Advances in Applied Mathematics 应用数学进展, **2018, 7(8), 1047-1056**
Published Online August 2018 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2018.78122

Research on the Conditional Distribution and the Conditional Digital Characteristics

Qi Shang, Yiqiang Wei

Taiyuan University of Technology, Jincheng Shanxi
Email: 2580104305@qq.com

Received: Aug. 4th, 2018; accepted: Aug. 20th, 2018; published: Aug. 27th, 2018

Abstract

In recent years, with the continuous observation and study of random phenomena, conditional distribution and conditional expectation have been widely used in various fields, but there are still a lot of shortcomings in the existing theories. In this paper, the conditional distribution and its digital characteristics are studied. Not only the definition of classical conditional distribution and extended conditional distribution is introduced, but also the density function form and the distribution function form of the full probability formula are given. In addition, by discussing the digital features of the conditional distribution, we prove and generalize the new properties of the conditional expectation, for example, the full probability form of the heavy expectation formula, and also give the definition of the conditional variance and the proof of its properties.

Keywords

Classical Conditional Distribution, Extended Conditional Distribution, Conditional Expectation, Conditional Variance

关于条件分布与条件数字特征的研究

尚　琦，魏毅强

太原理工大学，山西 晋城
Email: 2580104305@qq.com

收稿日期：2018年8月4日；录用日期：2018年8月20日；发布日期：2018年8月27日

摘　要

近年来，随着人们对随机现象的不断观察和研究，条件分布及条件期望在各个领域中得到了广泛的应用，但是现有理论仍存在很多不足的地方。本文对条件分布及其数字特征进行了研究，不仅介绍了经典条件

文章引用：尚琦，魏毅强．关于条件分布与条件数字特征的研究[J]．应用数学进展, 2018, 7(8): 1047-1056.
DOI: 10.12677/aam.2018.78122

分布和扩展条件分布的定义，而且给出了全概率公式的密度函数形式和分布函数形式。此外，通过对条件分布的数字特征进行探讨，我们证明并推广得到了有关条件期望的新的性质，例如，重期望公式的全概率形式，同时还给出了条件方差的定义和其性质的证明。

关键词

经典条件分布，扩展条件分布，条件期望，条件方差

1. 引言

概率论是研究复杂随机现象规律的有效方法和工具，近年来，随着社会上人们对随机现象的不断观察和研究，概率论的应用受到了越来越多的关注。事实上，在现实生活中很少存在单一的不受别的事件影响的情况，对于很多随机现象的研究往往有着一定的约束条件，比如，当今热门的大数据和微博热搜，因此，人们在利用概率论的理论知识去研究生活中的随机现象时往往离不开条件分布及其数字特征。

条件分布是概率论中最重要和最基本的概念之一，它描述了随机变量之间不独立时，已知其中一个随机变量发生的条件下另一随机变量的概率分布问题，在计算科学、信息网络等领域都有广泛的应用。条件期望刻画了已知一个随机变量发生的条件下另一随机变量取值的平均水平，它在随机过程和鞅的研究中是必不可少的工具，而且在统计、金融等领域发挥了极大的作用。

众所周知，在有关随机变量的问题上，我们往往需要计算期望、方差、分位数等特征数，这些特征数各从一个侧面描述了该分布的特征，例如，方差描述了随机变量取值的平均“波动”程度。那么在条件分布已知时，我们也可以计算得到条件方差等一系列特征数，从而更好的了解其分布特性。但是，现有理论中仅仅对条件期望的定义和性质做出了阐述，并未提及有关条件分布的其他特征数。

在已有的理论型文章中，文[1] [2]系统阐述了条件概率、条件分布与条件期望的定义和性质；文[3]从概率空间及事件域意义下解释了条件概率和条件分布的定义；文[4]对条件期望的性质及不同情形下的求法做出了详细分析；文[5]利用条件概率的定义，由随机变量分布函数的性质，给出了一般情形下条件分布函数的定义；文[6]探讨了条件分布的概念、与随机变量的独立性的关系及在条件期望中的应用；文[7]从几何角度揭示了边缘分布条件分布的几何意义；文[8]通过举例介绍当二维连续型随机变量的边缘密度函数为 0 时，相应的条件分布函数可能为连续型分布、离散型分布，还可能是非连续型非离散型分布。

鉴于此，本文对条件分布与条件数字特征进行了探讨，各个章节的具体安排如下：

第一章，我们简要的回顾了条件分布及其数字特征的研究背景和现状并介绍了文章的具体结构。

第二章，首先介绍了经典条件分布的定义，并且得到了全概率公式的密度函数形式和分布函数形式；然后，我们进一步给出了扩展条件分布的定义，即任一非零概率事件成立的条件下分布函数的一般性定义，并且通过例子给出了具体条件下的一些定义。

第三章，首先介绍了条件期望的定义和性质，并且在现有内容的基础上推广得到了某些新的性质，例如，重期望公式的全概率形式，同时给予了证明。此外，还定义了条件方差并证明了其性质。

第四章，我们提出了进一步需要研究的相关问题，对未来的研究工作做了展望。

本文的创新点在于对已有内容做了推广，定义了一般条件下的条件分布函数，给出了条件方差的定

义和性质，并且得到了全概率公式的分布函数形式和重期望公式的全概率形式，值得推广应用。

2. 条件分布

2.1. 经典条件分布

条件分布是研究变量之间相依关系的一个有力工具，它描述了随机变量之间不独立时，已知其中一个随机变量发生的条件下另一随机变量的概率分布问题。下面给出经典的条件分布函数、条件密度函数的定义及全概率公式的密度函数形式等内容。

设 (X,Y) 是二维离散型随机变量，其联合分布列为

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

定义 2.1.1：对一切使得 $P\left(Y=y_j\right)=p_{\cdot j}=\sum_{i=1}^{\infty}p_{ij}>0$ 的 y_j，称给定 $Y=y_j$ 条件下 X 的条件分布列和条件分布函数分别为

$$p_{i|j}=P\left(X=x_i\mid Y=y_j\right)=\frac{P\left(X=x_i,Y=y_j\right)}{P\left(Y=y_j\right)}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots \tag{2.1.1}$$

$$F\left(x\mid y_j\right)=\sum_{x_i\le x}P\left(X=x_i\mid Y=y_j\right)=\sum_{x_i\le x}p_{i|j} \tag{2.1.2}$$

设 (X,Y) 是二维连续型随机变量，其联合密度函数为 $f(x,y)$，边际密度函数为 $f_X(x)$，$f_Y(y)$。

定义 2.1.2 [1]：对一切使得 $f_Y(y)>0$ 的 y，给定 $Y=y$ 条件下 X 的条件分布函数和条件密度函数分别为

$$F(x\mid y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}\mathrm{d}u \tag{2.1.3}$$

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)} \tag{2.1.4}$$

注：当 $f_Y(y)=0$ 或 $f_X(x)=0$ 时，概率论中并未给出明确的定义。事实上，当 $f_Y(y)=0$ 或 $f_X(x)=0$ 时，条件分布可能为连续型分布、离散型分布和既非连续又非离散型分布，例如，设二维连续型随机变量 (X,Y) 的联合密度函数为 $f(x,y)=\begin{cases}4xy, & 0\le x\le 1,0\le y\le 1\\ 0, & \text{其他}\end{cases}$，求在 $Y=0$ 条件下 X 的条件分布。经计算 $f_X(x)=\begin{cases}2x, & 0\le x\le 1\\ 0, & \text{其他}\end{cases}$，$f_Y(y)=\begin{cases}2y, & 0\le y\le 1\\ 0, & \text{其他}\end{cases}$，由于 $f(x,y)=f_X(x)f_Y(y)$，所以 X 和 Y 相互独立，其条件分布就是无条件分布，故在 $Y=0$ 条件下 X 的条件密度为 $f_{X|Y}(x\mid 0)=f_X(x)=\begin{cases}2x, & 0\le x\le 1\\ 0, & \text{其他}\end{cases}$。

有了条件分布密度函数的概念，顺便给出连续随机变量场合下全概率公式和贝叶斯公式的密度函数形式，先将(2.1.4)式改写成

$$f(x,y)=f_Y(y)f_{X|Y}(x\mid y) \tag{2.1.5}$$

再对 $f(x,y)$ 求边际密度函数，即得如下定理：

定理 2.1.1 (全概率公式的密度函数形式)

$$f_X(x)=\int_{-\infty}^{\infty}f_Y(y)f_{X|Y}(x\mid y)\mathrm{d}y \tag{2.1.6}$$

有了联合密度函数 $f(x,y)$ 和边际密度函数 $f_X(x)$，很容易即可得到如下定理：

定理 2.1.2 (贝叶斯公式的密度函数形式)

$$f_{Y|X}(y|x)=\frac{f_Y(y)f_{X|Y}(x|y)}{\int_{-\infty}^{\infty}f_Y(y)f_{X|Y}(x|y)\mathrm{d}y} \tag{2.1.7}$$

根据全概率公式的密度函数形式，我们很容易可以推广到分布函数上，即得：

定理 2.1.3 (全概率公式的分布函数形式)

$$F_X(x)=\int_{-\infty}^{\infty}F(x|y)f_Y(y)\mathrm{d}y \tag{2.1.8}$$

证明：

$$\int_{-\infty}^{\infty}F(x|y)f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{x}f_{X|Y}(x|y)\mathrm{d}x\right)f_Y(y)\mathrm{d}y=\int_{-\infty}^{\infty}\int_{-\infty}^{x}f(x,y)\mathrm{d}x\mathrm{d}y=F_X(x)$$

结论得证。

2.2. 扩展条件分布

以上定义的条件分布及条件密度函数等都是在给定 $Y=y$ 条件下给出的，如果将所定义的条件延伸为任一事件的发生，分布函数、密度函数的定义会有什么样的变化呢？下面我们首先给出 B 事件发生条件下 X 的条件分布函数的一般性定义。

定义 2.2.1 由条件概率，记 $A=\{\omega:-\infty<X(\omega)\le x\}$，$B$ 为任一概率非零事件，即 $P(B)>0$，则称

$$F_{X|B}(x|B)=P(X\le x|B)=P(A|B)=\frac{P(AB)}{P(B)} \tag{2.2.1}$$

为 B 事件发生条件下 X 的条件分布函数。

下面通过举例给出几个具体条件下的定义。

例 2.2.1：设 (X,Y) 是随机变量，由定义 2.2.1 我们可进一步分连续和离散两种情况给出 $Y\le y$ 条件下 X 的条件分布函数和相应的条件密度函数。

当 (X,Y) 为二维连续型随机变量时，对一切使得 $F_Y(y)>0$ 的 y，称

$$F_{X|Y\le y}(x|Y\le y)=P(X\le x|Y\le y)=\frac{P(X\le x,Y\le y)}{P(Y\le y)}=\frac{\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v}{\int_{-\infty}^{\infty}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v}=\frac{\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v}{F_Y(y)} \tag{2.2.2}$$

为给定 $Y\le y$ 条件下 X 的条件分布函数。

原来的条件密度函数是利用取极限和中值定理得到的，这里的条件密度函数通过求导的方式来给出，如下：

$$f_{X|Y\le y}(x|Y\le y)=\frac{\partial F_{X|Y\le y}(x|Y\le y)}{\partial x}=\frac{\int_{-\infty}^{y}f(x,v)\mathrm{d}v}{\int_{-\infty}^{\infty}\int_{-\infty}^{y}f(u,v)\mathrm{d}u\mathrm{d}v}=\frac{\int_{-\infty}^{y}f(x,v)\mathrm{d}v}{F_Y(y)} \tag{2.2.3}$$

当 (X,Y) 为二维离散型随机变量时，对一切使得 $F_Y(y)>0$ 的 y，称

$$\begin{aligned}F_{X|Y\le y}(x|Y\le y)&=P(X\le x|Y\le y)=\frac{P(X\le x,Y\le y)}{P(Y\le y)}\\&=\frac{\sum_{x_i\le x}\sum_{y_j\le y}P(X=x_i,Y=y_j)}{\sum_{x_i}\sum_{y_j\le y}P(X=x_i,Y=y_j)}=\frac{\sum_{x_i\le x}\sum_{y_j\le y}P(X=x_i,Y=y_j)}{F_Y(y)}\end{aligned} \tag{2.2.4}$$

为给定 $Y\le y$ 条件下 X 的条件分布函数。

相应的条件分布列为

$$p_{x_i|Y\le y}=P\left(X=x_i\mid Y\le y\right)=\frac{P\left(X=x_i,Y\le y\right)}{P\left(Y\le y\right)}=\frac{\sum\limits_{y_j\le y}P\left(X=x_i,Y=y_j\right)}{\sum\limits_{x_i}\sum\limits_{y_j\le y}P\left(X=x_i,Y=y_j\right)}=\frac{\sum\limits_{y_j\le y}P\left(X=x_i,Y=y_j\right)}{F_Y\left(y\right)} \tag{2.2.5}$$

注意到，这里只给出了给定 $Y\le y$ 条件下 X 的条件分布函数以及连续场合的条件密度函数、离散场合的条件分布列，我们可以把这个事件定义为任意一个区间，例如，$Y>y$，在实际问题中，可以根据情况的不同进行定义。

例 2.2.2：设随机变量 X 的密度函数为 $f\left(x\right)=\begin{cases}2x, & 0<x<1\\0, & \text{其他}\end{cases}$，求 $a<X<b$ $\left(b-1<a<b\right)$ 条件下 X 的条件分布函数。

解：考虑到需满足 $P\left(a<X<b\right)>0$，这里分析三种情况：

1) 当 $a<0<b$ 时，

$$F\left(x\mid a<X<b\right)=P\left(X\le x\mid a<X<b\right)=\frac{\min\left\{P\left(X\le x\right),P\left(X<b\right)\right\}}{P\left(X<b\right)}=\begin{cases}0 & x<0\\ \dfrac{\min\left\{x^2,b^2\right\}}{b^2} & 0\le x<b\\ 1 & x\ge b\end{cases}$$

2) 当 $0<a<b<1$ 时，

$$F\left(x\mid a<X<b\right)=P\left(X\le x\mid a<X<b\right)=\frac{P\left(X\le x,a<X<b\right)}{P\left(a<X<b\right)}=\begin{cases}0 & x<a\\ \dfrac{\min\left\{x^2-a^2,b^2-a^2\right\}}{b^2-a^2} & a\le x<b\\ 1 & x\ge b\end{cases}$$

3) 当 $a<1<b$ 时，

$$F\left(x\mid a<X<b\right)=P\left(X\le x\mid a<X<b\right)=\frac{P\left(X\le x,X>a\right)}{P\left(X>a\right)}=\begin{cases}0 & x\le a\\ \dfrac{x^2-a^2}{1-a^2} & a<x<1\\ 1 & x\ge 1\end{cases}$$

以上便给出了 $a<X<b$ 条件下 X 的条件分布函数。

例 2.2.3：设 X 和 Y 是相互独立的随机变量，且 $X\sim P\left(\lambda_1\right)$，$Y\sim P\left(\lambda_2\right)$。在已知 $X+Y=n$ 的条件下，求 X 的条件分布函数。

解：$F\left(x\mid X+Y=n\right)=P\left(X\le x\mid X+Y=n\right)=\sum\limits_{k\le x}P\left(X=k\mid X+Y=n\right)$

又知在 $X+Y=n$ 条件下，X 服从二项分布 $b\left(n,p\right)$，其中 $p=\lambda_1/\left(\lambda_1+\lambda_2\right)$，则

$$F\left(x\mid X+Y=n\right)=\sum_{k\le x}C_n^k p^k\left(1-p\right)^{k-1}$$

其中 $p=\lambda_1/\left(\lambda_1+\lambda_2\right)$，上式便给出了 $X+Y=n$ 条件下 X 的条件分布函数。

3. 条件分布下的数字特征

3.1. 条件期望

条件期望在概率论及现实生活中用途广泛，近年来随着研究的深入，条件期望在很多领域中都得到

了应用，并取得了很好的效果。下文详细介绍了有关条件期望的定义、性质和重期望公式等，并基于上一章的内容，给出了重期望公式的全概率形式。

定义 3.1.1：设 (X,Y) 是二维连续型随机变量，对于一切使得 $f_Y(y)>0$ 的 y，若 $E(|X| \mid Y=y)=\sum_i |x_i| P(X=x_i \mid Y=y)<\infty$，则称

$$E(X \mid Y=y)=\sum_i x_i P(X=x_i \mid Y=y) \tag{3.1.1}$$

为给定 $Y=y$ 条件下 X 的条件数学期望。

定义 3.1.2：设 (X,Y) 是二维连续型随机变量，对于一切使得 $f_Y(y)>0$ 的 y，若 $E(|X| \mid Y=y)=\int_{-\infty}^{\infty}|x| f_{X|Y}(x \mid y)\mathrm{d}x<\infty$，则称

$$E(X \mid Y=y)=\int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\mathrm{d}x \tag{3.1.2}$$

为给定 $Y=y$ 条件下 X 的条件数学期望。

我们特别要强调的是：$E(X \mid Y=y)$ 是 y 的函数，对 y 的不同取值，$E(X \mid Y=y)$ 的取值也在发生变化，而 $E(X \mid Y)$ 是随机变量 Y 的函数，可以将 $E(X \mid Y=y)$ 看成是 $Y=y$ 时 $E(X \mid Y)$ 的一个取值。

性质 3.1.1：设 X,Y 是随机变量，$g(x)$，$h(y)$ 是实函数，且以下涉及的数学期望均存在，则

$$E\left(\left(a_0+\sum_{i=1}^n a_i X_i\right) \mid Y\right)=a_0+\sum_{i=1}^n a_i E(X_i \mid Y),\quad a_i \in R,\quad i=1,\cdots,n \tag{3.1.3}$$

$$E(h(Y) \mid Y)=h(Y) \tag{3.1.4}$$

$$E(g(X)h(Y) \mid Y)=h(Y)E(g(X) \mid Y) \tag{3.1.5}$$

$$\text{当 } X,Y \text{ 独立时，}\quad E(g(X) \mid Y)=Eg(X) \tag{3.1.6}$$

定理 3.1.1 [1] (重期望公式)：设 (X,Y) 是二维随机变量，且 $E(X)$ 存在，则

$$E(X)=E(E(X \mid Y)) \tag{3.1.7}$$

重期望公式的用途在于当计算 X 的均值较困难时，借助一个与 X 有关的量 Y，用 Y 的不同取值把 X 的取值划分成若干个小区域，先在小区域上求 X 的平均，再以此类平均求加权平均，即可得 X 的均值 EX。

上一章给出了给定 $Y \le y$ 条件下 X 的条件分布函数和密度函数，我们很容易联想到 Y 取值于某一区间条件下 X 的条件分布函数和密度函数，考虑到重期望公式在该条件下的表示，很容易即得：

定理 3.1.2 (重期望公式的全概率形式)：设 (X,Y) 是随机变量，且 $E(X)$ 存在，若存在 $y_1,\cdots,y_n$，且 $-\infty<y_1\le y_2\le\cdots\le y_n<\infty$，则

$$\begin{aligned}E(X)=&E(X \mid Y\le y_1)\cdot F_Y(y_1)+E(X \mid y_1<Y\le y_2)\cdot(F_Y(y_2)-F_Y(y_1))+\cdots\\&+E(X \mid y_{n-1}<Y\le y_n)\cdot(F_Y(y_n)-F_Y(y_{n-1}))+E(X \mid Y>y_n)\cdot(1-F_Y(y_n))\end{aligned} \tag{3.1.8}$$

其中，最简单的形式为

$$EX=E(X \mid Y\le y)\cdot F_Y(y)+E(X \mid Y>y)\cdot(1-F_Y(y)) \tag{3.1.9}$$

证明：这里仅给出(3.1.9)式的证明，(3.1.8)同理可证。

当 (X,Y) 为二维连续随机变量时，

$$\begin{aligned}E(X \mid Y\le y)\cdot F_Y(y)&=\int_{-\infty}^{\infty} x f_{X|Y\le y}(x \mid Y\le y)\mathrm{d}x\cdot F_Y(y)=\int_{-\infty}^{\infty} x\frac{\int_{-\infty}^{y} f(x,v)\mathrm{d}v}{F_Y(y)}\mathrm{d}x\cdot F_Y(y)\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{y} f(x,v)\mathrm{d}v\mathrm{d}x\end{aligned}$$

$$E(X \mid Y>y)\cdot(1-F_Y(y))=\int_{-\infty}^{\infty} x f_{X\mid Y>y}(x \mid Y>y)\mathrm{d}x\cdot(1-F_Y(y))=\int_{-\infty}^{\infty} x\frac{\int_y^{\infty} f(x,v)\mathrm{d}v}{1-F_Y(y)}\mathrm{d}x\cdot(1-F_Y(y))$$
$$=\int_{-\infty}^{\infty}\int_y^{\infty} f(x,v)\mathrm{d}v\mathrm{d}x$$

将结果代入(3.1.9)式可得

$$\begin{aligned}E(X \mid Y\le y)\cdot F_Y(y)+E(X \mid Y>y)\cdot(1-F_Y(y))&=\int_{-\infty}^{\infty}\int_{-\infty}^{y} f(x,v)\mathrm{d}v\mathrm{d}x+\int_{-\infty}^{\infty}\int_y^{\infty} f(x,v)\mathrm{d}v\mathrm{d}x\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,v)\mathrm{d}v\mathrm{d}x\\&=EX\end{aligned}$$

离散场合证明类似，故结论得证。

注：定理 3.1.2 是重期望公式的另一种表达形式，实际与重期望公式是一致的。当随机变量 Y 在不同区间取值时，若 X 有不同的分布形式，此时借助该定理去求解会更容易，若 X 有相同的分布形式，该定理只是通过分区间求期望再累加的方式进行求解，与单点求期望再积分或累加是一致的。

推论 3.1.1 设 (X,Y) 是随机变量，$g(x)$ 为实函数，若 $Eg(X)$ 存在，则

$$\begin{aligned}E(g(X))=&E(g(X)\mid Y\le y_1)\cdot F_Y(y_1)+E(g(X)\mid y_1<Y\le y_2)\cdot(F_Y(y_2)-F_Y(y_1))\\&+\cdots+E(g(X)\mid y_{n-1}<Y\le y_n)\cdot(F_Y(y_n)-F_Y(y_{n-1}))\\&+E(g(X)\mid Y>y_n)\cdot(1-F_Y(y_n))\end{aligned}\tag{3.1.10}$$

特别地，

$$\begin{aligned}E(X^k)=&E(X^k\mid Y\le y_1)\cdot F_Y(y_1)+E(X^k\mid y_1<Y\le y_2)\cdot(F_Y(y_2)-F_Y(y_1))\\&+\cdots+E(X^k\mid y_{n-1}<Y\le y_n)\cdot(F_Y(y_n)-F_Y(y_{n-1}))\\&+E(X^k\mid Y>y_n)\cdot(1-F_Y(y_n))\end{aligned}\tag{3.1.11}$$

此外，根据中心矩与原点矩的关系，同理可求得 X 的 k 阶中心矩。

性质 3.1.2 设 X，Y，Z 是随机变量，$g(x)$ 是实函数，且以下涉及的数学期望均存在，则

$$E\left[E(g(X)\mid Y)\right]=Eg(X)\tag{3.1.12}$$

$$E(XY)=E(YE(X\mid Y))\tag{3.1.13}$$

$$Cov(Y,E(X\mid Y))=Cov(X,Y)\tag{3.1.14}$$

$$E(XYZ)=E\Big(XE\big(YE(Z\mid(X,Y))\mid X\big)\Big)\tag{3.1.15}$$

证明：在此仅给出上述四条性质在连续场合的证明，离散场合的证明可类似得出。设二维连续随机变量 (X,Y) 的联合密度函数为 $f(x,y)$，$Y=y$ 条件下 X 的条件密度函数为 $f_{X|Y}(x\mid y)$。

1)

$$\begin{aligned}E\left[E(g(X)\mid Y)\right]&=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} g(x)f_{X|Y}(x\mid y)\mathrm{d}x\right)f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} g(x)\frac{f(x,y)}{f_Y(y)}\mathrm{d}x\right)f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)f(x,y)\mathrm{d}x\mathrm{d}y\\&=Eg(X)\end{aligned}$$

2)
$$\begin{aligned}E(XY)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{X|Y}(x|y)f_Y(y)\mathrm{d}x\mathrm{d}y\\&=\int_{-\infty}^{\infty}y\left(\int_{-\infty}^{\infty}xf_{X|Y}(x|y)\mathrm{d}x\right)f_Y(y)\mathrm{d}y\\&=\int_{-\infty}^{\infty}yE(X|Y=y)f_Y(y)\mathrm{d}y\\&=E(YE(X|Y))\end{aligned}$$

3)
$$\begin{aligned}Cov(Y,E(X|Y))&=E(YE(X|Y))-E(Y)E(E(X|Y))\\&=E(XY)-E(Y)E(X)=Cov(X,Y)\end{aligned}$$

4) 假设给定 $X=x$ 时，(Y,Z) 的条件密度函数为 $f_{(Y,Z)|X}((y,z)|x)$，给定 $X=x$ 和 $Y=y$ 时，Z 的条件密度函数为 $f_{Z|(X,Y)}(z|(x,y))$，且 $f_{(Y,Z)|X}((y,z)|x)=f_{Z|(X,Y)}(z|(x,y))f_{Y|X}(y|x)$ 成立。

由(2)可得 $E(XYZ)=E(XE(YZ|X))$

当 $X=x$ 时，

$$\begin{aligned}E(YZ|X=x)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yzf_{(Y,Z)|X}((y,z)|x)\mathrm{d}y\mathrm{d}z=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yzf_{Z|(X,Y)}(z|(x,y))f_{Y|X}(y|x)\mathrm{d}y\mathrm{d}z\\&=\int_{-\infty}^{\infty}y\left(\int_{-\infty}^{\infty}zf_{Z|(X,Y)}(z|(x,y))\mathrm{d}z\right)f_{Y|X}(y|x)\mathrm{d}y=\int_{-\infty}^{\infty}yE(Z|(X=x,Y=y))f_{Y|X}(y|x)\mathrm{d}y\\&=E(YE(Z|(X=x,Y))|X=x)\end{aligned}$$

故 $E(YZ|X)=E(YE(Z|(X,Y))|X)$

从而可得 $E(XYZ)=E(XE(YE(Z|(X,Y))|X))$

证毕。

定理 3.1.3 [2]：设 (X,Y) 是随机变量，且 $EX^2<\infty$，对于任何实函数 $g(x)$，都有

$$E(X-E(X|Y))^2\le E(X-g(Y))^2 \tag{3.1.16}$$

定理 3.1.3 给出了均方意义下，已知随机变量 Y 的条件下，X 的最优预测为 $E(X|Y)$，该定理可以解决一系列的预测问题，它在当前的经济发展中发挥了不可或缺的作用。

3.2. 条件方差

条件期望 $E(X|Y)$ 是给定 Y 条件下随机变量 X 的分布的一种位置特征数，但该位置特征数无法反映出给定 Y 条件下 X 取值的“波动”程度，以下定义了度量此种“波动”程度大小的条件方差，并对其性质给予了证明。

定义 3.2.1：设 (X,Y) 是二维随机变量，若 $EX^2<\infty$，则称

$$D(X|Y)=E\left[(X-E(X|Y))^2|Y\right] \tag{3.2.1}$$

为 Y 已知时 X 的条件数学方差，简称为条件方差。

性质 3.2.1：条件方差具有如下性质：

$$D(X|Y)=E(X^2|Y)-(E(X|Y))^2 \tag{3.2.2}$$

$$D(c|Y)=0,\quad \forall c\in R \tag{3.2.3}$$

$$D((aX+b)|Y)=a^2D(X|Y),\quad \forall a,b\in R \tag{3.2.4}$$

$$D(X) = ED(X\,|\,Y) + DE(X\,|\,Y) \tag{3.2.5}$$

证明：假设以下涉及的期望及方差均存在。

1)
$$\begin{aligned} D(X\,|\,Y) &= E\left[\left(X - E(X\,|\,Y)\right)^2 \,|\, Y\right] \\ &= E\left(X^2\,|\,Y\right) + \left(E(X\,|\,Y)\right)^2 - 2E\left(XE(X\,|\,Y)\,|\,Y\right) \\ &= E\left(X^2\,|\,Y\right) + \left(E(X\,|\,Y)\right)^2 - 2\left(E(X\,|\,Y)\right)^2 \\ &= E\left(X^2\,|\,Y\right) - \left(E(X\,|\,Y)\right)^2 \end{aligned}$$

2) $D(c\,|\,Y) = E\left[\left(c - E(c\,|\,Y)\right)^2 \,|\, Y\right] = E\left[(c-c)^2 \,|\, Y\right] = 0$

3)
$$\begin{aligned} D\left((aX+b)\,|\,Y\right) &= E\left[\left(aX + b - E\left((aX+b)\,|\,Y\right)\right)^2 \,|\, Y\right] \\ &= E\left[\left(aX - aE(X\,|\,Y)\right)^2 \,|\, Y\right] \\ &= a^2 E\left[\left(X - E(X\,|\,Y)\right)^2 \,|\, Y\right] \\ &= a^2 D(X\,|\,Y) \end{aligned}$$

4)
$$\begin{aligned} D(X) &= E(X - EX)^2 \\ &= E\left(X - E(X\,|\,Y) + E(X\,|\,Y) - EX\right)^2 \\ &= E\left(X - E(X\,|\,Y)\right)^2 + E\left(E(X\,|\,Y) - EX\right)^2 + 2E\left[\left(X - E(X\,|\,Y)\right)\left(E(X\,|\,Y) - EX\right)\right] \\ &= EX^2 + \left(E(X\,|\,Y)\right)^2 - 2E\left(XE(X\,|\,Y)\right) + E\left(E(X\,|\,Y) - E\left(E(X\,|\,Y)\right)\right)^2 \\ &\quad + 2E\left[XE(X\,|\,Y) - XEX - \left(E(X\,|\,Y)\right)^2 + EXE(X\,|\,Y)\right] \end{aligned}$$

$$\begin{aligned} &= EX^2 + \left(E(X\,|\,Y)\right)^2 - 2E\left(XE(X\,|\,Y)\right) + DE(X\,|\,Y) \\ &\quad + 2E\left(XE(X\,|\,Y)\right) - 2(EX)^2 - 2\left(E(X\,|\,Y)\right)^2 + 2(EX)^2 \\ &= EX^2 - \left(E(X\,|\,Y)\right)^2 + DE(X\,|\,Y) \\ &= E\left(E\left(X^2\,|\,Y\right)\right) - E\left(E(X\,|\,Y)\right)^2 + DE(X\,|\,Y) \\ &= ED(X\,|\,Y) + DE(X\,|\,Y) \end{aligned}$$

证毕。

4. 展望

本论文虽然在已有内容的基础上推广得到了一些结论，但随着条件分布及其数字特征的理论的深入研究，今后还有很多相关的内容值得我们去探讨。比如：分布函数和特征函数是一一对应的，那么条件分布函数的特征函数怎样去表示，它具有怎样的性质呢?希望有关条件分布的理论日益完善，并且得到越来越广泛的应用。

参考文献

[1] 茆诗松, 程依明, 濮晓龙. 概率论与数理统计教程[M]. 第 2 版. 高等教育出版社, 2011: 41-49, 193-205.

[2] 何书元. 概率论[M]. 北京: 北京大学出版社, 2006: 177-183.

[3] 叶仁玉, 桂春燕. 不同类型随机变量之间的条件分布及其应用[J]. 安庆师范学院学报(自然科学版), 2013, 19(2): 17-19.

[4] 魏艳华, 李艳颖, 王丙参. 条件期望的性质及求法[J]. 牡丹江大学学报, 2009, 18(9): 116-117.
[5] 江五元, 丁卫平. 二维随机变量条件分布函数教学的思考[J]. 数学理论与应用, 2013, 33(4): 126-128.
[6] 吕文华, 韩慧霞. 关于条件分布及其应用的教学研究[J]. 滁州学院学报, 2014, 16(2): 126-127.
[7] 胡晓华, 虞敏. 边缘分布条件分布的几何意义及推广[J]. 大学数学, 2014, 30(3): 81-86.
[8] 宁健荣, 周玲. 条件分布计算的几个问题研究[J]. 大学数学, 2016, 32(5): 61-66.