A Diversity Monitor with Known Errors for Process Variability Observed in Categorical Data

JUSTIN R CHIMKA, QIN HONG, and LAKSHMI JETTY

Department of Industrial Engineering, University of Arkansas, Fayetteville, AR, USA

(Received on November 19 2010; revised on March 15 and April 6, 2011)

Abstract: In this article we extend quality control research into nominal and ordinal data from simple monitors of location to those of variability. Given ordinal data more traditional process control relies on demerit systems that explicitly monitor central location but not the distribution spread. It is quantified here in terms borrowed from ecology. An established index of diversity and its standard error are the basis for a new quality control chart that we have also assessed with respect to error rates.

Keywords: binomial distribution, categorical data, demerits, diversity, quality control

1. Introduction

Ordinal and nominal data are measured on scales that distinguish between different categories, but numerical values assigned to them are arbitrary which calls into question the usefulness of most traditional engineering statistics. Furthermore theoretically appropriate methods to analyze ordinal and nominal scale variables lag behind the development of quality technology for interval and ratio data. When quality is measured on an ordinal scale demerits control charts are used to monitor the process average [1, 2, 3, 4, 5]. Relevant application contexts range from more traditional inspection and production to medication error severity [6, 7]. However there has not been previously available a chart to explicitly monitor the variability of a quality characteristic measured on an ordinal or nominal scale.

Assume a multinomial population has \( N \) categories with proportion \( \pi_j \) of category \( j \). If two observations are randomly chosen with replacement from the population, an index of diversity \( I(\pi) \) is the probability that two observations belong to different categories, with summation from \( j = 1 \) to \( N \) [8]:

\[
I(\pi) = 1 - \sum \pi_j^2
\]

(1)

Based on \( n \) observations of the categorical variable the standard error of \( I(\pi) \) is given by Agresti [9]:

\[
2 \left( \sum \pi_j^3 - \left( \sum \pi_j^2 \right)^2 \right) / n)^{1/2}
\]

(2)

So in practice this could lead to 3σ limits around a centerline function of \( \pi \):

\[
1 - \sum \pi_j^2 \pm 6 \left( \sum \pi_j^3 - \left( \sum \pi_j^2 \right)^2 \right) / n)^{1/2}
\]

(3)

2. Centerline

Consider a multinomial population having \( N = 4 \) categories with proportion \( \pi_j \) of category \( j \) given by the binomial distribution with \( t = 3 \) trials and probability of success \( p \) on each trial. We use the binomial distribution to systematically generate arbitrary yet realistic density functions associated with the ordinal variable. The categorical variable describes defects named Class A, Class B, Class C, and Class D: Class A is associated with three successes and represents a serious defect; Class B is associated with two successes and represents a moderate defect; Class C is associated with one success and represents a minor defect; Class D is associated with zero successes and represents practically no defects. Now we can systematically define the probability of observing each class defects by arbitrarily defining a single parameter \( p \), and for every binomial \( p \) we can find an index of diversity \( I \).

*Corresponding author’s email: jchimka@uark.edu
For example of binomial \((t = 3, p = 0.3)\): \(p_0 = 0.343, p_1 = 0.441, p_2 = 0.189, p_3 = 0.027\). The chance of observing a serious defect is 34.3%; the chance of observing a moderate defect is 44.1%; the chance of observing a minor defect is 18.9%; the chance of observing no defect is 2.7%. According to Equation (1) the index of diversity associated with the binomial \((3, 0.3)\) probabilities is \(I = 0.65142\). In other words given two opportunities to observe a defect the chance that will result in observing different class defects is approximately 65%. Diversity in this sense refers to the variability of a quality characteristic measured on an ordinal or nominal scale.

3. Control Limits

We can exploit the standard error of \(I\) by assuming a constant value for \(n\) observations (of the multinomial population) that will serve as sequential samples to monitor process variability. To extend our example of binomial \((3, 0.3)\) if \(n = 3\), according to Equation (2) the standard error is approximately 0.12319. Now according to Equation (3) the \(3\sigma\) limits are approximately \((0.28184, 1.021)\) in theory. In Table 1 we show the following for different binomial \(p\) assuming \(t = 3\), and \(n = 3\): defect class probabilities \(P(A), P(B), P(C)\) and \(P(D)\); index of diversity \(I\); approximate standard error \(SE(I)\), and approximate control limits LCL and UCL.

<table>
<thead>
<tr>
<th>(P)</th>
<th>(P(A))</th>
<th>(P(B))</th>
<th>(P(C))</th>
<th>(P(D))</th>
<th>(I)</th>
<th>(SE(I))</th>
<th>LCL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.729</td>
<td>0.243</td>
<td>0.027</td>
<td>0.001</td>
<td>0.40878</td>
<td>0.26394</td>
<td>(0.38304)</td>
<td>1.20060</td>
</tr>
<tr>
<td>0.2</td>
<td>0.512</td>
<td>0.384</td>
<td>0.096</td>
<td>0.008</td>
<td>0.58112</td>
<td>0.14727</td>
<td>0.13932</td>
<td>1.02292</td>
</tr>
<tr>
<td>0.3</td>
<td>0.343</td>
<td>0.441</td>
<td>0.189</td>
<td>0.027</td>
<td>0.65142</td>
<td>0.12319</td>
<td>0.28184</td>
<td>1.02100</td>
</tr>
<tr>
<td>0.4</td>
<td>0.216</td>
<td>0.432</td>
<td>0.288</td>
<td>0.064</td>
<td>0.67968</td>
<td>0.12777</td>
<td>0.29636</td>
<td>1.06300</td>
</tr>
<tr>
<td>0.5</td>
<td>0.125</td>
<td>0.375</td>
<td>0.375</td>
<td>0.125</td>
<td>0.68750</td>
<td>0.12500</td>
<td>0.31250</td>
<td>1.06250</td>
</tr>
</tbody>
</table>

For binomial \(p = 0.1\) and other binomial \(p < 0.2\) our diversity monitor is not practical, because there is no statistic \(0 < I < 1\) that would exist outside the control limits. Also in no case would we expect to reject the hypothesis of control due to exceedingly great diversity, because the statistic \(I\) will not be greater than 1. For most binomial \(p > 0.1\) the diversity monitor seems to include a practical lower control limit.

4. Diversity Statistic

Given \(n = 3\) observations in a sample of the categorical quality characteristic with respect to diversity there are only three possible outcomes: 1) each of the three observations shares the same category, 2) two observations share the same category, 3) there are no shared categories among observations. Now if we consider diversity strictly according to the sample of three observations the index of diversity will take on one of three possible values:

1. When each of the three observations shares the same category, the statistic \(I' = 0\), because there is no chance to observe different class defects in the sample.
2. When two observations share the same category, \(I' = 1 - (2 / 3)^2 - (1 / 3)^2 \approx 0.44444\).
3. Without shared categories among observations, \(I' = 1 - (1 / 3)^2 - (1 / 3)^2 - (1 / 3)^2 = 2/3\).

In reference to the practical lower control limits in Table 1 it is clear that when each of the three observations shares the same category, the hypothesis of control will be rejected. Otherwise we will not reject the hypothesis of control based on a sample of size \(n = 3\). Greater sample sizes would 1) increase computational intensity with number of possible outcomes associated with the diversity statistic, and 2) decrease its standard error. Figure 1 illustrates how the diversity monitor is sensitive to \(n\) observations of the categorical variable.

5. Errors

We can find the Type I errors associated with rows in Table 1 by computing the probabilities of observing three of the same categories. For example binomial \(p = 0.3\) the Type I error rate is \((0.343)^3 + (0.441)^3 + (0.189)^3 + (0.027)^3 \approx 0.13289\).
A defects, three Class B defects, three Class C defects, and three Class D defects. Table 2 shows approximate Type I error rates associated with 3σ limits for different binomial p.

![Figure 1: Standard Error of Diversity Index versus Observations of the Categorical Variable for Binomial p = 0.5, assuming t = 3]

It follows the probability of failing to detect a shift in binomial p is the probability of a Type II error, not observing three of the same categories, or the complementary probability $\beta = (1 - \alpha)$ associated with the new binomial p. Table 3 shows approximate Type II error rates associated with shifts to binomial $p'$ no matter what binomial p was assumed to find the practical lower control limit.

<table>
<thead>
<tr>
<th>Bin p</th>
<th>$P$ (Type I error)</th>
<th>Bin $p'$</th>
<th>$P$ (Type II error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.40179</td>
<td>0.1</td>
<td>0.59821</td>
</tr>
<tr>
<td>0.2</td>
<td>0.19173</td>
<td>0.2</td>
<td>0.80827</td>
</tr>
<tr>
<td>0.3</td>
<td>0.15289</td>
<td>0.3</td>
<td>0.86711</td>
</tr>
<tr>
<td>0.4</td>
<td>0.11485</td>
<td>0.4</td>
<td>0.88515</td>
</tr>
<tr>
<td>0.5</td>
<td>0.10938</td>
<td>0.5</td>
<td>0.89062</td>
</tr>
</tbody>
</table>

So we have established a diversity monitor with known errors for process variability in categorical data, and economic design assuming equal cost errors suggests the monitor is best for detecting increasingly dramatic departures from a symmetrical categorical random variable.

References


