概率论与数理统计

第三章：联合分布

李立颖 (lily@sustech.edu.cn)

南方科技大学数学系

2023 秋季

多维随机变量的背景

例

- 人的身高 H 与体重 W
- 某地区的气温 X, 气压 Y 与湿度 Z
- 射击中落点横向偏差与纵向偏差 Y

问

能不能将上述随机变量单独分别进行研究?

分析

- 一般人的身高 $H \sim \mathcal{N}(\cdot,\cdot)$, $W \sim \mathcal{N}(\cdot,\cdot)$. 但身高与体重之间有一定关系.
- 气象指标中的气温、气压与湿度也是相关联的
- 导弹射程误差与落点的横向偏差及纵向偏差都有关

由于同一对象的不同指标之间往往是有一定联系的, 所以应该把它们作为一个从中整体来看待.

二维随机变量的概念

定义

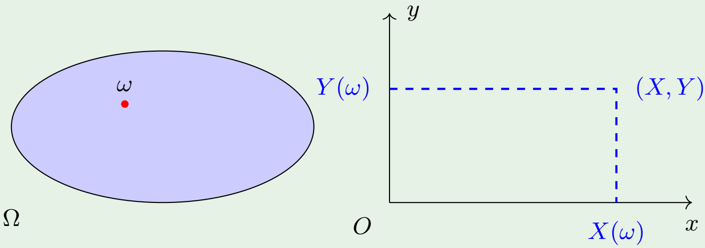
设 Ω 为样本空间, $X = X(\omega)$, $Y = Y(\omega)$, $\omega \in \Omega$, 是定义在 Ω 上的两个随机变量. 记

$$(X, Y) := \big(X(\omega), Y(\omega)\big), \quad \omega \in \Omega.$$

称 (X, Y) 为**二维随机变量 (向量)**.

注

一个试验产生的二维随机变量可视为向二维平面"投掷"一个"随机点".



累积分布函数

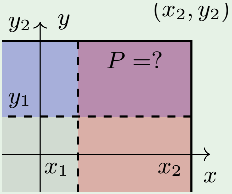
定义

设 (X,Y) 为二维随机变量. 对任意 $x,y\in\mathbb{R}$, 定义

$$F(x,y) := \mathsf{P}\Big(\{X\le x\}\cap\{Y\le y\}\Big) = \mathsf{P}(X\le x,\ Y\le y).$$

称 $F(x,y)$ 为二维随机变量 (X,Y) 的**累积分布函数**, 或称为 X 与 Y 的**联合累积分布函数**.

计算 $\mathsf{P}(X\in(x_1,x_2],\ Y\in(y_1,y_2])$



$$\mathsf{P}(X\in(x_1,x_2],\ Y\in(y_1,y_2]) = F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)\ge 0.$$

分布函数 $F(x,y)$ 的本质特征

单调性

对任意固定 x_0, $F(x_0,y)$ 是 y 的单调不减函数; 对任意固定 y_0, $F(x,y_0)$ 是 x 的单调不减函数.

无穷处极限

$0 \le F(x,y) \le 1$ 且

$$F(\infty,\infty)=1, \quad F(-\infty,y)=F(x,-\infty)=F(-\infty,-\infty)=0, \quad \forall x,y.$$

右连续性

对任意 x,y, $F(x,\cdot)$, $F(\cdot,y)$ 是右连续的.

二维单调性

$$F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1)=\mathsf{P}(X\in(x_1,x_2],\ Y\in(y_1,y_2])\ge 0.$$

二维单调性

注

最后一条性质不能由前三条性质推出.

反例

$$F(x,y)=\begin{cases}1, & x+y\geq -1,\\ 0, & x+y<-1.\end{cases}$$

则 $F(1,1)-F(-1,1)-F(1,-1)+F(-1,-1)=1-1-1+0<0.$

n 维随机变量

定义

设 $X_1, X_2, \ldots, X_n$ 是定义在样本空间 Ω 上的 n 个随机变量. 称

$$(X_1, X_2, \ldots, X_n)$$

为 n **维随机变量** 或 n **维随机向量**.

分布函数

称 n 元函数

$$F(x_1, x_2, \ldots, x_n) = \mathsf{P}(X_1 \le x_1,\ X_2 \le x_2, \ldots, X_n \le x_n)$$

为 n 维随机向量 $(X_1, X_2, \ldots, X_n)$ 的**分布函数** 或 $X_1, X_2, \ldots, X_n$ 的**联合分布 (函数)**.

二维随机变量的基本分类

- 二维离散型随机变量
- 二维连续型随机变量

边缘分布 (marginal distribution)

如果 (X,Y) 是一个二维随机变量, 则它的分量 X (或者 Y) 是一维随机变量. 因此 X 或 Y 也有分布.

定义

称 X 或 Y 的分布为 X 或 Y 关于二维随机变量 (X,Y) 的**边缘分布 (边际分布)**.

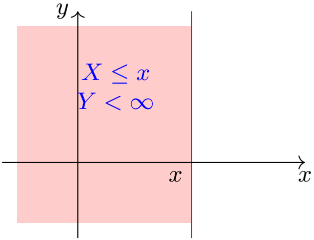
边际分布函数

- 二维随机变量的整体概率特性: $(X, Y) \sim F(x, y)$.
- 两个一维随机变量的概率特性: $X \sim F_X(x)$, $Y \sim F_Y(y)$.

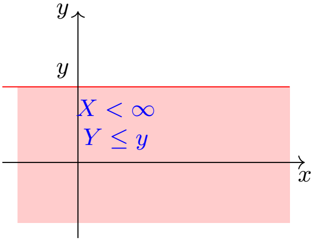
定义

称 $F_X(x)$、$F_Y(y)$ 为 (X, Y) 关于 X、Y 的**边际分布 (函数)**.

联合分布与边际分布的关系



$$F_X(x) = F(x, \infty)$$



$$F_Y(y) = F(\infty, y)$$

性质

随机变量的边际分布完全由它们的联合分布决定. **反之不然.**

例题

设随机变量 (X,Y) 的联合分布函数为

$$F(x,y)=A\left(B+\arctan\frac{x}{2}\right)\left(C+\arctan\frac{y}{2}\right),\quad -\infty<x,y<\infty,$$

其中 A,B,C 为常数.

1. 确定 A,B,C.
2. 求 X 和 Y 的边缘分布函数.
3. 求 $\mathsf{P}(X>2)$.

解

$$\begin{aligned} F(\infty,\infty)&=A(B+\pi/2)(C+\pi/2)=1\\ F(-\infty,\infty)&=A(B-\pi/2)(C+\pi/2)=0\\ F(\infty,-\infty)&=A(B+\pi/2)(C-\pi/2)=0. \end{aligned}$$

因此 $B=C=\pi/2,\ A=\dfrac{1}{\pi^2}$.

解

$$F_X(x) = F(x,\infty) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{x}{2}, \quad -\infty < x < \infty,$$
$$F_Y(y) = F(\infty,y) = \frac{1}{2} + \frac{1}{\pi}\arctan\frac{y}{2}, \quad -\infty < y < \infty,$$

$$\mathsf{P}(X > 2) = 1 - \mathsf{P}(X \le 2) = 1 - F_X(2) = 1 - \left(\frac{1}{2} + \frac{1}{\pi}\arctan\frac{2}{2}\right) = \frac{1}{4}.$$

二维离散型随机变量

回顾: 二维随机变量分类

- 二维离散型随机变量
- 二维连续型随机变量

定义

设随机变量 (X,Y) 的所有可能的取值为 (x_i,y_j), $i,j=1,2,\ldots$, 取值的概率为

$$\mathsf{P}(X=x_i,\ Y=y_j)=p(x_i,y_j):=p_{i,j},\quad i,j=1,2,\ldots$$

则称 (X,Y) 为**离散型随机变量**, 称 $p_{i,j}$ 为它的**(联合) 频率函数 (joint frequency function)**.

频率函数的基本性质

本质特征

设随机变量 (X,Y) 的频率函数为

$$p_{ij} = \mathsf{P}(X = x_i,\ Y = y_j), \quad i,j = 1,2,\ldots$$

则

- $p_{ij} \geq 0,\ i,j = 1,2,\ldots$
- $\displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{ij} = 1.$

频率函数的表格表示法

$Y \backslash X$	x_1	x_2	$\cdots$	x_i	$\cdots$
y_1	p_{11}	p_{21}	$\cdots$	p_{i1}	$\cdots$
y_2	p_{12}	p_{22}	$\cdots$	p_{i2}	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_j	p_{1j}	p_{2j}	$\cdots$	p_{ij}	$\cdots$

例题

袋中装有 2 只白球及 3 只黑球. 现进行无放回的摸球, 定义随机变量如下:

$$X = \begin{cases} 1, & \text{第一次摸出白球,} \\ 0, & \text{第一次摸出黑球,} \end{cases} \quad Y = \begin{cases} 1, & \text{第二次摸出白球,} \\ 0, & \text{第二次摸出黑球.} \end{cases}$$

求 (X, Y) 的频率函数.

解

$$\mathsf{P}(X=0,\ Y=0) = \mathsf{P}(Y=0 \mid X=0)\cdot\mathsf{P}(X=0) = \frac{2}{4}\cdot\frac{3}{5} = \frac{6}{20},$$
$$\mathsf{P}(X=0,\ Y=1) = \mathsf{P}(Y=1 \mid X=0)\cdot\mathsf{P}(X=0) = \frac{2}{4}\cdot\frac{3}{5} = \frac{6}{20},$$
$$\mathsf{P}(X=1,\ Y=0) = \mathsf{P}(Y=0 \mid X=1)\cdot\mathsf{P}(X=1) = \frac{3}{4}\cdot\frac{2}{5} = \frac{6}{20},$$
$$\mathsf{P}(X=1,\ Y=1) = \mathsf{P}(Y=1 \mid X=1)\cdot\mathsf{P}(X=1) = \frac{1}{4}\cdot\frac{2}{5} = \frac{2}{20}.$$

例题

有一个射击游戏，参加游戏的人先掷一下骰子，若出现点数为 X，则射击 X 次. 设某人击中目标概率为 $p = 0.9$. 记击中目标的次数为 Y. 求 (X, Y) 的频率函数.

分析

X 的取值为 $1, 2, \ldots, 6$, Y 的取值为 $0, 1, \ldots, X$. 当 $X = i$ 时, $Y \sim \mathrm{Bin}(i, p)$, $i = 1, 2, \ldots, 6$. 我们应当用乘法公式计算概率!

解

由乘法公式及如上分析,

$$\begin{aligned}\mathsf{P}(X = i,\ Y = j) &= \mathsf{P}(Y = j \mid X = i)\mathsf{P}(X = i) \\ &= \frac{1}{6}\binom{i}{j}p^j(1-p)^{i-j}, \quad 0 \le j \le i,\ 1 \le i \le 6.\end{aligned}$$

边际频率函数

设 (X,Y) 的频率函数为 $\mathsf{P}(X=x_i,\ Y=y_j)=p_{ij},\ i,j=1,2,\ldots$. 则随机变量 X 的频率函数是

$$\mathsf{P}(X=x_i)=\sum_{j=1}^{\infty}p_{ij}:=p_{i\cdot},\quad i=1,2,\ldots$$

同理 Y 的频率函数是

$$\mathsf{P}(Y=y_j)=\sum_{i=1}^{\infty}p_{ij}:=p_{\cdot j},\quad j=1,2,\ldots$$

定义

称数列 $\{p_{i\cdot}\}$ 为 (X,Y) 关于 X 的**边际频率函数**, 称数列 $\{p_{\cdot j}\}$ 为 (X,Y) 关于 Y 的**边际频率函数 (marginal frequency function)**.

边际频率函数的性质

- 它是一维随机变量的频率函数
- 它可通过二维随机变量的频率函数计算得到.

例子

设随机变量 X 从 1, 2, 3, 4 中等可能取值, 又设随机变量 Y 从 $1,2,\ldots,X$ 中等可能取值. 求 X,Y 的联合频率函数以及边际频率函数.

解

当 $X=i$ 时, Y 以 $\dfrac{1}{i}$ 的概率等可能取到 1 至 i. 因此由乘法公式,

$$\mathsf{P}(X=i,\ Y=j)=\mathsf{P}(Y=j\mid X=i)\cdot\mathsf{P}(X=i)=\frac{1}{i}\cdot\frac{1}{4},\quad 1\le j\le i.$$

因此 X,Y 的频率函数及边际频率函数为

$Y \backslash X$	1	2	3	4	$p_{\cdot j}=\sum_{i=1}^4 p_{ij}$
1	1/4	1/8	1/12	1/16	25/48
2	0	1/8	1/12	1/16	13/48
3	0	0	1/12	1/16	7/48
4	0	0	0	1/16	3/48
$p_{i\cdot}=\sum_{j=1}^4 p_{ij}$	1/4	1/4	1/4	1/4	

n 维离散型随机变量的边际频率函数

设 $X_1, \ldots, X_n$ 的联合频率函数为

$$\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n) = p(x_1, \ldots, x_n).$$

边际频率函数

随机变量 X_1 的**边际频率函数**是

$$\mathsf{P}_{X_1}(x_1) = \sum_{x_2, \ldots, x_n} p(x_1, x_2, \ldots, x_n).$$

随机变量 X_1, X_2 的**二维边际频率函数** 是

$$p_{X_1 X_2}(x_1, x_2) = \sum_{x_3, \ldots, x_n} p(x_1, x_2, \ldots, x_n).$$

多项分布 (multinomial distribution): 二项分布的推广

定义

假设进行 n 次独立试验, 每次试验有 r 种可能的结果, 各自出现的概率是 $p_1,\ldots,p_r$. 令 N_i 是 n 次试验中出现第 i 种试验结果的所有次数, 其中 $i=1,\ldots,r$. 则 $N_1,\ldots,N_r$ 的联合频率函数是

$$p(n_1,\ldots,n_r)=\binom{n}{n_1\cdots n_r}p_1^{n_1}p_2^{n_2}\cdots p_r^{n_r},$$

称为**多项分布**.

两种计算 N_i 边际频率函数的方法

- 将联合频率函数关于其它的 n_j 求和;
- N_i 可解释为 n 次试验中**成功的次数**, 故 $N_i\sim\mathrm{Bin}(n,p_i)$, 因此

$$p_{N_i}(n_i)=\binom{n}{n_i}p_i^{n_i}(1-p_i)^{n-n_i}.$$

例子

箱子里装有 4 只白球和 2 只黑球. 在其中随机地取两次, 每次取一只. 分别考虑 "有放回抽样" 与 "无放回抽样". 定义

$$X=\begin{cases}0, & \text{第一次为黑球},\\ 1, & \text{第一次为白球}.\end{cases}\quad Y=\begin{cases}0, & \text{第二次为黑球},\\ 1, & \text{第二次为白球}.\end{cases}$$

求 X, Y 的联合分布律和边缘分布律.

解

有放回抽样:

Y \ X	0	1	$p_{i\cdot}$
0	1/9	2/9	1/3
1	2/9	4/9	2/3
$p_{\cdot j}$	1/3	2/3	1

无放回抽样:

Y \ X	0	1	$p_{i\cdot}$
0	1/15	4/15	1/3
1	4/15	6/15	2/3
$p_{\cdot j}$	1/3	2/3	1

例题

袋中有 1 个红球, 2 个黑球, 3 个白球. 现有放回地取两次, 每次取一球, 以 X, Y, Z 分别表示取球所得红、黑、白球个数. 求: $\mathsf{P}(X=1\mid Z=0)$, $\mathsf{P}(X=1,\ Z=0)$, 以及 (X,Y) 的分布.

解

$$\mathsf{P}(X=1\mid Z=0)=\mathsf{P}(\mathrm{Bin}(2,1/3)=1)=\frac{1}{3}\times\frac{2}{3}+\frac{2}{3}\times\frac{1}{3}=\frac{4}{9}$$

$$\mathsf{P}(X=1,\ Z=0)=\frac{1}{6}\times\frac{2}{6}+\frac{2}{6}\times\frac{1}{6}=\frac{1}{9}.$$

X,Y 的取值均为 0, 1, 2. 我们只计算几个例子:

$$\mathsf{P}(X=0,\ Y=0)=\frac{3}{6}\times\frac{3}{6},\qquad \text{两个均为白球}$$

$$\mathsf{P}(X=0,\ Y=1)=\frac{2}{6}\times\frac{3}{6}\times 2=\frac{1}{3},\qquad \text{黑白或白黑}$$

$$\mathsf{P}(X=1,Y=2)=0,\qquad \text{总数超 2 只, 不可能}$$

$$\mathsf{P}(X=2,Y=0)=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36},\qquad \text{两只均为红球.}$$

作业

- P76: 3
- 补充题
  1. 把一枚均匀硬币抛掷三次, 设 X 为三次抛掷中正面出现的次数, 而 Y 为正面出现次数与反面出现次数之差的绝对值. 求 (X, Y) 的频率函数.
  2. 设 X 的分布为 $\mathsf{P}(X=-1)=\mathsf{P}(X=0)=\mathsf{P}(X=1)=\frac{1}{3}$. 令 $Y=X^2$. 求 (X, Y) 的联合频率函数及边缘频率函数.
  3. 设随机变量 Y 服从参数为 1 的指数分布, 随机变量
$$X_k=\begin{cases}0, & Y \leq k, \\ 1, & Y>k,\end{cases} \quad k=1,2.$$
求二维随机变量 (X_1, X_2) 的联合频率函数及边缘频率函数.

二维连续型随机变量

定义

设随机变量 (X, Y) 的联合分布函数 (joint cdf) 为

$$F(x,y) = \mathsf{P}(X \le x,\ Y \le y).$$

若存在非负可积函数 $f(x,y) \ge 0$ 使得

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dudv, \quad (x,y) \in \mathbb{R}^2,$$

则称 (X, Y) 为**二维连续型随机变量**, 称 $f(x,y)$ 为**概率密度函数 (密度函数、密度、联合密度函数)**.

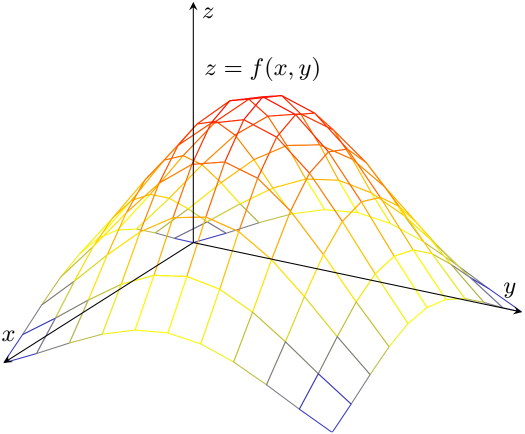
注

由微积分知识, $F(x,y)$ 是连续函数.

密度函数的基本性质

本质特征

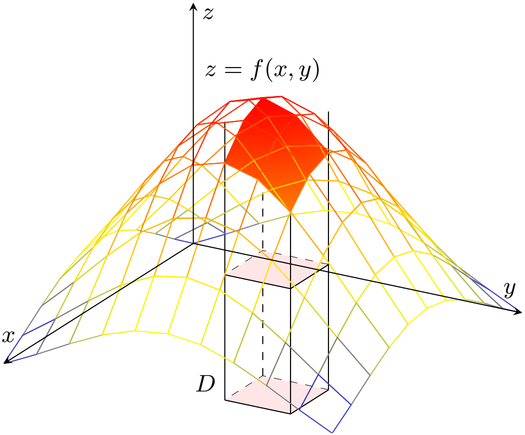
- $f(x, y) \geq 0, \ \forall (x, y) \in \mathbb{R}^2,$
- $\displaystyle\int_{\mathbb{R}}\int_{\mathbb{R}} f(u, v)\, dudv = 1.$ **几何意义：曲面** $z = f(x, y)$ **与** xOy **平面围成的体积为 1**.



区域概率

对任意 (由逐段光滑曲线围成的) 区域 $D \subset \mathbb{R}^2$,

$$\mathsf{P}\big((X,Y) \in D\big) = \iint_D f(x,y)\,dxdy.$$



分布函数与密度函数

性质

在 $f(x,y)$ 的连续点处，有

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = f(x,y).$$

注

在 $f(x,y)$ 的连续点处，由导数的定义有

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y} = \lim_{\Delta x \to 0^+,\ \Delta y \to 0^+} \frac{\mathsf{P}(x < X \le x + \Delta x,\ y < Y \le y + \Delta y)}{\Delta x \Delta y}.$$

密度函数的意义

$$\mathsf{P}(x < x \le x + \Delta x,\ y < Y \le y + \Delta y) \approx f(x,y)\Delta x \Delta y.$$

例题

设随机变量 (X,Y) 的概率密度为

$$f(x,y) = \mathbb{1}_{\{x>0,\ y>0\}} k e^{-(2x+y)}.$$

1. 确定常数 k;
2. 求分布函数 $F(x,y)$;
3. 计算概率 $\mathrm{P}(Y \le X)$.

解

1. 由归一性,

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(u,v)\,dudv = k\int_0^{\infty}\int_0^{\infty} e^{-(2x+y)}\,dxdy$$
$$= k\left[\int_0^{\infty} e^{-2x}\,dx\right]\cdot\left[\int_0^{\infty} e^{-y}\,dy\right] = \frac{k}{2}.$$

因此 $k=2$.

$$f(x,y) = \mathbb{1}_{\{x>0,\ y>0\}} k e^{-(2x+y)}.$$

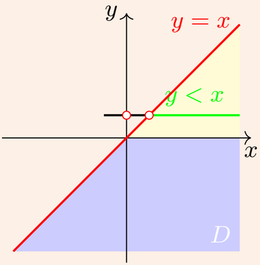
解

❷ 显然，分布函数只有在 $x>0$, $y>0$ 时才非零. 我们有

$$\begin{aligned} F(x,y) &= \mathbb{1}_{\{x>0,\ y>0\}} \int_0^x \int_0^y 2e^{-(2u+v)}\,du dv \\ &= \mathbb{1}_{\{x>0,\ y>0\}} (1-e^{-2x})(1-e^{-y}). \end{aligned}$$

❸ 记 $D = \{(x,y) \mid y \le x,\ x,y \in (-\infty,\infty)\}$. 于是

$$\begin{aligned} \mathsf{P}(Y \le X) &= \mathsf{P}\big((X,Y) \in D\big) \\ &= \iint_D f(x,y)\,dxdy \\ &= \iint_{D\cap\{x,y>0\}} 2e^{-(2x+y)}\,dxdy \\ &= \int_0^\infty dy \int_y^\infty 2e^{-(2x+y)}\,dx = \frac{1}{3}. \end{aligned}$$



例题

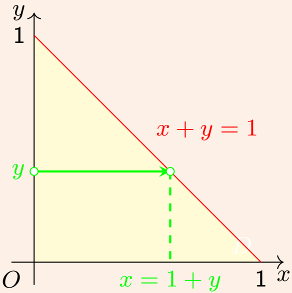
设随机变量 (X,Y) 的密度函数为

$$f(x,y) = \mathbb{1}_{\{x>0,\ y>0\}} 2e^{-(2x+y)}.$$

计算概率 $\mathsf{P}(X+Y\le 1)$.

解

$$\begin{aligned}\mathsf{P}(X+Y\le 1) &= \iint_{\{x+y\le 1\}} f(x,y)\,dxdy \\ &= \iint_{\{x+y\le 1, x>0, y>0\}} 2e^{-(2x+y)}\,dxdy \\ &= \int_0^1 dy \int_0^{1-y} 2e^{-(2x+y)}\,dx \\ &= 1-2e^{-1}+e^{-2}.\end{aligned}$$



约定

记号

$$(X,Y) \sim F(x,y)$$

表示二维随机变量 (X,Y) 的分布函数为 $F(x,y)$.

$$(X,Y) \sim f(x,y)$$

表示二维随机变量 (X,Y) 的概率函数为 $f(x,y)$, 即

- 离散型随机变量 $f(x,y)$ 表示频率函数 (PMF);
- 连续型随机变量 $f(x,y)$ 表示密度函数 (PDF).

n 维随机变量的记号

$$(X_1, X_2, \ldots, X_n) \sim F(x_1, x_2, \ldots, x_n), \quad (X_1, X_2, \ldots, X_n) \sim f(x_1, x_2, \ldots, x_n).$$

设 (X,Y) 的分布函数和密度函数分别为 $F(x,y)$, $f(x,y)$.

X 的边际分布

随机变量 X 的分布函数为

$$F_X(x) = \mathsf{P}(X \le x) = \mathsf{P}(X \le x,\ Y < \infty) = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f(u,y)\,dy \right] du$$

因此随机变量 X 的密度函数为

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad x \in \mathbb{R}.$$

Y 的边际分布

$$F_Y(y) = \int_{-\infty}^{y} \left[\int_{-\infty}^{\infty} f(x,v)\,dx \right] dv, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

定义

称 $f_X(x)$ $(f_Y(y))$ 为 (X,Y) 关于 X (Y) 的**边际密度 (函数)**.

注

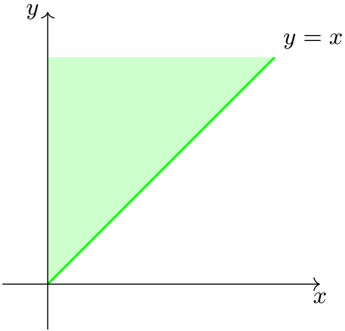
已知联合密度可以求得边际密度.

例子

设 (X,Y) 的概率密度是

$$f(x,y)=\begin{cases} e^{-y}, & x>0, y>x, \\ 0, & \text{其它}. \end{cases}$$

求 (X,Y) 关于 X 和 Y 的边际密度函数.



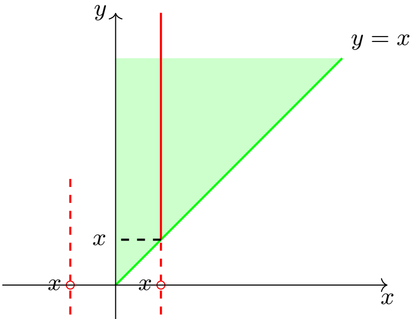
X 的边际分布

解

$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$. 当 $x \le 0$ 时, 显然 $f_X(x) = 0$. 当 $x > 0$ 时, 由图知

$$f_X(x) = \int_x^{\infty} e^{-y}\,dy = e^{-y}|_{\infty}^{x} = e^{-x}, \quad x \ge 0.$$

因此 $f_X(x) = \mathbb{1}_{\{x\ge 0\}} e^{-x}$ (指数分布).



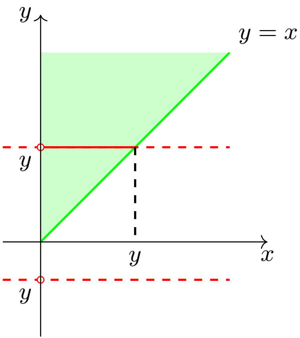
Y 的边际分布

解

当 $y \le 0$ 时, $f_Y(y) = 0$. 当 $y > 0$ 时,

$$f_Y(y) = \int_0^y e^{-y}\,dx = ye^{-y}.$$

故 $f_Y(y) = \mathbb{1}_{\{y>0\}} ye^{-y}$.



例题

设 (X,Y) 的联合密度为

$$f(x,y)=\begin{cases}6, & x^2\le y\le x,\\ 0, & \text{其它}.\end{cases}$$

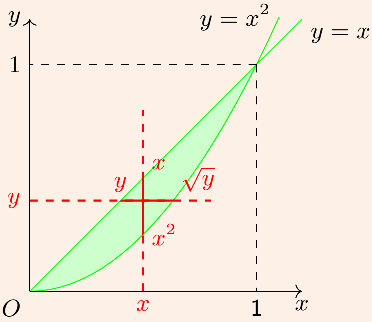
求边际密度 $f_X(x)$, $f_Y(y)$.

解

由图知 $0\le X,Y\le 1$. 因此

$$f_X(x)=\mathbb{1}_{(0,1)}(x)\int_{x^2}^{x}6\,dy=\mathbb{1}_{(0,1)}(x)6x(1-x),$$

$$f_Y(y)=\mathbb{1}_{(0,1)}(y)\int_{y}^{\sqrt{y}}6\,dx=6(\sqrt{y}-y).$$



n 维连续型变量的边际密度

定义

设 X, Y, Z 的联合密度函数为 $f(x, y, z)$, 则随机变量 X 的一维边际密度函数是

$$f_X(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z)\, dydz,$$

随机变量 X 和 Y 的二维边际密度函数是

$$f_{XY}(x, y) = \int_{-\infty}^{\infty} f(x, y, z)\, dz.$$

联合分布的非唯一性

Farlie–Morgenstern 族

设 $F(x)$ 和 $G(y)$ 都是一维连续型分布函数. 则对任意的 α, $|\alpha| \le 1$,

$$H(x,y) = F(x)G(y)\Big[1 + \alpha\Big(1 - F(x)\Big)\Big(1 - G(y)\Big)\Big]$$

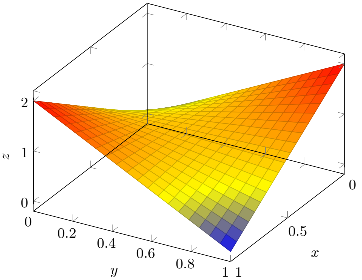
是二维连续型分布函数, 且边际分布为

$$H(x,\infty) = F(x), \quad H(\infty,y) = G(y).$$

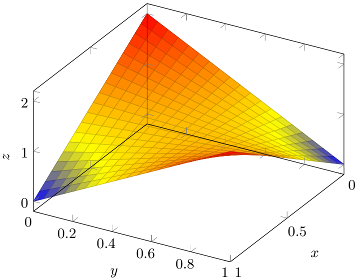
注

按这种方式, 可以构造给定边际分布的无数个不同的二维联合分布.

$F = G = \mathrm{U}(0,1)$ 时 Farlie–Morgensten 族



(a) $\alpha = -1$: $h(x,y) = 2x + 2y - 4xy$



(b) $\alpha = 1$: $h(x,y) = 2 - 2x - 2y + 4xy$

连接函数 (copula)

定义

称那些使得边际分布为均匀分布的联合累积分布函数为**连接函数**, 记为 $C(u,v)$.

性质

1. $C(u,v)$ 关于每个变量都是非降的
2. $\mathsf{P}(U \le u) = C(u,1) = u$, $C(1,v) = v$.
3. 限定连续型连接函数, 即 $C(u,v)$ 具有密度:

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v) \ge 0.$$

由连接函数构造两个分布的耦合 (coupling)

引理

- 假设 X 和 Y 是分布函数分别为 $F_X(x)$ 和 $F_Y(y)$ 的连续型随机变量，则 $U = F_X(X)$ 和 $V = F_Y(Y)$ 是均匀分布 $\mathrm{U}(0,1)$.
- 对于连接函数 $C(u,v)$，考虑定义联合分布

$$F_{XY}(x,y) = C\big(F_X(x), F_Y(y)\big),$$

则其边际分布为 $F_X(x)$ 和 $F_Y(y)$，且密度为

$$f_{XY}(x,y) = c\big(F_X(x), F_Y(y)\big) f_X(x) f_Y(y).$$

注

由两个边际分布和任意连接函数，可以构造出相同边际分的联合分布，即：**边际函数不能决定联合分布**，两个变量的相依性由连接函数控制.

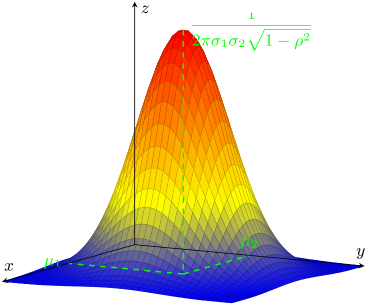
二维正态分布

定义

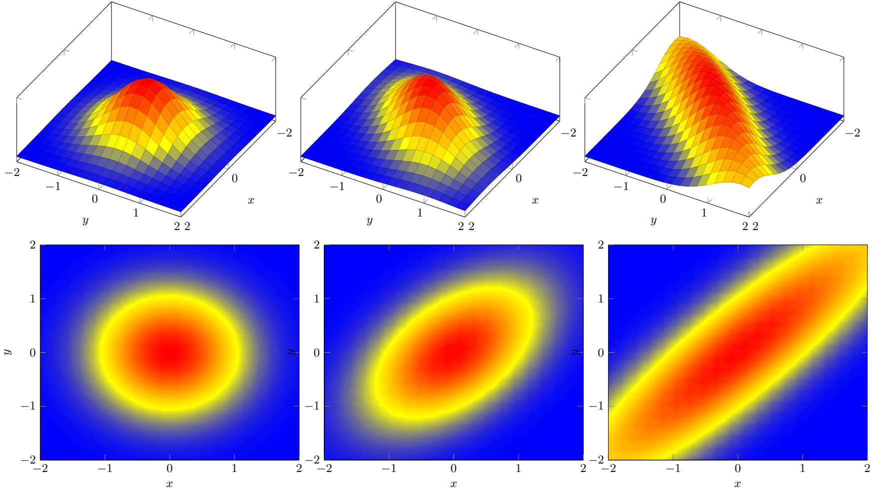
若 X, Y 的联合密度为

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]}$$

则称 (X,Y) 服从参数为 $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的**二维正态分布**, 记为 $(X,Y) \sim \mathcal{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$. 其中 $\mu_1,\mu_2 \in \mathbb{R}$, $\sigma_1,\sigma_2 > 0$, $|\rho| < 1$.

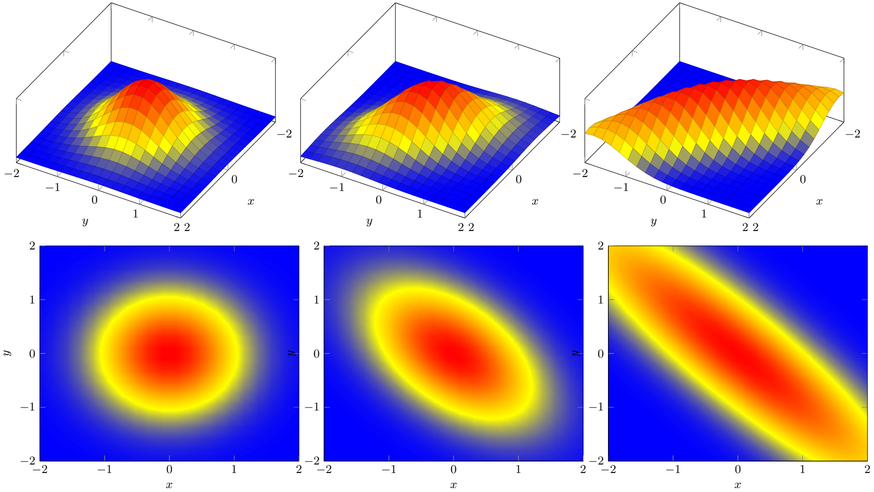


$\mu_1 = \mu_2 = 0,\ \sigma_1 = \sigma_2 = 1$ 的二维正态分布



(a) $\rho = 0$ (b) $\rho = 0.5$ (c) $\rho = 0.9$

$\mu_1 = \mu_2 = 0,\ \sigma_1 = \sigma_2 = 1$ 的二维正态分布

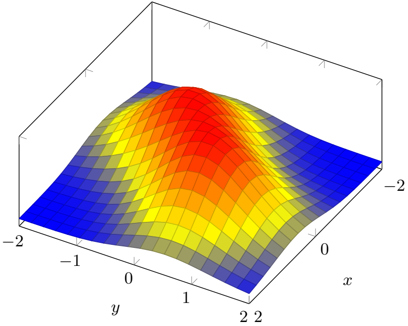


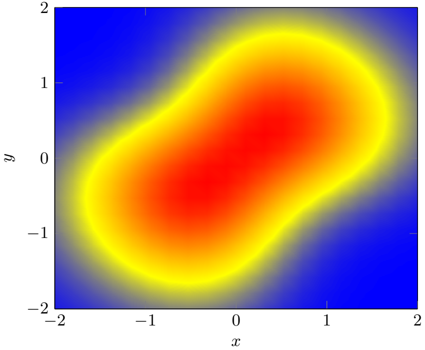
(a) $\rho = 0$ (b) $\rho = -0.5$ (c) $\rho = -0.9$

取边际密度都为 $\mathcal{N}(0,1)$ 的正态分布, 利用密度为 $c(u,v)=2-2u-2v+4uv$ 的连接函数, 则二维联合密度为

$$f(x,y)=\Big[2-2\Phi(x)-2\Phi(y)+4\Phi(x)\Phi(y)\Big]\cdot\varphi(x)\varphi(y).$$

此函数具有边际正态, 但不是二维正态密度.





均匀分布

定义

设有界区域 G 的面积为 A. 若随机变量 (X,Y) 的联合密度函数为

$$f(x,y)=\begin{cases}\dfrac{1}{A}, & (x,y)\in G,\\ 0, & \text{其它}.\end{cases}$$

则称 (X,Y) 服从区域 G 上的**均匀分布**, 记作$(X,Y)\sim \mathrm{U}(G)$.

性质

若 $(X,Y)\sim \mathrm{U}(G)$, 则对任意 $G_1\subset G$,

$$\mathsf{P}\big((X,Y)\in G_1\big)=\frac{|G_1|}{|G|}.$$

这就是二维的几何概型!

例题

设 $(X,Y) \sim \mathrm{U}(G)$, 其中

$$G = \{(x,y) \mid 0 \le y \le x,\ 0 \le x \le 1\}.$$

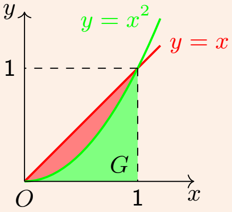
求

1. $f(x,y)$;
2. $\mathsf{P}(Y > X^2)$;
3. (X,Y) 在平面上的落点到 y 轴距离小于 0.3 的概率.

解

1. 显然 $|G| = \int_0^1 x\,dx = \frac{1}{2}$, 故 $f(x,y) = \mathbb{1}_G(x,y)\cdot 2$.

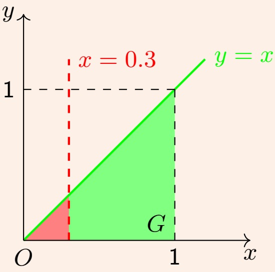
$$\mathsf{P}(Y > X^2) = \int_0^1 dx \int_{x^2}^{x} 2\,dy = \frac{1}{3}.$$



2.

解

3. $\mathsf{P}(|X| < 0.3) = \mathsf{P}(-0.3 < X < 0.3) = \int_0^{0.3} dx \int_0^x 2\,dy = 0.09.$



例题

设随机变量 (X,Y) 服从区域 D 上的均匀分布，其中 $D=\{(x,y)\mid x\geq 0, y\geq 0, x+\frac{y}{2}\leq 1\}$. 试求随机变量 (X,Y) 的边际密度函数.

解

区域 D 的面积为 $A=1$, 所以联合密度函数为 $f(x,y)=\mathbb{1}_D(x,y)$.
当 $x<0$ 或 $x>1$ 时, $f_X(x)=0$. 当 $0\leq x\leq 1$ 时,

$$f_X(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\int_0^{2(1-x)} 1\,dy=2(1-x).$$

因此, $f_X(x)=\mathbb{1}_{(0,1)}(x)\cdot 2(1-x)$.
同理, 随机变量 Y 的边际密度函数为

$$f_Y(y)=\mathbb{1}_{(0,2)}(y)\int_0^{1-\frac{y}{2}} 1\,dx=\mathbb{1}_{(0,2)}(y)\cdot(1-y/2).$$

- P76: 5, 6, 7, 8
- 补充题:
  1. 设二维随机变量 (X,Y) 的联合分布函数为

$$F(x,y)=\begin{cases} k(1-e^{-x})(1-e^{-y}), & x>0,\ y>0, \\ 0, & \text{其它}. \end{cases}$$

     求边际密度函数以及 $\mathrm{P}(1<X<3,\ 1<Y<2)$.
  2. 设二维随机变量 (X,Y) 的密度函数为

$$f(x,y)=\begin{cases} x+y, & 0<x,y<1, \\ 0, & \text{其它}. \end{cases}$$

     (1) 求边际密度函数;
     (2) 求 $\mathrm{P}(X>Y)$;
     (3) 求 $\mathrm{P}(X<0.5)$.

概念的引入

事件的独立性

$$A, B \text{ 相互独立} \Leftrightarrow A \text{ 和 } B \text{ 之间没有任何关系}$$
$$\Leftrightarrow \mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B).$$

问

怎样定义随机变量 X、Y 之间的独立性?

分析

若 X 和 Y 相互 "独立", 从直观上看, X 和 Y 取任何值之间应是没有任何关系的, 即 $\forall x, y \in \mathbb{R}^1$, 两个事件

$$\{X \le x\}, \quad \{Y \le y\}$$

应该相互独立, 即

$$\mathsf{P}(X \le x,\ Y \le y) = \mathsf{P}(X \le x) \cdot \mathsf{P}(Y \le y) \Leftrightarrow F(x, y) = F_X(x) \cdot F_Y(y).$$

随机变量的独立性

定义

设 $(X, Y) \sim F(x, y)$, $X \sim F_X(x)$, $Y \sim F_Y(y)$, 若对任意的 $x, y \in (-\infty, \infty)$, 有

$$\mathsf{P}(X \leq x,\ Y \leq y) = \mathsf{P}(X \leq x) \cdot \mathsf{P}(Y \leq y),$$

即 $F(x, y) = F_X(x) \cdot F_Y(y)$. 称随机变量 X, Y **相互独立**.

注

两个随机变量相互独立, 联合分布函数等于两个边缘分布函数的乘积.

性质

若 X, Y 相互独立. 对任意的 $x_1 < x_2, y_1 < y_2$, 有

$$\mathsf{P}(x_1 < X \le x_2,\ y_1 < Y \le y_2) = \mathsf{P}(x_1 < X \le x_2) \cdot \mathsf{P}(y_1 < Y \le y_2),$$

即 $\{x_1 < X \le x_2\}$ 与 $\{y_1 < Y \le y_2\}$ 相互独立.

证明

$$\begin{aligned}
&\mathsf{P}(x_1 < X \le x_2,\ y_1 < Y \le y_2) \\
&= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \\
&= F_X(x_2)F_Y(y_2) - F_X(x_1)F_Y(y_2) - F_X(x_2)F_Y(y_1) + F_X(x_1)F_Y(y_1) \\
&= \Big[F_X(x_2) - F_X(x_1)\Big] \cdot \Big[F_Y(y_2) - F_Y(y_1)\Big] \\
&= \mathsf{P}(x_1 < X \le x_2) \cdot \mathsf{P}(y_1 < Y \le y_2).
\end{aligned}$$

X 与 Y 独立的直观意义

X 的取值与 Y 的取值是相互独立、互不相干的.

独立性的定义

定义

设 (X,Y) 的频率函数为

$$\mathsf{P}(X=x_i,\ Y=y_i)=p_{ij},\ i,j=1,2,\dots$$

则 X,Y 相互独立等价于对任意 $i,j=1,2,\dots$ 有

$$\mathsf{P}(X=x_i,\ Y=y_j)=\mathsf{P}(X=x_i)\cdot\mathsf{P}(Y=y_j)$$

即

$$X \text{ 与 } Y \text{ 独立} \Leftrightarrow p_{ij}=p_{i\cdot}\cdot p_{\cdot j}.$$

例题

甲袋中有 3 个红球, 2 个白球; 乙袋中有 4 个红球, 5 个白球. 从甲、乙两袋中各任取两球, 记 X, Y 分别表示取到白球的个数, 问 X, Y 是否独立?

分析

由于从两袋中取球是相互独立的过程, 所以 X, Y 的取值是相互独立、互不相干的, 故 X, Y 相互独立.

判断随机变量的独立性的方法

- 按定义判断
- 从直观背景判断

例题

设随机变量 X 从 1, 2, 3, 4 四个数中等可能取值. 又设随机变量 Y 从 1 到 X 中等可能取值. 问 X, Y 是否独立?

解

(X, Y) 的频率函数及边际频率函数为

$Y \backslash X$	1	2	3	4	$p_{\cdot j} = \sum_{i=1}^{4} p_{ij}$
1	1/4	1/8	1/12	1/16	25/48
2	0	1/8	1/12	1/16	13/48
3	0	0	1/12	1/16	7/48
4	0	0	0	1/16	3/48
$p_{i\cdot} = \sum_{j=1}^{4} p_{ij}$	1/4	1/4	1/4	1/4	

可以直接验证, 对任意 $i, j = 1, 2, 3, 4$, $\mathsf{P}(X = i,\ Y = j) \neq \mathsf{P}(X = i) \cdot \mathsf{P}(Y = j)$. 因此 X, Y 独立. **从直观上看, X 和 Y 也不独立.**

例子

设 (X, Y) 的频率函数为

$X \backslash Y$	1	2	3
1	1/8	a	1/24
2	b	1/4	1/8

问:

1. a, b 应满足什么条件?
2. 若 X, Y 独立, 求 a, b.

分析

第一问是考察归一化条件, 第二问是考察独立性条件.

解

1. 由归一化条件, $\sum_{i,j} p_{ij} = 1 \Rightarrow a + b = 1 - (1/8 + 1/24 + 1/4 + 1/8) = \frac{11}{24}$, $a \geq 0$, $b \geq 0$.

解

2. 由 X,Y 的独立性

$$a = \mathsf{P}(X=2,\ Y=1) = \mathsf{P}(X=2)\cdot\mathsf{P}(Y=1) = (a+1/4)(1/8+a+1/24)$$
$$b = \mathsf{P}(X=1,\ Y=2) = \mathsf{P}(X=1)\cdot\mathsf{P}(Y=2) = (b+1/8)(b+1/4+1/8).$$

求解这两个二次方程, 得到 $a=1/12$ 或 $a=1/2$, $b=1/8$ 或 $b=3/8$. 再利用归一化条件:

$$a+b=\frac{11}{24} \Rightarrow a=\frac{1}{12},\ b=\frac{3}{8}.$$

独立性的条件

定义

设 (X,Y) 为连续型随机变量，且

$$(X,Y) \sim f(x,y), \quad X \sim f_X(x), \quad Y \sim f_Y(y).$$

则 (X,Y) 相互独立当且仅当在 $f(x,y), f_X(x), f_Y(y)$ 的连续点处有

$$f(x,y) = f_X(x) \cdot f_Y(y).$$

分析

(X,Y) 相互独立表明对任意的 x,y,

$$\mathsf{P}(X \le x,\ Y \le y) = \mathsf{P}(X \le x)\mathsf{P}(Y \le y)$$
$$\Leftrightarrow \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\,dudv = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)\,dudv.$$

独立性的推论

定义

X 和 Y 独立 $\Leftrightarrow$ 对任意 x 和 y 有

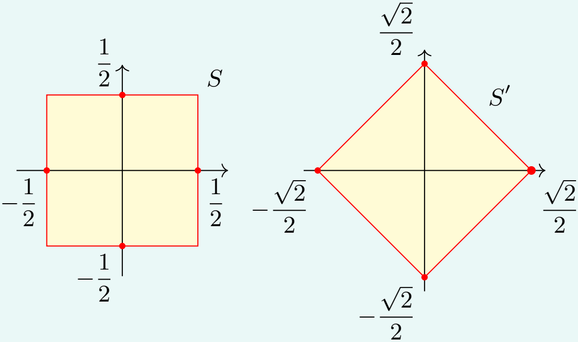
$$f(x,y) = f_X(x) \cdot f_Y(y).$$

推论

二维随机变量 (X,Y) 相互独立，则边缘分布完全确定联合分布.

例子

设 (X,Y) 服从正方形 $S=\{(x,y)\mid -1/2\le x,y\le 1/2\}$ 上的均匀分布. 又设 S' 为 S 旋转 45°, (X',Y') 为 S' 上的均匀分布.



独立性验证

- 易见 $f_X(x)=\mathbb{1}_{(-1/2,1/2)}(x)$, $f_Y(y)=\mathbb{1}_{(-1/2,1/2)}(y)$, 因此 $f_{XY}(x,y)=f_X(x)f_Y(y)\Rightarrow X,Y$ 独立.
- $f_{X'}(0.5)>0$, $f_{Y'}(0.5)>0$, 但 $f_{X'Y'}(0.5,0.5)=0$, 因此 X',Y' 不独立.

例题

设 (X,Y) 的密度函数为

$$f(x,y)=\begin{cases}2e^{-(2x+y)}, & x>0,\ y>0\\ 0, & \text{其它}.\end{cases}$$

问 X,Y 是否独立?

分析

先利用边际密度的公式计算出 $f_X(x)$, $f_Y(y)$, 再验证独立性条件.

解

X,Y 的边际密度函数分别为

$$f_X(x)=\mathbb{1}_{(0,\infty)}(x)\int_{-\infty}^{\infty}2e^{-(2x+y)}\,dy=\mathbb{1}_{(0,\infty)}(x)2e^{-2x},$$

$$f_Y(y)=\mathbb{1}_{(0,\infty)}(y)=\int_{-\infty}^{\infty}2e^{-(2x+y)}\,dx=\mathbb{1}_{(0,\infty)}(y)e^{-y}.$$

因此 $f_{XY}(x,y)=f_X(x)f_Y(y)$, 故 X,Y 相互独立.

问

若 (X,Y) 的密度函数能分解为 $f(x,y)=g(x)h(y)$, 其中 $g(x)\geq 0$, $h(y)\geq 0$. 问 X,Y 是否独立?

答

要注意 $f(x,y)$ 、$f(x)$ 和 $g(y)$ 的定义域!

例题

若 (X,Y) 的概率密度为

$$f(x,y)=\begin{cases}2, & 0<x<y,\ 0<y<1,\\ 0, & \text{其它}.\end{cases}$$

则 X,Y 是否相互独立?

解

$$f_X(x)=\int_x^1 2\,dy=2(1-x),\quad 0<x<1$$
$$f_Y(y)=\int_0^y 2\,dx=2y,\quad 0<y<1.$$

由于在面积不为 0 的区域 $D=\{(x,y)\mid y<x<1,\ 0<y<1\}$ 上, $f(x,y)=0\neq f_X(x)f_Y(y)$, 故 X 和 Y 不独立

注

密度函数形式可分离, 但支撑区域不可分离.

Farlie–Morgenstern 族的独立性

Farlie–Morgenstern 族

设 $F(x)$ 和 $G(y)$ 都是一维连续型分布函数. 则对任意的 α, $|\alpha| \le 1$,

$$H(x,y) = F(x)G(y)\Big[1 + \alpha\Big(1 - F(x)\Big)\Big(1 - G(y)\Big)\Big]$$

是二维连续型分布函数, 且边际分布为

$$H(x,\infty) = F(x), \quad H(\infty,y) = G(y).$$

注

只有当 $\alpha = 0$ 时, X 和 Y 才是相互独立的. 此时, $H(x,y) = F(x)G(y)$, 即 H 分解成了边际分布 $F(x)$ 和 $G(y)$ 的乘积.

例题

在某一分钟内，信号进入收信机是等可能的．若收到两个互相独立的信号的时间间隔小于 0.5 秒时，信号将相互产生干扰．求两信号相互干扰的概率．

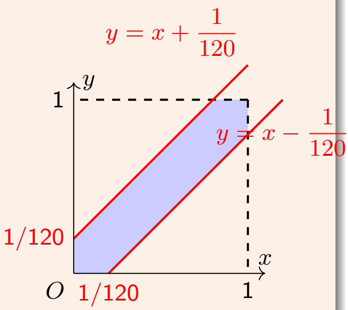
解

设两信号进入收信机的时间分别为 X, Y 分钟．则 $X, Y \sim \mathrm{U}(0,1)$．因为 X, Y 独立，故联合密度为

$$f(x,y) = f_X(x) f_Y(y) = \begin{cases} 1, & 0 < x < 1,\ 0 < y < 1, \\ 0, & \text{其它}, \end{cases}$$

即 $(0,1)^2$ 上的均匀分布．故两信号相互干扰的概率为

$$\begin{aligned} &\mathrm{P}(|X - Y| < \frac{1}{120}) \\ &= \int_{(x,y)\in(0,1)^2,\ |x-y|<\frac{1}{120}} 1\, dxdy \\ &= 1 - \left(1 - \frac{1}{120}\right)^2 \approx 0.016. \end{aligned}$$



二维正态分布的独立性

二维正态分布回顾

$(X,Y) \sim \mathcal{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的密度函数为

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]},$$

其中 $\mu_1,\mu_2 \in \mathbb{R}$, $\sigma_1,\sigma_2 > 0$, $|\rho| < 1$.

重要结论

若 $(X,Y) \sim \mathcal{N}(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$, 则

$$X \sim \mathcal{N}(\mu_1,\sigma_1^2), \quad Y \sim \mathcal{N}(\mu_2,\sigma_2^2).$$

定理

X,Y 相互独立 $\Leftrightarrow \rho = 0$.

边际分布

设 n 维随机变量 $(X_1, X_2, \ldots, X_n)$ 的分布函数为

$$F(x_1, x_2, \ldots, x_n) = \mathsf{P}(X_1 \le x,\ X_2 \le x, \ldots, X_n \le x_n).$$

一维边际分布

X_i 的边际分布函数为

$$\begin{aligned} F_{X_i}(x_i) &= \mathsf{P}(X_1 < \infty, \ldots, X_{i-1} < \infty, X_i \le x_i, X_{i+1} < \infty, \ldots, X_n < \infty) \\ &= F(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty). \end{aligned}$$

二维边际分布

X_1, X_2 的联合分布是

$$F_{X_1 X_2}(x_1, x_2) = \mathsf{P}(X_1 \le x_1,\ X_2 \le x_2,\ X_3 < \infty, \ldots, X_n < \infty)$$
$$= F(x_1, x_2, \infty, \ldots, \infty).$$

注

类似地可定义三维、四维等高维边际分布.

随机向量的独立性

分量的独立性

对任意 $x_1, x_2, \ldots, x_n \in \mathbb{R}^1$, 若

$$F(x_1, x_2, \ldots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n),$$

则称 $X_1, X_2, \ldots, X_n$ 相互独立.

向量的独立性

设 $(X_1, \ldots, X_m) \sim F_1(x_1, \ldots, x_m)$, $(Y_1, \ldots, Y_n) \sim F_2(y_1, \ldots, y_n)$. 若

$(X_1, \ldots, X_m; Y_1, \ldots, Y_n) \sim F(x_1, \ldots, x_m; y_1, \ldots, y_n) = F_1(x_1, \ldots, x_m) \cdot F_2(y_1, \ldots, y_n)$,

则称 $(X_1, \ldots, X_m)$ 与 $(Y_1, \ldots, Y_n)$ 相互独立.

注

从直观上看: 随机向量的独立性是指各随机向量的取值是相互独立、互不相干的.

独立随机变量的函数的独立性

定理

设 $(X_1,\ldots,X_m)$, $(Y_1,\ldots,Y_n)$ 相互独立, 则

1. $\forall 1\le i\le m, 1\le j\le n$, X_i 和 Y_j 相互独立.
2. 对任意 m 元 (连续) 函数 h 与 n 元 (连续) 函数 g,

$$h(X_1,\ldots,X_m),\quad g(Y_1,\ldots,Y_n)$$

   相互独立.

注

$h:(x_1,\ldots,x_m)\mapsto x_i$ 也是一个连续函数.

应用

$X_1,\ldots,X_m$ 和 $Y_1,\ldots,Y_n$ 是两组独立的数据, h,g 是对两组数据的处理. 则处理后的结果依然独立.

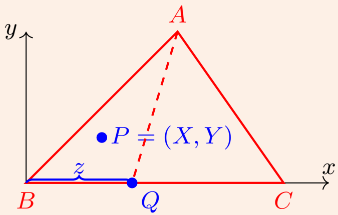
例题

设在 $\triangle ABC$ 内部任取一点 P, 在底边 BC 边上任取一点 Q. 求直线 PQ 与线段 AB 相交的概率.

解

如图建立坐标系. 依题意, 点 P 服从 $\triangle ABC$ 上的均匀分布, 点 Q 服从区间 $(0, BC)$ 上的均匀分布, 它们的概率密度分别为

$$f_{XY}(x,y) = \frac{1}{S_{\triangle ABC}} \mathbb{1}_{\triangle ABC}(x,y),$$
$$f_Z(z) = \frac{1}{BC} \mathbb{1}_{(0,BC)}(z).$$



由独立性, $f(x,y,z) = f_{XY}(x,y) f_Z(z)$.
线段 AB 与直线 PQ 相交当且仅当 $P \in \triangle ABQ$. 因此所求概率为

$$\begin{aligned}
&\mathsf{P}((X,Y) \in \triangle ABQ,\ 0 < z < BC) \\
&= \int_0^{BC} f_Z(z)\, dz \int_{(x,y)\in\triangle ABQ} f_{XY}(x,y)\, dxdy \\
&= \frac{1}{BC} \int_0^{BC} dz \cdot \frac{z}{BC} = \frac{1}{BC^2} \cdot \frac{BC^2}{2} = \frac{1}{2}.
\end{aligned}$$

- P77: 19
- 补充题:
  1. 一个袋中有 5 个球, 其中两个白球 3 个黑球,
     (1) 先后有放回地任取两球 (一共取两球);
     (2) 先后无放回地任取两球 (一共取两球);

     假设 X 和 Y 分别表示第一次和第二次取到白球的数量, 求 (X, Y) 的联合频率函数及边缘频率函数, 并讨论独立性.
  2. 在一个以原点为圆心、半径为 R 的圆内随机选取一点, 令 (X, Y) 表示这一点的分布, 则 (X, Y) 服从

$$f(x,y) = \begin{cases} c, & x^2+y^2 \le R^2, \\ 0, & \text{其它}. \end{cases}$$

     (1) 求 c
     (2) 求边缘密度函数
     (3) 讨论 X 和 Y 的独立性.

实例

考虑南科大的全体学生，从其中随机抽取一个学生，分别以 X 和 Y 表示其身高和体重. 则 X 和 Y 都是随机变量，它们都有一定的概率分布.

额外的限制条件

现在若限制 $1.7 < X < 1.8$ (米)，在这个条件下去求 Y 的**条件分布**. 这就意味着要从该校的学生中把身高在 1.7 米和 1.8 米之间的那些人都挑出来，然后在挑出的学生中求其体重的分布.

讨论

条件分布与不加条件的分布会**不一样**. 如：在条件分布中体重取大值的概率会显著增加.

条件分布

条件概率回顾

$$\mathsf{P}(A \mid B) = \frac{\mathsf{P}(AB)}{\mathsf{P}(B)}, \quad \mathsf{P}(B) > 0.$$

条件分布的“定义”

设 (X, Y) 为二维随机变量. 对任意 $y \in \mathbb{R}^1$, 考虑条件概率

$$\mathsf{P}(X \le x \mid Y = y), \quad x \in \mathbb{R}^1$$

这可视为在 $\{Y = y\}$ 条件下, 随机变量 X 的概率分布 — **条件分布**.

问

能否由条件概率定义计算

$$\mathsf{P}(X \le x \mid Y = y) = \frac{\mathsf{P}(X \le x,\ Y = y)}{\mathsf{P}(Y = y)}?$$

设二维离散型随机变量 (X, Y) 的频率函数为 $\mathsf{P}(X = x_i,\ Y = y_j) = p_{ij},\ i, j = 1, 2, \ldots$

条件概率公式的应用

由条件概率公式, $\{Y = y_j\}$ 发生的条件下, $\{X = x_i\}$ 发生的概率为

$$\mathsf{P}(X = x_i \mid Y = y_j) = \frac{\mathsf{P}(X = x_i, Y = y_j)}{\mathsf{P}(Y = y_j)} = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \ldots.$$

同理, 在 $\{X = x_i\}$ 发生的条件下, $\{Y = y_j\}$ 发生的条件概率为

$$\mathsf{P}(Y = y_j \mid X = x_i) = \frac{\mathsf{P}(Y = y_j, X = x_i)}{\mathsf{P}(X = x_i)} = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1, 2, \ldots$$

定义

对于固定的 j, 若 $\mathsf{P}(Y = y_j) = p_{\cdot j} > 0$, 则称

$$p_{X|Y}(x_i \mid y_j) = \mathsf{P}(X = x_i \mid Y = y_j) = \frac{p_{ij}}{p_{\cdot j}}, \quad i = 1, 2, \ldots$$

为 $Y = y_j$ 的条件下, 随机变量 X 的**条件 (conditional) 频率函数**.
对于固定的 i, 若 $\mathsf{P}(X = x_i) = p_{i\cdot} > 0$, 则称

$$p_{Y|X}(y_j \mid x_i) = \mathsf{P}(Y = y_j \mid X = x_i) = \frac{p_{ij}}{p_{i\cdot}}, \quad j = 1, 2, \ldots$$

为在 $X = x_i$ 的条件下, 随机变量 Y 的**条件 (conditional) 频率函数.**

例题

设随机变量 X 从 1, 2, 3, 4 四个数中等可能取值, 又设随机变量 Y 从 $1 \sim X$ 中等可能取值. 问当 Y 取到数字 3 时, X 取四个数字的可能性各是多少?

解

由

$Y \backslash X$	1	2	3	4	$p_{\cdot j} = \sum_{i=1}^{4} p_{ij}$
1	1/4	1/8	1/12	1/16	25/48
2	0	1/8	1/12	1/16	13/48
3	0	0	1/12	1/16	7/48
4	0	0	0	1/16	3/48
$p_{i\cdot} = \sum_{j=1}^{4} p_{ij}$	1/4	1/4	1/4	1/4	

在 $Y = 3$ 的条件下, X 取到四个数字的概率为

$X = k$	1	2	3	4
$p_{X\|Y}(k, 3)$	0	0	4/7	3/7

条件频率函数的性质

性质

1. $\mathsf{P}(X = x_i \mid Y = y_j) \geq 0,\ i = 1, 2, \ldots$
2. $\sum_{i=1}^{\infty} \mathsf{P}(X = x_i \mid Y = y_j) = \sum_{i=1}^{\infty} \frac{p_{ij}}{p_{\cdot j}} = \frac{1}{p_{\cdot j}} \sum_{i=1}^{\infty} p_{ij} = 1.$

注

这两条性质说明，条件频率函数也是一种频率函数.

例题

设

$$F_X(x) = \begin{cases} 0, & x < 1, \\ 0.3, & 1 \le x < 2, \\ 1, & x \ge 2, \end{cases} \quad F_Y(y) = \begin{cases} 0, & y < 0, \\ 0.4, & 0 \le y < 1, \\ 1, & y \ge 1. \end{cases}$$

已知 $\mathrm{P}(X = 1, Y = 0) = 0.1$. 求

1. 联合分布律
2. 当 $Y = 0$ 时, X 的条件分布律 $\mathrm{P}(X = k \mid Y = 0)$;
   $p_{X|Y}(1 \mid 0) = 0.25, p_{X|Y}(2 \mid 0) = 0.75.$
3. $Y = 0$ 时 X 的条件分布函数.

回顾

$\mathrm{P}(X = x_0) = F(x_0) - F(x_0 - 0).$

解

$X \backslash Y$	0	1	$p_{i\cdot}$
1	0.1	0.2	0.3
2	0.3	0.4	0.7
$p_{\cdot j}$	0.4	0.6	

二维随机连续型随机变量的条件概率密度

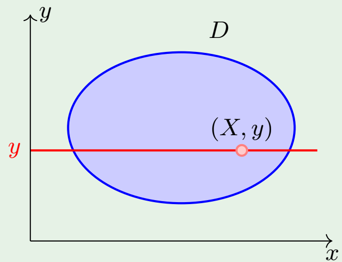
问题

设 (X,Y) 的概率密度函数为 $f(x,y)$. 考虑在 $\{Y=y\}$ 已发生的条件下, $\{X\le x\}$ 发生的条件概率

$$\mathsf{P}(X\le x \mid Y=y),\quad x\in\mathbb{R}^1.$$

背景解释

- (X,Y) 在区域 D 上具有密度 $f(x,y)$
- (X,Y) 限制在直线时可视为一维随机变量, $\mathsf{P}(X\le x\mid Y=y)$ 为此一维随机变量的分布函数.



如何定义 $\mathsf{P}(X \le x \mid Y = y)$?

问题

对于连续型随机变量 Y, $\mathsf{P}(Y = y) = 0$.

解决方案

用 $\{Y \in (y-\varepsilon, y+\varepsilon)\}$ 代替 $\{Y = y\}$, 并令 $\varepsilon \downarrow 0$.

假设密度函数是连续的, 由积分中值定理,

$$
\begin{aligned}
\mathsf{P}(X \le x \mid Y \in (y-\varepsilon, y+\varepsilon)) &= \frac{\mathsf{P}(X \le x, y-\varepsilon < Y < y+\varepsilon)}{\mathsf{P}(y-\varepsilon < Y < y+\varepsilon)} \\
&= \frac{\int_{-\infty}^{x} \int_{y-\varepsilon}^{y+\varepsilon} f(u,v)\, dv du}{\int_{y-\varepsilon}^{y+\varepsilon} f_Y(v)\, dv} \\
&= \frac{2\varepsilon \int_{-\infty}^{x} f(u, y_{\varepsilon,u})\, du}{2\varepsilon f_Y(\tilde{y}_\varepsilon)}, \quad y_{\varepsilon,u}, \tilde{y}_\varepsilon \in (y-\varepsilon, y+\varepsilon) \\
&\to \int_{-\infty}^{x} \frac{f(u,y)}{f_Y(y)}\, du, \quad \varepsilon \downarrow 0.
\end{aligned}
$$

连续型条件分布与概率

定义

设 (X,Y) 的概率密度为 $f(x,y)$，若对于固定的 y，(X,Y) 关于 Y 的边际密度 $f_Y(y)>0$，则称

$$\frac{f(x,y)}{f_Y(y)} := f_{X|Y}(x \mid y), \quad -\infty < x < \infty,$$

为在 $Y=y$ 的条件下，X 的**条件密度 (conditional density)**. 称

$$F_{X|Y}(x \mid y) := \int_{-\infty}^{x} f_{X|Y}(u \mid y)\, du, \quad -\infty < x < \infty,$$

为在 $Y=y$ 条件下，X 的**条件分布 (函数)**.

类似地，定义

$$f_{Y|X}(y \mid x) := \frac{f(x,y)}{f_X(x)}, \quad -\infty < y < \infty$$

$$F_{Y|X}(y \mid x) := \int_{-\infty}^{y} f_{Y|X}(v \mid x)\, dv, \quad -\infty < y < \infty.$$

连续情形的全概率公式

定理

Y 的边际密度可表示为

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

证明

由 $f_{Y|X}(y \mid x) := \dfrac{f(x,y)}{f_X(x)}$, $-\infty < y < \infty$, 我们有

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

两边对 x 在 $(-\infty, \infty)$ 上积分即可.

注

这表明联合密度可以用边际密度和条件密度表示.

条件密度的性质

1. $f_{X|Y}(x \mid y) \geq 0,$
2. $\displaystyle\int_{-\infty}^{\infty} f_{X|Y}(u \mid y)\,du = \int_{-\infty}^{\infty} \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(u,y)\,du = 1.$

注

这两条性质说明，条件密度也是一种密度.

与独立性的关系

事件独立性与条件概率的关系

A, B 相互独立 $\Leftrightarrow \mathsf{P}(A \mid B) = \mathsf{P}(A)$, $\mathsf{P}(B \mid A) = \mathsf{P}(B)$.

随机变量独立性与条件密度的关系

$$X, Y \text{ 相互独立} \Leftrightarrow f(x, y) = f_X(x) f_Y(y), \text{a.e.}$$

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)} = f_X(x), \text{a.e.}$$

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} = f_Y(y), \text{a.e.}$$

平面上的均匀分布

定义

设 G 是平面上的有界区域, 其面积为 A. 若 (X, Y) 的概率密度为

$$f(x,y) = \begin{cases} \dfrac{1}{A}, & (x,y) \in G, \\ 0, & \text{其它}. \end{cases}$$

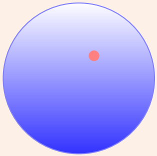
则称 (X, Y) 服从区域 G 上的均匀分布.

实际背景

若随机点 (X, Y) 在平面区域 G 上“等可能”取值, 则 (X, Y) 服从 G 上均匀分布.

例

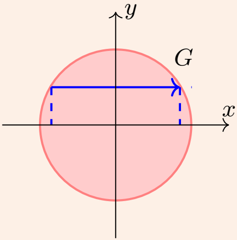
设雷达的圆形屏幕半径为 1, 当用雷达捕捉目标时, 可认为目标出点在点 (X, Y) 在屏幕上服从圆域 $G: x^2 + y^2 \le 1$ 上的均匀分布.



例题

设 (X,Y) 服从圆域 $G: x^2+y^2 \le 1$ 上的均匀分布，求条件概率密度 $f_{X|Y}(x \mid y)$.

解



(X,Y) 的密度及 Y 的边际密度分别为

$$f(x,y) = \mathbb{1}_G(x,y)\cdot\frac{1}{\pi},$$
$$f_Y(y) = \mathbb{1}_{(-1,1)}(y)\cdot\frac{2}{\pi}\sqrt{1-y^2}.$$

故当 $-1<y<1$ 时有

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \dfrac{1}{\pi f_Y(y)} = \dfrac{1}{2\sqrt{1-y^2}}, & |x| \le \sqrt{1-y^2}, \\ 0, & \text{其它}. \end{cases}$$

二维正态分布

设 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]},$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

二维正态条件分布

$$f_{Y|X}(y \mid x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} e^{-\frac{1}{2}\frac{\left[y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right]^2}{\sigma_2^2(1-\rho^2)}}$$

$$\sim \mathcal{N}(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1), \sigma_2^2(1-\rho^2)).$$

给定 X 时，Y 的条件密度是一维 (单变量) 正态分布.

边际分布

随机变量边际分布由联合分布决定

$$F_X(x) = F(x, \infty), \quad F_Y(y) = F(\infty, y).$$

边际频率/密度函数

离散型随机变量边际频率函数	连续型随机变量边际密度
$\mathsf{P}(X = x_i) = \sum_{j=1}^{\infty} p_{ij} := p_{i\cdot},\ i = 1, 2, \ldots$	$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy,\ x \in \mathbb{R}$
$\mathsf{P}(Y = y_j) = \sum_{i=1}^{\infty} p_{ij} := p_{\cdot j},\ j = 1, 2, \ldots$	$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx,\ y \in \mathbb{R}$

条件分布

定义

$$\mathsf{P}(X \le x \mid Y = y), \quad x \in \mathbb{R}$$

可视为 $\{Y = y\}$ 发生的条件下随机变量 X 的概率分布 (思考: $\mathsf{P}(Y = y) = 0$ 时怎么理解?)

条件频率/密度函数

离散型随机变量的条件频率函数	连续型随机变量的条件概率密度
$\mathsf{P}(X = x_i \mid Y = y_j) = \dfrac{p_{ij}}{p_{\cdot j}},\ p_{\cdot j} > 0$	$f_{X\mid Y}(x, y) := \dfrac{f(x, y)}{f_Y(y)},\ f_Y(y) > 0,\ x \in \mathbb{R}$
$\mathsf{P}(Y = y_j \mid X = x_i) = \dfrac{p_{ij}}{p_{i\cdot}},\ p_{i\cdot} > 0$	$f_{Y\mid X}(y \mid x) := \dfrac{f(x, y)}{f_X(x)},\ f_X(x) > 0,\ y \in \mathbb{R}$

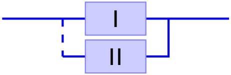
作业

- P77: 1, 9, 10, 15
- 补充题
  1. 高随机变量 X 在区间 $(0,1)$ 内服从均匀分布, 在 $X=x$ $(0<x<1)$ 的条件下, 随机变量 Y 在区间 $(0,x)$ 内服从均匀分布, 求
     (1) X 和 Y 的联合密度函数;
     (2) Y 的密度函数;
     (3) $\mathrm{P}(X+Y>1)$.
  2. 设二维随机变量 (X,Y) 的概率密度为
     $$f(x,y)=\begin{cases} e^{-y}, & 0<x<y, \\ 0, & \text{其它}. \end{cases}$$
     (1) 求边际密度函数并讨论独立性;
     (2) 求条件概率密度 $f_{X|Y}(x \mid y)$ 和 $f_{Y|X}(y \mid x)$.

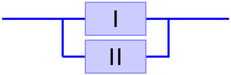
实际背景

设有两个部件 I 和 II, 其工作寿命分别为 X, Y.

冷冗余系统　部件 I 坏了, 换上备用部件 II 继续工作.
系统寿命: $X + Y$.



热冗余系统　部件 I 、II 并联同时工作, 仅当两个部件都损坏时, 整个系统才失效.
系统寿命: $\max(X, Y)$.



串联系统　部件 I、II 串联同时工作, 只要有一个部件损坏, 整个系统都失效.
系统寿命: $\min(X, Y)$.

问题

怎么确定上述各系统的寿命?

问题

若 $(X,Y) \sim f(x,y)$，怎样求

$$X+Y, \quad \max(X,Y), \quad \min(X,Y)$$

的分布?

一般化

设 $z=g(x,y)$ 是一个二元函数，怎样求随机变量 $Z=g(X,Y)$ 的分布?

思路

$$\begin{aligned} F_Z(z) = \mathsf{P}(Z \le z) = \mathsf{P}\Big(g(X,Y) \le z\Big) \\ = \int_{(x,y):g(x,y)\le z} f(x,y)\,dxdy \\ = \cdots = \int_{-\infty}^{z} f_Z(u)\,du. \end{aligned}$$

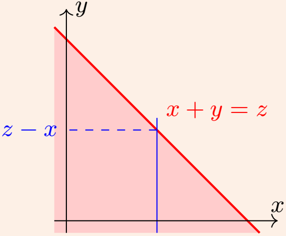
则 $Z \sim f_Z(z)$.

定理

若 $(X,Y)\sim f(x,y)$, 则 $f_Z(z)\sim\int_{-\infty}^{\infty} f(x,z-x)\,dx$.

证明

$$
\begin{aligned}
F_Z(z) &= \mathsf{P}(Z\le z)=\mathsf{P}(X+Y\le z)\\
&\iint_{x+y\le z} f(x,y)\,dxdy\\
&=\int_{-\infty}^{\infty}dx\int_{-\infty}^{z-x} f(x,y)\,dy\\
&\xlongequal{u=x+y}\int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x,u-x)\,dudx\\
&=\int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f(x,u-x)\,dx\right]du
\end{aligned}
$$



独立随机变量的和

$Z=X+Y$ 密度函数

$(X,Y)\sim f(x,y)$, 则 $Z=X+Y$ 的密度函数为

$$f_Z(z)=\int_{-\infty}^{\infty} f(z-y,y)\,dy=\int_{-\infty}^{\infty} f(x,z-x)\,dx.$$

独立随机变量

若 X,Y 相互独立, 则 $Z=X+Y$ 的密度函数为

$$f_Z(z)=\int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\,dy=\int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx.$$

卷积公式 (convolution)

$$(f*g)(x)=\int_{-\infty}^{\infty} f(x-t)g(t)\,dt=\int_{-\infty}^{\infty} f(t)g(x-t)\,dt.$$

例题

设随机变量 X,Y 相互独立, 且 $X\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{N}(0,1)$, 求 $Z=X+Y$ 的密度函数.

解

由独立性及卷积公式, 有

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{(z-x)^2}{2}}\,dx \\
&= \frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{\infty} e^{-(x-z/2)^2}\,dx \\
&= \frac{1}{2\pi}e^{-\frac{z^2}{4}}\int_{-\infty}^{\infty} e^{-t^2}\,dt \\
&= \frac{1}{2\pi}e^{-\frac{z^2}{4}}\sqrt{\pi} = \frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{z^2}{2(\sqrt{2})^2}}.
\end{aligned}$$

易见 $X+Y\sim\mathcal{N}(0,2)$.

独立正态分布的和

两个正态分布的和

设 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ 相互独立, 则

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

一般结果

一般地, 若 $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ 相互独立, $i = 1, 2, \ldots, n$, 则

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n \sim \mathcal{N}\Big(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\Big)$$

即: **独立正态分布的和依然服从正态分布**.

例题

某电气设备中的两个部件存在接触电阻 R_1、R_2. 两个部件的工作状态是相互独立的，概率密度均为

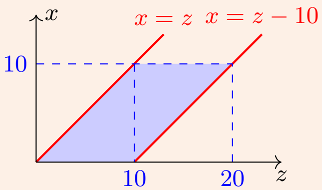
$$f(x)=\begin{cases}\dfrac{10-x}{50}, & 0\le x\le 10,\\ 0, & \text{其它}.\end{cases}$$

求 R_1、R_2 串联后的总电阻 $R=R_1+R_2$ 的概率密度.

解

由卷积公式，$f_R(z)=\displaystyle\int_{-\infty}^{\infty} f(x)f(z-x)\,dx$. 被积函数是非零区域是

$0<x<10,\ 0<z-x<10 \Leftrightarrow 0<x<10, z-10, x<z.$



因此

$$f_R(z)=\begin{cases}\displaystyle\int_0^z \frac{10-x}{50}\cdot\frac{10-(z-x)}{50}\,dx=\frac{600z-60z^2+z^3}{15000}, & 0\le z<10,\\ \displaystyle\int_{z-10}^{10} \frac{10-x}{50}\cdot\frac{10-(z-x)}{50}\,dx=\frac{(20-z)^3}{15000}, & 10\le z<20,\\ 0, & \text{其它}.\end{cases}$$

例题

设 X, Y 相互独立且都服从参数为 λ 的指数分布, 求随机变量 $Z = X + Y$ 的概率密度.

解

由卷积公式, Z 的密度函数为

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \\
&= \begin{cases} \displaystyle\int_0^z \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)}, & z > 0, \\ 0, & z \le 0 \end{cases} \\
&= \begin{cases} \lambda^2 z e^{-\lambda z}, & z > 0, \\ 0, & z \le 0 \end{cases} \sim \Gamma(2, \lambda).
\end{aligned}
$$

注

由泊松流的知识, n 个独立 $\mathrm{Exp}(\lambda)$ 的和为 $\Gamma(n, \lambda)$.

离散卷积公式

设 X, Y 相互独立, 其频率函数分别为

$$\mathsf{P}(X=i)=p_i,\ i=0,\pm1,\pm2,\dots,\quad \mathsf{P}(Y=j)=q_j,\ j=0,\pm1,\pm2,\dots$$

令 $Z=X+Y$. 则我们有

离散卷积公式

$$\begin{aligned}\mathsf{P}(Z=k)&=\sum_{i\in\mathbb{Z}}\mathsf{P}(X=i)\mathsf{P}(Y=k-i)=\sum_{i\in\mathbb{Z}}p_iq_{k-i}\\&=\sum_{i\in\mathbb{Z}}\mathsf{P}(X=k-i)\mathsf{P}(Y=i)=\sum_{i\in\mathbb{Z}}p_{k-i}q_i,\quad k=0,\pm1,\pm2,\dots\end{aligned}$$

比较: 连续型卷积公式

$$f_Z(z)=\int_{\mathbb{R}}f_X(z-y)f_Y(y)\,dy=\int_{\mathbb{R}}f_X(x)f_Y(z-x)\,dx.$$

泊松分布的和

设 X, Y 独立, 且 $X \sim \pi(\lambda_1)$, $Y \sim \pi(\lambda_2)$. 则 $Z = X + Y \sim \pi(\lambda_1 + \lambda_2)$.

由离散卷积公式,

$$\begin{aligned}
\mathsf{P}(Z = k) &= \sum_{i=0}^{k} \mathsf{P}(X = k - i) \cdot \mathsf{P}(Y = i) \\
&= \sum_{i=0}^{k} \frac{\lambda_1^{k-i}}{(k-i)!} e^{-\lambda_1} \frac{\lambda_2^i}{i!} e^{-\lambda_2} \\
&= e^{-\lambda_1 - \lambda_2} \sum_{i=0}^{k} \frac{1}{i!(k-i)!} \lambda_1^{k-i} \lambda_2^i \\
&= e^{-(\lambda_1 + \lambda_2)} \frac{(\lambda_1 + \lambda_2)^k}{k!}.
\end{aligned}$$

注

泊松流可以**合并及细分**.

例题

设二维随机变量 (X,Y) 的概率分布为

X \ Y	-1	1	2
-1	1/4	1/6	1/8
0	1/4	1/8	1/12

求 $X+Y$, XY 的概率分布.

解

根据 (X,Y) 的联合分布可得

P	1/4	1/4	1/6	1/8	1/8	1/12
(X,Y)	(-1,-1)	(-1,0)	(1, -1)	(1,0)	(2, -1)	(2,0)
$X+Y$	-2	-1	0	1	1	2
XY	1	0	-1	0	-2	0

合并相同的项可得 $X+Y$ 和 XY 的分布列.

直接法

例题

设 X、Y 独立同分布，其密度函数为

$$f(x) = \mathbb{1}_{(0,\infty)}(x)e^{-x}.$$

求 $Z = X/Y$ 的密度.

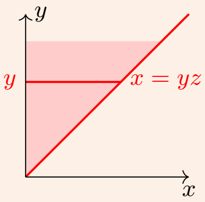
解

当 $z \geq 0$ 时，Z 的分布函数为

$$F_Z(z) = \mathsf{P}(X/Y \leq z) = \iint_{x/y \leq z, x>0, y>0} e^{-(x+y)}\,dxdy$$

$$= \int_0^\infty dy \int_0^{yz} e^{-(x+y)}\,dx = \int_0^\infty e^{-y}(1-e^{-yz})\,dy$$

$$= 1 - \frac{1}{1+z}.$$

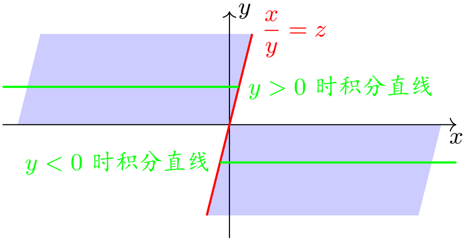


因此 $f_Z(z) = \mathbb{1}_{(0,\infty)}(z)\dfrac{1}{(1+z)^2}$.

一般情形

设 $(X,Y) \sim f(x,y)$，则 $Z = \dfrac{X}{Y}$ 的分布函数为

$$F_Z(z) = \mathsf{P}(X/Y \leq z) = \iint_{x/y \leq z} f(x,y)\,dxdy$$



$$F_Z(z) = \int_0^{\infty} dy \int_{-\infty}^{yz} f(x,y)\,dx + \int_{-\infty}^{0} dy \int_{yz}^{\infty} f(x,y)\,dx.$$

由于积分区域不是矩形，计算积分以及求导比较繁琐.

回顾

二重积分的变量替换

若连续可微分函数

$$\begin{cases} x = x(u,v) \\ y = y(u,v) \end{cases}$$

把平面 Oxy 上区域 Ω 单值映射到平面 $O'uv$ 上的区域 Ω', 其 Jacobi 式为

$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \neq 0,$$

则

$$\iint_{\Omega} f(x,y)\,dxdy = \iint_{\Omega'} f\big(x(u,v), y(u,v)\big)\,|J|\,dudv.$$

为了处理 $z = \dfrac{x}{y}$, 我们的变换函数必须包含 $u = \dfrac{x}{y}$.

公式法

令 $x/y = u$, $y = y$, 则变换 $(x,y) \mapsto (u,y)$ 的 Jacobi 式为

$$\frac{\partial(x,y)}{\partial(u,y)} = \begin{vmatrix} y & u \\ 0 & 1 \end{vmatrix} = y.$$

因此

$Z = X/Y$ 的分布与密度函数

$$F_Z(z) = \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} f(uy,y)|y|\,dy \right] du,$$

$$f_Z(z) = \int_{-\infty}^{\infty} f(zy,y)|y|\,dy.$$

独立的情形

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(zy) f_Y(y)|y|\,dy.$$

例题

设 X, Y 独立且同服从正态分布 $\mathcal{N}(0,1)$, 求随机变量 $Z = Y/X$ 的概率密度.

解

我们利用公式法求解.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{|x|}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{x^2 z^2}{2}}\, dx \\ &= \frac{1}{\pi} \int_0^{\infty} x e^{-x^2 \cdot \frac{z^2+1}{2}}\, dx \\ &\xlongequal{u=x^2} \frac{1}{2\pi} \int_0^{\infty} e^{-u \cdot \frac{z^2+1}{2}} \\ &= \frac{1}{\pi(z^2+1)}, \quad z \in \mathbb{R}. \end{aligned}$$

注

Z 服从 Cauchy 分布.

例子

设 ξ,η 为独立同分布，均服从指数分布 $f(x)=\mathbb{1}_{(0,\infty)}(x)e^{-x}$. 试求 $(U,V)=(\xi+\eta,\xi/\eta)$ 的联合密度，并证明 U,V 独立

思路

我们要利用换元

$$x+y=\tilde{u}, x/y=\tilde{v} \Leftrightarrow x=\frac{\tilde{u}\tilde{u}}{1+\tilde{v}}, y=\frac{\tilde{u}}{1+\tilde{v}},\quad J=\frac{\partial(x,y)}{\partial(\tilde{u},\tilde{v})}=-\frac{\tilde{u}}{(1+\tilde{v})^2}.$$

解

$$\begin{aligned} F_{UV}(u,v) &= \mathsf{P}(\xi+\eta\le u,\ \xi/\eta\le v) \\ &= \iint_{x+y\le u, x/y\le v, x\ge 0, y\ge 0} e^{-(x+y)}\,dxdy = \int_0^u\int_0^v e^{-\tilde{u}}|J|\,d\tilde{u}d\tilde{v}. \end{aligned}$$

因此 $f_{UV}(u,v)=\dfrac{ue^{-u}}{(1+v)^2}\mathbb{1}_{(0,\infty)^2}(u,v)$. 易见 U,V 相互独立. (其实 $V=\dfrac{1-X}{X}$, $X\sim \mathrm{U}(0,1)$.)

二维正态布的线性变换

例子

设 X_1, X_2 相互独立且服从标准正态分布 $\mathcal{N}(0,1)$, 且 $Y_1 = X_1$, $Y_2 = X_1 + X_2$. 则 $(Y_1, Y_2) \sim \mathcal{N}(0, 0, 1, 2, \frac{1}{\sqrt{2}})$.

推广

两个独立标准正态分布随机变量 (更一般地, 二维联合正态分布) 的 (非奇异) 线性变换服从二元正态分布.

解释

正态分布的密度函数为 $Ce^{-Q(x,y)}$ 的形式, 其中 Q 为一个 (椭圆型) 二次型, 经过线性变换后依然得到一个椭圆二次型.

随机变量的其它函数

例子

设 X, Y 独立且均服从 $\mathcal{N}(0,\sigma^2)$, 求 $Z=\sqrt{X^2+Y^2}$ 的概率分布.

解

当 $z \geq 0$ 时, Z 的分布函数为

$$\begin{aligned} F_Z(z) = \mathsf{P}(\sqrt{X^2+Y^2} \leq z) &= \int_{x^2+y^2\leq z^2} f_X(x) f_Y(y)\, dxdy \\ &= \frac{1}{2\pi\sigma^2} \iint_{x^2+y^2\leq z^2} e^{-\frac{x^2+y^2}{2\sigma^2}}\, dxdy \\ &= \frac{1}{2\pi\sigma^2} \int_0^z d\rho \int_0^{2\pi} \rho e^{-\frac{\rho^2}{2\sigma^2}}\, d\varphi \\ &= 1 - e^{-\frac{z^2}{2\sigma^2}}. \end{aligned}$$

因此, $f_Z(z) = \mathbb{1}_{(0,\infty)}(z) \dfrac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}}$ (**Rayleigh 分布**).

- P79: 43, 44, 51, 52, 57
- 补充题
  1. 假设 X 和 Y 是两个独立的随机变量，均服从标准正态分布 $\mathcal{N}(0,1)$. 令 $U = X + Y$, $V = X - Y$. 求 U 和 V 的边缘密度函数及联合密度函数，并讨论独立性.
  2. 设二维连续型随机变量 (X, Y) 的概率密度函数为

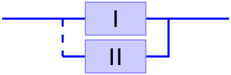
$$f(x,y) = \begin{cases} 1, & 0 < x < 1,\ 0 < y < 2x, \\ 0, & \text{其它}. \end{cases}$$

     (1) 求边缘密度函数;
     (2) 求 $Z = 2X - Y$ 的概率密度函数;
     (3) 求 $\mathrm{P}(Y < \frac{1}{2} \mid X < \frac{1}{2})$.

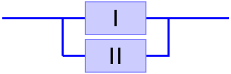
实际背景

设有两个部件 I 和 II, 其工作寿命分别为 X, Y.

冷冗余系统 部件 I 坏了, 换上备用部件 II 继续工作.
系统寿命: $X+Y$.



热冗余系统 部件 I 、II 并联同时工作, 仅当两个部件都损坏时, 整个系统才失效.
系统寿命: $\max(X, Y)$.



串联系统 部件 I、II 串联同时工作, 只要有一个部件损坏, 整个系统都失效.
系统寿命: $\min(X, Y)$.

问题

怎么确定上述各系统的寿命?

极值 $\max(X,Y)$, $\min(X,Y)$ 的分布

极大值

$$\begin{aligned}F_{\max}(z) &= \mathsf{P}(\max(X,Y) \le z) = \mathsf{P}(X \le z, Y \le z)\\ &= \mathsf{P}(X \le z)\cdot \mathsf{P}(Y \le z) = F_X(z)\cdot F_Y(z).\end{aligned}$$

极小值

$$\begin{aligned}F_{\min}(z) &= \mathsf{P}(\min(X,Y) \le z)\\ &= 1-\mathsf{P}(\min(X,Y) > z) = 1-\mathsf{P}(X > z, Y > z)\\ &= 1-\mathsf{P}(X > z)\cdot \mathsf{P}(Y > z)\\ &= 1-\big[1-F_X(z)\big]\cdot\big[1-F_Y(z)\big]\end{aligned}$$

多个随机变量

一般情形

$$
\begin{aligned}
F_{\max}(z) &= \mathsf{P}\Big(\max(X_1, X_2, \ldots, X_n) \leq z\Big) \\
&= F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z) \\
F_{\min}(z) &= \mathsf{P}\Big(\min(X_1, X_2, \ldots, X_n)\Big) \\
&= 1 - \prod_{i=1}^{n}\Big[1 - F_{X_i}(z)\Big].
\end{aligned}
$$

同分布情形

$X_1, X_2, \ldots, X_n \sim F(x)$ 且独立，则

$$
F_{\max}(z) = F^n(z), \quad F_{\min}(z) = 1 - \Big[1 - F(z)\Big]^n.
$$

$\max(X,Y)$ 与 $\min(X,Y)$ 的密度函数

设 $X, Y \sim f(x)$ 且独立，则

极大值

$$F_{\max}(z) = F^2(z) \Rightarrow f_{\max}(z) = 2f(z)F(z) = 2f(z)\int_{-\infty}^{z} f(t)\,dt.$$

极小值

$$F_{\min}(z) = 1 - \big[1 - F(z)\big]^2 \Rightarrow f_{\min}(z) = 2f(z)\big[1 - F(z)\big] = 2f(z)\int_{z}^{\infty} f(t), dt.$$

n 个独立同分布随机变量的极值密度

$$f_{\max}(z) = nf(z)\big[F(z)\big]^{n-1}, \quad f_{\min}(z) = nf(z)\big[1 - F(z)\big]^{n-1}.$$

推广

n 个独立随机变量的极大值与极小值

设 $X_1, X_2, \ldots, X_n$ 是 n 个相互独立的随机变量，其分布函数分别为 $F_{X_i}(x)$, $i = 1, 2, \ldots, n$. 则 $M = \max(X_1, \ldots, X_n)$ 的分布函数为

$$F_M(z) = F_{X_1}(z) \cdots F_{X_n}(z),$$

$N = \min(X_1, \ldots, X_n)$ 的分布函数为

$$F_N(z) = 1 - \big[1 - F_{X_1}(z)\big] \cdots \big[1 - F_{X_n}(z)\big].$$

同分布的情况

当 $X_1, \ldots, X_n$ 的有相同的分布函数 $F(z)$ 时，有

$$F_M(z) = \big[F(z)\big]^n, \quad F_N(z) = 1 - \big[1 - F(z)\big]^n.$$

以上结论对任何类型的随机变量都成立. 当 $X_1, \ldots, X_n$ 是连续型的时候，可以进一步利用分布函数求出密度函数.

指数分布的极小值

体育馆的大屏幕由信号处理机和显示屏构成，它们的寿命分别为 X, Y. 若它的概率密度分别为

$$f_X(x) = \mathbb{1}_{(0,\infty)}(x)\alpha e^{-\alpha x}, \quad f_Y(y) = \mathbb{1}_{(0,\infty)}(y)\beta e^{-\beta y}, \qquad \alpha, \beta > 0.$$

试求大屏幕的寿命 Z 的概率密度.

思路

显然 $Z = \min(X, Y)$. 注意到对于参数为 λ 的指数分布，有 $\mathsf{P}(\mathrm{Exp}(\lambda) \geq t) = e^{-\lambda t}$, $t \geq 0$.

解

由独立性有

$$\mathsf{P}(Z \geq t) = \mathsf{P}(\min(X, Y) \geq t) = \mathsf{P}(X \geq t)\mathsf{P}(Y \geq t) = e^{-(\alpha+\beta)t}, \quad t \geq 0.$$

因此独立指数分布的极小值仍然是指数分布.

例题

设某种电子管的寿命 (以天计) 近似服从 $\mathcal{N}(1195, 15^2)$. 随机地选取 3 只, 求

1. 其中没有一只寿命超过 1210 天的概率;
2. 其中没有一只寿命小于 1210 天的概率.

思路

设 X_1, X_2, X_3 为 3 只电子管的寿命, 则所求为 $M = \max(X_1, X_2, X_3)$, $N = \min(X_1, X_2, X_3)$ 相关的概率.

解

已知 $X_i \sim \mathcal{N}(1195, 15^2)$ 且相互独立.

1. 所求为 $\mathsf{P}(M \leq 1210) = F_M(1210) = \left[F(1210)\right]^3 = \left[\Phi\left(\frac{1210-1195}{15}\right)\right]^3 = \left[\Phi(1)\right]^3 \approx 0.5955.$
2. 所求为 $\mathsf{P}(N \geq 1210) = 1 - F_N(1210) = \left[1 - F(1210)\right]^3 = \left[1 - \Phi(1)\right]^3 \approx 0.004.$

例题

设 X, Y 为独立的、参数为 α, β 的指数分布. 求 $M = \max(X, Y)$, $S = X + Y$ 的密度函数.

解

易见 $M \geq 0$. 当 $t \geq 0$ 时, 由**独立性**, M 的分布函数为

$$F_M(t) = \mathsf{P}(M \leq t) = \mathsf{P}(X \leq t) \cdot \mathsf{P}(Y \leq t) = (1 - e^{-\alpha t})(1 - e^{-\beta t}), t \geq 0.$$

因此 M 的密度函数为

$$f_M(t) = \mathbb{1}_{(0,\infty)}(t) F_M'(t) = \mathbb{1}_{(0,\infty)}(t)(\alpha e^{-\alpha t} + \beta e^{-\beta t} - (\alpha + \beta) e^{-(\alpha+\beta)t}).$$

又 $S \geq 0, X, Y \geq 0$. 当 $t \geq 0$ 时, 由卷积公式, Z 的密度函数为

$$\begin{aligned} f_Z(t) &= \int_{-\infty}^{\infty} f_X(t-y) f_Y(y)\, dy = \int_0^t f_X(t-y) f_Y(y)\, dy \\ &= \int_0^t \alpha e^{-\alpha(t-y)} \beta e^{-\beta y}\, dy = \alpha\beta e^{-\alpha t} \int_0^t e^{-(\beta-\alpha)y}\, dy \\ &= \frac{\alpha\beta}{\beta - \alpha}[e^{-\alpha t} - e^{-\beta t}]. \end{aligned}$$

微元法求 $Z = \max(X_1, \ldots, X_n)$ 的密度

假设 $X_1, X_2, \ldots, X_n$ 独立. 对充分小的区间 $(z, z+dz)$,

$$\begin{aligned}\mathsf{P}(Z \in (z, z+dz)) &\approx \mathsf{P}(\exists i: X_i \in (z, z+dz) \text{ 且 } X_j < z, \forall j \neq i) \\ &= \sum_{i=1}^{n} \mathsf{P}(X_i \in (z, z+dz) \text{ 且 } X_j < z, \forall j \neq i) \\ &= n\Big[f_X(z)dz\Big] \cdot \Big[F_X(z)\Big]^{n-1}.\end{aligned}$$

而左边根据定义为 $f_Z(z)dz$. 比较等式两边得

$$f_Z(z)dz = n f_X(z)\Big[F_X(z)\Big]^{n-1}.$$

定义

顺序统计量

设 $X_1, X_2, \ldots, X_n \sim f(x)$ 是独立同分布的连续型随机变量. 将 $X_1, X_2, \ldots, X_n$ 由小到大排列为 $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$.

顺序统计量 $(X_{(1)}, \ldots, X_{(n)})$.

最小值 $X_{(1)} = \min(X_1, \ldots, X_n)$.

最大值 $X_{(n)} = \max(X_1, \ldots, X_n)$.

若 $n = 2m+1$ 为奇数, 称 $X_{(m+1)}$ 为中位数.

注

对连续型随机变量 X 与 Y, $\mathsf{P}(X = Y) = 0$. 因此顺序统计量以概率 1 是唯一决定的.

问

如何求 $X_{(k)}$ 的密度?

顺序统计量的密度

$$\begin{aligned}
\mathsf{P}(X_{(k)} \in (z, z+dz)) \approx \mathsf{P}\Big(&\exists i, j_1, \ldots, j_{k-1}, l_1, \ldots, l_{n-k} : \\
&X_i \in (z, z+dz), \\
&X_{j_m} < z, \forall 1 \le m \le k-1, \\
&X_{l_m} \ge z+dz, \forall 1 \le m \le n-k\Big) \\
= \sum_{i, j_1, \ldots, j_{k-1}, l_1, \ldots, l_{n-k}} & F^{k-1}(z) \cdot f(z)dz \cdot (1 - F(z+dz))^{n-k} \\
= \frac{n!}{(k-1)!1!(n-k)!} & F^{k-1}(z) \cdot f(z)dz \cdot (1 - F(z+dz))^{n-k}
\end{aligned}$$

顺序统计量的密度

$$f_{X_{(k)}}(z) = \frac{n!}{(k-1)!(n-k)!} f(z) F^{k-1}(z) \big[1 - F(z)\big]^{n-k}.$$

均匀分布的顺序统计量

回顾

连续型随机变量 X 可以表为 $X = F_X^{-1}(U)$, $U \sim \mathrm{U}[0,1]$. 注意到 F_X^{-1} 为增函数, 因此不改变随机变量间的顺序. 因此若我们能了解 n 个独立 $\mathrm{U}[0,1]$ 随机变量的顺序统计量, 也能借此研究一般连续型随机变量的顺序统计量.

均匀分布的顺序统计量

设 $U_i \sim \mathrm{U}[0,1]$ 相互独立, 则 $U_{(k)}$ 的密度为

$$\frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k} \sim \mathrm{Beta}(k, n-k+1).$$

Beta 分布

$X \sim \mathrm{Beta}(\alpha, \beta)$ 当且仅当

$$f(u) = \mathbb{1}_{(0,1)}(u) \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1}(1-u)^{\beta-1}$$

顺序统计量的联合密度

问题

设 $U = \max(X_1, \dots, X_n)$, $V = \min(X_1, \dots, X_n)$. 求 U 和 V 的联合密度.

微元法

$$
\begin{aligned}
&\mathsf{P}(V \in (v, v+dv),\ U \in (u, u+du)) \\
=&\ \mathsf{P}(\exists i, j : X_i \in (v, v+dv),\ X_j \in (u, u+du),\ X_k \in (v+dv, u), \forall k \neq i, j) \\
=&\ n(n-1) \cdot f(v)dv \cdot f(u)du \cdot \Big[F(u) - F(v+dv)\Big]^{n-2}
\end{aligned}
$$

U, V 联合密度函数

$$f(u,v) = \mathbb{1}_{\{u \geq v\}}(u,v) \cdot n(n-1) f(v) f(u) [F(u) - F(v)]^{n-2}.$$

对于均匀分布,

$$f(u,v) = \mathbb{1}_{\{1 \geq u \geq v \geq 0\}}(u,v) n(n-1)(u-v)^{n-2}.$$

- P81: 70
- 补充题:
  1. 从 1, 2, 3 中一次任取两个数. 第一个数记为 X, 第二个数记为 Y. 记 $Z = \max(X, Y)$. 求 (X, Y) 和 (X, Z) 的联合频率函数和边际频率函数.
  2. 设 X 和 Y 是两个独立的随机变量, 服从 $\mathcal{N}(0, 1)$. 令 $Z = \min(X, Y)$. 求 Z 的分布函数.