Expert Systems with Applications 37 (2010) 5498–5506

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

An optimal bivariate Poisson field chart for controlling high-quality manufacturing processes

Surath Aebtarm, Nizar Bouguila *

Concordia Institute for Information Systems Engineering, Faculty of Engineering and Computer Science, Concordia University, Montreal, Qc, Canada H3G 2W1

ARTICLE INFO

Keywords:
Attribute chart
Control chart
Count data
Poisson distribution
Poisson field
Bivariate Poisson

ABSTRACT

Shewhart C-chart is a widely accepted control chart for monitoring number of defects in a given process. This chart is based on normal approximations. Normal assumption is, however, impractical in many cases especially for count data. This assumption becomes stronger when correlation between characteristics exists. In this article, we propose an optimal bivariate Poisson field chart to monitor two correlated characteristics of count data for both industrial and non-industrial purposes. Our chart is based on optimization of bivariate Poisson confidence interval and projection of bivariate Poisson data in Poisson field. The performance of our proposed algorithm is presented using both real case study and simulations. The experimental results demonstrate improved performances regarding visualization and false alarm rate.

1. Introduction

To deal with number of defects, C-chart is the most widely used tool in statistical process control. There are many works that have attempted to improve C-chart (Kittlitz, 2006; Quesenberry, 1991; Ryan, 1989; Ryan & Schwertman, 1997; Schwertman & Ryan, 1997; Tsai, Lin, & Wu, 2006; Winterbottom, 1993). However, when numbers of defects are very low such as in high quality processes, traditional C-chart becomes an unsuitable tool. In this case, instead of focusing on numbers of defects or fractions of nonconforming items, it is better to focus on numbers of conforming items between the occurrences of nonconforming items, which is generally referred to "interevent counts". In Goh (1987), Goh introduced CCC-chart which can be used for monitoring interevent counts, by pointing out the effect of low fraction of nonconforming (i.e where small probability of nonconforming occurred). Indeed, in this case, normal approximation becomes out of reality. By using actual numerical examples, Xie and Goh presented some applications of CCC-chart, and suggested some methods for decision making in high yield processes (Xie & Goh, 1993, 1992). In Nelson (1994), numbers of interevent counts are transformed by simple power transformation from exponential distribution to Weibull distribution. Then, normal approximation is applied to construct control limits. In He, Xie, Goh, and Tsui (2006), authors applied generalized Poisson distribution to model over-dispersed data, and suggested the use of CCC-chart for high quality process monitoring. Furthermore, in Kaminsky, Benneyan, Davis, and Burke (1992), He et al. (2006), the authors agreed that hypothesis test or histogram should be conducted in initial state of constructing the charts for high yield process. However, although monitoring interevent counts is preferable, to observe various types of defects simultaneously, multiple C-charts and CCC-charts are needed.

Monitoring two or more types of correlated characteristics in high quality processes still leave room for improvement. Lowery and Montgomery pointed out that multivariate control charts perform better to signal out of control alarms than univariate charts, since correlation between variables is taken into account (Lowery & Montgomery, 1995). They also suggested that univariate charts are only suitable for diagnosing process behavior. In Bersimis, Psarakis, and Panaretos (2006), the authors pointed out four conditions that every control chart needs to satisfy: "Is the process in-control?", "Is out-of control state pointed out?", "Is relationship between two or more variables taken into account?", and "what is the problem that out of-control signal actually tells?". According to Bersimis et al. (2006), there are many alternative charts which are based on improving χ^2 and T^2 charts for continuous data. However, for discrete variables, few multivariate attribute charts such as Jolayemi (1999), Lu, Xie, Goh, and Lai (1998), Patel (1973), Skinner, Montgomery, and Runger (2003) have been proposed. In Patel (1973), Patel presented his multivariate control chart for both binomial and Poisson data. For multiple defects, he presented multivariate Hotelling-like chart where time dependency between variables is considered. However, this chart is not practical to apply in nearly zero defect processes, since it considers normal assumption, and requires complicated steps to be constructed (Akhavan & Abbasi, 2006; Chiu & Kuo, 2008). In Lu et al. (1998), to deal with multi-attribute variables, improved Mnp-chart is presented by

* Corresponding author.
E-mail addresses: s_aebta@encs.concordia.ca (S. Aebtarm), bouguila@ciise.concordia.ca (N. Bouguila).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.02.060

considering correlation between characteristics. Not only this chart shows improved results compared with univariate p-chart, but it is also simple. Moreover, Jolayemi has proposed another enhanced Mnp-chart for multiple independent discrete variables by proposing simple designing of optimal Mnp-chart (Jolayemi, 1999). Nonetheless, both Mnp-charts are not practical for high quality processes, since they are constructed under normal assumption. Skinner et al. (2003) have suggested to use generalized linear model (GLM) to construct attribute control chart for multiple counts where input and output variables are measurable. By observing the residuals, generalized linear model-based control charts are more effective to monitor multi-count data than C-chart. Furthermore, the results show effective performance in the case of over-dispersion. However, inputs and outputs are not measurable in every process. Besides, GLM-based charts require multiple charts to observe multivariate variables. In Akhavan and Abbasi (2006), the authors suggested two transformations for multivariate Poisson distribution. For the first transformation, they applied bisection method to find the proper power of the root transformation of each attribute characteristics. The second transformation is Normal distribution To Anything (NORTA) inverse transformation method. After acquiring almost zero skew distribution from both transformations, χ^2 control chart is applied. According to Akhavan and Abbasi (2006), NORTA inverse transformation method shows robust performance when dealing with correlated multivariate Poisson data. Moreover, it needs less complex steps than other charts. In Chiu and Kuo (2008), the authors presented the use of multivariate Poisson sum probability density function to define the control limits of multivariate Poisson sum chart (MPSUM chart). By their chart, monitoring multiple attribute characteristics can be done in a single chart. However, in high quality processes, numbers of defects are very low, and correlation between pairwises of two characteristics is often crucial. According to our knowledge, there are no works that have provided a chart which effectively monitors correlated characteristics. Moreover, none of the charts is mainly concerned with the illustration of how pairwises of characteristic spread which can reflect process behavior.

In this article, we propose an optimal bivariate Poisson field chart for monitoring two correlated characteristics of count data. The basic concept is defining the optimal limits of bivariate Poisson distribution and illustrating data in Poisson field. This chart provides satisfied rate of false alarms, and illustrate original values of two characteristics and changes of correlation between them. In Section 2, the basic concepts of bivariate Poisson distribution are briefly discussed. In Section 3, the basic principals of an optimal bivariate Poisson field chart are introduced. Finally, real case study and simulations are presented to illustrate the effectiveness of our control chart in Section 4, and this paper is concluded in Section 5.

2. Bivariate Poisson distribution and its estimation

Bivariate Poisson distribution (BP), which was firstly introduced by Campbell (1934), is often used for modeling pairwises of correlated count data. Let random variables X_1, X_2, X_3 are unobserved variables which follow independent Poisson distributions with parameters $\lambda_1, \lambda_2, \lambda_3$. Then, $X = X_1 + X_3$ and $Y = X_2 + X_3$ are observed pairwises which follows jointly a bivariate Poisson distribution $\mathrm{BP}(\lambda_1, \lambda_2, \lambda_3)$ with joint probability density function (Holgate, 1964; Johnson, Kotz, & Balakrishnan, 1997; Kocherlakota & Kocherlakota, 1992):

$$P_{\mathrm{BP}}(X = x, Y = y|\lambda_1, \lambda_2, \lambda_3) = e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\lambda_3}{\lambda_1 \lambda_2}\right)^i \quad (1)$$

We have, also:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad (2)$$

where n is total number of samples. Moreover, the marginal distribution of X and Y with means $\lambda_1 + \lambda_3$ and $\lambda_2 + \lambda_3$ satisfies the following recurrence relations (Holgate, 1964):

$$xP(x,y) = \lambda_1 P(x-1,y) + \lambda_3 P(x-1,y-1) \quad (3)$$
$$yP(x,y) = \lambda_2 P(x,y-1) + \lambda_3 P(x-1,y-1) \quad (4)$$

For maximum likelihood estimation, if Eq. (1) is differentiated with respect to parameters λ_1, λ_2, and λ_3, from recurrence relation in (3) and (4), the differential-different equations are given by Holgate (1964):

$$\frac{\partial P(x,y)}{\partial \lambda_1} = P(x-1,y) - P(x,y) \quad (5)$$
$$\frac{\partial P(x,y)}{\partial \lambda_2} = P(x,y-1) - P(x,y) \quad (6)$$

Table 1
Probabilistic Poisson field. $p_{00}, p_{10}, p_{01}, p_{11}, \ldots, p_{r_{max}, s_{max}}$ are probabilities of each pairwise on probabilistic Poisson field. r_{max} and s_{max} are numbers of rows and columns, respectively, where all probabilities are equal to zero.

x	y							
	0	1	2	3	.	.	.	s_{max}
0	p_{00}	p_{01}	p_{02}	p_{03}	.	.	.	0
1	p_{10}	p_{11}	p_{12}	p_{13}	.	.	.	0
2	p_{20}	p_{21}	p_{22}	p_{23}	.	.	.	0
3	p_{30}	p_{31}	p_{32}	p_{33}	.	.	.	0
.	.	.	.	.	.	.	.	0
.	.	.	.	.	.	.	.	0
.	.	.	.	.	.	.	.	0
r_{max}	0	0	0	0	0	0	0	0

Table 2
The schedule of events in Calit2 building from 07/24/05 to 11/05/05.

Dates of events (month/day/year)	Starting time (hour:minute)	Finishing time (hour:minute)
7/26/2005	11:00	14:00
7/29/2005	8:00	11:00
8/2/2005	15:30	16:30
8/4/2005	16:30	17:30
8/5/2005	8:00	11:00
8/9/2005	11:00	14:00
8/9/2005	8:00	16:00
8/10/2005	8:00	16:00
8/12/2005	8:00	11:00
8/16/2005	11:00	14:00
8/18/2005	8:00	17:00
8/18/2005	18:00	20:30
8/19/2005	8:00	11:00
8/23/2005	11:00	14:00
08/26/05	08:00	11:00
08/30/05	16:00	18:00
09/01/05	14:00	16:30
09/15/05	08:30	10:00
09/21/05	09:00	14:00
09/22/05	14:00	14:30
10/03/05	15:30	17:00
10/04/05	12:00	15:00
10/07/05	09:00	10:30
10/10/05	16:30	19:00
10/14/05	09:00	10:30
10/19/05	22:00	23:30
10/21/05	09:00	10:30
10/23/05	21:00	22:30
10/24/05	08:00	12:00
10/24/05	16:00	21:00