Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 182 (2007) 1188–1202
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Stochastics and Statistics

Bayesian inspection model with the negative binomial prior in the presence of inspection errors

Young H. Chun *, Robert T. Sumichrast

Department of Information Systems and Decision Sciences, E. J. Ourso College of Business, Louisiana State University, Baton Rouge, LA 70803-6316, United States

Received 17 February 2006; accepted 2 September 2006
Available online 13 December 2006

Abstract

One of the basic assumptions in Bayesian inspection models is that we have some prior knowledge about the number of defects in a certain product or software system. The prior knowledge can be often described as a probability distribution (e.g., Poisson distribution). In the paper, we propose three conditions that should be put forth as desirable properties for a prior probability distribution of the number of defects in the product. We review various prior probability distributions and test if they meet those conditions. The negative binomial distribution is found to be the only one that satisfies all the desirable conditions. With the negative binomial prior, we analyze the effects of various parameters on the Bayesian estimate of the number of undetected errors still remaining in the product.

Keywords: Statistical quality control; Inspection errors; Bayesian inspection; Negative binomial distribution

1. Introduction

Suppose that a certain complex product or a software system is being inspected for any defects, errors, or non-conformities. Let Z denote the total number of defects in the product, where $Z = 0, 1, \ldots, \infty$. Due to inspection errors, all the defects are not detected during the inspection process. Based on the number of defects x discovered during the inspection process, we want to estimate: (i) the unknown number of defects, Z, and (ii) the unknown number of defects still remaining in the product, $Z - x$.

The estimation problem is an important aspect of quality control in manufacturing, in which each product is inspected via visual or machine-aided techniques. When the inspector is unable to identify every defect with certainty, estimation of the number of undiscovered defects in a complex product (e.g., circuit board, automobile, airplane, or mobile home) is the first step in setting quality assurance levels for the product.

Estimating the number of undiscovered errors is also an important task in software engineering. In software reliability, the defects correspond to program errors, faults, or "bugs" that can be detected and removed by a

* Corresponding author. Tel.: +1 225 578 2506.
E-mail address: chun@lsu.edu (Y.H. Chun).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.09.081

programmer, using some combination of visual scanning of program codes and experimental running of the program on typical input. The accurate estimation of the number of undetected errors "not only helps certify the application-readiness of the software, but also provides an indication of the effort that will be needed for customer support and for the upgrading of future program releases." (Jewell, 1985)

One of the widely used estimation methods is the Bayesian method, in which the inspector's prior knowledge on Z is represented by a probability mass function. The prior distribution of Z may be a specific member of a family of subjective prior distributions. Under reasonably general conditions, for example, Dorris and Foote (1978, p. 186) claimed that "the distribution of the number of defects per inspection unit is Poisson." Other popular choices are the non-informative or improper distribution, compound Poisson distribution, negative binomial distribution, Neyman distribution, and so on.

In the paper, we propose three major conditions that should be put forth as desirable properties for a prior distribution of the number of defects. We then examine various prior distributions proposed in the open literature and check if they meet those conditions. Surprisingly enough, some of the most widely used probability distributions, such as the Poisson distribution, fail to satisfy all the three conditions. For example, consider the Rallis and Landsdowne (2001) inspection plan in which the number of errors in a software document is described as a Poisson distribution. In such a case, the number of undetected errors still remaining in the software document is independent of the number of errors discovered during the inspection. Rallis and Landsdowne (2001, p. 1056) asserts that "this result seems remarkable because it gives a circumstance in which the statistical confidence from a Bayesian analysis is actually independent of all observed data."

The remarkable result is simply attributed to their choice of the Poisson distribution, which fails to satisfy one of the three conditions proposed in this paper. The use – or misuse – of such distributions has caused some confusion and misunderstanding among quality professionals about inspection plans for attributes data. This paper is intended to identify the prior probability distribution that satisfies all the desirable conditions and clear up the confusion and misunderstanding about the usefulness of Bayesian inspection plans for count data.

In the following section, we review several types of Bayesian inspection plans considered in the literature. We propose in Section 3 three conditions that should be put forth as desirable properties for a prior probability distribution of the number of defects. In Section 4, we then review various prior distributions and test if they meet those conditions. We recommend the negative binomial distribution which is the only one that satisfies all the desirable conditions, and analyze the effects of its parameter values on the Bayesian estimate of the number of undetected errors in Section 5. Concluding remarks are presented in Section 6. The proof of a proposition is given in Appendix to improve the readability of the paper.

2. Bayesian inspection plan

There are two types of attribute or count data in quality control that deserve special attentions. One is the number of "defective items" in a production lot, and the other is the number of "defects" in a product. Nowadays, the terminology, "non-conforming items" and "non-conformities," have become more popular. In the restricted technical sense of the word, however, we call them defective items and defects throughout this paper. In the paper, we are concerned with the problem of inspecting a product and estimating the number of "defects" in the product, which should be contrasted with the problem of sampling items from a lot and estimating the number of "defective items" in the lot.

In the latter case, each item is either defective or non-defective, and we are interested in the total number of defective items Z in the lot, where $Z = 0, 1, \ldots, N$. In acceptance sampling, we take a sample of n items from the lot and only inspect the items in the sample. Based on the number of defective items x in this sample, we can estimate the lot quality and make a decision regarding the lot. This decision is either to accept or to reject the lot. Sometime we refer to this decision as lot sentencing.

Such an acceptance sampling plan has been treated extensively in the Bayesian framework (Starbird, 1994), in which the total number of defective items Z in the lot is described as a prior distribution $P[Z]$. Some of the most popular prior distributions for the number of defective items are the discrete uniform distribution, binomial distribution, two-point binomial distribution, and beta-binomial (or Polya) distribution. Mood (1943) and Hald (1960, 1981) studied various types of prior distribution in the context of acceptance sampling, but they did not consider inspection errors.