Extension of Two-Dimensional Discrete Random Variables Conditional Distribution

Feixue Huang
Department of Economics, Dalian University of Technology
Dalian 116024, China
E-mail: software666@163.com

Cheng Li
Department of Economics, Dalian University of Technology
Dalian 116024, China

This work is supported by the National Science Foundation under Grant No. 70890083;the funds project under the Ministry of Education of the PRC for young people who are devoted to the researches of humanities and social sciences under Grant No. 09YJC790025;software + X research of Dalian University of Technology under Grant No. 842301.

Abstract

Conditional distribution reflects the dependency link among random variables, but two-dimensional random variables Conditional Distribution has some limitations. In order to rich the content of conditional distribution this paper gives the extension of conditional distribution under discrete random variables and some examples. This article obtains the extension strictly in accordance with the definition of two-dimensional random variables. So it can get conditional distributions after changing the condition and get conditional distributions that are extended into n-dimensional random variables, thereby enriching the contents of the conditional distribution.

Keywords: Discrete Random Variables, Two-dimensional Random Variables, Conditional Distribution

1. Introduction

Two-dimensional random variables conditional distribution is the distribution of one variable when another variable is a fixed value. The relationships of two-dimensional random variables (X, Y) are mainly divided into two types: independence and dependence. The more evaluations X and Y have, the more conditional distributions will be there. Each conditional distribution describes one state-specific distribution from a side. So the contents of conditional distribution are richer and its applications are broader. For in many issues values of concerned random variables tend to influence each other, this makes conditional distribution as a powerful tool for studying dependencies among variables. Conditional distribution for random variables derives from conditional probability for random events, so there is a close relationship between the two and the approaches to handle them are the same, but conditional distribution is more complex to deal with.

In recent years, the main research direction in China is the research of conditional eigenvalues and the extension and application of conditional distribution. Conditional eigenvalues are mainly pointed to conditional expectation, which is expectation under conditional distribution. In this area the main contributions are Wang Cheng, Zou Hailei gave the definition and researched the characteristics of random variables conditional particular function based on measurement and integral theory; Zhang Mei used minimum mean-squared error to solve one kind problems of best prediction and took examples to analyze the application of conditional expectation in practical prediction problems; Xu hui and Wu Guogeng educed total probability formula of discrete and continuous random variables based on conditional expectation and indicative function I_A of random events A. For extension of conditional distribution, primacy researches are Cheng Feiyue generalized the condition distribution of the poisson process arrival time; Yang Jingping, etc. investigated the marginal recursive equations on excess-of-loss reinsurance treaty under the assumption that the number of claims belongs to the family consisting of Poisson, binomial and negative binomial, and that the severity distribution hand bounded continuous density function; Hu Duanping gave a express formula of distribution for elliptically contoured matrix distributions and showed that condition distribution of elliptically contoured matrix distribution is elliptically contoured distribution yet. In the above study, they focused on the application of the conditional distribution showing the research significance of the conditional distribution, but there is few studies researching the nature of the conditional distribution.

In comparison, the studies of conditional distribution abroad are more in-depth. For example, Rodney C.L. Wolff, etc studied the methods of evaluating conditional distribution function; Jushan Bai investigated the dynamic model of testing parametric conditional distributions Peter Hall and Qiwei Yao discussed approximating conditional distribution function using dimension reduction Bruce E. Hansen studied nonparametric estimation of smooth conditional distributions[12]; Persi Diaconis and Bernd Sturmfels analyzed conditional distributions using algebraic algorithms for sampling. They have shown that the researches of conditional distribution are multi-faceted and more complex while make against undergraduate teaching.

Therefore, this paper begins to discuss and analyze from the basic content of conditional distribution and educes general formulas with certain conditions on the basis of the definition of conditional distribution. It first starts form the two-dimensional random variable conditional distribution and changes the given conditions to obtain the extensions of conditional distribution and then gives extensions of conditional distribution when there are three-dimensional random variables. This paper is to solve the conditional distribution of multidimensional random variables under the given conditions and its results can be used for teaching, expending the knowledge of the conditional distribution and facilitating people's calculations.

2. Extension of discrete random variables conditional distribution

Extension 2.1 Set X and Y for the discrete random variables and X,Y are independent. Known the distribution series of X and Y, under the given condition of $X+Y=n$ the conditional distribution of X is

$$P\left(X=k \middle| X+Y=n\right)=\frac{P\left(X=k,X+Y=n\right)}{P\left(X+Y=n\right)}=\frac{P\left(X=k\right)P\left(Y=n-k\right)}{P\left(X+Y=n\right)} \tag{1}$$

Example2.1.1 I) basal example: In the case of two-dimensional random variables, X and Y are independent and $X \sim P(\lambda_1), Y \sim P(\lambda_2)$. Given the condition of $X+Y=n$ to solve the conditional distribution of X.

To solve: For the sum of the independent Poisson variables is still Poisson variable, viz. $X+Y \sim P(\lambda_1+\lambda_2)$, so

$$\begin{aligned}P\left(X=k \middle| X+Y=n\right)&=\frac{P\left(X=k,X+Y=n\right)}{P\left(X+Y=n\right)}=\frac{P\left(X=k\right)P\left(Y=n-k\right)}{P\left(X+Y=n\right)}\\
&=\frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}\cdot\frac{\lambda_2^{n-k}}{(n-k)!}e^{-\lambda_2}}{\frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}}=\frac{n!}{k!(n-k)!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n}\\
&=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}, k=0,1,\cdots,n.\end{aligned} \tag{2}$$

That is under the condition of $X+Y=n$, X subjects to binomial distribution $b(n,p)$, thereinto $p=\frac{\lambda_1}{\lambda_1+\lambda_2}$.

II) Two-dimensional discrete random variables (X,Y) subject to trinomial distribution $M(n,p_1,p_2,p_3)$. Given the condition of $X=i$ to solve the conditional distribution of Y.

To solve: (X,Y) subject to trinomial distribution $M(n,p_1,p_2,p_3)$, then theirs joint distribution is

$$P\left(X=i,Y=j\right)=\frac{n!}{i!j!(n-i-j)!}p_1^i p_2^j\left(1-p_1-p_2\right)^{n-i-j}, i,j=1,2,\cdots,n,i+j\le n. \tag{3}$$

For the marginal distribution of multinomial distribution is still multinomial distribution and the marginal distribution of trinomial distribution is binomial, so $X \sim b(n,p_1)$, $Y \sim b(n,p_2)$.

$$P\left(y=j \mid x=i\right)=\frac{P\left(x=i,y=j\right)}{P\left(x=i\right)}=\frac{\frac{n!}{i!\,j!\left(n-i-j\right)!}p_1^{\,i}p_2^{\,j}\left(1-p_1-p_2\right)^{n-i-j}}{\frac{n!}{i!\left(n-i\right)!}p_1^{\,i}\left(1-p_1\right)^{n-i}}$$

$$=\frac{\left(n-i\right)!}{j!\left(n-i-j\right)!}\frac{p_2^{\,j}\left(1-p_1-p_2\right)^{n-i-j}}{\left(1-p_1\right)^{n-i}}$$

$$=\binom{n-i}{j}\left(\frac{p_2}{1-p_1}\right)^{j}\left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-i-j}. \tag{4}$$

That is under the condition of $X=i$, Y subjects to binomial distribution $b\left(n-i,p\right)$, thereinto $p=\frac{p_2}{1-p_1}$.

Extension2.2 Set X,Y,Z for the discrete random variables and X,Y,Z are mutual independent. Known the distribution series of X,Y,Z, under the given condition of $X+Y+Z=n$ the conditional distribution of X is

$$P\left(X=k \mid X+Y+Z=n\right)=\frac{P\left(X=k,X+Y+Z=n\right)}{P\left(X+Y+Z=n\right)}=\frac{P\left(X=k\right)P\left(Y+Z=n-k\right)}{P\left(X+Y+Z=n\right)}. \tag{5}$$

Example2.2.1 I) X,Y,Z are mutual independent, and $X\sim P\left(\lambda_1\right)$, $Y\sim P\left(\lambda_2\right)$, $Z\sim P\left(\lambda_3\right)$.

1) Given the condition of $Y+Z=n$ to solve the conditional distribution of X.

To solve: For the sum of the independent Poisson variables is still Poisson variable, viz. $Y+Z\sim P\left(\lambda_2+\lambda_3\right)$, so

$$P\left(X=k \mid Y+Z=n\right)=\frac{P\left(X=k,Y+Z=n\right)}{P\left(Y+Z=n\right)}=P\left(X=k\right)\sim P\left(\lambda_1\right). \tag{6}$$

That is X still subjects to $P\left(\lambda_1\right)$.

2) Given the condition of $X+Y+Z=n$ to solve the conditional distribution of X.

To solve: For the sum of the independent Poisson variables is still Poisson variable, viz. $X+Y+Z\sim P\left(\lambda_1+\lambda_2+\lambda_3\right)$, so

$$P\left(X=k \mid X+Y+Z=n\right)=\frac{P\left(X=k,X+Y+Z=n\right)}{P\left(X+Y+Z=n\right)}=\frac{P\left(X=k\right)P\left(Y+Z=n-k\right)}{P\left(X+Y+Z=n\right)}$$

$$=\frac{\frac{\lambda_1^k}{k!}e^{-\lambda_1}\cdot\frac{\left(\lambda_2+\lambda_3\right)^{n-k}}{\left(n-k\right)!}e^{-\left(\lambda_2+\lambda_3\right)}}{\frac{\left(\lambda_1+\lambda_2+\lambda_3\right)^{n}}{n!}e^{-\left(\lambda_1+\lambda_2+\lambda_3\right)}}=\frac{n!}{k!\left(n-k\right)!}\frac{\lambda_1^{\,k}\left(\lambda_2+\lambda_3\right)^{n-k}}{\left(\lambda_1+\lambda_2+\lambda_3\right)^{n}}$$

$$=\binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}\right)^{k}\left(\frac{\lambda_2+\lambda_3}{\lambda_1+\lambda_2+\lambda_3}\right)^{n-k},k=0,1,\cdots,n. \tag{7}$$

That is under the condition of $X+Y+Z=n$, X subjects to binomial distribution $b\left(n,p_1\right)$, thereinto $p_1=\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}$.

In a similar way, under the condition of $X+Y+Z=n$, Y subjects to binomial distribution $b\left(n,p_2\right)$, thereinto $p_2=\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}$; Z subjects to binomial distribution $b\left(n,p_3\right)$, thereinto $p_3=\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$.

II) $X_1,X_2,\cdots,X_n$ are mutual independent and $X_i\sim P\left(\lambda_i\right),i=1,2,\cdots,n$. Given the condition of $X_1+\cdots+X_n=m$ to solve the conditional distribution of X_i.

To solve: $$P\left(X_i = k \middle| \sum_{j=1}^{n} X_j = m\right) = \frac{P\left(X_i = k, \sum_{j=1}^{n} X_j = m\right)}{P\left(\sum_{j=1}^{n} X_j = m\right)}$$

$$= \frac{P(X_i = k)\,P\left(\sum_{\substack{j=1\\ j\neq i}}^{n} X_j = m-k\right)}{P\left(\sum_{j=1}^{n} X_j = m\right)} = \frac{\dfrac{\lambda_i^k}{k!} e^{-\lambda_i} \cdot \dfrac{\left(\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_j\right)^{m-k}}{(m-k)!} e^{-\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_j}}{\dfrac{\left(\sum_{j=1}^{n}\lambda_j\right)^{m}}{m!} e^{-\sum_{j=1}^{n}\lambda_j}}$$

$$= \frac{m!}{k!(m-k)!} \frac{\lambda_i^k \left(\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_j\right)^{m-k}}{\left(\sum_{j=1}^{n}\lambda_j\right)^{m}} = \binom{m}{k} \left(\frac{\lambda_i}{\sum_{j=1}^{n}\lambda_j}\right)^{k} \left(\frac{\sum_{\substack{j=1\\ j\neq i}}^{n}\lambda_j}{\sum_{j=1}^{n}\lambda_j}\right)^{m-k}, k = 0,1,\cdots,n. \qquad (8)$$

That is under the condition of $X_1 + \cdots + X_n = m$, X_i subjects to binomial distribution $b(n, p_i)$, thereinto $p_i = \frac{\lambda_i}{\sum_{j=1}^{n}\lambda_j}$.

When $\lambda_1 = \lambda_2 = \cdots = \lambda_n$, $p_1 = p_2 = \cdots = p_n = \frac{1}{n}$, $X_i \sim b\left(n, \frac{1}{n}\right), i = 1,2,\cdots,n$.

The solution of the above two extensions is relatively simple so it is omitted.

3. Conclusions

This paper mainly discusses conditional distributions of multidimensional random variables and its related examples given certain conditions in the case of discrete situation. It changes the original condition of one fixed variable into more complex conditions, for example the condition that the sum of two variables is fixed in two-dimensional situation.

In addition, in condition of discrete random variables this paper extends two-dimensional into three and n-dimensional random variables and gives the conditional distributions. For example, in the case of three-dimension, we can get one variable's conditional distribution given the sum of the other two or the three variables fixed. It is the same in multi-dimension. This article gets the above results strictly according to two-dimensional random variables conditional distribution.

Conditional distribution can be applied in the life and work to resolve practical problems. The application of using conditional distribution theory to carry out scientific analysis and calculations with real data is an important reflection of the usefulness of conditional distribution. In addition, for there are more and more discussions about multidimensional random variables in reality this paper extends the conditional distribution to provide new train of thought for the research in some extent and it aims at enriching the content of conditional distribution, deepening the understanding of it and applying it well in practical.

References

Bruce E. Hansen. (2009). Nonparametric Estimation of Smooth Conditional Distributions [EB/OL]. (2004-05)[7-30]. http://www.ssc.wisc.edu/~bhansen/papers/cdf.pdf.

Chen, Fei-yue. (2006). Generalization of the Condition Distribution of the Poisson Process Arrival Time. *Journal of Chang Sha University of Electric Power (Natural Science)*, 21(2):75-79.

Hu, Duan-ping. (2001). The Condition Distribution of Elliptically Contoured Matrix Distributions. *Communication on Applied Mathematics and Computation*, 15(2):41-44.

Jushan Bai. (2003). Testing Parametric Conditional Distributions of Dynamic Models. *The Review of Economics and Statistics*, 85(3):531-549.

Liang, Yi. (1998). Conditional Probability and Conditional Distribution. *Journal of Higher Correspondence Education* (Natural Sciences),(5):29-31,52.

Mao, Shi-song, Cheng, Yi-ming & Pu, Xiao-long. (2004). *Probability Theory and Mathematical Statistics Tutorial*. Beijing: Advanced Education Press.

Persi Diaconis & Bernd Sturmfels. (1998). Algebraic algorithms for sampling from conditional distributions. *The Annals of statistics*, 26(1):363-397.

Peter Hall & Qiwei Yao. (2005). Approximating Conditional Distribution Functions Using Dimension Reduction. *The Annals of statistics*, 33(3):1404-1421

Rodney C.L. Wolff, Qiwei Yao & Peter Hall. (1999). Methods for Estimating a Conditional Distribution Function. *Journal of the American Statistical Association*, 94(445):154-163.

Wang, Cheng & Xou, Hai-lei. (2008). Some Properties of Conditional Characteristic function. *Journal of China Jiliang University*, 19(4):360-362.

Xu, Hui & Wu, Guo-gen. (2004). Full Probability Formula Based on the Conditional Expectation and its Appllications. *Journal of East China Institute of Technology*, 27(2):193-195.

Yang, Jing-ping, Wang, Xiao-qian & Cheng, Shi-hong. (2006). Conditional Recursive Equations on Excess-of-Loss Reinsurance. *Applied Mathematics and Mechanics,* 27(8):931-939.

Zhang, Mei. (2006). Using Conditional Expectation to solve Problems of Best Prediction. *Journal of Shananxi Institute of Education*, 22(2):83-84.