Lecture 6: Joint Distributions

Wen Dingzhu

School of Information Science and Technology (SIST)
ShanghaiTech University

wendzh@shanghaitech.edu.cn

April 3, 2023

Outline

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

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

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

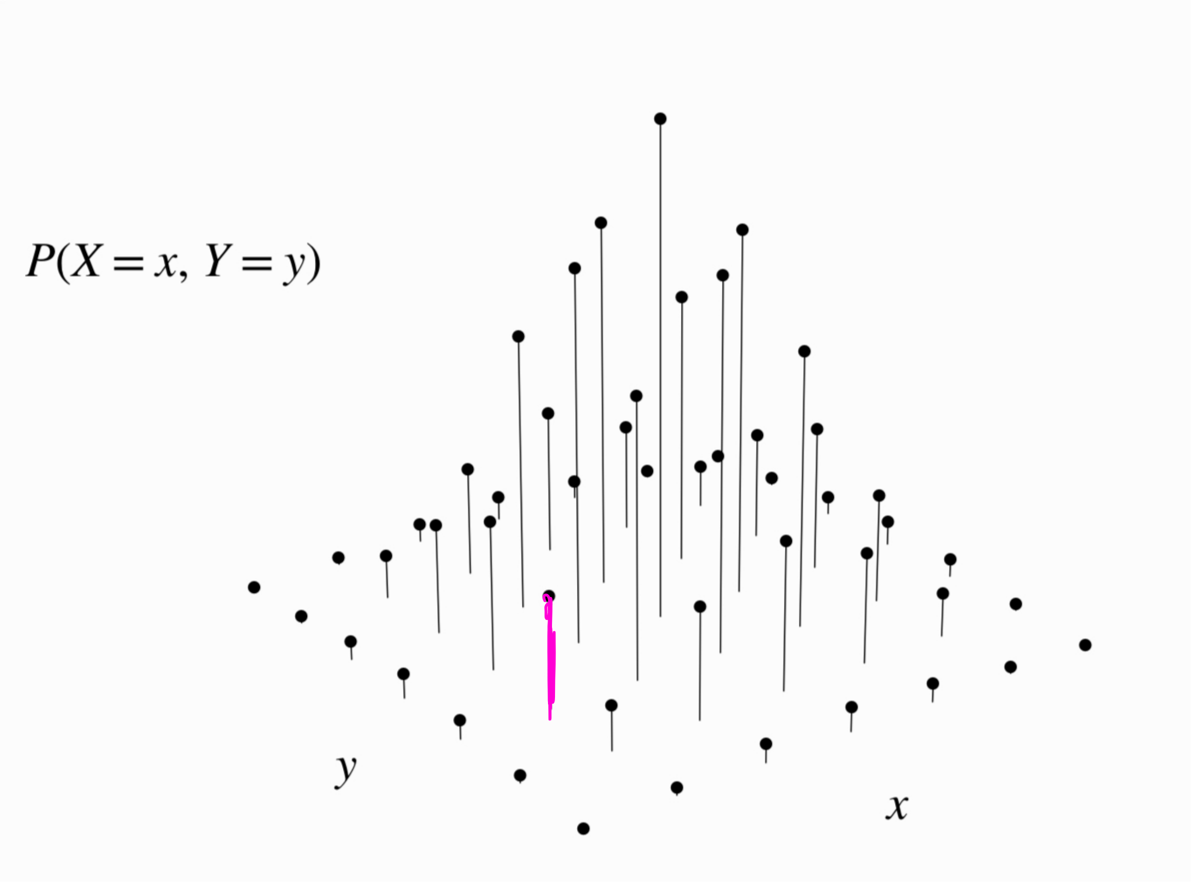
Definition

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

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

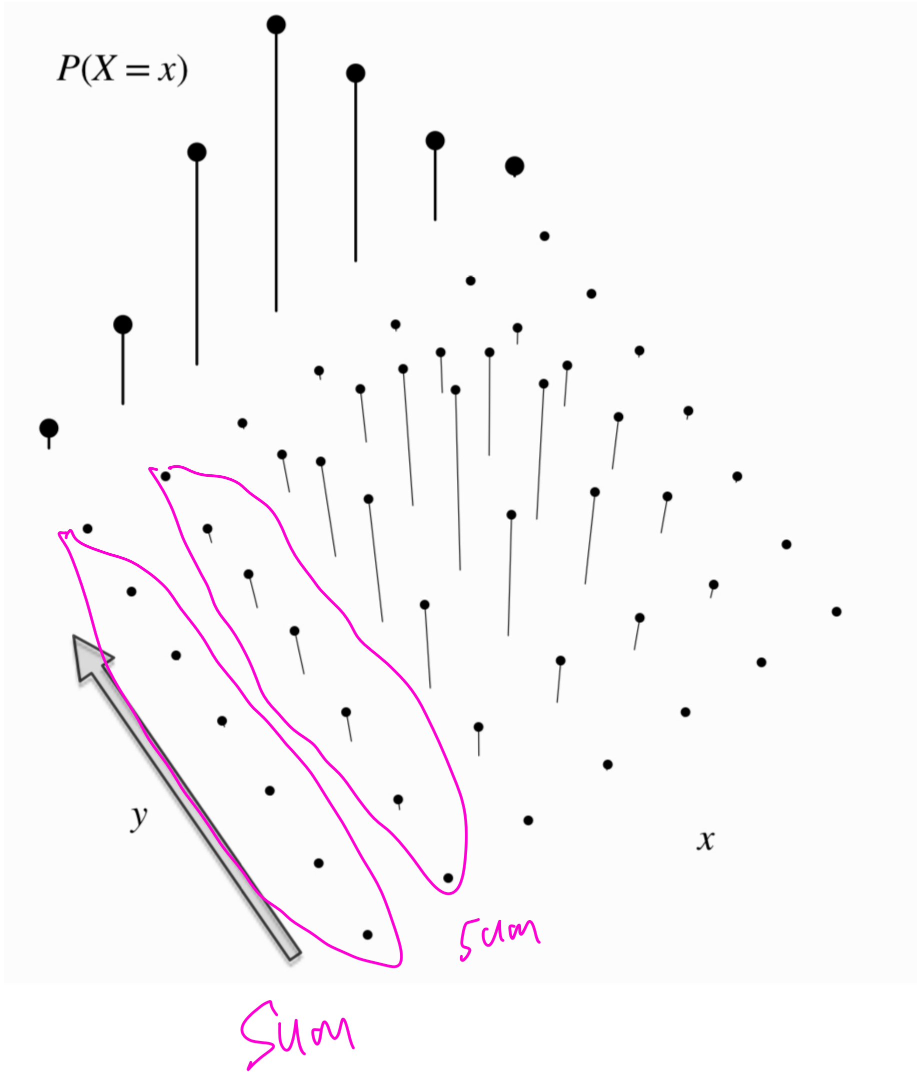
Definition

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

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

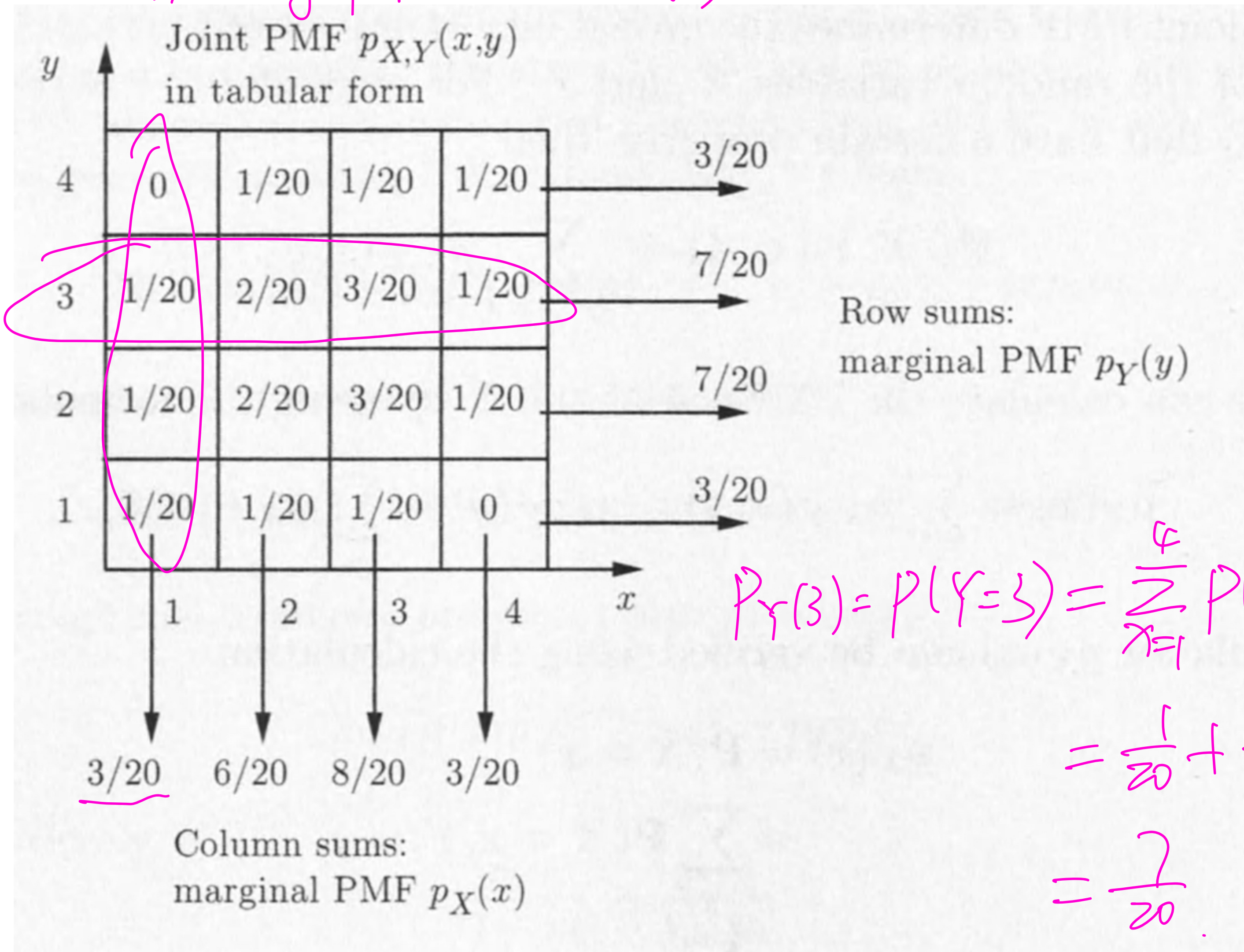
Marginal PMF

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



Example

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



y \ x	1	2	3	4	Row sums $p_Y(y)$
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 $p_X(x)$	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. $\begin{cases} \sum_x P_{X|Y}(x|y) = 1. \\ P_{X|Y}(x|y) \geq 0 \end{cases}$

Definition

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

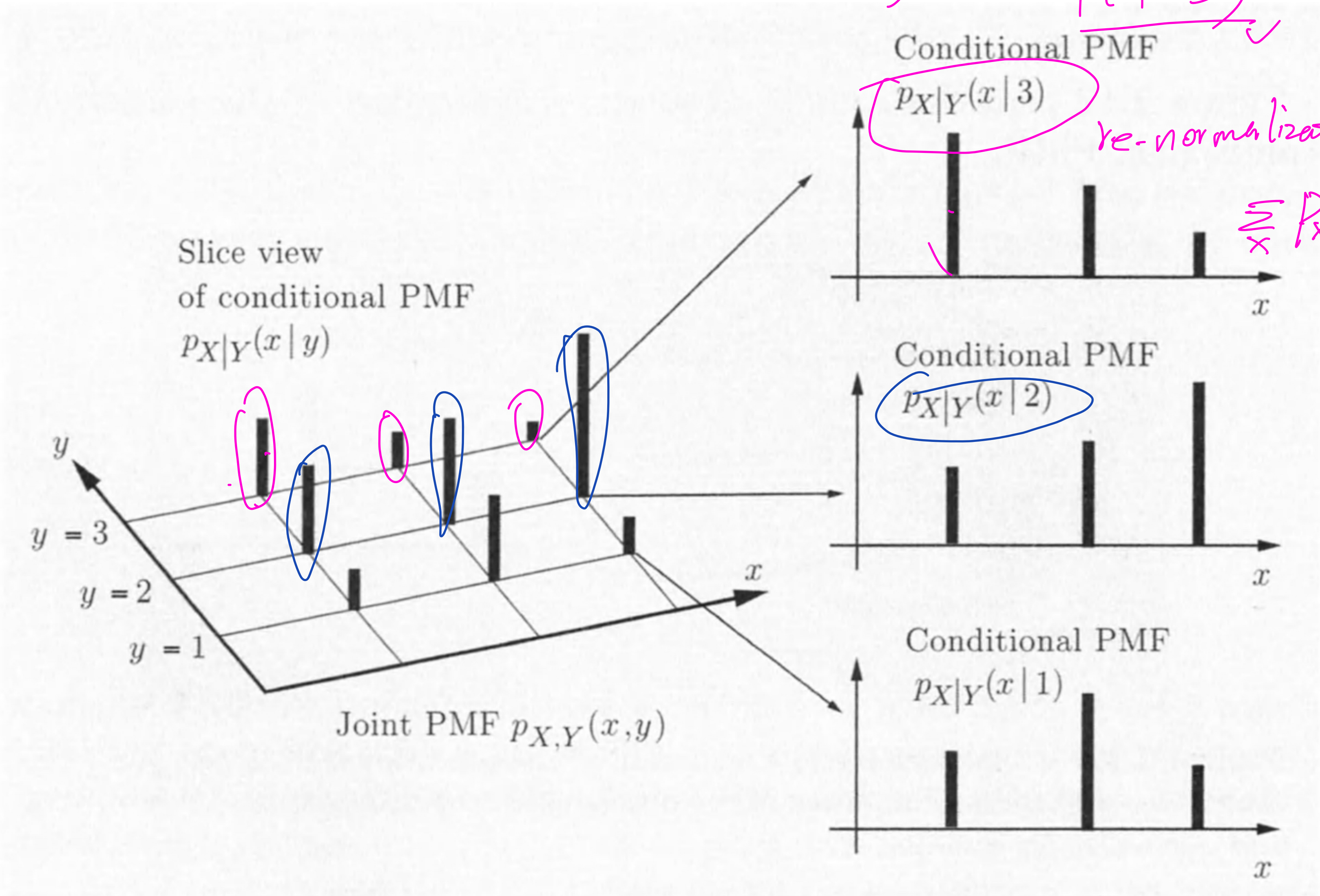
$$P_{X|Y}(x|y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

Conditional PMF

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

re-normalization

$\sum_x p_{X|Y}(x|3) = 1.$



Independence of Discrete R.V.s

Definition

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

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

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

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 $P(X = x) > 0$.

Example: Chicken-egg

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

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

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

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

$P(A,B|C)$

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

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

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

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

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

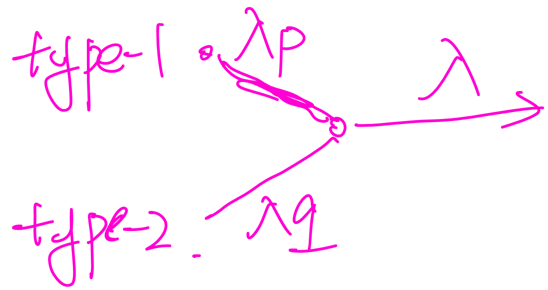
Solution

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

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

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

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

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

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

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

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:

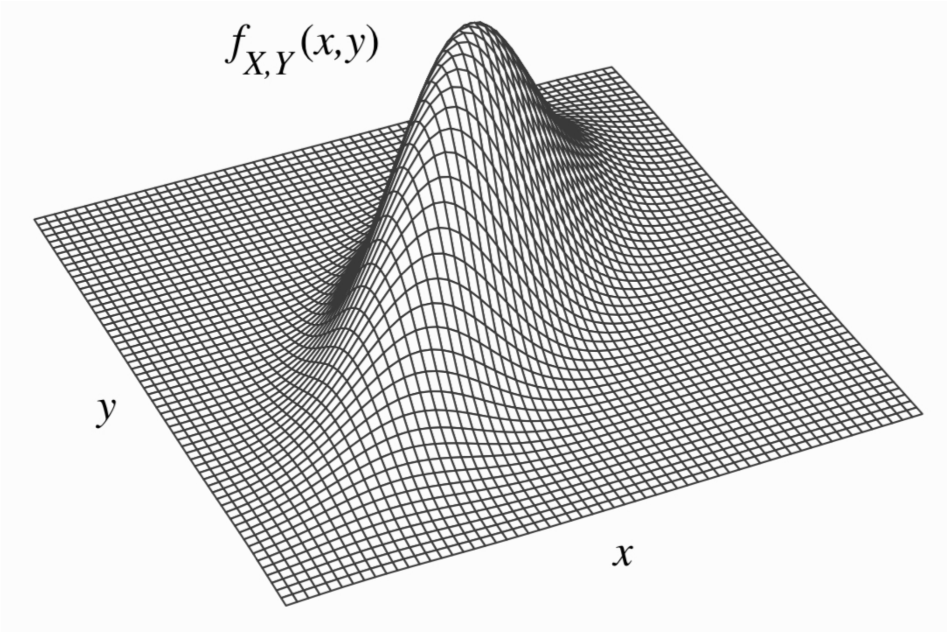
$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

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

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

Joint PDF

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



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

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

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

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

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

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

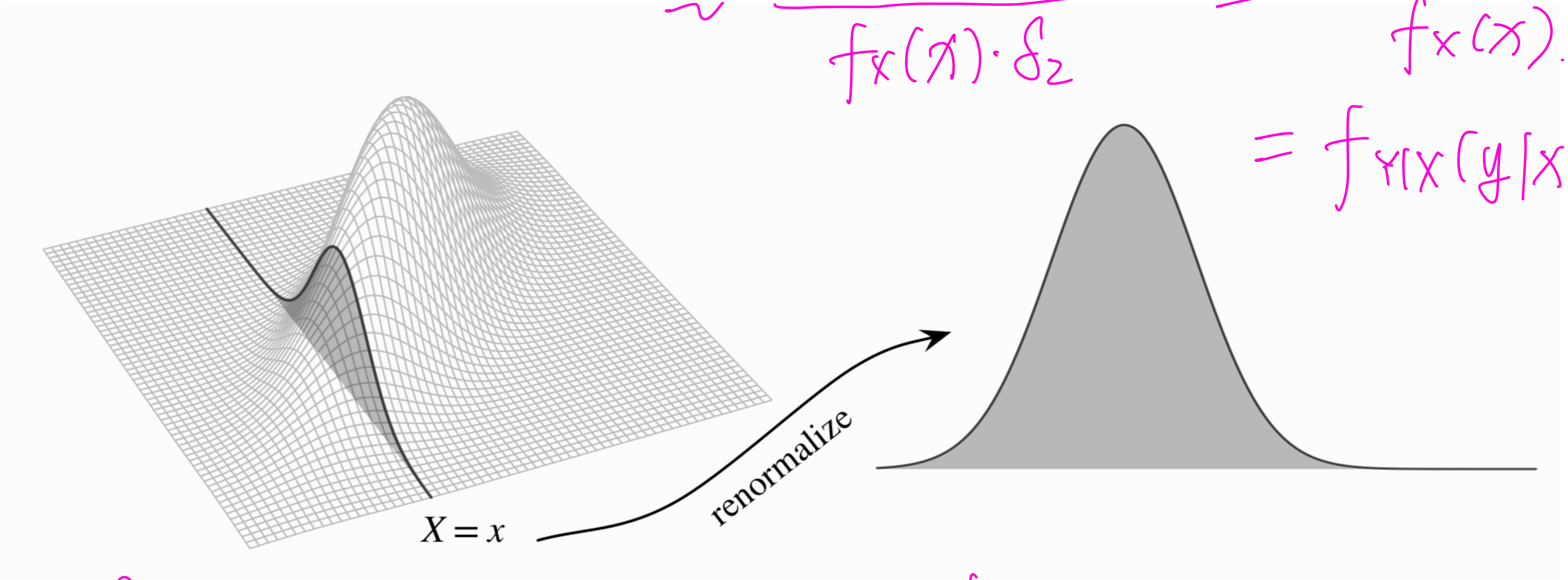
Conditional PDF

$\delta_1, \delta_2 \simeq 0$

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

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

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



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

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

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

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

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

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

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 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

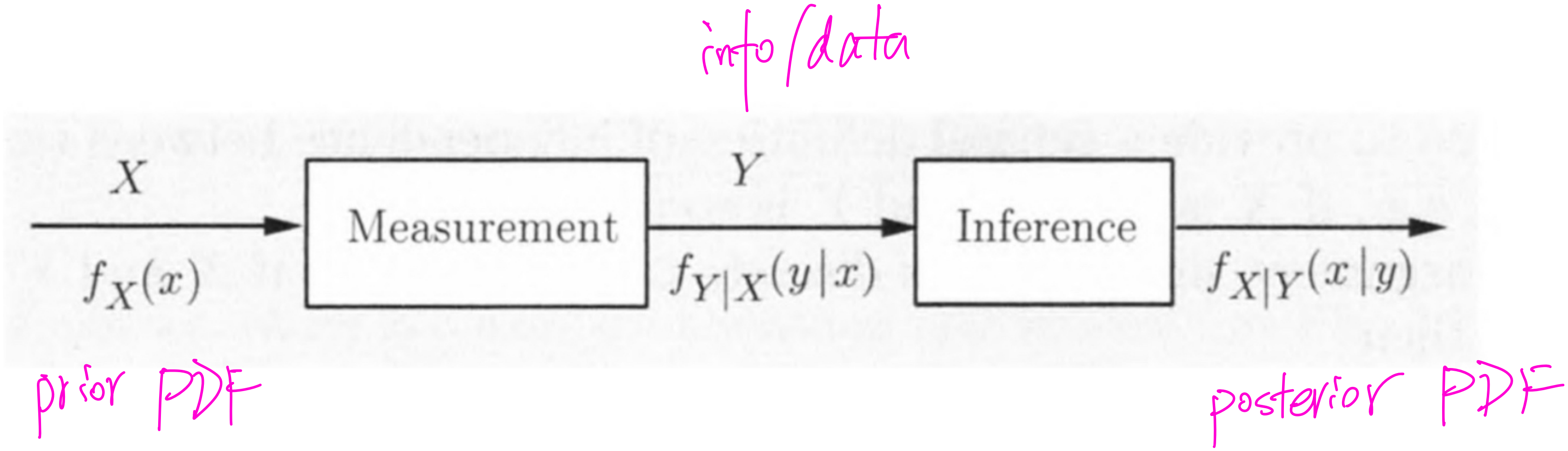
For continuous r.v.s X and Y,

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

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

Proof

Bayes' Rule: Inference Perspective



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

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,

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

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

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

General LOTP

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

Proof

Example

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?

Example: Comparing Exponentials of Different Rates

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

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

Theorem

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 non-negative 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

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

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

Expected Distance between Two Normals

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

Outline

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

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

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

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

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

Uncorrelated

Theorem

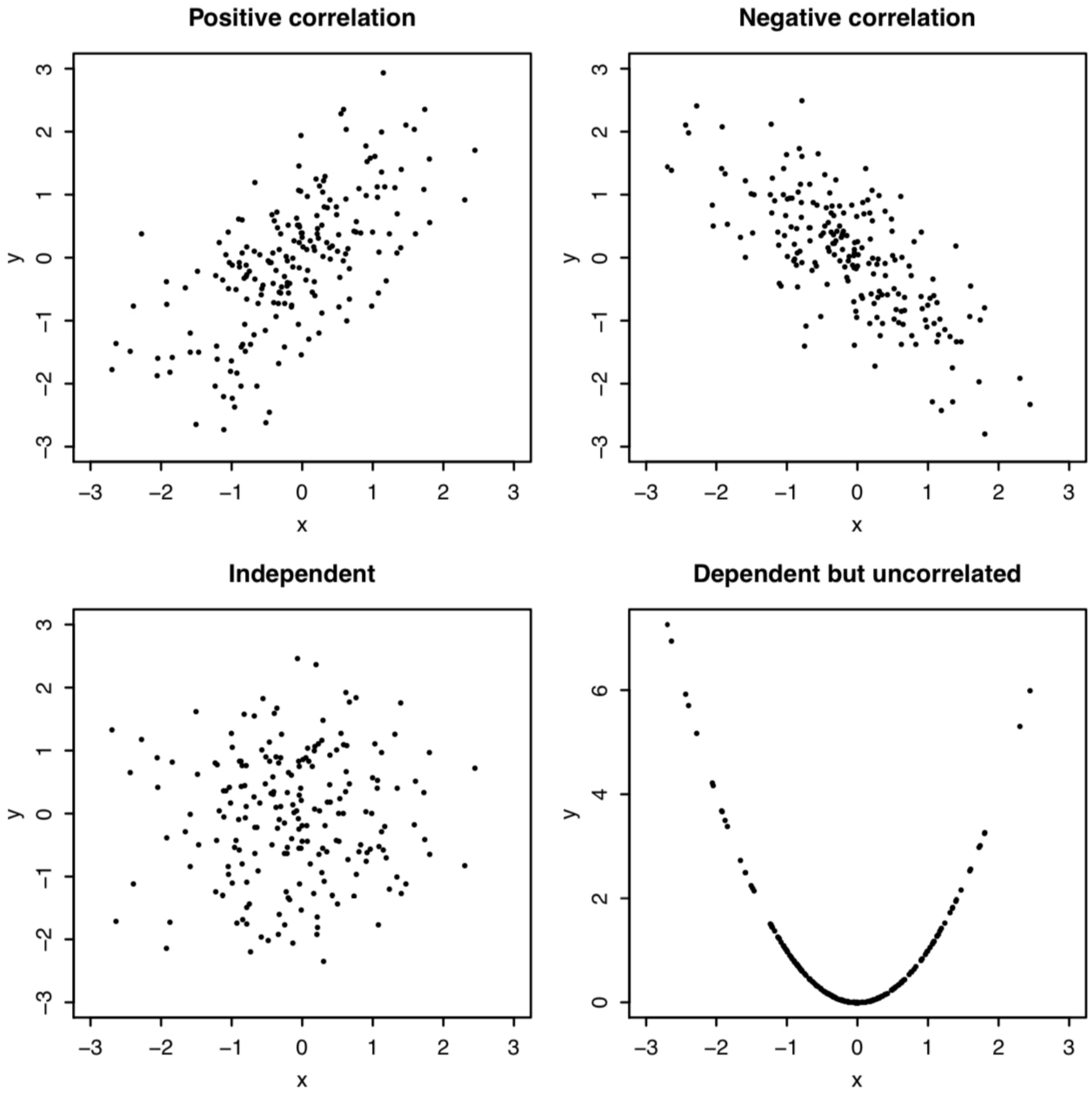
If X and Y are independent, then they are uncorrelated.

Uncorrelated $\nRightarrow$ 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



Correlation Bounds

Theorem

For any r.v.s X and Y,

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

Example: Exponential Max and Min

Let X and Y be i.i.d. $\mathrm{Expo}(1)$ r.v.s. Find the correlation between $\max(X, Y)$ and $\min(X, Y)$.

Solution

Outline

Story

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 non-negative 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 $p = (p_1, \ldots, p_k)$. We write this as $X \sim \mathrm{Mult}_k(n, p)$.

Multinomial Joint PMF

Theorem

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

Theorem

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

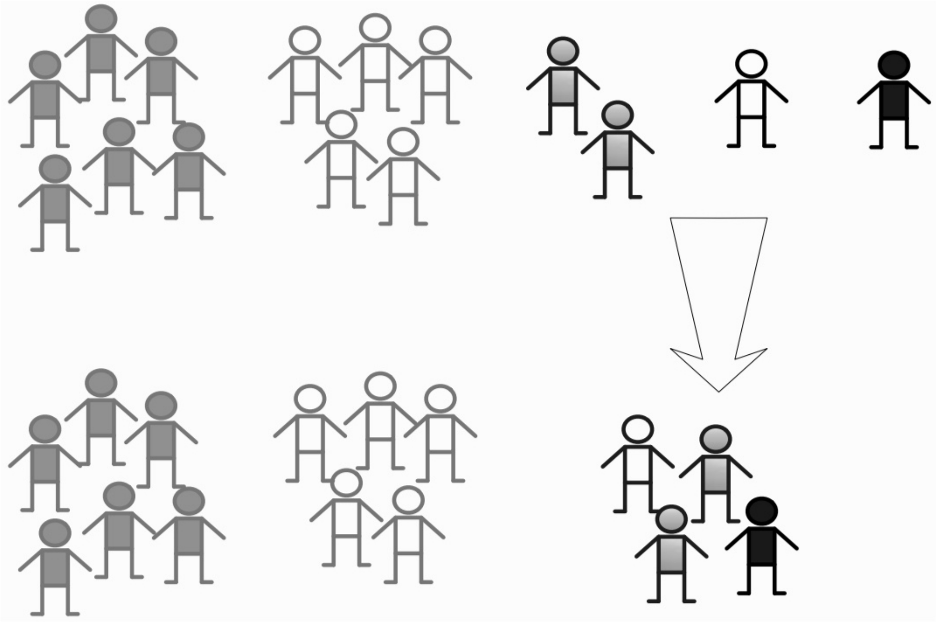
Multinomial lumping

Theorem

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

Theorem

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

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

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

Covariance in A Multinomial

Theorem

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

Proof

Outline

Multivariate Normal Distribution

Definition

A random vector $X = (X_1, \ldots, X_k)$ 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

$$t_1 X_1 + \ldots + t_k X_k$$

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

Actual MVN

Theorem

Theorem

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

theorem

Theorem

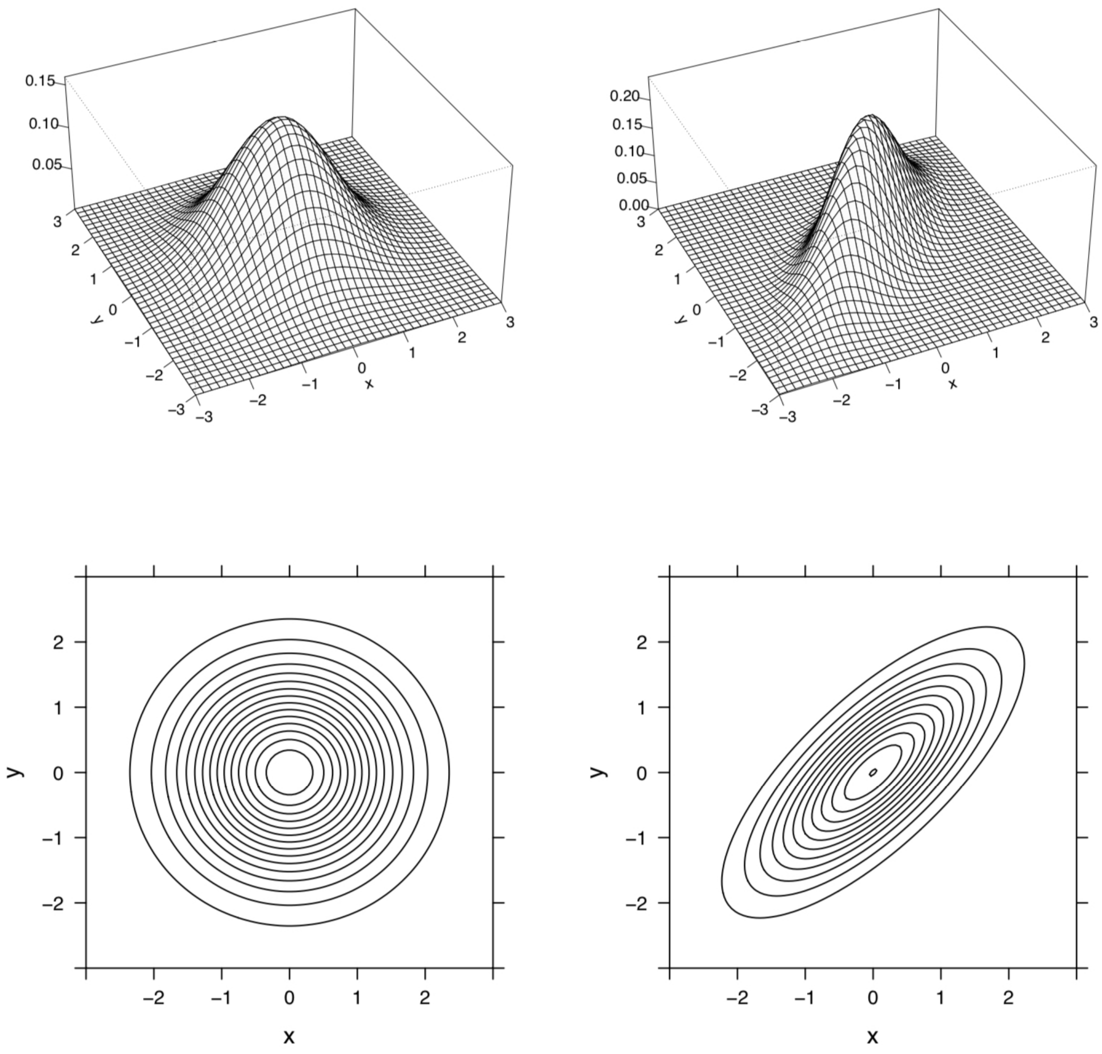
If $X = (X_1, \ldots, X_n)$ *and* $Y = (Y_1, \ldots, Y_m)$ *are MVN vectors with* X *independent of* Y*, then the concatenated random vector* $W = (X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ *is Multivariate Normal.*

parameters of MVN

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 $\mathrm{Cov}(X_i, X_j)$.

joint PDF of Bivariate Normal Distributions



Joint MGF

Definition

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

$$M(t) = E(e^{t'X}) = E(e^{t_1 X_1 + \cdots + t_k X_k}).$$

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 $X \sim \mathrm{MVN}$ can be written as $X = (X_1, X_2)$, where X_1 and X_2 are subvectors, and every component of X_1 is uncorrelated with every component of X_1 is uncorrelated with every component of X_2, then X_1 and 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

Independence of Sum and Difference

Let $X, Y \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Find the joint distribution of $(X+Y, X-Y)$.

Independence of Sample Mean and Sample Variance

Let $X_1, \ldots, X_n$ be i.i.d. $\mathcal{N}(\mu, \sigma^2)$, with $n \geq 2$. Define

$$\bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n),$$

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2.$$

The sample mean $\bar{X}_n$ has expectation μ (the true mean) and the sample variance S_n^2 has expectation σ^2 (the true variance). Show that $\bar{X}_n$ and S_n^2 are independent by applying MVN ideas and results to the vector $X_n, X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n$.

proof

Proof

Bivariate Normal generation

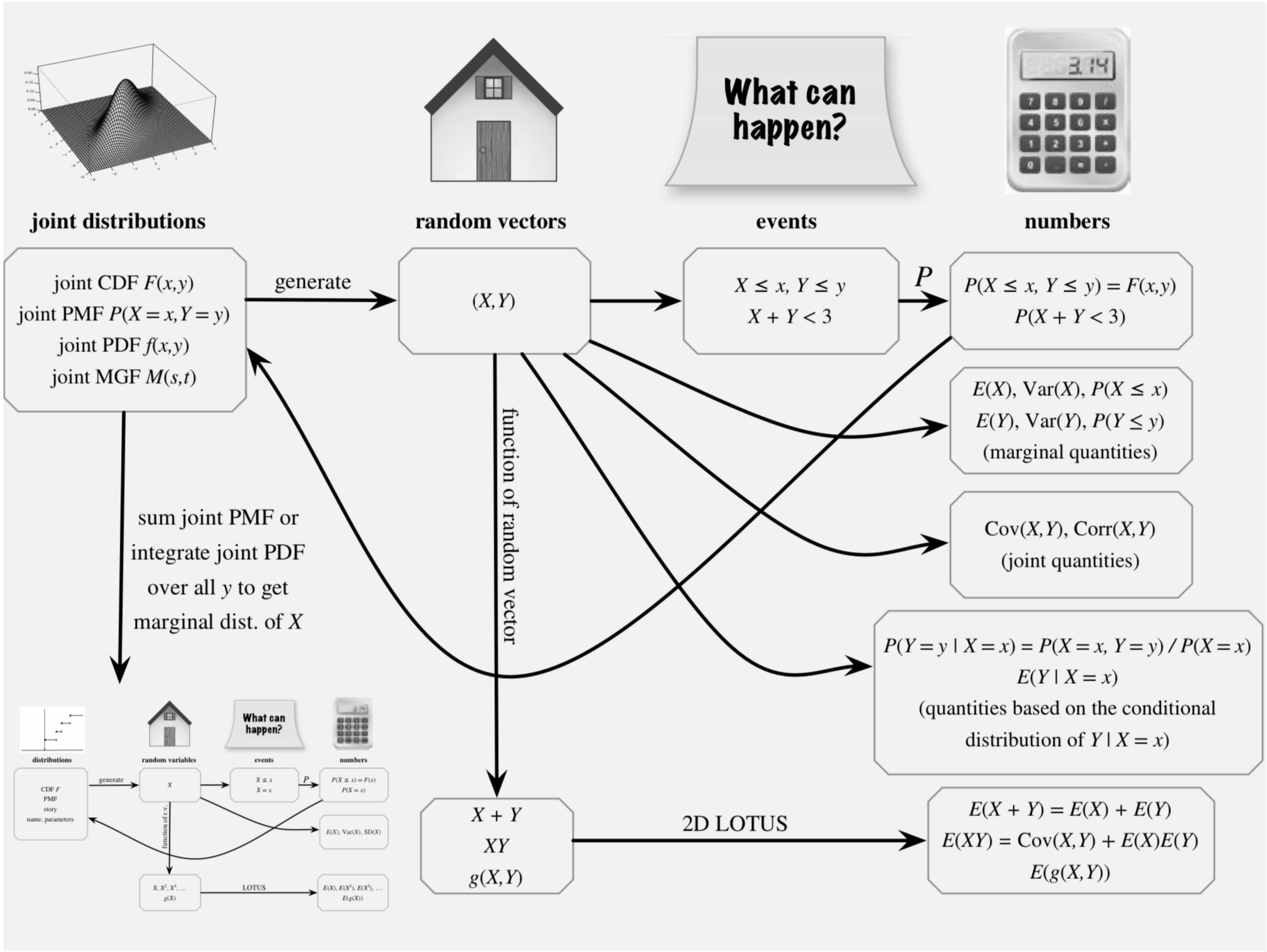
Suppose that we have access to i.i.d. r.v.s $X, Y \sim \mathcal{N}(0,1)$, but want to generate a Bivariate Normal (Z, W) with $\mathrm{Corr}(Z, W) = \rho$ and Z, W marginally $\mathcal{N}(0,1)$, for the purpose of running a simulation. How can we construct Z and W from linear combinations of X and Y?

Solution

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**