Probability and Statistics
Tutorial 8

Siyi Wang

Southern University of Science and Technology

11951002@mail.sustech.edu.cn

November 10, 2020

Outline

Review

1. Function of Random Vector (X, Y)
   - Discrete Case: $Z = h(X, Y)$,
     $P(Z = z) = P(h(X, Y) = z) = \sum_{(x,y):h(x,y)=z} P(X = x, Y = y)$.
   - Continuous Case: $Z = h(X, Y)$
     - Method 1
       - Compute $F_Z(z) = P(Z \leq z)$
       - Compute $f_Z(z) = F'_Z(z)$
     - Method 2
       - Consider $(Z, Y) = (h(X, Y), Y)$
       - Then, using Jacobian we have

命题 3.6.1 在上述假设下，对于某些 (x, y) 满足 $u = g_1(x, y), v = g_2(x, y)$ 的 (u, v) 点，U 和 V 的联合密度是

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J^{-1}(h_1(u, v), h_2(u, v))|$$

否则，取 0.

Review

对于一般情形，假设 X 和 Y 是连续型随机变量，通过如下变换投影到 U 和 V 上：

$$u = g_1(x, y)$$

$$v = g_2(x, y)$$

并且存在逆变换

$$x = h_1(u, v)$$

$$y = h_2(u, v)$$

假设 g_1 和 g_2 具有连续偏导数，并且对任意的 x 和 y, 雅可比行列式

$$J(x, y) = \det \begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} = \left(\frac{\partial g_1}{\partial x}\right)\left(\frac{\partial g_2}{\partial y}\right) - \left(\frac{\partial g_2}{\partial x}\right)\left(\frac{\partial g_1}{\partial y}\right) \neq 0$$

Review

2. Convolution Formula
   - For two independent r.v.s X and Y, ve the pdf of $x+y$ is $f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$.
3. Order Statistics $X_{(1)}, X_{(2)}, .., X_{(n)}$
   - $F_{X_{(n)}}(t) = (F_X(t))^n$
   - $F_{X_{(1)}}(s) = 1 - (1 - F_X(s))^n$
   - $f_{X_{(1)}, X_{(n)}}(s, t) = n(n-1)f(s)f(t)[F(t) - F(s)]^{n-2}$
   - $f_{X_{(k)}} = \frac{n!}{(k-1)!(n-k)!} f(x)[F(x)]^{k-1}[1 - F(x)]^{n-k}$.

Review

4. Expectation

- Discrete Case: $EX = \sum_{i=1}^{\infty} n_i P(X = n_i)$.
  $Eh(X) = \sum_{i=1}^{\infty} h(n_i) P(X = n_i)$.
- Continuous Case: $EX = \int_{-\infty}^{\infty} x f_X(x) dx$. $Eh(X) = \int_{-\infty}^{\infty} h(x) f_X(x) dx$.
- $E[I_A(X)] = P(X \in A)$.

Homework

43. 令 U_1 和 U_2 是 $[0,1]$ 上相互独立的均匀随机变量. 计算并画出 $S = U_1 + U_2$ 的密度函数.

Solution

43. Solution. $f_{U_1}(u_1) = \mathbb{1}_{\{0<u_1<1\}}$, $f_{U_2}(u_2) = \mathbb{1}_{\{0<u_2<1\}}$.

$$f_U(u) = \int_{-\infty}^{+\infty} f_{U_1}(x) f_{U_2}(u-x)\,dx$$

$$= \begin{cases} 0, & u \leq 0 \\ \int_0^u dx = u, & u \in (0,1] \\ \int_{-1+u}^{1} dx = 2-u, & u \in (1,2] \\ 0, & u > 2. \end{cases}$$

44. 假设 X 和 Y 是独立的离散随机变量，并且取值 0,1 和 2 时的概率都是 $\frac{1}{3}$. 计算 $X+Y$ 的频率函数.

Homework

Solution

44. Solution,

$X+Y$	0	1	2	3	4
$\mathbb{P}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

51. 令 X 和 Y 具有联合密度函数 $f(x,y)$，$Z=XY$. 证明：Z 的密度函数为

$$f_Z(z)=\int_{-\infty}^{\infty} f\left(y,\frac{z}{y}\right)\frac{1}{|y|}\mathrm{d}y$$

Solution

51. Proof.

$$\begin{cases} Z = XY \\ X = X \end{cases} \rightsquigarrow \begin{cases} X = X \\ Y = Z/X \end{cases}$$

$$f_{Z,X}(z,x) = \left| \det \begin{pmatrix} 0 & 1 \\ \frac{1}{x} & -\frac{z}{x^2} \end{pmatrix} \right| f_{X,Y}\left(x, \frac{z}{x}\right)$$

$$= \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right)$$

Hence, $f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|y|} f_{X,Y}\left(y, \frac{z}{y}\right) dy$. $\square$

52. 计算两个独立均匀随机变量商的密度.

Solution

52. Solution. $X \sim U(a,b) \quad Y \sim U(c,d)$

$$\begin{cases} Z = \frac{X}{Y} \\ Y = Y \end{cases} \Rightarrow \begin{cases} X = ZY \\ Y = Y \end{cases}$$

$$f_{Z,Y}(z,y) = \left| \det \begin{pmatrix} y & z \\ 0 & 1 \end{pmatrix} \right| f_{X,Y}(zy, y)$$

$$= |y| f_{X,Y}(zy,y) = |y| \mathbb{1}_{\{zy \in (a,b)\}} \mathbb{1}_{\{y \in (c,d)\}}$$

We consider special case $a=c=0$, $b=d=1$ here.

$$f_Z(z) = \begin{cases} \int_0^1 y\,dy = \frac{1}{2}, & z \in (0,1) \\ \int_0^{\frac{1}{z}} y\,dy = \frac{1}{2z^2}, & z \in (1,\infty) \\ 0, & \text{otherwise.} \end{cases}$$

Homework

57. 假设 Y_1 和 Y_2 服从二元正态分布，具有参数 $\mu_{Y_1} = \mu_{Y_2} = 0, \sigma^2_{Y_1} = 1, \sigma^2_{Y_2} = 2$，且 $\rho = 1/\sqrt{2}$. 找出线性变换 $x_1 = a_{11}y_1 + a_{12}y_2, x_2 = a_{21}y_1 + a_{22}y_2$，使得 x_1 和 x_2 是独立的标准正态随机变量. (提示：见 3.6.2 节的例 3.6.2.3.)

Solution

57. Solution. $f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{2\pi}\exp\left[-\left(y_1^2-y_1y_2+\frac{y_2^2}{2}\right)\right]$

We have
$$\begin{cases} y_1=\dfrac{1}{a_{11}a_{22}-a_{12}a_{21}}(a_{22}x_1-a_{12}x_2) \\ y_2=\dfrac{1}{a_{11}a_{22}-a_{12}a_{21}}(-a_{21}x_1+a_{11}x_2) \end{cases}$$

$$|\det J|=\frac{1}{|a_{11}a_{22}-a_{12}a_{21}|}$$ Let $D=|a_{11}a_{22}-a_{12}a_{21}|$

$$f_{X_1,X_2}(x_1,x_2)=(2\pi)^{-1}D\exp\left[-D^2\left[(a_{22}x_1-a_{12}x_2)^2-(a_{22}x_1-a_{12}x_2)(-a_{21}x_1+a_{11}x_2)+\frac{1}{2}(-a_{21}x_1+a_{11}x_2)^2\right]\right]=(2\pi)^{-1}\exp\left[-\left(\frac{1}{2}x_1^2+\frac{1}{2}x_2^2\right)\right]$$

That is,
$$\begin{cases} D=|a_{11}a_{22}-a_{12}a_{21}|=1 & ① \\ a_{22}^2+a_{22}a_{21}+\frac{1}{2}a_{21}^2=\frac{1}{2} & ② \\ a_{12}^2+a_{12}a_{11}+\frac{1}{2}a_{11}^2=\frac{1}{2} & ③ \\ -2a_{22}a_{12}-a_{22}a_{11}-a_{12}a_{21}-a_{11}a_{21}=0 & ④ \end{cases}$$

Homework

Solution

$$EXX^T = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$Y = AX$$

$$\mathbb{E}\, A X X^T A^T = A \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} A^T$$

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ \hline 1 & 0 & & \\ 0 & 1 & & \end{array}\right) \sim\!\!> \left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ \hline 1 & -1 & & \\ 0 & 1 & & \end{array}\right)$$

$$A = \boxed{\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}}$$

Homework

1. 假设 X 和 Y 是两个独立的随机变量，服从标准正态分布 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。

Solution

补1.
$$\begin{cases} U = X+Y \\ V = X-Y \end{cases} \Rightarrow \begin{cases} X = \frac{1}{2}(U+V) \\ Y = \frac{1}{2}(U-V) \end{cases}$$

$$f_{U,V}(u,v) = \frac{1}{2} \cdot \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2} \cdot \frac{1}{4}\left[(u+v)^2 + (u-v)^2\right]\right)$$

$$= \frac{1}{4\pi} \exp\left(-\frac{u^2}{4} - \frac{v^2}{4}\right)$$

Hence, U, V 独立 无关

$$f_U(u) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{u^2}{4}\right)$$

$$f_V(v) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{v^2}{4}\right)$$

Homework

2. 设二维连续随机变量(X,Y)的概率密度为

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

(1)求边缘密度函数;

(2)Z=2X-Y 的概率密度函数;

(3)P(Y<1/2|X<1/2).

ENI

Solution

补2. (1)

$$f_X(x)=\begin{cases}\int_0^{2x} dy = 2x, & x\in(0,1)\\ 0, & \text{otherwise}\end{cases}$$

$$f_Y(y)=\begin{cases}\int_{\frac{y}{2}}^{1} dx = 1-\frac{y}{2}, & y\in(0,2)\\ 0, & \text{otherwise.}\end{cases}$$

(2) $\mathbb{P}(2X-Y\leqslant z)$

$$=\begin{cases}0, & z\leqslant 0\\ \int_0^{\frac{z}{2}}\int_0^{2x} dydx+\int_{\frac{z}{2}}^{1}\int_{2x-z}^{2x} dydx = z-\frac{z^2}{4}, & z\in(0,2]\\ 0, & z>2\end{cases}$$

Solution

Then,

$$f_Z(z) = \left(1 - \frac{z}{2}\right) I_{(0,2)}(z).$$

(3) $$\mathbb{P}\left(Y < \tfrac{1}{2} \mid X < \tfrac{1}{2}\right) = \frac{\mathbb{P}\left(Y < \frac{1}{2}, X < \frac{1}{2}\right)}{\mathbb{P}\left(X < \frac{1}{2}\right)} \quad (*)$$

$$\mathbb{P}\left(Y < \tfrac{1}{2}, X < \tfrac{1}{2}\right) = S_{\square} = \left(\tfrac{1}{2} + \tfrac{1}{4}\right) \times \tfrac{1}{2} \times \tfrac{1}{2} = \frac{3}{16}$$

$$\mathbb{P}\left(X < \tfrac{1}{2}\right) = S_{\Delta} = \tfrac{1}{2} \times 1 \times \tfrac{1}{2} = \frac{1}{4}$$

$$(*) = \frac{3}{4}.$$

Homework

70. 令 $X_1, X_2, \cdots, X_n$ 是独立的连续随机变量，每个变量具有累积分布函数 F. 证明：$X_{(1)}$ 和 $X_{(n)}$ 的联合累积分布函数是

$$F(x, y) = F^n(y) - [F(y) - F(x)]^n, \quad x \leqslant y$$

Solution

70. Proof. Consider $(*)$: $\mathbb{P}(X_{(1)} > x, X_{(n)} \leq y)$

On the one hand,

$$(*) = \mathbb{P}(x < X_i \leq y, \; i=1,\cdots,n)$$
$$= (F(y) - F(x))^n$$

On the other hand,

$$(*) = \mathbb{P}(X_{(n)} \leq y) - \mathbb{P}(X_{(n)} \leq y, X_{(1)} \leq x)$$
$$= \mathbb{P}(X_{(n)} \leq y) - F(x,y).$$

Since $\mathbb{P}(X_{(n)} \leq y) = \mathbb{P}(X_i \leq y, \; i=1,\cdots,n)$
$$= (F(y))^n.$$

Thus, $(F(y) - F(x))^n = (F(y))^n - F(x,y)$

i.e., $F(x,y) = (F(y))^n - (F(y) - F(x))^n$.

Homework

1. 从 1,2,3 中一次任取两个数，第一个数为 X，第二个为 Y， 记 Z=max(X,Y), 求 (X,Y) 和 (X,Z) 的联合频率函数和边缘频率函数。

Solution

补1. Solution.

$X \backslash Y$	1	2	3	$P(X)$
1	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
2	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
3	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
$P(Y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$X \backslash 8$	1	2	3	$P(X)$
1	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
2	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
3	0	0	$\frac{1}{3}$	$\frac{1}{3}$
$P(Y)$	0	$\frac{1}{3}$	$\frac{2}{3}$	

Homework

2. 设 X 和 Y 是两个相互独立的随机变量，服从 N(0,1)，令 Z=min(X,Y)，求 Z 的分布函数。

Solution

补2. $\mathbb{P}(Z > z) = (1 - F(z))^2$

$$F_Z(z) = \mathbb{P}(Z \le z) = 1 - (1 - F(z))^2$$

$$= 2F(z) - F(z)^2. \quad \square$$

1. St. Petersburg Paradox

The paradox [edit]

A casino offers a game of chance for a single player in which a fair coin is tossed at each stage. The initial stake begins at 2 dollars and is doubled every time heads appears. The first time tails appears, the game ends and the player wins whatever is in the pot. Thus the player wins 2 dollars if tails appears on the first toss, 4 dollars if heads appears on the first toss and tails on the second, 8 dollars if heads appears on the first two tosses and tails on the third, and so on. Mathematically, the player wins 2^k dollars, where k is a positive integer equal to the number of tosses. What would be a fair price to pay the casino for entering the game?

To answer this, one needs to consider what would be the average payout: with probability $\frac{1}{2}$, the player wins 2 dollars; with probability $\frac{1}{4}$ the player wins 4 dollars; with probability $\frac{1}{8}$ the player wins 8 dollars, and so on. The expected value is thus

$$\begin{aligned} E &= \frac{1}{2}\cdot 2 + \frac{1}{4}\cdot 4 + \frac{1}{8}\cdot 8 + \frac{1}{16}\cdot 16 + \cdots \\ &= 1 + 1 + 1 + 1 + \cdots \\ &= +\infty \,. \end{aligned}$$

Assuming the game can continue as long as the coin toss results in heads and in particular that the casino has unlimited resources, this sum grows without bound and so the expected win for repeated play is an infinite amount of money. Considering nothing but the expected value of the net change in one's monetary wealth, one should therefore play the game at any price if offered the opportunity. Yet, in published descriptions of the game, many people expressed disbelief in the result. Martin Robert quotes Ian Hacking as saying "few of us would pay even $25 to enter such a game" and says most commentators would agree.[4] The paradox is the discrepancy between what people seem willing to pay to enter the game and the infinite expected value.

Further Reading

1. St. Petersburg Paradox

Expected utility theory

The classical resolution of the paradox involved the explicit introduction of a utility function, an expected utility hypothesis, and the presumption of diminishing marginal utility of money. In Daniel Bernoulli's own words:

> The determination of the value of an item must not be based on the price, but rather on the utility it yields.... There is no doubt that a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain the same amount.

Using a utility function, e.g., as suggested by Bernoulli himself, the logarithmic function $u(x) = \ln(x)$ (known as "log utility"[2]), the expected utility of the lottery (for simplicity assuming an initial wealth of zero) becomes finite:

$$EU = \sum_{k=1}^{\infty} p_k \cdot u(2^{k-1}) = \sum_{k=1}^{\infty} \frac{\ln(2^{k-1})}{2^k} = \ln 2 = u(2) < \infty$$

(This particular utility function suggests that the lottery is as useful as 2 dollars.)
Before Daniel Bernoulli published, in 1728, another Swiss mathematician, Gabriel Cramer, found already parts of this idea (also motivated by the St. Petersburg Paradox) in stating that the mathematicians estimate money in proportion to its quantity, and men of good sense in proportion to the usage that they may make of it.

Thank you!