

广义边缘密度函数相关理论及应用

任浩林 袁永生 王启明 张志分

(河海大学理学院, 江苏 南京 210098)

摘要 对二维连续型随机变量的联合密度函数,在一定要求下变上限积分得到新函数,依据密度函数性质进行规范化,从而构造出一个广义的边缘密度函数.分析该边缘密度函数广义形式与通常形式之间的关系,探讨密度函数广义形式下的多个性质.针对具体的广义形式,得出了多个新结果,并将密度函数广义形式推广到 n 维,最后讨论了这一广义形式在舍选抽样中的应用.

关键词 广义形式;边缘密度函数;条件密度;舍选抽样法

中图分类号 O211.9 **文献标识码** A **文章编号** 1000-198X(2006)03-0341-04

n 维连续型随机变量的一维边缘密度函数是联合密度函数在 R^{n-1} 上的积分,这一边缘密度函数满足一维随机变量密度函数的各种性质^[1-2].如果让联合密度函数在 R^{n-1} 中的指定区域内变上限积分,则可得到新的函数,将这一函数规范化得到边缘密度函数的广义形式,本文将讨论其特性和应用.

1 问题的推导

1.1 广义边缘密度函数定义

设 $f(x, y)$ 为二维连续型随机变量 (X, Y) 的联合密度函数,构造如下积分

$$f(x) = \int_{-\infty}^{H(x)} f(x, y) dy \quad f(y) = \int_{-\infty}^{H(y)} f(x, y) dx$$

其中 $H(x), H(y)$ 为黎曼可积函数,且使 $f(x), f(y)$ 均不恒为零(大于零).定义

$$g(x) = \frac{f(x)}{\int_{-\infty}^{+\infty} f(x) dx} \quad g(y) = \frac{f(y)}{\int_{-\infty}^{+\infty} f(y) dy}$$

称 $g(x), g(y)$ 分别为随机变量 (X, Y) 关于 X 和 Y 的广义边缘密度函数,满足密度函数的性质,并称

$$\alpha(x) = \int_{-\infty}^x \frac{f(x)}{\int_{-\infty}^{+\infty} f(x) dx} dx \quad \alpha(y) = \int_{-\infty}^y \frac{f(y)}{\int_{-\infty}^{+\infty} f(y) dy} dy$$

为随机变量 (X, Y) 关于 X 和 Y 的广义边缘分布函数.

1.2 广义边缘密度函数与通常的边缘密度函数的关系

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{+\infty} f(x, y) dy = \int_{-\infty}^{H(x)} f(x, y) dy + \int_{H(x)}^{+\infty} f(x, y) dy = \\ &g(x) \int_{-\infty}^{+\infty} f(x) dx + \int_{H(x)}^{+\infty} f(x, y) dy \end{aligned} \quad (1)$$

同理

$$f_Y(y) = g(y) \int_{-\infty}^{+\infty} f(y) dy + \int_{H(y)}^{+\infty} f(x, y) dx \quad (2)$$

若令 $y = -z$, 有

$$\int_{H(x)}^{+\infty} f(x, y) dy = \int_{-\infty}^{-H(x)} f(x, -z) dz = \int_{-\infty}^{-H(x)} f(x, -y) dy \quad (3)$$

进一步,若 $f(x, y)$ 关于 y 为偶函数,即 $f(x, -y) = f(x, y)$, 此时有

$$\begin{aligned}
 \int_{H(x)}^{+\infty} f(x, y) dy &= \int_{-\infty}^{-H(x)} f(x, y) dy \\
 f_X(x) &= f(x) + \int_{-\infty}^{-H(x)} f(x, y) dy = f(x) + \left[\int_{-\infty}^{H(x)} f(x, y) dy - \int_{-H(x)}^{H(x)} f(x, y) dy \right] = \\
 &= f(x) + \left[f(x) - \int_{-H(x)}^{H(x)} f(x, y) dy \right] = 2g(x) \int_{-\infty}^{+\infty} f(x) dx - 2 \int_0^{H(x)} f(x, y) dy \quad (4)
 \end{aligned}$$

1.3 广义边缘分布函数与通常条件概率的关系

设二维连续型随机变量 (X, Y) 的联合密度函数为 $f(x, y)$ 则有

$$P(X \leq z | Y \leq H(x)) = \frac{P(X \leq z, Y \leq H(x))}{P(Y \leq H(x))} = \frac{\int_{-\infty}^z \int_{-\infty}^{H(x)} f(x, y) dy dx}{\int_{-\infty}^{+\infty} \int_{-\infty}^{H(x)} f(x, y) dy dx} = \int_{-\infty}^z \frac{f(x)}{\int_{-\infty}^{+\infty} f(x) dx} dx = G(z)$$

因此有

$$P(X \leq z | Y \leq H(x)) = G(z) \quad (5)$$

1.4 广义边缘密度的一些性质

1.4.1 随机变量 X 和 Y 独立式的性质

随机变量 X 和 Y 独立的通常定义:联合分布 $F(x, y)$ 等于 X 和 Y 的通常边缘分布 $F_X(x)$ 和 $F_Y(y)$ 的乘积.连续型随机变量 X 和 Y 相互独立的等价定义是联合密度几乎处处等于通常边缘密度的乘积^[3].即

$$f(x, y) \stackrel{\text{a.e.}}{=} f_X(x) f_Y(y)$$

当随机变量 X 和 Y 相互独立时,按照1.1中广义边缘密度函数定义有

$$\begin{aligned}
 f_X(x) &= \int_{-\infty}^{H(x)} f(x, y) dy = \int_{-\infty}^{H(x)} f_X(x) f_Y(y) dy = f_X(x) \int_{-\infty}^{H(x)} f_Y(y) dy = \\
 &= f_X(x) F_Y(H(x)) = g(x) \int_{-\infty}^{+\infty} f(x) dx
 \end{aligned}$$

而

$$\begin{aligned}
 \int_{-\infty}^{+\infty} f(x) dx &= \int_{-\infty}^{+\infty} \int_{-\infty}^{H(x)} f(x, y) dy dx = \int_{-\infty}^{+\infty} f_X(x) \int_{-\infty}^{H(x)} f_Y(y) dy dx = \\
 &= \int_{-\infty}^{+\infty} f_X(x) F_Y(H(x)) dx = E_X[F_Y(H(x))]
 \end{aligned}$$

则

$$g(x) E_X[F_Y(H(x))] = f_X(x) F_Y(H(x)) \quad (6)$$

同理可得

$$g(y) E_Y[F_X(H(y))] = f_Y(y) F_X(H(y)) \quad (7)$$

这里 $E_X[F_Y(H(x))]$ 为 $F_Y(H(x))$ 的期望, $E_Y[F_X(H(y))]$ 为 $F_X(H(y))$ 的期望.则

$$g(x) g(y) E_X[F_Y(H(x))] E_Y[F_X(H(y))] = f_X(x) f_Y(y) F_Y(H(x)) F_X(H(y)) \quad (8)$$

当 $H(\cdot)$ 为常函数时,有

$$E_X[F_Y(H(x))] E_Y[F_X(H(y))] = F_Y(H(x)) F_X(H(y))$$

此时

$$g(x) g(y) = f_X(x) f_Y(y) \quad (9)$$

广义边缘密度的乘积等于通常边缘密度的乘积,这说明了当 X 与 Y 相互独立时,联合密度函数确定上的不唯一性.

1.4.2 广义边缘密度下条件密度的新形式

记 $g(y|x)$ 为已知 $X=x$ 发生的条件下 Y 的广义条件密度函数,定义

$$g(y|x) = \frac{f(x, y)}{g(x)} = \frac{f(x, y)}{f_X(x)} \int_{-\infty}^{+\infty} f(x) dx \quad (10)$$

$$E[E(Y|X)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y g(y|x) dy g(x) dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{y f(x, y)}{g(x)} g(x) dy dx = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y f(x, y) dx dy \quad (11)$$

由式(11)可知,不论边缘密度是什么形式,条件期望的期望均为无条件期望.这与条件期望理论中的重期望定理结论一致^[4].

1.4.3 广义边缘密度的一些具体形式

设二维随机变量 (X, Y) 的密度函数为

$$f(x,y)=\begin{cases} 1/A & (x,y)\in G \\ 0 & \text{其他} \end{cases} \quad G=\{(x,y)|a\leq x\leq b,c\leq y\leq d\}$$

式中 a, b, c, d 为常数, $A=(b-a)(d-c)$.

设 $F_X(x), F_Y(y)$ 是 (X, Y) 关于 X 和 Y 的边缘分布函数, $F(x, y)$ 是 (X, Y) 的分布函数, 这里 $F(x, y) = F_X(x)F_Y(y)$, 即随机变量 X 与 Y 相互独立.

a. 当 $a < x < b$ 且 $c < H(x) < d$; $c < y < d$ 且 $a < H(y) < b$ 时, 有

$$g(x)=\frac{H(x)-c}{\int_a^b H(x)dx - \alpha(b-a)} \quad g(y)=\frac{H(y)-a}{\int_c^d H(y)dy - \alpha(d-c)}$$

若约定 $x < a$ 或 $x > b$ 时, $g(x) = 0$; $y < c$ 或 $y > d$ 时 $g(y) = 0$. 则广义边缘密度函数下的分布函数有

$$\alpha(x)=\begin{cases} 0 & x < a \\ \frac{\int_a^x (H(u)-c)du}{\int_a^b H(u)du - \alpha(b-a)} & a \leq x \leq b \\ 1 & x > b \end{cases} \quad \alpha(y)=\begin{cases} 0 & y < c \\ \frac{\int_c^y (H(u)-a)du}{\int_c^d H(y)dy - \alpha(d-c)} & c \leq y \leq d \\ 1 & y > d \end{cases}$$

则当 $y \geq d, x \geq b$ 时, 有

$$\alpha(x)\alpha(y)=1=F(x,y)$$

当 $x < a$ 时, 有

$$\alpha(x)\alpha(y)=0=F(x,y)$$

当 $y < c$ 时, 有

$$\alpha(x)\alpha(y)=0=F(x,y)$$

其他情况下都有

$$\alpha(x)\alpha(y)\neq F(x,y)$$

即此时不能以广义边缘分布函数的乘积与联合分布函数的比较验证常见的随机变量独立性.

b. 当 $a < x < b$ 且 $H(x) \geq d$; $c < y < d$ 且 $H(y) \geq b$ 时, 有

$$g(x)=\frac{d-c}{(d-c)(b-a)}=\frac{1}{b-a} \quad g(y)=\frac{b-a}{(b-a)(d-c)}=\frac{1}{d-c}$$

此时

$$\alpha(x)=\int_a^x \frac{1}{b-a}dx \quad \alpha(y)=\int_c^y \frac{1}{d-c}dy$$

即在 $H(x) \geq d, H(y) \geq b$ 时, 无论 x 和 y 取什么值, 都有

$$\alpha(x)\alpha(y)=F(x,y)$$

1.4.4 推广到 n 维形式

设 $f(x_1, x_2, \dots, x_n)$ 为 n 维随机变量的联合密度函数, 令

$$f(x_n)=\bigoplus_G f(x_1, x_2, \dots, x_n)dx_1dx_2\dots dx_{n-1}$$

$$G=\{(x_1, x_2, \dots, x_{n-1})|-\infty < x_1 \leq H(x_1), \dots, -\infty < x_{n-1} \leq H(x_{n-1})\}$$

其中 $H(\cdot)$ 为任意函数并只在使 $f(x_1, x_2, \dots, x_n)$ 大于零时取值, 按照概率密度的性质规范化, 可得如下广义密度函数:

$$g(x_n)=\frac{\bigoplus_G f(x_1, x_2, \dots, x_n)dx_1dx_2\dots dx_{n-1}}{\int_{-\infty}^{+\infty} \bigoplus_G f(x_1, x_2, \dots, x_n)dx_1dx_2\dots dx_{n-1}dx_n}$$

广义边缘密度函数的许多特征与函数 $H(\cdot)$ 的性质有密切的关系, 在实际应用中, 可以根据需要选取适当的 $H(\cdot)$, 从而得出对于解决问题有用或重要的部分.

2 应用实例

舍选抽样法是产生伪随机数的一种抽样方法, 它不对所产生的随机数都录用, 而是建立一个检验条件, 利用这一检验条件进行舍选, 得到所需的随机数. 由于其灵活、计算简单、使用方便而得到较为广泛的应用[5-6].

令 $L = \frac{1}{\int_{-\infty}^{+\infty} f(x) dx}$ 为规范化常量, 则广义边缘密度函数

$$g(z) = Lf(z) = L \int_{-\infty}^{H(z)} f(z, y) dy \quad (12)$$

设随机变量 Z 的密度函数为 $g(z)$, 则由式(5)知

$$\frac{d}{dz} P(X \leq z | Y \leq H(X)) = g(z) \quad (13)$$

即此时 Z 的密度函数可理解为, 在 $Y \leq H(X)$ 条件下 X 的分布函数的导数.

因此对随机变量 Z 的舍选抽样过程为 (a) 产生随机向量 $(X, Y) \sim f(x, y)$ (b) 若 $Y \leq H(X)$, 则令 $Z = X$ 输出 Z , 否则转 (a). 抽样过程通过检验条件 $Y \leq H(X)$ 对 X 的抽样值进行舍选, 从而控制随机变量 Z 的抽样.

在实际应用中通常构造 Z 的密度函数为广义边缘密度函数的形式, 其中随机变量 X, Y 相互独立.

这样就可以通过对 X, Y 抽样来间接地对 Z 抽样, 使抽样更简洁^[7-8].

3 结 语

本文对二维联合密度在指定范围积分, 得出了一个广义的边缘密度函数, 并给出了一些重要性质以及这一广义密度函数形式下的条件分布, 进一步指出这种分布在舍选抽样法中的应用. 若随机变量 Z 分布复杂, 难以分析它的性质, 则可以考虑通过寻找适合函数 $H(\cdot)$, 将其密度函数表示为某已知二维分布的广义边缘密度函数形式, 通过分析这一已知二维分布的性质来间接地分析 Z 的性质. 对于在直接分析 Z 的性质比较困难的情况下, 间接分析不失为一条很好的途径.

参考文献:

- [1] JOHN A R. Mathematical statistics and data analysis[M]. Peking: China Machine Press, 2003: 80-100.
- [2] 复旦大学数学系. 概率论与数理统计[M]. 上海: 上海科学技术出版社, 1960: 55-68.
- [3] 魏宗舒. 概率论与数理统计[M]. 北京: 高等教育出版社, 1983: 102-140.
- [4] 孟庆生, 何才法, 孙志盈. 概率统计简明教程[M]. 南京: 河海大学出版社, 1993: 43-47.
- [5] SOBOLO I M. The Monte Carlo method[M]. Moscow: Mir Publishers, 1975: 112-140.
- [6] 高惠璇. 统计计算[M]. 北京: 北京大学出版社, 1995: 121-166.
- [7] 吴宇新, 余松煜. 广义高斯分布随机变量生成算法[J]. 数据采集与处理, 1999, 14(4): 443-446.
- [8] 张杰, 杨振海. 变换区域舍选法在贝塔分布随机数中的应用[J]. 北京工业大学学报, 2003, 29(3): 359-363.

Relative theory of general marginal density function and its application

REN Hao-lin, YUAN Yong-sheng, WANG Qi-ming, ZHANG Zhi-fen

(College of Sciences, Hohai University, Nanjing 210098, China)

Abstract For the joint density function of 2-D continuous random variables, a new function can be obtained by upper limit-variable integral, and then, by normalization of the new function according to the property of density functions, a general marginal density function was constructed. The relationship between the general form and the conventional form of the density function was analyzed, and the properties of the density function with the general form were discussed. Furthermore, some new conclusions were drawn for a particular general form, and the general form was extended to the n-dimensional general marginal density function. Finally, the application of the general form to acceptance-rejection sampling was discussed.

Key words general form; marginal density function; condition density; acceptance-rejection sampling