5 多维随机变量及其分布

实际问题中很多随机现象是由两个或多个随机因素造成的, 需用多个随机变量描述. 例如:

i) 导弹攻击点的坐标(经度、纬度)

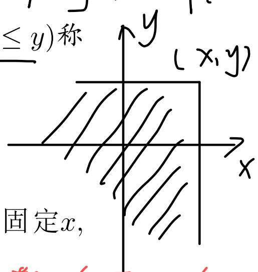
ii) 学生的高考成绩(语文、数学、英语等)

定义5.1. 设随机试验的样本空间是Ω, $X = X(\omega)$, $Y = Y(\omega)$, 定义在Ω上的随机变量. 由它们构成的向量(X, Y)称为二维随机变量.

二维随机向量又称为二维随机变量; 将(X, Y)看作一个整体, 不能分开看待; 几何上, (X, Y)可看作平面上的随机点.

5.1 基本概念与性质

F(x,y) 为 X≤x 且 Y≤y 的分布.



定义5.2 (分布函数)**.** 设(X, Y)是一个二维随机变量, 对任意实数对(x, y), $F(x, y) = P(X \le x, Y \le y)$称为二维随机变量$(X, Y)$的分布函数, 或称为随机变量$X$和$Y$的联合分布函数.

几何意义: $F(x, y)$表示点(X, Y)落入以(x, y)为右上定点的无穷矩形的概率.

分布函数性质

1) 分布函数$F(x, y)$对每个变量单调不减: 固定y, 当$x_1 > x_2$时, 有$F(x_1, y) \ge F(x_2, y)$; 固定x, 当$y_1 > y_2$时, 有$F(x, y_1) \ge F(x, y_2)$.

由于已知三个非全无穷量, 因此无论什么题目我们都已有至少三个条件.

2) 分布函数$F(x, y) \in [0, 1]$且对任意$x, y \in \mathbb{R}$, $F(-\infty, y) = 0$, $F(x, -\infty) = 0$, $F(-\infty, -\infty) = 0$, $F(+\infty, +\infty) = 1$.

3) 分布函数$F(x, y)$关于每个变量右连续.

由分布函数推导概率的公式: 对比一维情况 $F_X(x) = P(X \le x)$

$$P(x_1 < X \le x_2, y_1 < Y \le y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1).$$ 类似容斥

随机变量(X, Y)的联合分布函数为$F(x, y)$, 如果我们将随机变量X和Y单独分别看到, 则其分别为随机变量, 下面研究随机变量X和Y的分布函数$F_X(x)$和$F_Y(y)$.

定义5.3. 对二维随机变量(X, Y), 其联合分布函数为$F(x, y)$, 称

$$F_X(x) = P(X \le x) = P(X \le x, y < +\infty) = F(x, +\infty) = \lim_{y \to +\infty} F(x, y),$$

为随机变量X的边缘分布函数.

求解边缘分布
令 $x \to +\infty$ / $y \to +\infty$ 即可.

类似定义随机变量Y的边缘分布函数为:

$$F_Y(y) = P(Y \le y) = P(Y \le y, x < +\infty) = F(+\infty, y) = \lim_{x \to +\infty} F(x, y).$$

例5.1. 二维随机变量(X,Y)的联合分布函数为$F(x,y)=A(B+\arctan\frac{x}{2})(C+\arctan\frac{y}{3})(x,y\in\mathbb{R})$. 求随机变量$X$与$Y$的边缘分布函数以及概率$P(Y>3)$.

解. 由分布函数的性质有

$$\begin{aligned} 1 &= F(+\infty,+\infty)=A(B+\frac{\pi}{2})(C+\frac{\pi}{2}) \\ 0 &= F(x,-\infty)=A(B+\arctan\frac{x}{2})(C-\frac{\pi}{2}) \\ 0 &= F(-\infty,y)=A(B-\frac{\pi}{2})(C+\arctan\frac{y}{3}) \end{aligned}$$

使用$F(x,-\infty)=0$
时利用性质之思想.

求解可得

$$C=\frac{\pi}{2},B=\frac{\pi}{2},A=\frac{1}{\pi^2}$$

从而得到$F(x,y)=\frac{1}{\pi^2}(\frac{\pi}{2}+\arctan\frac{x}{2})(\frac{\pi}{2}+\arctan\frac{y}{3})$, 进一步得到

$$F_X(x)=\lim_{y\to\infty}\frac{1}{\pi^2}(\frac{\pi}{2}+\arctan\frac{x}{2})(\frac{\pi}{2}+\arctan\frac{y}{3})=\frac{1}{\pi}(\frac{\pi}{2}+\arctan\frac{x}{2})$$

同理可得

$$F_Y(y)=\frac{1}{\pi}(\frac{\pi}{2}+\arctan\frac{y}{3})$$

最后得到

$$P(Y>3)=1-P(Y\le 3)=1-F_Y(3)=1-\left(\frac{1}{2}+\frac{1}{\pi}\arctan 1\right)=\frac{1}{4}.$$

□

独立性

定义5.4. 若X,Y为随机变量, 对任意$x,y\in\mathbb{R}$, 事件$X\le x$和$Y\le y$相互独立, 即

$$P(X\le x,Y\le y)=P(X\le x)\cdot P(Y\le y)\quad\Leftrightarrow\quad F(x,y)=F_X(x)F_Y(y)$$

称X与Y相互独立.

独立性性质影响到
后续期望等性质
的计算.

定理5.1. 设随机变量X与Y相互独立, $f(x)$和$g(y)$是连续或分段连续函数, 则$f(X)$与$g(Y)$也相互独立.

例如: 随机变量X与Y相互独立, 则X^2与Y^3相互独立, $\sin X$与$\cos Y$相互独立.

5.2 二维离散型随机变量

定义5.5. 若二维随机变量(X,Y)的取值是有限个或无限可列的, 称(X,Y)为二维离散型随机变量.

设离散型随机变量X,Y的取值分别为$(x_i,y_i),i=1,2,\cdots,j=1,2,\cdots$, 称

$$p_{ij}=p(X=x_i,Y=y_j)$$

为(X,Y)的联合分布列. 由分布列的性质可知$p_{ij}\ge 0$和$\sum_{i,j}p_{ij}=1$.

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

已知随机变量(X,Y)的联合分布列，讨论随机变量X和Y各自的分布列，即边缘分布列. 首先讨论X的边缘分布列

$$P(X=x_i)=\sum_{j=1}^{\infty}P(X=x_i,Y=y_j)=\boxed{\sum_{j=1}^{\infty}p_{ij}}=p_{i\cdot}$$

将y的所有情况x取i的概率叠加.

同理可得随机变量Y的边缘分布列

$$P(Y=y_j)=\sum_{i=1}^{\infty}P(X=x_i,Y=y_j)=\sum_{i=1}^{\infty}p_{ij}=p_{\cdot j}$$

为什么称之为"边缘"？

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

例5.2. 现有1,2,3三个数，X表示从这三个数中随机地抽取一个整数，Y表示从1到X中随机抽取一个数. 求(X,Y)的联合分布列和边缘分布列.

解. 由题意可知随机变量X和Y的取值为1,2,3，当$X=1$时，$Y=1$；当$X=2$时，Y等可能性取1，2；当$X=3$时，Y等可能性取1，2，3. 于是有

$X \backslash Y$	1	2	3	$p_{i\cdot}$
1	1/3	0	0	1/3
2	1/6	1/6	0	1/3
3	1/9	1/9	1/9	1/3
$p_{\cdot j}$	11/18	5/18	1/9	1

□

定义5.6. 对离散型随机变量(X,Y), 若对所有(x_i,y_j), 有

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j), \quad 即 \quad p_{ij}=p_{i\cdot}p_{\cdot j}$$

称随机变量X与Y独立.

习题. 对二维离散随机变量(X,Y), 如果对任何(x_i,y_j), 有$F(x_i,y_j)=F_X(x_i)F_Y(y_j)$ 成立, 则有

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j).$$

定理5.2. 若X和Y独立, 则对任意集合$A\in\mathbb{R},B\in\mathbb{R}$, 有事件$X\in A$和$Y\in B$独立.

例5.3. 设离散型X,Y独立, 求解(X,Y)的联合分布律

$X \backslash Y$	y_1	y_2	y_3	$p_{i\cdot}$
x_1		$1/8$		
x_2	$1/8$			
$p_{\cdot j}$	$1/6$			

例5.4. 将两个球A,B放入编号为$1,2,3$的三个盒子中, 令X: 放入1号盒的球数, Y: 放入2号盒的球数. 判断X和Y是否独立.

解. 由题意可知

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

由此可到$P(X=2,Y=2)=0\neq P(X=2)P(Y=2)$, 所以$X$和$Y$不独立. □

5.3 二维连续型随机变量

定义5.7. 设二维随机变量的分布函数为$F(x,y)$, 如果存在二元非负可积函数$f(x,y)$, 使得对任意实数(x,y)有$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$, 称$(X,Y)$为二维连续型随机变量, 称$f(x,y)$称为二维随机变量$(X,Y)$的概率密度, 或称为随机变量$X$和$Y$的联合概率密度.

根据定义, 很容易得到概率密度函数$f(x,y)$的如下性质:

非负性 1) $p(x,y)\geq 0$.

规范性 2) $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=1$.

性质 3) $f(x,y)$在(x,y)连续, 则$f(x,y)=\frac{\partial^2F(x,y)}{\partial x\partial y}$.

计算概率

4) G为平面上的一个区域, 则点(X,Y)落入G的概率为

$$P((X,Y)\in G)=\int\int_G p(x,y)dxdy$$

.

例5.5. 设二维随机变量(X,Y)的概率密度

$$f(x,y)=\begin{cases} ce^{-(3x+4y)} & x>0,y>0 \\ 0 & \text{其它} \end{cases}$$

求$P(0<X<1,0<Y<2)$.

解. 有概率密度的性质可知

$$1=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}ce^{-(3x+4y)}dxdy=\frac{c}{12},$$

由此可得$c=12$. 进一步可得

$$P(0<X<1,0<Y<2)=12\int_0^1\int_0^2 e^{-(3x+4y)}dxdy=(1-e^{-3})(1-e^{-8}).$$

□

习题. 二维随机变量(X,Y)密度函数$f(x,y)=\begin{cases} x^2+axy & 0\le x\le 1, 0\le y\le 2 \\ 0 & \text{其它} \end{cases}$. 求$P(X+Y\ge 1)$.

给定概率密度$f(x,y)$, 下面考虑随机变量X和Y的边缘概率密度, 其定义如下:

定义5.8. 二维随机变量(X,Y)的概率密度为$f(x,y)$, 则随机变量X和Y的边缘概率密度为

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy \qquad f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx.$$

另一个变量从负无穷到正无穷积分

边缘概率密度的定义完全由边缘分布的定义导出, 即由前面的知识可知X的边缘分布函数为

$$F_X(x)=F(x,+\infty)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(t,y)dt=\int_{-\infty}^{x}(\int_{-\infty}^{+\infty}f(t,y)dy)dt.$$

从而得到X的边缘密度为

$$f_X(x)=F_X'(x)=\int_{-\infty}^{+\infty}f(x,y)dy.$$

例5.6. 设(X,Y)的概率密度$f(x,y)=\begin{cases} 8xy & 0\le x\le y\le 1 \\ 0 & \text{其它} \end{cases}$. 求$P(X\le\frac{1}{2})$.

解. 当$0\le x\le 1$, $f_X(x)=4x(1-x^2)$. 所以$P(X\le\frac{1}{2})=\int_0^{\frac{1}{2}}4x(1-x^2)dx=\frac{7}{16}$. □

关于连续随机变量的独立性, 有如下性质:

☆ 独立性: $F(x,y) = F_X(x) \cdot F_Y(y)$

$\Rightarrow f(x,y) = f_X(x) f_Y(y)$

定理5.3. 设二维随机变量(X,Y)的概率密度为$f(x,y)$, 随机变量X和Y的边缘概率密度为$f_X(x)$和$f_Y(y)$, 如果随机变量X和Y独立, 则有

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

Proof. 由二维连续型随机变量独立性定义有$F(x,y) = F_X(x)F_Y(y)$, 其等价于

$$\int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy = \int_{-\infty}^{x} f_X(x)dx \int_{-\infty}^{y} f_Y(y)dy = \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(u)f_Y(v)dudv$$

对上式两边同时求导有$f(x,y) = f_X(x)f_Y(y)$. □

例5.7. 设二维随机变量的密度函数$f(x,y) = \begin{cases} cxe^{-y} & 0 < x < y < +\infty \\ 0 & \text{其它} \end{cases}$. 问$X$与$Y$是否独立.

解.

$$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = c\int_{0}^{+\infty} dy \int_{0}^{y} xe^{-y}dx = c.$$

当$x > 0$时:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_{x}^{+\infty} xe^{-y}dy = xe^{-x}.$$

所以

$$f_X(x) = \begin{cases} xe^{-x} & x > 0 \\ 0 & \text{其它} \end{cases}$$

. 当$y > 0$时:

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

□

例5.8. 设X与Y相互独立, X服从$[-1,1]$均匀分布, Y服从参数为$\lambda = 2$的指数分布, 求$P(X+Y \le 1)$.

解. $f_X(x) = \begin{cases} \frac{1}{2} & x \in [-1,1] \\ 0 & \text{其它} \end{cases}$, $f_Y(y) = \begin{cases} 2e^{-2y} & y \ge 0 \\ 0 & \text{其它} \end{cases}$. 因为$X,Y$相互独立, 所以

$$f(x,y) = \begin{cases} e^{-2y} & -1 \le x \le 1, y \ge 0 \\ 0 & \text{其它} \end{cases}.$$

所以

$$P(X+Y \le 1) = \int_{-1}^{1} dx \int_{0}^{1-x} e^{-2y}dy = \frac{3}{4} + \frac{1}{4}e^{-4}.$$

. □

二维正态分布

二维正态分布$(X,Y)\sim\mathcal{N}(\mu,\Sigma)$, 其中

σ_x^2, σ_y^2 即方差.
ρ 满足 $|\rho|<1$

期望矢量

$$\mu=(\mu_x,\mu_y)^\top,\quad \Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}$$

Σ是一个正定的矩阵, 即要求$|\rho|<1$. 正太分布的密度函数为

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sqrt{|\Sigma|}}\exp\left(-\frac{1}{2}(\xi-\mu)^\top\Sigma^{-1}(\xi-\mu)\right) \qquad \xi=(x,y)^\top \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho}{\sigma_x\sigma_y}(x-\mu_x)(y-\mu_y)\right]\right)\end{aligned}$$

其中

☆

$$|\Sigma|=(1-\rho^2)\sigma_x^2\sigma_y^2,\quad \Sigma^{-1}=\frac{1}{(1-\rho^2)\sigma_x^2\sigma_y^2}\begin{pmatrix}\sigma_y^2 & -\rho\sigma_x\sigma_y\\ -\rho\sigma_x\sigma_y & \sigma_x^2\end{pmatrix}=\frac{1}{1-\rho^2}\begin{pmatrix}1/\sigma_x^2 & -\rho/\sigma_x\sigma_y\\ -\rho/\sigma_x\sigma_y & 1/\sigma_y^2\end{pmatrix}.$$

定理5.4. 对二维正态分布, $\xi=(X,Y)\sim\mathcal{N}(\mu,\Sigma)$, 其中

$$\mu=(\mu_x,\mu_y)^\top,\quad \Sigma=\begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y\\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix},$$

则边缘分布$X\sim\mathcal{N}(\mu_x,\sigma_x^2)$, $Y\sim\mathcal{N}(\mu_y,\sigma_y^2)$. 边缘分布仍然满足正态分布的分布.

Proof. 根据正太分布的概率密度定义以及边缘密度公式有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\frac{e^{-\frac{1}{2(1-\rho^2)}(\frac{x-\mu_x}{\sigma_x})^2}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}e^{-\frac{1}{2(1-\rho^2)}[(\frac{y-\mu_y}{\sigma_y})^2-2\rho\frac{x-\mu_x}{\sigma_x\sigma_y}(y-\mu_y)]}dy$$

令$t=\frac{y-\mu_y}{\sigma_y}$, 所以$dy=\sigma_y dt$, 则有

$$f_X(x)=\frac{e^{-\frac{1}{2(1-\rho^2)}(\frac{x-\mu_x}{\sigma_x})^2}}{2\pi\sqrt{1-\rho^2}\sigma_x\sigma_y}\int_{-\infty}^{+\infty}e^{-\frac{1}{2(1-\rho^2)}[t^2-2\rho t(x-\mu_x)^2/\sigma_x]}dy=\frac{e^{-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2}}{\sqrt{2\pi}\sigma_x}.$$

从而完成证明. □

上述定理说明正太分布的边缘分布还是正太分布.

定理5.5. 若二维随机变量$(X,Y)\sim\mathcal{N}(\mu,\Sigma)$, 则$X$与$Y$独立的充要条件为$\Sigma=\begin{pmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{pmatrix}$.

Proof. 必要性证明: 当$\Sigma=\begin{pmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{pmatrix}$时, 由上一定理可得:

$$f(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e^{-\frac{1}{2}[(\frac{x-\mu_x}{\sigma_x})^2+(\frac{y-\mu_y}{\sigma_y})^2]}=f_X(x)f_Y(y).$$

充分性证明: 利用定义反证即可. □

多维正太分布

假设n维随机变量$X=(X_1,X_2,\cdots,X_n)^\top$服从正太分布

$$X\sim\mathcal{N}(\mu,\sigma)\qquad \mu\in\mathbb{R}^n,\quad \Sigma\in\mathbb{R}^{n\times n}\text{ 正定},$$

其密度函数为

$$f(x_1,\ldots,x_n)=\frac{1}{(2\pi)^{n/2}|\Sigma|}\exp\left(-\frac{1}{2}(\xi-\mu)^\top\Sigma^{-1}(\xi-\mu)\right)$$

其中$\xi=(x_1,\ldots,x_n)^\top$.

定理5.6. 如果正太分布$\xi=(X,Y)\sim\mathcal{N}(\mu,\sigma)$, 其中

$$\mu=(\mu_x,\mu_y)^\top,\quad \mu_x=(\mu_{x_1},\mu_{x_2},\cdots,\mu_{x_n}),\quad \mu_y=(\mu_{y_1},\mu_{y_2},\cdots,\mu_{y_m}),\quad \Sigma=\begin{pmatrix}\sum_{xx} & \sum_{yx}\\ \sum_{xy} & \sum_{yy}\end{pmatrix},$$

那么,

- 边缘分布$X\sim N(\mu_x,\Sigma_{xx})$和$Y\sim N(\mu_y,\Sigma_{yy})$
- X与Y独立$\Leftrightarrow\sum=\begin{pmatrix}\sum_{xx} & 0\\ 0 & \sum_{yy}\end{pmatrix}$.

例：$X\sim N(\mu,\sigma^2)$，$Y\sim N(\mu,\sigma^2)$，X与Y独立，$Z_1=\alpha X+\beta Y$，$Z_2=\alpha X-\beta Y$

求$\rho_{Z_1Z_2}$

解：$\rho_{Z_1Z_2}=\dfrac{Cov(Z_1,Z_2)}{\sqrt{Var(Z_1)Var(Z_2)}}$

$Cov(Z_1,Z_2)$计算：① 若用$Cov(Z_1,Z_2)=E[Z_1Z_2]-\underbrace{E[Z_1]E[Z_2]}_{\text{容易计算}}$

$$E[Z_1Z_2] = E[\alpha^2X^2-\beta^2Y^2]$$

$\downarrow$

计算$E[X^2]$略烦。

② $Cov(Z_1,Z_2)=Cov(\alpha X+\beta Y,\ \alpha X-\beta Y)$

$=\alpha^2 Cov(X,X)-\beta^2 Cov(Y,Y)$

$=\alpha^2 Var(X)-\beta^2 Var(Y)$

研究：如果X, Y是随机变量，那么关于X、Y的函数 Z=g(X, Y)，它作为随机变量的分布情况.
一般情况下同时要求g连续.

5.4 多维随机变量函数的分布

离散型求解方法

本节研究的问题: 已知(X,Y)的分布, 求$Z=g(X,Y)$的分布.

如果$X_1,X_2,\cdots,X_n$是n维离散型随机变量, 那么$Z=g(X_1,X_2,\cdots,X_n)$为一维随机变量, 其分布列可以通过如下两步求得:

i) 对$X_1,X_2,\cdots,X_n$的各种取值, 计算Z对应的取值;

ii) 对相同的Z值, 合并其概率.

例5.9. 设(X,Y)的联合分布列为

$X \backslash Y$	0	1	2
0	1/4	1/6	1/8
1	1/4	1/8	1/12

求$Z_1=X+Y, Z_2=XY$的分布列.

解. 通过简单计算、合并可得Z_1和Z_2的分布列分别为:

Z_1	0	1	2	3
P	1/4	5/12	1/4	1/12

Z_2	0	1	2
P	19/24	1/8	1/12

□

连续型求解方法

对连续随机变量(X,Y), 其联合密度为$f(x,y)$, 如何求解$Z=g(X,Y)$的密度函数. 主要的解决思路是分布函数法, 即:

i) 求$Z=g(X,Y)$的分布函数

$$F_Z(z)=P(Z\le z)=P(g(x,y)\le z)=\int\int_{g(x,y)\le z} f(x,y)dxdy.$$

ii) 求Z的密度函数

$$f_Z(z)=F_Z'(z).$$

5.4.1 极大极小分布

极大极小分布

设随机变量X,Y相互独立, 其分布函数分别为$F_X(x),F_Y(y)$, 求$Z_1=\max(X,Y)$, $Z_2=\min(X,Y)$的分布函数. 求解Z_1的分布函数为

$$\begin{aligned}F_{Z_1}(z_1)&=P(Z_1\le z_1)\\&=P(\max(X,Y)\le z_1)=P(X\le z_1,Y\le z_1)\\&=P(X\le z_1)P(Y\le z_1)=F_X(z_1)F_Y(z_1).\end{aligned}$$

独立性

如果独立: $f_Z(z_1)=F_X(z_1)f_Y(z_1)+f_X(z_1)F_Y(z_1)$

讨论：若 X, Y 不独立。
设 $Z = \max(f(X,Y), g(X,Y))$
$F_Z(z) = P(\max(f(x,y), g(x,y)) \le z)$
$= \iint_D h(x,y)\,dxdy$
其中，$h(x,y)$ 为 Z 的概率密度，D 为 $\max(f,g) \le z$ 的区域。需分类讨论！

进一步求解Z_2的分布函数为

$$
\begin{aligned}
F_{Z_2}(z_2) &= P(Z_2 \le z_2) \\
&= P(\min(X,Y) \le z_2) = 1 - P(\min(X,Y) > z_2) \\
&= 1 - P(X \ge z_2)P(Y > z_2) = 1 - (1 - F_X(z_2))(1 - F_Y(z_2)).
\end{aligned}
$$

上述结论可以进一步推广到n个独立的随机变量: 设$X_1, X_2, \cdots, X_n$为n个独立的随机变量, 其分布函数分别为$F_{X_i}(x)$, 则$Y = \max(X_1, X_2, \cdots, X_n)$, $Z = \min(X_1, X_2, \cdots, X_n)$的分布函数分别为

$$
F_Y(y) = F_{X_1}(y)F_{X_2}(y)\cdots F_{X_n}(y), \quad F_Z(z) = 1 - (1 - F_{X_1}(z))(1 - F_{X_2}(z))\cdots(1 - F_{X_n}(z)).
$$

特别地, 当$X_1, X_2, \cdots, X_n$ 独立同分布时, $F_Y(y) = (F_{X_1}(y))^n$, $F_Z(z) = 1 - (1 - F_{X_1}(z))^n$. 根据分布函数, 可以求得概率密度. 同时，$F_{Z_1}(z)$ 和 F_{Z_2} 满足 $F_{Z_2} = 1 - F_{Z_1} + F_X + F_Y$.

例5.10. 假设随机变量X与Y相互独立, 以及$X \sim e(\alpha)$, $Y \sim e(\beta)$, 求$Z_1 = \max(X,Y)$, $Z_2 = \min(X,Y)$的概率密度.

解. 由指数函数的定义可得随机变量X和Y的概率密度为

$$
f_X(x) = \begin{cases} \alpha e^{-\alpha x} & x \ge 0 \\ 0 & x \le 0 \end{cases} \qquad f_Y(y) = \begin{cases} \beta e^{-\beta y} & y \ge 0 \\ 0 & y \le 0 \end{cases}.
$$

求解Z_1的分布函数为

$$
F_{Z_1}(z_1) = F_X(z_1)F_Y(z_1) = \int_{-\infty}^{z_1} f_X(t)dt \int_{-\infty}^{z_1} f_Y(t)dt.
$$

当$z_1 \le 0$时由$F_{Z_1}(z_1) = 0$; 当$z_1 > 0$时

$$
F_{Z_1}(z_1) = \int_0^{z_1} f_X(t)dt \int_0^{z_1} f_Y(t)dt = \int_0^{z_1} \alpha e^{-\alpha t}dt \int_0^{z_1} \beta e^{-\beta y}dt = (1 - e^{-\alpha z_1})(1 - e^{-\beta z_1}).
$$

两边对z_1求导可得其概率密度为

$$
f_{Z_1}(z_1) = \begin{cases} \alpha e^{-\alpha z_1} + \beta e^{-\beta z_1} - (\alpha + \beta)e^{-(\alpha+\beta)z_1} & z_1 \ge 0 \\ 0 & z_1 < 0. \end{cases}
$$

同理可求得随机变量Z_2的分布函数和概率密度分别为

$$
F_{Z_2}(z_2) = \begin{cases} 1 - e^{-(\alpha+\beta)z_2} & z_2 \ge 0 \\ 0 & z_2 < 0 \end{cases} \qquad f_{Z_2}(z_2) = \begin{cases} (\alpha + \beta)e^{-(\alpha+\beta)z_2} & z_2 \ge 0 \\ 0 & z_2 < 0. \end{cases}
$$

□

5.4.2 和的分布$Z = X + Y$

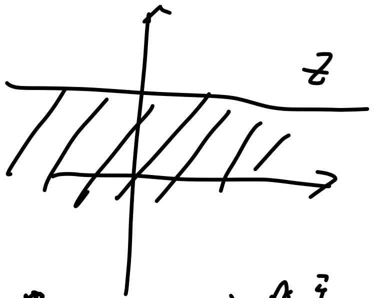
和的分布

引理5.1. 设二维连续随机变量(X,Y)的联合密度为$f(x,y)$, 则$Z = X + Y$的概率密度为

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

解. 首先求解分布函数

$$
\begin{aligned}
F_Z(z) &= P(Z \le z) = P(X+Y \le z) \\
&= \iint_{x+y\le z} f(x,y)dxdy = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x,y)dy \\
&= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z} f(x,u-x)du \qquad (\text{变量替换} u = y+x) \\
&= \int_{-\infty}^{z} \left(\int_{-\infty}^{+\infty} f(x,u-x)dx \right) du
\end{aligned}
$$



变上限求导，这里的积分换序由于是右图的形状，故积分换序能成立。

两边对z求导数可得$Z=X+Y$的概率密度为

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

□

卷积公式

定理5.7 (卷积公式). 如果连续随机变量X与Y独立，则$Z=X+Y$的密度函数为

(实际上是和分布+独立性的结论)

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

如果离散随机变量X与Y独立，其分布列分别为$a_i = P(X=i)$ $(i=0,1,\cdots)$和$b_j = P(Y=j)$ $(j=0,1,\cdots)$，则$Z=X+Y$的分布列为

$$P(Z=X+Y=k) = \sum_{i=0}^{k} a_i b_{k-i}.$$

和的分布性质

B分布.

定理5.8. 若随机变量$X \sim B(n_1,p)$和$Y \sim B(n_2,p)$独立，则$Z=X+Y \sim B(n_1+n_2,p)$.

(可推广到n个的情形)

由此定理可推出: 如果$X_i \sim Ber(p) = B(1,p)$，那么$\sum_{i=1}^{n} X_i \sim B(n,p)$.

Proof. 由卷积公式可得

$$
\begin{aligned}
P[Z=k] &= \sum_{i=0}^{k} P[X=i]P[Y=k-i] \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^i (1-p)^{n_1-i} \binom{n_2}{k-i} p^{k-i}(1-p)^{n_2-(k-i)} \\
&= p^k (1-p)^{n_1+n_2-k} \sum_{i=0}^{k} \binom{n_1}{i}\binom{n_2}{k-i} \\
&= \binom{n_1+n_2}{k} p^k (1-p)^{n_1+n_2-k}.
\end{aligned}
$$

□

P分布 (泊松)

定理5.9. 若随机变量$X \sim P(\lambda_1)$和$Y \sim P(\lambda_2)$相互独立，则$Z=X+Y \sim P(\lambda_1+\lambda_2)$.

Proof. 由泊松分布的定义可知: 当$i \geq 0$和$j \geq 0$时, 有

$$P(X=i)=\frac{\lambda_1^i}{i!}e^{-\lambda_1} \qquad P(Y=j)=\frac{\lambda_2^j}{j!}e^{-\lambda_2}.$$

根据卷积公式有

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

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

正态分布

定理5.10. 若随机变量$X\sim\mathcal{N}(\mu_1,\sigma_1^2)$和$Y\sim\mathcal{N}(\mu_2,\sigma_2^2)$相互独立, 则$X+Y\sim\mathcal{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

Proof. 首先证明$Z=X-\mu_1+Y-\mu_2\sim\mathcal{N}(0,\sigma_1^2+\sigma_2^2)$, 其中$X-\mu_1\sim\mathcal{N}(0,\sigma_1^2)$和$Y-\mu_2\sim\mathcal{N}(0,\sigma_2^2)$. 根据卷积公式有

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)dx = \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} e^{-\frac{x^2}{2\sigma_1^2}-\frac{(z-x)^2}{2\sigma_2^2}}dx \\ &= \frac{1}{2\pi\sigma_1\sigma_2}\int_{-\infty}^{+\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(x-\frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\right)^2-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}dx \\ &= \frac{e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}\frac{\sqrt{\sigma_1^2+\sigma_2^2}}{\sqrt{2\pi}\sigma_1\sigma_2}\int_{-\infty}^{+\infty} e^{-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(x-\frac{\sigma_1^2 z}{\sigma_1^2+\sigma_2^2}\right)^2}dx \\ &= \frac{e^{-\frac{z^2}{2(\sigma_1^2+\sigma_2^2)}}}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}}. \end{aligned}$$

$= \sqrt{2\pi} \cdot \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}$

由此可得$X-\mu_1+Y-\mu_2\sim\mathcal{N}(0,\sigma_1^2+\sigma_2^2)$, 进一步证明$X+Y\sim\mathcal{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$. □

习题. 若随机变量$X\sim e(\lambda_1)$和$Y\sim e(\lambda_2)$相互独立, 求$Z=X+Y$的分布函数和概率密度.

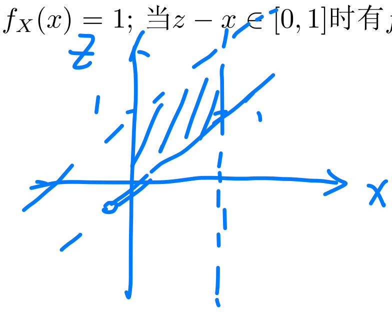
U分布

例5.11. 设随机变量$X\sim U(0,1)$和$Y\sim U(0,1)$相互独立, 求$Z=X+Y$的概率密度. (开始出现对定义域的讨论!)

解. 由卷积公式可得

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

由$X\sim U(0,1)$和$Y\sim U(0,1)$可知: 当$x\in[0,1]$时有$f_X(x)=1$; 当$z-x\in[0,1]$时有$f_Y(z-x)=1$, 即积分区域为$\{x\in[0,1], z-x\in[0,1]\}$. 由此可得:



- 当 $z\leq 0$ 或 $z\geq 2$ 时, 有$f_Z(z)=0$;
- 当 $z\in(0,1)$ 时, 有$f_Z(z)=\int_0^z 1dx=z$;
- 当$z\in[1,2)$ 时, 有$f_Z(z)=\int_{z-1}^1 dx=2-z$.

要对z进行讨论, 先固定z(划横线), 然后看在这个横线上 x的可行域的范围。 □

例5.12. 设随机变量$X \sim U(0,1)$和$Y \sim e(1)$相互独立，求$Z = X + Y$的概率密度.

解. 由卷积公式有

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

由于$X \sim U(0,1)$和$Y \sim e(1)$可知$f_X(x)f_Y(z-x) \neq 0$的区域为$\{x \in [0,1], z \geq x\}$. 因此有:

- 当$z \leq 0$时有$f_Z(z) = 0$;
- 当$0 \leq z \leq 1$时有$f_Z(z) = \int_0^z e^{-(z-x)} dx = e^{-z}(e^z - 1) = 1 - e^z$;
- 当$z \geq 1$时，$f_Z(z) = \int_0^1 e^{-(z-x)} dx = e^{-z}(e^1 - 1) = (e-1)e^{-z}$.

同上分析

对于连续型，使用卷积公式＋ $z-x$ 图像

对于离散型，使用 $\sum_{i=0}^{z} p_i q_{z-i}$

5.4.3 随机变量的乘/除法分布

设二维连续连续随机变量(X,Y)的联合概率密度为$f(x,y)$，则随机变量$Z = XY$的概率密度为

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx.$$

随机变量$Z = Y/X$的概率密度为

$$f_{Y/X}(z) = \int_{-\infty}^{+\infty} |x| f(x, xz) dx.$$

这里给出除法$Z = Y/X$概率密度的详细证明(同理证明$Z = XY$). 考虑分布函数

$$\begin{aligned}
F_{Y/X}(z) &= P(Y/X \leq z) = \int\int_{y/x \leq z} f(x,y) dxdy \\
&= \int\int_{x<0, y \geq zx} f(x,y) dxdy + \int\int_{x>0, y \leq zx} f(x,y) dxdy \\
&= \int_{-\infty}^{0} dx \int_{zx}^{+\infty} f(x,y) dy + \int_0^{+\infty} dx \int_{-\infty}^{xz} f(x,y) dy.
\end{aligned}$$

变量替换$t = y/x$有

$$\begin{aligned}
F_{Y/X}(z) &= \int_{-\infty}^{0} dx \int_{z}^{-\infty} x f(x,tx) dt + \int_0^{+\infty} dx \int_{-\infty}^{z} x f(x,tx) dt \\
&= \int_{-\infty}^{0} \int_{-\infty}^{z} (-x) f(x,tx) dtdx + \int_0^{+\infty} \int_{-\infty}^{z} x f(x,tx) dtdx \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{z} |x| f(x,tx) dtdx = \int_{-\infty}^{z} dt \int_{-\infty}^{+\infty} |x| f(x,tx) dx.
\end{aligned}$$

最后求导得到概率密度.

5.5 多维随机变量的数学特征

5.5.1 多维随机变量的期望

定理5.12. 对离散型二维随机变量(X,Y), 设随机变量$Z=g(X,Y)$, 则

$$E[Z]=E[g(X,Y)]=\sum_{i,j}g(x_i,y_j)p_{ij};$$

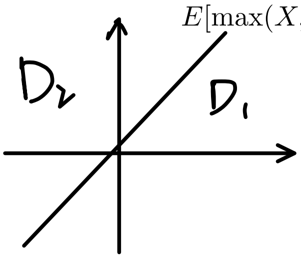
对连续型二维随机变量(X,Y), 设联合概率密度为$f(x,y)$, 以及随机变量$Z=g(X,Y)$, 则

$$E[Z]=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f(x,y)dxdy.$$

trivial.

例5.13. 设$X\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{N}(0,1)$且X与Y独立, 求$E[\max(X,Y)]$. → 也可以通过上一讲的max来做，但不如使用通用公式!

解. 根据独立性定义得到随机变量X和Y的联合概率密度为$f(x,y)=f_X(x)f_Y(y)=\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}$



$$\begin{aligned}E[\max(X,Y)] &= \int\int_{D_1}xf(x,y)dxdy+\int\int_{D_2}yf(x,y)dxdy\\ &= \int_{-\infty}^{+\infty}dy\int_y^{+\infty}xf(x,y)dx+\int_{-\infty}^{+\infty}dx\int_x^{+\infty}yf(x,y)dy\\ &= 2\int_{-\infty}^{+\infty}dy\int_y^{+\infty}xf(x,y)dx=\frac{1}{\pi}\int_{-\infty}^{+\infty}dy\int_y^{+\infty}xe^{-\frac{x^2+y^2}{2}}dx\\ &= \frac{1}{\pi}\int_{-\infty}^{+\infty}e^{-y^2}dy=\frac{1}{\sqrt{\pi}}\end{aligned}$$

最后一个等式成立是因为$\int_{-\infty}^{+\infty}e^{-y^2}dy=\sqrt{\pi}$. □

习题. 在长为1的线段上任取两点X,Y, 求$E[\min(X,Y)], E[|X-Y|]$.

定理5.13. *1)* 对任意随机变量X,Y和常数a,b, 有$E[aX+bY]=aE[X]+bE[Y]$.
2) 对独立随机变量X和Y, 有$E[XY]=E[X]E[Y]$, 对任意函数h,g有$E[h(X)g(Y)]=E[h(X)]E[g(Y)]$;
3) 对非独立随机变量X和Y, 有*Cauchy-Schwartz*不等式$E[XY]\le\sqrt{E[X^2]E[Y^2]}$.
4) 对独立随机变量X和Y, $E[\frac{X}{Y}]=E[X]\cdot E[\frac{1}{Y}]$.

Proof. 设X,Y的联合概率密度为$f(x,y)$, 则

$$\begin{aligned}E[aX+bY] &= \int\int(ax+by)f(x,y)dxdy\\ &= a\underbrace{\int\int xf(x,y)dxdy}_{=E[X]}+b\int\int yf(x,y)dxdy=aE(X)+bE(Y).\end{aligned}$$

注意区别.

若X与Y独立, 则有

$$\begin{aligned}E[XY] &= \int\int xyf(x,y)dxdy=\int\int xf_X(x)yf_Y(y)dxdy\\ &= \underbrace{\int_{-\infty}^{+\infty}xf_X(x)dx}_{=E[X]}\int_{-\infty}^{+\infty}yf_Y(y)dy=E(X)E(Y).\end{aligned}$$

若X与Y不独立, 对任意$t\in\mathbb{R}$有$E[(X+tY)^2]\ge0$成立, 即任意$t\in\mathbb{R}$, → 构造一个非负变量.

$$t^2E[Y^2]+E[X^2]+2tE[XY]\ge0.$$

因此有$\Delta=4[E(XY)]^2-4E[X^2]E[Y^2]\le0$, 即$E(XY)\le\sqrt{E(X^2)E(Y^2)}$. □

5.5.2 协方差

定理5.14. 对随机变量X与Y, 有 $Var(X-Y) = Var(X)+Var(Y)-2E[(X-E(X))(Y-E(Y))]$

$$Var(X+Y) = Var(X) + Var(Y) + 2E[(X-E(X))(Y-E(Y))].$$

特别地, 当X与Y独立时, 有

$$Var(X+Y) = Var(X) + Var(Y).$$

Proof. 令$Z = X+Y$, 则

第一定义

$$\begin{aligned} Var(Z) &= E[(Z-EZ)^2] = E[(X-EX+Y-EY)^2] \\ &= E(X-EX)^2 + E(Y-EY)^2 + 2E[(X-EX)(Y-EY)] \\ &= Var(X) + Var(Y) + 2E[(X-EX)(Y-EY)]. \end{aligned}$$

若X与Y独立, 则$2E[(X-EX)(Y-EY)] = 0$, 所以$Var(X+Y) = Var(X) + Var(Y)$. □

定义5.9. 定义随机变量X和Y的协方差为

$$Cov(X,Y) = E[(X-E(X))(Y-E(Y))] = E(XY) - E(X)E(Y).$$ 计算方法

根据协方差定义和定理5.14有 能推.

$$Cov(X,X) = Var(X) \quad 和 \quad Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y).$$

下面研究协方差的性质.

性质5.1. 对任意常数c, 有$Cov(X,c) = 0$; $Cov(X,Y) = Cov(Y,X)$. (交换律)

性质5.2. 对任意常数a和b, 随机变量X和Y, 有

$$Cov(aX,bY) = abCov(X,Y) = Cov(abX,Y), \quad Cov(X+a,Y+b) = Cov(X,Y).$$

Proof. 根据协方差定义有

$$\begin{aligned} Cov(aX,bY) &= E[(aX-E(aX))(bY-E(bY))] = abE[(X-E(X))(Y-E(Y))]; \\ Cov(X+a,Y+b) &= E[(X+a-E(X+a))(Y+b-E(Y+b))] = E[(X-E(X))(Y-E(Y))]. \end{aligned}$$

$Cov(X+a, Y+b) = Cov(X,Y)$ □

性质5.3. 对任意随机变量X_1, X_2, Y, 有

$$Cov(X_1+X_2,Y) = Cov(X_1,Y) + Cov(X_2,Y).$$ 线性叠加性.

Proof. 我们有

$$
\begin{aligned}
Cov(X_1+X_2,Y) &= E[(X_1+X_2-E(X_1)-E(X_2))(Y-E(Y))] \\
&= E[(X_1-E(X_1))(Y-E(Y))]+E[(X_2-E(X_2))(Y-EY)] = Cov(X_1,Y)+Cov(X_2,Y).
\end{aligned}
$$

□

由此性质可进一步得到: 对随机变量$X_1, X_2, \ldots, X_n$和$Y_1, Y_2, \ldots, Y_m$, 有

$$
Cov\left(\sum_i^n X_i, \sum_j^m Y_j\right) = \sum_i^n \sum_j^m Cov(X_i, Y_j),
$$

以及进一步有

$Var(X)=Cov(X,X)$

$$
Var\left(\sum_{i=1}^n X_i\right) = Cov\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j).
$$

性质5.4. 随机变量X与Y独立 $\Rightarrow Cov(X,Y)=0$; 但反之不成立.

Proof. 若X与Y独立, 则

$$
Cov(X,Y) = E[(X-E(X))(Y-E(Y))] = E[X-E(X)]E[Y-E(Y)] = 0.
$$

反之不成立, 例如随机变量X的分布列为

X	-1	0	1
P_i	$1/3$	$1/3$	$1/3$

当$X\neq 0$时随机变量$Y=0$, 否则$Y=0$; 则X与Y不独立. 同时有

$$
Cov(X,Y) = E[(X-E(X))(Y-E(Y))] = E[XY]-E(X)E(Y) = 0.
$$

□

性质5.5. 对任意随机变量X与Y, 有

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

等号成立的充要条件是$Y=aX+b$ (即X与Y之间有线性关系).

可直接得到.

Proof. 由Cauchy-Schwartz不等式有

$$
\begin{aligned}
Cov(X,Y) &= E[(X-E(X))(Y-E(Y))] \\
&\leq \sqrt{E[(X-E(X))^2]E[(Y-E(Y))^2]} = \sqrt{Var(X)Var(Y)}.
\end{aligned}
$$

下面证明等号成立的充要条件. 如果$Y=aX+b$, 则

$$
Cov(X,Y) = Cov(X, aX+b) = aVar(X), \quad Var(Y) = a^2Var(X),
$$

所以

$$Cov^2(X,Y) = a^2Var^2(X) = Var(X)a^2Var(X) = Var(X)Var(Y).$$

另一方面, 若$(Cov(X,Y))^2 = Var(X)Var(Y)$, 则

$$(E[(X-EX)(Y-EY)])^2 = E(X-EX)^2E(Y-EY)^2,$$

设

$$\begin{aligned} f(t) &= E[t(X-EX)-(Y-EY)]^2 \\ &= t^2E[X-E(X)]^2 - 2tE[(X-E(X))(Y-E(Y))] + E[Y-E(Y)]^2 \end{aligned}$$

根据一元二次方程的性质

$$\Delta = 4(E[(X-EX)(Y-EY)])^2 - 4E(X-EX)^2E(Y-EY)^2 = 0,$$

得到$f(t)=0$恰有一重根t_0. 由此得到

$$f(t_0) = 0 = E[(t_0(X-EX)-(Y-EY))^2]$$

大于等于0的数的加权和为0
则每个数均为0

因为$[t_0(X-EX)-(Y-EY)]^2 = 0$, 所以$Y = t_0(X-E(X)) + E(Y) = aX + b$. □

习题. 随机变量X与Y独立, 且$Var(X)=6$和$Var(Y)=3$, 求$Var(2X \pm Y) =?$

习题. 随机变量$X \sim P(2)$, $Y \sim \mathcal{N}(-2,4)$, X与Y独立, 则$E[(X-Y)^2] =?$

根据性质 5.5可知

$$\frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \le 1.$$

等号成立的充要条件是X与Y存在线性相关. 上式一定程度上反应了随机变量X和Y的线性相关程度, 由此引入一个新概念: 相关系数.

定义5.10. 设X和Y为二维随机变量, 如果$Var(X), Var(Y)$存在且不为0, 则称

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

为X与Y的相关系数, 简记ρ.

注1) 这里使用相关系数而不是$Cov(X,Y)$, 主要是规范$|\rho_{XY}| \le 1$, $Cov(X,Y)$受数值大小影响.

注2) 相关系数$|\rho_{XY}| \le 1$: 若$\rho > 0$, X与Y正相关; 若$\rho < 0$, X与Y负相关; $|\rho_{XY}| = 1$的充要条件为X与Y有线性关系$Y = aX + b$. 本质上ρ_{XY}刻画了X, Y的线性相关程度, 又称为"线性相关系数".

注3) 相关系数$\rho = 0$称X与Y不相关(线性不相关). 独立$\Rightarrow$不相关, 不相关$\not\Leftarrow$ 独立.

注4) 随机变量X与Y不相关, 仅表示X与Y之间无线性关系, 还可能存在其他关系. 例如: $X \sim U[-\frac{1}{2}, \frac{1}{2}]$, $Y = cosX$. 易有$E(X) = 0$,

高阶相关性.

$$\begin{aligned} Cov(X,Y) &= E[(X-E(X))(Y-E(Y))] \\ &= E[X\cdot\cos(X) - XE(\cos(X))] = E[X\cdot\cos(X)] = \int_{-\frac{1}{2}}^{\frac{1}{2}} x\cdot\cos(x)dx = 0. \end{aligned}$$

等价性条件)

定理5.15. 随机变量X, Y不相关的等价性如下:

$$\rho_{XY}=0 \Longleftrightarrow Cov(X,Y)=0 \Longleftrightarrow E(X,Y)=E(X)E(Y) \Longleftrightarrow Var(X\pm Y)=Var(X)\pm Var(Y).$$

二维正态分布:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$$

1) $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$.

2) $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$, 参数$\rho$为$X_1$与$X_2$的相关系数.

3) X_1与X_2独立$\Longleftrightarrow \rho = 0$(不相关).

① 1) 2) 都对 二维正态分布 成立.

② 由此，可将二维正态分布的协差写为 $\begin{pmatrix} Cov(V_1,V_1) & Cov(V_1,V_2) \\ Cov(V_1,V_2) & Cov(V_2,V_2) \end{pmatrix}$

习题. 随机变量(X, Y)联合概率密度为

$$f(x,y)=\begin{cases} \frac{1}{8}(x+y) & 0 \le x \le 2, 0 \le y \le 2 \\ 0 & \text{其它} \end{cases}$$

求$Cov(X,Y), Var(X+Y)$.

习题. 随机变量$X \sim \mathcal{N}(\mu, \sigma^2)$, $Y \sim \mathcal{N}(\mu, \sigma^2)$, 且$X$与$Y$相互独立. 求$Z_1 = \alpha X + \beta Y$和$Z_2 = \alpha X - \beta Y$的相关系数($\alpha, \beta$不全为0).

5.5.3 随机向量的数学期望与协方差阵

定义5.11. 设$X = (X_1, X_2, \cdots, X_n)'$, 则随机向量的期望

☆ $$E(X) = (E(X_1), E(X_2), \cdots, E(X_n))',$$

称随机变量X的协方差矩阵为

☆ $$Cov(X) = \Sigma \begin{pmatrix} Cov(X_1,X_1) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & \cdots & Cov(X_2,X_n) \\ \vdots & & \vdots \\ Cov(X_n,X_1) & \cdots & Cov(X_n,X_n) \end{pmatrix}.$$

定理5.16. 随机变量X的协方差阵是对称非负定的矩阵.

多维正态分布$X = (X_1, X_2, \ldots, X_n)' \sim N(\mu, \Sigma)$, 则有

$$\mu = (E[X_1], E[X_2], \ldots, E[X_n])' \qquad ☆\ \Sigma = [Cov(X_i, X_j)]_{n\times n}$$

其概率密度为:

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)).$$

多维随机变量具有如下性质:

1) $X=(X_1,X_2,\cdots,X_n)'\sim\mathcal{N}(\mu,\Sigma)$, 则每个变量$X_i$的边缘分布是正态分布.

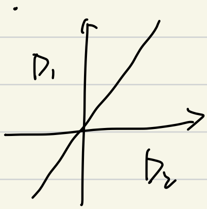
2) $X=(X_1,X_2,\cdots,X_n)'\sim\mathcal{N}(\mu,\Sigma)$, 则$\sum_{i=1}^{n}a_iX_i$是正态分布.

3) $X=(X_1,X_2,\cdots,X_n)'\sim\mathcal{N}(\mu,\Sigma)$, 则$X_1,X_2,\cdots,X_n$相互独立$\Longleftrightarrow X_1,X_2,\cdots,X_n$相互不相关.

例： $X \sim N(0,1)$ $Y \sim N(0,1)$ 且 X 与 Y 独立。求 $E[\min(X,Y)]$

$\because$ x 与 y 独立，故 $f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$

$\therefore$ $E[\min(X,Y)] = \frac{1}{2\pi} \iint\limits_{D_1} x e^{-\frac{x^2+y^2}{2}} dxdy + \frac{1}{2\pi} \iint\limits_{D_2} y e^{-\frac{x^2+y^2}{2}} dxdy$



$$= \frac{1}{\pi} \iint\limits_{D_1} x e^{-\frac{x^2+y^2}{2}} dxdy$$

$$= \frac{1}{\pi} \int_{-\infty}^{+\infty} dy \int_{-\infty}^{y} x e^{-\frac{x^2+y^2}{2}} dx$$

$$= -\frac{1}{\pi} \int_{-\infty}^{+\infty} dy \cdot e^{-\frac{x^2+y^2}{2}} \Big|_{-\infty}^{y}$$

$$= -\frac{1}{\pi} \int_{-\infty}^{+\infty} e^{-y^2} dy$$

$$= -\frac{1}{\sqrt{\pi}}$$ $\longrightarrow$ 就是负的

例： X 与 Y 独立，$X \sim P(2)$，$Y \sim N(-2,4)$，$E((X-Y)^2)$。

$$E[(X-Y)^2] = E(X^2) + E(Y^2) - 2E(X)E(Y)$$

$$= E(X^2) - 2E(X)E(Y) + E(Y^2)$$

可用方差计算

5.6 条件分布与条件期望

前面学过随机事件的条件概率, 即在事件B发生的条件下事件A发生的条件概率

$$P(A|B) = P(AB)/P(B).$$

相关概念可推广到随机变量: 在给定随机变量Y取值条件下, 求X的概率分布, 即条件分布.

首先考虑离散型随机变量:

定义5.12. 设(X,Y)为二维离散型随机变量, 其分布列为$\{p_{ij}\}$. 若Y的边缘分布$P(Y=y_j)=p_{\cdot j}>0$, 称

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

为在$Y=y_j$条件下随机变量X的条件分布列.

类似可定义在$X=x_i$条件下随机变量Y的条件分布列. 条件分布是一种概率分布, 具有分布的一切性质, 例如:

- 非负性: $P(X=x_i|Y=y_j)\geq 0$;
- 规范性: $\sum_{i=1}^{\infty} P(X=x_i|Y=y_j)=1$.

例5.14. 一射手进行射击, 击中目标的概率为p, 射击进行到击中两次目标为止. X表示首次击中目标所进行的射击次数, Y表示第二次射中目标所进行的射击次数, 求X和Y的联合分布和条件分布.

解. 随机变量$X=m$表示首次击中目标射击了m次, $Y=n$表示第二次次击中目标射击了n次, 则X和Y的联合分布列为:

$$P\{X=m,Y=n\}=p^2(1-p)^{n-2}\qquad 1\leq m<n<\infty.$$

由此可得X的边缘分布列为

$$P\{X=m\}=\sum_{n=m+1}^{\infty}P\{X=m,Y=n\}=\sum_{n=m+1}^{\infty}p^2(1-p)^{n-2}=p^2\frac{(1-p)^{m-1}}{1-(1-p)}=p(1-p)^{m-1}.$$

同理得到随机变量Y的边缘分布列为

$$P\{Y=n\}=\sum_{m=1}^{n-1}P\{X=m,Y=n\}=\sum_{m=1}^{n-1}p^2(1-p)^{n-2}=(n-1)p^2(1-p)^{n-2}(n=2,3,\cdots).$$

因此, 当$n=2,3,\cdots$时, 随机变量X在$Y=n$条件下的分布列为:

$$P\{X=m|Y=n\}=\frac{P\{X=m,Y=n\}}{P\{Y=n\}}=\frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}}=\frac{1}{n-1}\quad m=1,2,\cdots,n-1.$$

当$m=1,2,3,\cdots$时, 随机变量Y在$X=m$条件下的分布列为:

$$P\{Y=n|X=m\}=\frac{P\{X=m,Y=n\}}{P\{X=m\}}=\frac{p^2(1-p)^{n-2}}{p(1-p)^{m-1}}=p(1-p)^{n-m-1}\quad n=m+1,m+2,\cdots$$

□

对于连续型随机变量(X,Y), 对任意x,y, 有$P(X=x)=0$和$P(Y=y)=0$成立, 因此不能利用条件概率得到条件分布. 下面给出条件概率的定义:

定义5.13. 设连续型X和Y的联合概率密度为$f(x,y)$, 随机变量Y的边缘概率密度为$f_Y(y)$, 对任意y, $f_Y(y)>0$, 称

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

($f_{X|Y}(x|y)$) 本质上仍为二元函数

在$Y=y$条件下随机变量X的条件概率密度. 称

$$F_{X|Y}(x|y)=P\{X\le x|Y=y\}=\int_{-\infty}^{x} f_{X|Y}(u|y)du$$

为$Y=y$条件下X的条件分布函数.

类似定义在$X=x$条件下随机变量Y的条件概率密度和分布函数分别为

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)} \qquad F_{Y|X}(y|x)=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv.$$

下面解释条件概率的含义, 这里以$f_{X|Y}=\frac{f(x,y)}{f_Y(y)}$为例, 首先分布函数有

$$F_{X|Y}(x|y)=P\{X\le x|Y=y\}=\lim_{\epsilon\to0+}P\{X\le x|y\le Y\le y+\epsilon\}=\lim_{\epsilon\to0+}\frac{P\{X\le x,y\le Y\le y+\epsilon\}}{P\{y\le Y\le y+\epsilon\}}.$$

根据积分中值定理有 → 考虑 $Y\in[y,y+\delta]$ 的一个小区域

$$\frac{P\{X\le x,y\le Y\le y+\epsilon\}}{P\{y\le Y\le y+\epsilon\}}=\frac{\int_{-\infty}^{x}\int_{y}^{y+\epsilon}f(u,v)dudv}{\int_{y}^{y+\epsilon}f_Y(u)dv}=\frac{\epsilon\int_{-\infty}^{x}f(u,y+\theta_1\epsilon)du}{\epsilon f_Y(y+\theta_2\epsilon)}$$

其中$\theta_1,\theta_2\in(0,1)$. 当$\epsilon\to0^+$时, 有

$$F_{X|Y}(x|y)=P\{X\le x|Y=y\}=\lim_{\epsilon\to0+}P\{X\le x|y\le Y\le y+\epsilon\}=\frac{\int_{-\infty}^{x}f(u,y)du}{f_Y(y)}=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du,$$

由此可得条件概率密度$f_{X|Y}(x|y)=f(x,y)/f_Y(y)$. 下面给出条件概率的性质:

引理5.2 (乘法公式)**.** 对于随机变量X和Y, 有

$$\begin{aligned} f(x,y) &= f_X(x)f_{Y|X}(y|x), \quad (f_X(x)>0),\\ f(x,y) &= f_Y(y)f_{X|Y}(x|y), \quad (f_Y(y)>0). \end{aligned}$$

引理5.3. 如果随机变量X和Y相互独立, 则有

$$f_{Y|X}(y|x)=f_Y(y) \qquad f_{X|Y}(x|y)=f_X(x).$$

由条件概率可判别随机变量(X,Y)是否独立. 下面看几个例子:

例5.15. 设(X,Y)的概率密度

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

求$P(X>1|Y=y)$.

解. 首先求解随机变量Y的边缘分布为

$$f_Y(y)=\int_0^{+\infty}\frac{e^{-x/y}e^{-y}}{y}dx=e^{-y}[-e^{-\frac{x}{y}}]_0^{+\infty}=e^{-y}\qquad (y>0).$$

进而得到在$Y=y$条件下X的条件概率密度为$f_{X|Y}(x|y)=e^{-x/y}/y$. 最后得到

$$P(X>1|Y=y)=\int_1^{\infty}\frac{e^{-x/y}}{y}dx=-e^{-x/y}|_1^{\infty}=e^{-\frac{1}{y}}.$$

这里应该解读为条件密度

□

例5.16. 设$X\sim U(0,1)$, 当观察到$X=x$的条件下, 随机变量$Y\sim U(x,1)$. 求Y的概率密度.

解. 根据题意可知$X\sim U(0,1)$, 在随机变量$X=x$的条件下$Y\sim U(x,1)$, 即$f_{Y|X}(y|x)=\frac{1}{1-x}$. 根据条件概率乘积公式有

$$f(x,y)=f_X(x)f_{Y|X}(y|x)=\begin{cases}\frac{1}{1-x} & 0<x<y<1,\\ 0 & \text{其它}.\end{cases}$$

根据联合分布求解随机变量X的边缘分布

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\int_0^y\frac{1}{1-x}dx=-\ln(1-y) & y>0,\\ 0 & \text{其它}.\end{cases}$$

□

习题. 设(X,Y)的概率密度为

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

求$f_{X|Y}(x|y)$.

定理5.17. 多维正太分布的条件分布是正太分布.

(正态分布的边缘分布、条件分布、和函数分布也是正态分布)

Proof. 为简单起见, 这里仅仅给出二维正太分布的详细证明. 设$X=(X_1,X_2)\sim\mathcal{N}(\mu,\Sigma)$, $\mu=(\mu_1,\mu_2)$, $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$, 在$X_2=x_2$的条件下证明$X_1\sim\mathcal{N}(\mu_1-\sigma_1^2\rho(x_2-\mu_2)/\sigma_2^2,\sigma_1^2(1-\rho^2))$. 首先给出二维正太分布的联合分布

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}]}.$$

以及随机变量X_2的边缘分布为$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. 于是得到条件概率

$$\begin{aligned} f_{X_1|X_2}(x_1|x_2) &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \rho^2\frac{(x_2-\mu_2)^2}{\sigma_2^2}]} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{x_1-\mu_1}{\sigma_1} + \frac{\rho(x_2-\mu_2)}{\sigma_2}]^2} \\ &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{\sigma_1^2(1-\rho)^2}[x_1-\mu_1+\sigma_1^2\rho^2(x_2-\mu_2)/\sigma_2^2]^2}. \end{aligned}$$

因此在$X_2 = x_2$条件下, $X_1 \sim \mathcal{N}(\mu_1 - \sigma_1^2\rho(x_2-\mu_2)/\sigma_2^2, \sigma_1^2(1-\rho^2))$. □

5.6.1 条件期望

定义5.14. 若X, Y为离散型随机变量,

→条件概率与X的加权和

$$E[X|Y=y] = \sum_{i=1}^{\infty} x_i P(X = x_i|Y=y).$$

若X, Y为连续型随机变量,

$$E[X|Y=y] = \int_{-\infty}^{+\infty} x f(x|y) dx.$$

→条件概率密度与X的加权积分

注: $E[X|Y=y]$是y的函数. 对条件期望, 有如下重要性质:

定理5.18. 对离散随机变量$X_1, X_2, \cdots, X_n$及常数$c_1, c_2, \cdots, c_n$, 有

(线性性质)

$$E[\sum_{i=1}^{n} c_i X_i|Y=y] = \sum_{i=1}^{n} c_i E[X_i|Y=y].$$

定理5.19 (全期望公式, law of total expectation). 对随机变量X和$\boxed{\text{事件}A}$有

$$E[X] = E[X|A]P(A) + E[X|\bar{A}](1 - P(A))$$

其中事件$\bar{A}$为事件A的补.

Proof. 利用全概率公式有

$P(A) = P(A|B) + P(A|\bar{B})$, 好处是是$P(A)$的概率难于直接求解时, 可引用$B$, B提供额外条件

$$\begin{aligned} E[X] &= \sum x_i P(X = x_i) = \sum_i x_i[P(X = x_i, A) + P(X = x_i, \bar{A})] \\ &= \sum_i x_i P(X = x_i|A)P(A) + \sum_i x_i P(X = x_i|\bar{A})P(\bar{A}) \\ &= P(A)\sum_i x_i P(X = x_i|A) + P(\bar{A})\sum_i x_i P(X = x_i|\bar{A}) \\ &= P(A)E[X|A] + P(\bar{A})E[X|\bar{A}]. \end{aligned}$$

□

全期望公式对应于全概率公式的期望版本, 在很多应用中有重要的性质。该定理有一个关于随机变量的定理:

推广: 若$A_1, A_2, \cdots A_n$是Ω的一个划分

则$E[X] = \sum_{i=1}^{n} E[X|A_i] \cdot P(A_i)$

定理5.20. 对随机变量X, Y, 有

$E[X] = E_Y[E_X[X|Y]] = \sum_{Y=y_i} P[Y=y_i]E[X|Y=y_i]$

$$E[X] = E[E(X|Y)] = \sum_{Y=y_j} P[Y=y_j]E[X|Y=y_j].$$

Proof. 利用全概率公式有

$$\begin{aligned}
E[X] &= \sum_i x_i P(X=x_i) = \sum_i \sum_j x_i P(X=x_i, Y=y_j) \\
&= \sum_i \sum_j x_i P(X=x_i, Y=y_j) \\
&= \sum_j P(Y=y_j) \sum_i x_i \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} \\
&= \sum_j P(Y=y_j) \sum_i x_i P(X=x_i | Y=y_j) \\
&= \sum_j P(Y=y_j) E[X|Y=y_j].
\end{aligned}$$

□

通过全期望公式, 可以证明Markov不等式:

定理5.21 (Markov不等式)**.** 设随机变量$X \geq 0$, 对任意$\epsilon > 0$, 有

$$P(X \geq \epsilon) \leq \frac{E(X)}{\epsilon}.$$

Proof. 利用全期望公式考虑随机事件$X \geq \epsilon$, 有

$$E[X] = E[X|X \geq \epsilon]P(X \geq \epsilon) + E[X|X \leq \epsilon]P(X \leq \epsilon) \geq Pr(X \geq \epsilon)\epsilon$$

从而完成证明. □

利用Markov不等式可以推导Chebyshev不等式:

定理5.22 (Chebyshev不等式)**.** 设随机变量X的均值为μ, 则

$$P(|X-\mu| > \epsilon) \leq \frac{Var(X)}{\epsilon^2}.$$

Proof. 根据Markov不等式有

$$P(|X-\mu| > \epsilon) = P((X-\mu)^2 \geq \epsilon^2) \leq \frac{E(X-\mu)^2}{\epsilon^2} = \frac{Var(X)}{\epsilon^2}.$$

□

推论：对$\forall x$，若$g(x)$单调增，则

$$P(X \geq \varepsilon) = P(g(x) \geq g(\varepsilon)) \leq \frac{E[g(x)]}{g(\varepsilon)}$$

$$P(X-\mu \geq \varepsilon) \leq \frac{1}{1+\left(\frac{\varepsilon}{\sqrt{Var(X)}}\right)^2} = \frac{Var(X)}{Var(X)+\varepsilon^2}$$