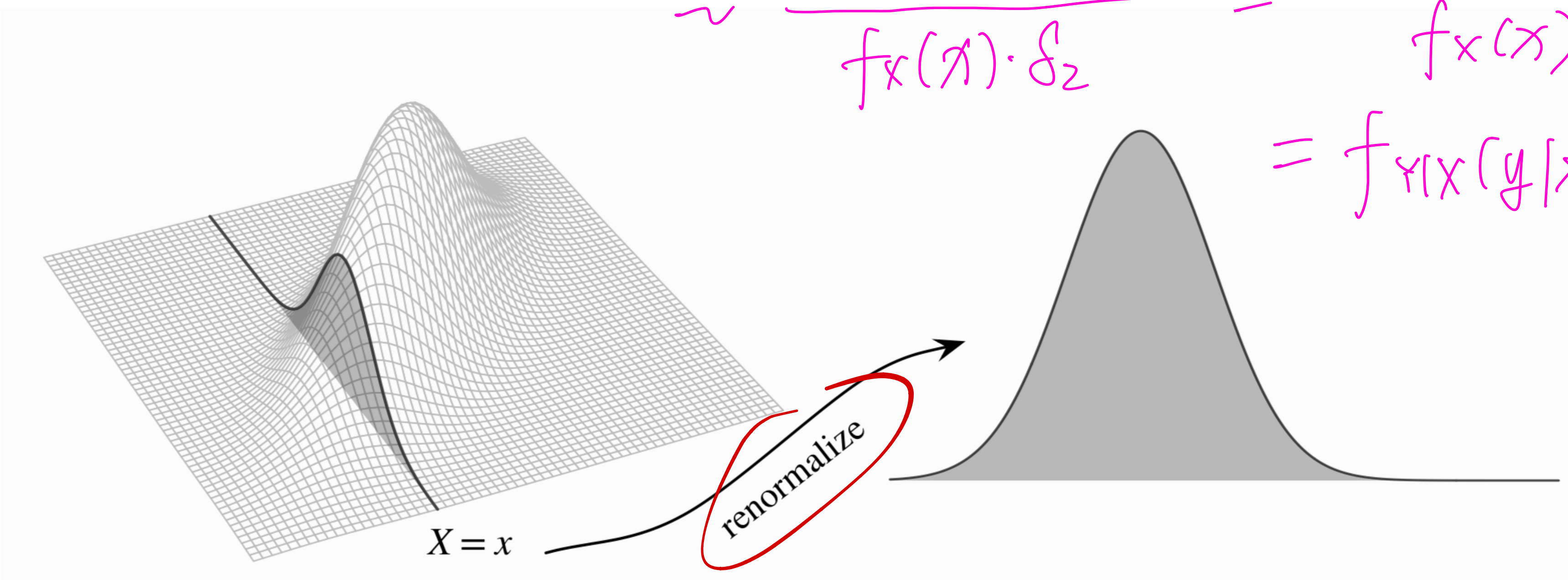
Conditional PDF

$\delta_1, \delta_2 \simeq 0$

① $$P(y \le Y \le y+\delta_1 \mid x \le X \le x+\delta_2) = \frac{P(y \le Y \le y+\delta_1, x \le X \le x+\delta_2)}{P(x \le X \le x+\delta_2)}$$

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

$$= f_{Y|X}(y|x)\,\delta_1$$



② let $\delta_2 \to 0$. $P(y \le Y \le y+\delta_1 \mid X=x) = f_{Y|X}(y|x)\cdot\delta_1$.

③ $A \subseteq \mathbb{R}$, $P(Y \in A \mid X=x) = \int_A f_{Y|X}(y|x)\,dy$.

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 For $0<x<1$, $0<y<1$, we have

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f(x,y)}{\int_0^1 f(x,y)\,dx} = \frac{\frac{12x(2-x-y)}{5}}{\int_0^1 \frac{12x(2-x-y)}{5}\,dx}$$

The joint PDF of X and Y is given by

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

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

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

Example

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

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

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

先PDF
后CDF

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

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 By definition of conditional PDF: $f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$

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

Theorem

For continuous r.v.s X and Y,

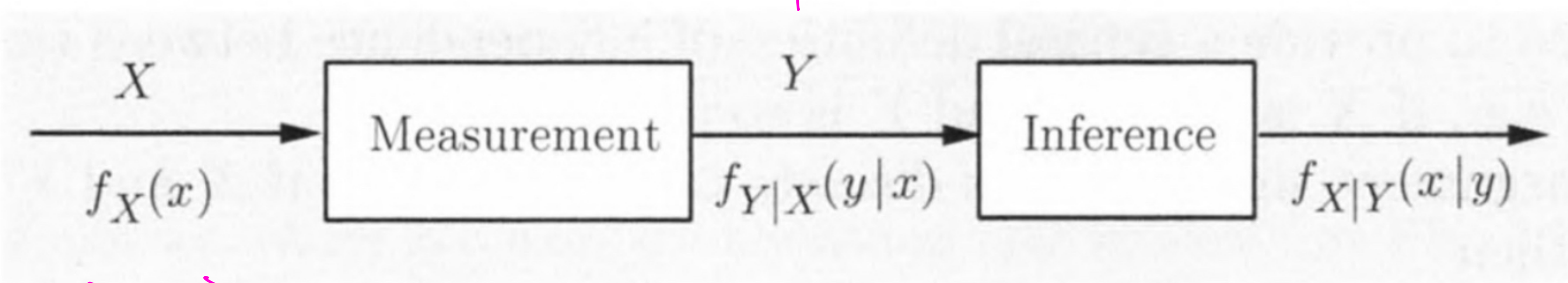
$$① \; f_{Y|X}(y|x) = \frac{\cancel{f_{Y|X}}^{f_{X|Y}}(x|y) f_Y(y)}{f_X(x)}$$

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

Proof

Bayes' Rule: Inference Perspective

info/data



prior PDF

posterior PDF

Example

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

② Conditional PDF: $f_{\Lambda|Y}(\lambda|y) = \dfrac{f_\Lambda(\lambda)\cdot f_{Y|\Lambda}(y|\lambda)}{f_Y(y)} = \dfrac{2\cdot\lambda e^{-\lambda y}}{\int_1^{\frac{3}{2}} 2te^{-ty}dt}$

A light bulb produced by the GE company is known is known to have an exponentially 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?

$$f_{\Lambda|Y}(\lambda|y) = \frac{2\lambda e^{-\lambda y}}{\int_1^{\frac{3}{2}} 2te^{-ty}dt}$$

③ $f_Y(y) = \int_{-\infty}^{+\infty} f_\Lambda(t) f_{Y|\Lambda}(y|t)\,dt$. LOTP.

General Bayes' Rule

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

涉及到 proof · $\varepsilon \Rightarrow$ 小区间

Proof

① X discrete, Y continuous, $f_{Y|X}(y|X=x)$.

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

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

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

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

Proof

② $P(Y=y \mid X=x)$, Y discrete, X continuous,

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

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

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

Outline

1. Discrete Multivariate R.V.s
2. Continuous Multivariate R.V.s
3. Covariance and Correlation
4. Multinomial Distribution
5. Multivariate Normal

① event → prob. → conditional prob.

prior / posterior.

② random variable. (discrete / continuous)

single : distribution. Story

⇓

two random variables (independence)

⇓

≥ 3 r.v.s

joint distribution

Marginal distribution

Multivariate Distribution

③ Conditional Expectation → estimation / learning (prediction) (Signal processing

→ Statistics classical → Monte Carlo

Bayesian Markov chain

- 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

3 Key themes: 三大观念、

① Decomposition / Integration

分解与合成

First-step

Conditioning

prob / expectation

② Transformation 变换

(invariance)

PGF

MGF

③ Approximation

近似

Asymptotic approximation 渐近近似

CLT : $n \to \infty$

eg. 中心极限定理

Finite-size approximation

1 Discrete Multivariate R.V.s

2 Continuous Multivariate R.V.s

3 Covariance and Correlation

4 Multinomial Distribution

5 Multivariate Normal

General Bayes' Rule

$f_X(x) \cdot 2\varepsilon \approx P(X \in (x-\varepsilon, x+\varepsilon))$

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

class label

feature

$f_{Y|X}(y|x)$: Discriminative AI.

$f_{X,Y}(x,y)$: Generative AI

Proof

$X_1, \dots X_n$

Stochastic processes

Stochastic information of $X_1, \dots, X_n$

Markov : → Markov chain

Martingale : → Martingale

→ Branching processes

→ Poisson process.

General LOTP

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

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

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

$f_X(x) \cdot 2\varepsilon$

Proof

① $P(X=x) = \int_{-\infty}^{+\infty} P(X=x \mid Y=y)\, f_Y(y)\, dy$

X: discrete ; Y: Continuous

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

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

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

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

$$= P(X=x) \cdot 1 = P(X=x)$$

条件概率也是概率
— PDF ——— PDF
— 期望 ——— 期望

Example

① $Y = N + S$; conditioning on $S = s$ ($s = 1$ or -1)

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

discrete ↑ continuous ↑

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

高斯白噪声

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?

$\mu = 0$ $\sigma = 1$

LOTP

③ $f_Y(y) = f_{Y|S}(y|1) \cdot P(S=1) + f_{Y|S}(y|-1) \cdot P(S=-1)$ $(S=-1)$, $N(-1,1)$

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

$P(S=1) = p$ prior.

④ $P(S=1 \mid Y=y) = \frac{p \cdot e^y}{p \cdot e^y + (1-p)e^{-y}} = \frac{p}{p + (1-p)e^{-2y}} = \begin{cases} > p & y > 0 \\ = p & y = 0 \\ < p & y < 0 \end{cases}$

posterior.

$y > 0$
$e^{-2y} < 1$
$p + (1-p)e^{-2y}$
$< p + 1 - p = 1$

Example: Comparing Exponentials of Different Rates

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

① $P(T_1 < T_2) \overset{\text{LOTP}}{=} \int_0^\infty P(T_1 < T_2 | T_2 = t) \cdot f_{T_2}(t) \cdot dt$

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

独立

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

$f_{T_2}(t) = \lambda e^{-\lambda t_2}$

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

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

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

T_1, T_2 独立 连续

② $P(T_1 = T_2) = 0$

$\boxed{P(T_1 < T_2) = P(T_1 = \min(T_1, T_2))}$

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

independent.

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

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

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

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

$\ne f_X(x) f_Y(y)$

~~X~~, 定义域不能耦合

domain (x,y) couple

$f_X(x) = \begin{cases} 4x(1-x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$, $f_Y(y) = \begin{cases} 4y^3 & , 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

Theorem

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

pre. decouple

domain of x and y

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

$(-\infty, +\infty)$

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.)

$\Longleftrightarrow$

$F_X(x) = g(x)$

$F_Y(y) = h(y)$

Proof

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

$c>0$ is a constant.

normalization

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

renormalization key idea:

$$\Rightarrow f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_{-\infty}^{\infty} cg(x)\cdot\tfrac{1}{c}h(y)dy$$

$$= cg(x)\cdot \underbrace{\int_{-\infty}^{\infty}\tfrac{1}{c}h(y)dy}_{1} = cg(x).$$

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

③ $$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \int_{-\infty}^{\infty} cg(x)\cdot\frac{h(y)}{c}dx$$

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

if g is a valid PDF, $\int_{-\infty}^{\infty} g(x)dx = 1$ $\Rightarrow c=1$ $\Rightarrow \int_{-\infty}^{\infty} h(y)dy = 1$ $\Rightarrow h$ is a valid PDF.

④ $f_{X,Y}(x,y) = cg(x)\cdot\frac{1}{c}h(y) = f_X(x)\cdot f_Y(y)$ $\Rightarrow X, Y$ are independent.

2D LOTUS

No LOTUS: $(X,Y) \longrightarrow g(X,Y) \longrightarrow E[g(X,Y)]$

LOTUS: $(X,Y) \longrightarrow E[g(X,Y)]$

Theorem

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

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

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

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

Expected Distance between Two Uniforms

① $E(|X-Y|) \overset{\text{LOTUS}}{=} \int_0^1 \int_0^1 |x-y| f_X(x) \cdot f_Y(y) \, dy = \int_0^1 \int_0^1 |x-y| \, dx dy$

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

②

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

$\frac{2}{3}$ $\frac{1}{3}$

$M = \max(X,Y)$; $L = \min(X,Y)$;

$M + L = X + Y.$

$\Rightarrow E(M+L) = E(X+Y) \Rightarrow E(M) + E(L) = E(X) + E(Y)$

$= \frac{1}{2} + \frac{1}{2} = 1$;

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

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

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

Expected Distance between Two Normals

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

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

eg. find dis joint PDF $(X+Y, X-Y)$

$(\frac{X}{Y}, \frac{Y}{X})$

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

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

矩生成函数 $\Rightarrow$ $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$

$\Rightarrow X_1 + X_2 \sim \boxed{N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)}$

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

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

$= \sqrt{\frac{2}{\pi}}$.

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

Outline

Covariance 协方差

$Var(X) = E[(X - EX)^2]$

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:

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

Key Properties of Covariance

$Var(X+Y) \overset{①}{=} Cov(X+Y, X+Y)$

$\overset{④}{=} Cov(X,X) + Cov(X,Y)$

$+ Cov(Y,X) + Cov(Y,Y)$

$\overset{①}{=} Var(X) + Var(Y)$

$+ Cov(X,Y)$

$+ Cov(Y,X)$

$\overset{②}{=} Var(X) + Var(Y)$

$+ 2Cov(X,Y)$

- ① $Cov(X, X) = Var(X)$.
- ② $Cov(X, Y) = Cov(Y, X)$.
- $Cov(X, c) = 0$ for any constant c.
- $Cov(a \cdot X, Y) = a \cdot Cov(X, Y)$ for any constant a.
- $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.
- ④ $Cov(X + Y, Z + W) =$
  $Cov(X, Z) + Cov(X, W) + Cov(Y, Z) + Cov(Y, W)$.
- ☆ $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$. ; proof
- For n r.v.s $X_1, \dots, X_n$,

若变量前有一，可以提出

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

类似平方项展开

Proof

Correlation

(线性)相关系数

Definition

Linear Correlation

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

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

归一化

(This is $\boxed{\text{undefined}}$ 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.

不相关(不存在线性相关关系)

Uncorrelated

X, Y are independent.

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

$$= \underbrace{E[(X-EX)]}_{0} \cdot \underbrace{E[Y-EY]}_{0} = 0$$

✓

Theorem

If X and Y are independent, then they are $\underbrace{E[X]-E[X]}_{0}$ uncorrelated.

仅、线性不相关

×

除、高斯

独立 $\Rightarrow \mathrm{Cov}(X,Y)=0$

Uncorrelated $\not\Rightarrow$ Independent

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

线性不相关（存在非线性关系 X^2）

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

1°. $E(X) = 0$; $\Rightarrow E(X) \cdot E(Y) = 0$

2°. $E(XY) = E(X^3) = 0$

3° $\Rightarrow$ $Cov(X,Y) = 0$.

X, Y are linear uncorrelated (potentially nonlinear correlated)

but, X, Y are dependent.

$E(X^3)$ $\qquad g(x)$

$= \int_{-\infty}^{+\infty} x^3 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0$

$g(-x) = -g(x)$

$\int_{-\infty}^{\infty} g(x)dx = \int_0^{\infty} g(x)dx + \int_{-\infty}^{0} g(x)dx$

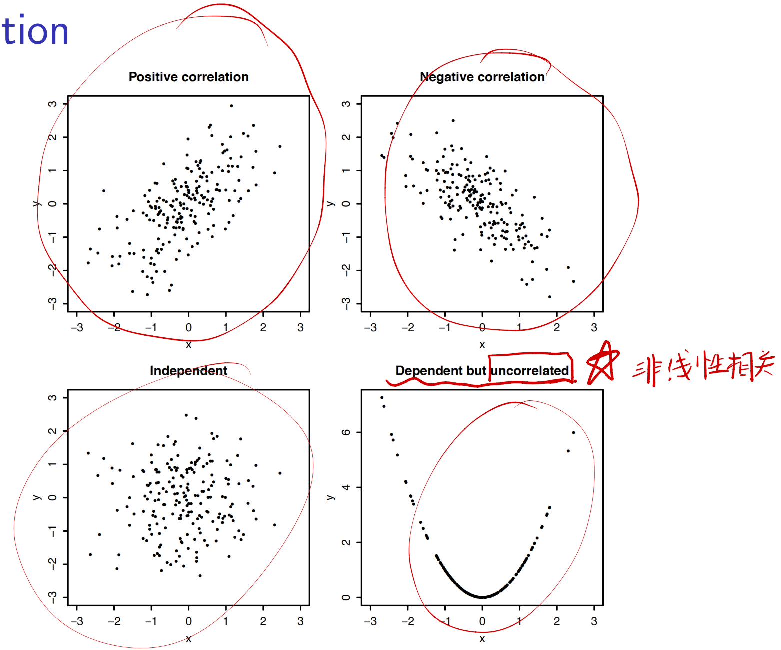
$\int_{\infty}^{0} g(x)dx$

$-\int_0^{\infty} g(x)dx$

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

Cauchy-Schwarz Inequality

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

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

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

1°. $\underbrace{f(t)}_{\ge 0} = E[(X-tY)^2] = E[X^2 - 2tXY + t^2Y^2]$

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

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

Theorem

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

不等版 Cauchy

For any r.v.s X and Y,

负相关 正相关

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

2° $X - E[X]$, $Y - E[Y]$ $\xrightarrow[C.S.]{}$ $E^2[(X-EX)(Y-EY)] \le E[(X-EX)^2] \cdot E[(Y-EY)^2]$

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

$\vec{a} \cdot \vec{b} \le \|a\| \cdot \|b\|$

$|\cos\theta| = \dfrac{\vec{a} \cdot \vec{b}}{\|a\| \cdot \|b\|} \le 1$

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

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

Outline

Story

$K=2$ → Binomial.

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

$P = (P_1, P_2)$ $(P_1, 1-P_1)$

$P_1 + P_2 = 1$

$P_1 =$

Multinomial Joint PMF

Method 1°. $\binom{n}{n_1}\cdot\binom{n-n_1}{n_2}\cdot\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{k-1}}{n_k} = \frac{n!}{n_1!\cdots n_k!}$

Method 2°. $\frac{n!}{n_1!\cdot n_2!\cdots n_k!}$ → 去重（无顺序）

Theorem

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

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

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

Proof

Multinomial Marginals

Successful event : object → j th category.

Theorem

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

$\Downarrow$

定义落在j处为成功

Multinomial Lumping

聚合

event A = " fall into i^{th} category"

B = " ... j^{th} ... "

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ $P(A \cap B) = 0$

$= P(A) + P(B) = p_i + p_j$

视落在 i/j 处为成功

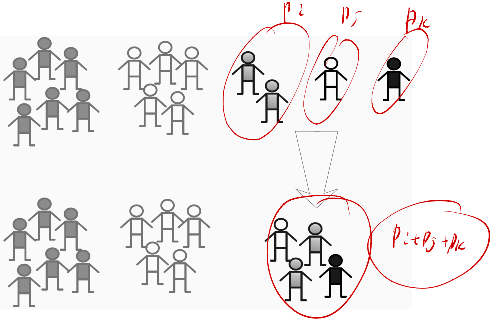
(X_i, X_j 互斥)

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*

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

Multinomial Lumpling



Multinomial Conditioning

1°. Given n_1 objects in Category 1, the remaining $n-n_1$ objects landing into categories $2, \dots k$: $\text{Multi}_{k-1}(n-n_1, \underline{p'})$

2°. $\underset{j\neq 1}{p_j'}$ = Prob ($\underbrace{\text{landing into category } j}_{A}$ | $\underbrace{\text{not landing into category 1}}_{B}$)

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

$A \cap B = A$

剩余概率
renormalization

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

$1 - p_1$

= $\dfrac{\text{Prob.(landing into Category } j)}{\text{Prob.(not landing into Category 1)}}$

$p_2' + \cdots + p_k' = 1$

$= \dfrac{p_j}{1-p_1} = \dfrac{p_j}{p_2 + \cdots + p_k} = p_j'$

renormalization.

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 $(X_1, ..., X_k) \sim \mathrm{Mult}_k(n, \mathbf{p})$, *where* $\mathbf{p} = (p_1, ..., p_k)$. *For* $i \neq j$, $\mathrm{Cov}(X_i, X_j) = -np_ip_j$. < 0

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

Multivariate Normal Distribution

多元正态分布

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

X_j 正态分布， ⇒ 多元 ⋯

任意线性组合还是正态分布

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

const ⇒ Variance 为 0 的正态分布

to have a Normal distribution for any choice of constants $t_1, ..., t_k$. 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).

尽管

二元正态分布

Non-example of MVN

$X=(X_1,\dots X_k)$ 若 $X_j \sim N(\mu_j,\sigma_j^2)$

$X \neq MVN$

S: 随机信号

① $X \sim N(0,1)$, $S = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$; S and X are independent.

② $Y = S\cdot X \sim N(0,1)$:

$\forall y \in \mathbb{R}$, $P(Y\le y) = P(S\cdot X \le y)$

LOTP

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

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

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

$P(X\ge -y)\cdot\frac{1}{2}$

$P(X\le y)\cdot\frac{1}{2}$

$= P(X\le y)$

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

③ $P(X+Y=0)$

$= P(S=-1) = \frac{1}{2}$.

X+Y is NOT Continuous.

$\Rightarrow$ Normal.

$(X,Y) \neq MVN$

Actual MVN

MGF

Sum of independent Normal $\longrightarrow$ Normal.

$\boxed{\text{i.i.d的 Normal之和 还是Normal}}$

① $Z, W \overset{\text{i.i.d.}}{\sim} N(0,1)$

② $(Z+2W, 3Z+5W)$ is a MVN.

$\forall t_1, t_2 \in R$, $t_1(Z+2W) + t_2(3Z+5W)$

$= (t_1+3t_2)\underline{Z} + (2t_1+5t_2)\underline{W}$

$\sim$ Normal.

Theorem

$(X_1, X_2, X_3) \xrightarrow{MVN} \forall t_1, t_2, t_3 \in R.$

$t_1X_1 + t_2X_2 + t_3X_3 \sim$ Normal.

Theorem

Let $t_3 = 0$, $\forall t_1, t_2 \in R$.

If (X_1, X_2, X_3) is Multivariate Normal, then so is the subvector (X_1, X_2). $t_3 = 0$

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.* 级联随机向量

Parameters of MVN

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

Joint PDF

$= \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{ -\frac{1}{2} \left((x,y)\Sigma^{-1}(x,y)^T \right) \right\}$

$\rho=0$ $f_{X,Y}(x,y) = h(x) \cdot g(y)$

X and Y are independent!

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 $Cov(X_i, X_j)$.

Standard Bivariate Normal. (X, Y)

$X \sim N(0,1)$

$Y \sim N(0,1)$

$\mathrm{Corr}(X,Y) = \rho \in (-1,1)$

协方差矩阵

mean vector $(0, 0)$

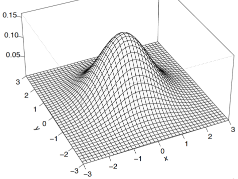
Covariance matrix $\Sigma = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(Y,X) & Var(Y) \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$

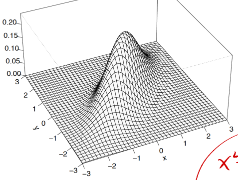
Joint PDF of Bivariate Normal Distributions

$\rho = 0$

$\rho = 0.75 > 0$

Joint PDF





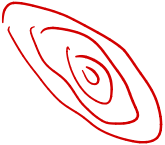
$x_1 = \frac{1}{\sqrt{2}}(x+y)$

$y_1 = \frac{1}{\sqrt{2}}(x-y)$

$\Rightarrow \frac{x_1^2}{1+\rho} + \frac{y_1^2}{1-\rho} = C > 0$

$\frac{x^2+y^2-2\rho xy}{1-\rho^2} = C > 0$

负相关

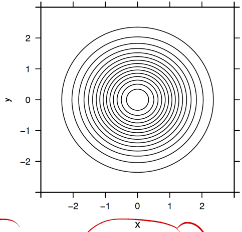


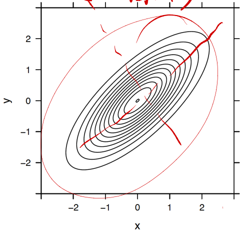
独立

contour

Corr > 0

(正相关)





$\rho > 0$

$\Downarrow$

$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

$\rho = 0$; $x^2+y^2 = C > 0$

Joint MGF

① $W \sim$ Normal. MGF. $= e^{E[tW] + \frac{1}{2}Var(tW)}$

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

② $(X_1, \dots X_k)$ MVN. ⇒ Joint MGF. $E[e^{t_1X_1 + \dots + t_kX_k}]$

$= \exp\{t_1E[X_1] + \dots + t_kE[X_k] + \frac{1}{2}Var(t_1X_1 + \dots + t_kX_k)\}$

Definition

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

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

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

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 Recall fact: $X \sim N(\mu, v)$, $\overset{\text{MGF}}{M_X(t)} = e^{\mu t + \frac{1}{2}v^2t^2}$

② Bivariate Normal (X, Y), $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$

$Corr(X, Y) = \rho$.

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

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

③ Joint MGF
$M_{X,Y}(s,t) = E[e^{sX+tY}] = Exp\{s\mu_1 + t\mu_2 + \frac{1}{2}Var(sX+tY)\}$

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

④ $\rho = 0$ $\Rightarrow M_{X,Y}(s,t) = Exp\{s\mu_1 + t\mu_2 + \frac{1}{2}s^2\sigma_1^2 + \frac{1}{2}t^2\sigma_2^2\}$

uncorrelated

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

$\boxed{M_{X,Y}(s,t) = M_X(s)M_Y(t)}$ $= M_X(s) \cdot M_Y(t)$ $\Rightarrow$ X and Y are independent.

Bivariate Normal Generation 二元正态生成

1°. $Z = aX + bY$. $(a,b,c,d) \in \mathbb{R}$? $\Rightarrow$ (Z,W) bivariate Normal.

$W = cX + dY$ $\qquad \underbrace{}_{}$ $N(0,1)$ $\qquad$ $Z, W \sim N(0,1)$

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

2°. $E(Z) = E(aX+bY) = a\underbrace{E[X]}_{0} + b\underbrace{E[Y]}_{0} = 0$;

$E[W] = 0$;

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 ?

3°. $Var(Z) = Var(aX+bY) = Var(aX) + Var(bY) = a^2 Var(X) + b^2 Var(Y)$

$= a^2 + b^2 = 1$.

$Var(W) = Var(cX+dY) = c^2 + d^2 = 1$

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

4°. $Corr(Z,W) = Cov(Z,W) = \rho$ $\Rightarrow Cov(aX+bY, cX+dY) = \rho$

$\Rightarrow ac\,Cov(X,X) + bd\,Cov(Y,Y) + 0 = \rho$ $\qquad \Rightarrow ac + bd = \rho$

Solution

5° $\begin{cases} a^2+b^2=1 \\ c^2+d^2=1 \\ ac+bd=\rho \end{cases}$

有一组解即可 ⇒ 取 $b=0$

Find one solution is enough.

$b=0$; $\Rightarrow a^2=1 \Rightarrow$ pick $a=1$; $c=\rho$.

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

6°. $Z = aX+bY = \underline{X}$.

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

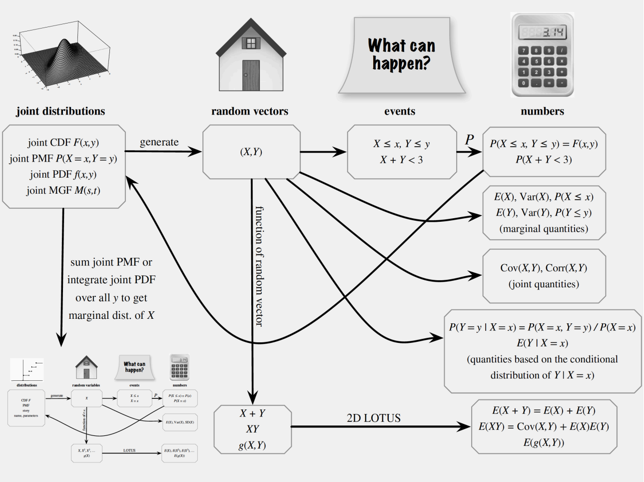
$\underline{(Z,W)}$ is the desired bivariate Normal distribution

$\underline{\text{Joint PDF ?}}$ $(Z,W) = \underline{g(X,Y)}$.

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

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