

Overdispersion at the Binomial and Multinomial Distribution

Emese Vágó / Sándor Kemény / Zsolt Láng

Received 2011-01-23

Abstract

Overdispersion is a widely discussed phenomenon in case of binomial and Poisson distributed data. We analysed multinomial data with varying multinomial parameters as a part of attribute gauge study. In this situation overdispersion is expected to occur. However in case of nonrepeated observations it is not present. We investigated and proved the “lack” of overdispersion in case of binomial and multinomial variables if there are no repeated observations. The explanation is a “compensation effect”, which only occurs in case of binomial and multinomial variables and does not occur for Poisson distributed data.

Keywords

Underdispersion · overdispersion · attribute gauge · clinical trial · binomial · multinomial and Poisson distributions

1 Introduction

The most frequent design in medical statistics is the clinical trial. A part of the patients selected for study is exposed to a treatment, to which some of them react, others do not. The other part of patients is exposed to the control (what may be another treatment or lack of treatment), and again some of them react, others do not. The typical question is if the proportion of reacting patients to the treatment significantly differs from that to the control. In hypothesis test context the question is if the binomial parameter of the treated population is equal to that of the control population. Agresti [1] hints that neither the results of the treated patients nor the results of the control patients are from binomial populations. The reason obviously is that e.g. the treated patients are not homogeneous, the distribution of outcomes is a mixture of binomial distributions. This would lead to the well-known phenomenon of overdispersion. Many literature sources deal with the diagnostics and modelling of overdispersion such as Hinde, J. and Demetrio, C.G.B. [2], Cox, D. R. [3], Ganio, L. M. and Schaffer D. W. [4] and Barron, D. N. [5].

A multinomial example is the attribute examination of product items. A part of these items is conforming, another part is non-conforming, the outcome of the assignment is that the item is accepted or rejected. This outcome is fourfold: a good item is either accepted or rejected, the same possibilities hold for a bad item. Again the outcomes (items and decisions) do not stem from a single multinomial distribution, one has to face overdispersion here as well.

2 How we got surprised (the research topic)

Our research topic is the analysis of attribute measurement systems (Gauge R&R).

When using a single limit attribute measurement system, there are two different decision categories: we can reject or accept the measured item. Though the decision is categorical there is often a continuous variable in the background. The probability of acceptance depends on this continuous variable (reference value). E.g. the geometric size of a part must be above the specification limit in order to be conforming. Another example for an attribute type decision is the sensory evaluation of ampoules.

Emese Vágó

Department of Chemical and Environmental Process Engineering, BME, Budapest H-1521, Hungary
e-mail: evago@mail.bme.hu

Sándor Kemény

Department of Chemical and Environmental Process Engineering, BME, Budapest H-1521, Hungary
e-mail: kemeny@mail.bme.hu

Zsolt Láng

Department of Biomathematics and Informatics Faculty of Veterinary Science, Szent István University, Budapest 2103, Hungary
e-mail: lang.zsolt@aotk.szie.hu

In practice the operator investigates the ampoules visually and decides if there are any solid particles in them. If there are solid particles, the ampoule is not useable.

In both examples there is a continuous characteristic behind the attribute type decision. The geometric size can be measured with a coordinate measurement machine, but in practice this kind of measurement would be too expensive and time consuming for performing it on shop floor. Thus they use a simple go/no-go gauge instead. In the ampoule example the continuous characteristic behind the decision is the size of the solid particles in the ampoule. If they are large enough they can be observed by human eye. In the following the continuous characteristic behind the decision is referred as size or reference value.

The probability of accepting a part depends on the reference value. The “very big parts” will be nearly always accepted, the “very small parts” will be nearly always rejected. There is a reference value interval (grey zone), in which the probability of acceptance strongly depends on the reference value. The connection between the probability of acceptance and the reference value is called gauge performance curve. Its mathematical form is assumed to be given by Eq. (1).

$$p(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \quad (1)$$

x : reference value

p : probability of acceptance

α, β : model parameters

The graphical form of the gauge performance curve is presented in Fig. 1. The different curves belong to different α and β model parameters. The figure demonstrates that the position and shape (or slope) of the curve depends on α and β . This connection is discussed detailed by Vágo and Kemény [6].

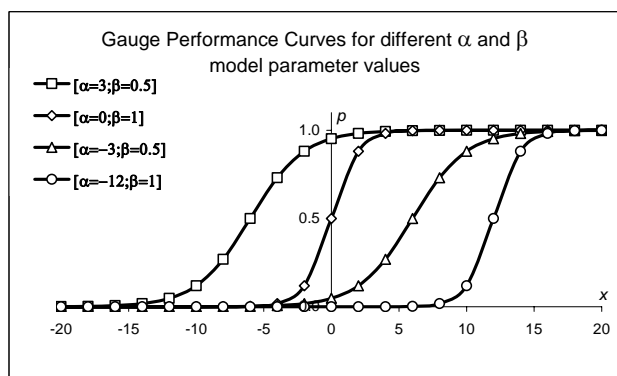


Fig. 1. Gauge Performance Curve

The reference value (x) differs from part to part. There are two sources of variation in the outcome: the difference between items and the random fluctuation of the result of the trial on a specific item. The two sources together result in an excess variation as compared with the variation obtained for identical items having parameters of marginal distributions. The phenomenon is termed overdispersion. It is expected that the marginal distribution of outcomes (accept/reject) may not be binomial, that is

the variance may not be given by the parameter obtained from the expected value.

Our purpose was to characterize the measurement system, more specifically to quantify the probability of wrong decision. These are that an accepted part is bad or a rejected part is good.

In the typical design of attribute gauge study (Automotive Industry Action Group’s measurement systems analysis (AIAG MSA) Manual [7]) several (n : say 50) items are selected in a random way from a large population of products, each of them is rated several (k) times. The outcomes of these ratings form the four categories of the distribution, they are summarized in a 2x2 table (Table 1).

Tab. 1. Outcome Categories

	Accept	Reject
Good Part	N_1	N_2
Bad Part	N_3	N_4

N_1, N_2, N_3, N_4 denote the number of occurrences in the appropriate cell. E.g. it happened N_1 times that a good part received “accept” decision. Thus the sum of cell counts (N_1, N_2, N_3, N_4) is equal to the total number of decisions ($k \cdot n$). In Table 1 cells #2 and #3 represent the wrong decisions. Again as items are different, excess variation is expected as compared with the multinomial distribution.

When the variance of estimators is calculated the covariance matrix of $\bar{N} = (N_1, N_2, N_3, N_4)$ is of interest. Simulation studies were performed in order to quantify the covariance matrix with different k number of repeated ratings, compared with the covariance matrix of a single multinomial distribution with average parameters. It was surprisingly found that if $k=1$ (single rating of all sampled items) there was no excess variation, that is no overdispersion was found.

Mathematical reasoning and proofs are given in the forthcoming sections.

3 Overdispersion at the binomial and at the Poisson case

The typical example in textbooks is the coin tossing.

If y_i denotes the outcome of the i -th experiment (1 in case of occurrence, 0 otherwise), and μ is the probability of the event (success, e.g. head), the expected value and variance of number of successes (number of heads obtained) ($y = \sum_i^n y_i$) in n experiments are

$$E(y) = n\mu$$

$$\text{Var}(y) = n\mu(1 - \mu)$$

In some situations the items on which the experiments are performed are not identical. E.g. when the effect of a medical treatment is investigated, the patients are not the same, the product

items qualified are not the same, etc. The corresponding picture is tossing coins taken from a population of different coins (Box, Hunter and Hunter [8]). (When we make a new experiment a new sample is taken from the coins.) In these cases there are two sources of variation: the random fluctuation of the binomial parameters of coins and the random fluctuation of the result of the trial on a specific item. The two sources together result in an excess variation as compared with the variation obtained for identical items termed overdispersion.

According to the theorem of total variance the resulting variance is expressed as ([1, p 8],):

$$\text{Var}(y) = E[\text{Var}(y|\boldsymbol{\mu})] + \text{Var}[E(y|\boldsymbol{\mu})] \quad (2)$$

where $\boldsymbol{\mu}$ is the vector of size n of binomial parameters.

Two cases will be distinguished: all n items in the sample are checked once or with k repetitions.

3.1 All Items are checked once

The picture is that a sample is taken from different (biased) coins and the coins are tossed, once each.

The first term

The y_i outcomes at μ_i binomial parameters are independent, thus

$$\text{Var}(y|\boldsymbol{\mu}) = \sum_i^n \text{Var}(y_i|\mu_i) = \sum_i^n \mu_i(1-\mu_i)$$

The μ_i binomial parameters come from the same population:

$$\begin{aligned} E[\text{Var}(y|\boldsymbol{\mu})] &= E\left[\sum_i^n \mu_i(1-\mu_i)\right] = \\ &= n[E(\mu)][1-E(\mu)] - n\text{Var}(\mu) \end{aligned}$$

The first product would be the variance if all elements were identical characterized with $E(\mu)$ binomial parameter.

The second term

$$\begin{aligned} E(y|\boldsymbol{\mu}) &= E\left(\sum_i^n y_i|\mu_i\right) = \sum_i^n \mu_i \\ \text{Var}[E(y|\boldsymbol{\mu})] &= \text{Var}\left(\sum_i^n \mu_i\right) = n\text{Var}(\mu) \end{aligned}$$

The sum of the two terms:

$$\begin{aligned} \text{Var}(y) &= \\ &= n[E(\mu)][1-E(\mu)] - n\text{Var}(\mu) + n\text{Var}(\mu) \\ &= n[E(\mu)][1-E(\mu)] \end{aligned}$$

The surprising result is that if a random sample of elements is taken with replacement or from an infinite population, and all exposed once, there is no overdispersion.

Agresti ([1, pp. 493-493]), used this phenomenon implicitly when modelling binary matched pairs. Here the items within each group have different binomial parameters. However, the marginals of each group are considered as binomial with binomial parameter equal to the expected value of the binomial parameters within the group.

It is worth remarking the special case when both the population and the sample contain the same n elements. The corresponding picture is tossing each different coin once. In a new experiment the same coins are tossed. In this case the only source of variation is the random fluctuation of the result of the trial on a specific item.

In this case the second term of Eq. (2) is missing as the items are fixed, thus their binomial parameter is not random. The first term:

$$\begin{aligned} E[\text{Var}(y|\boldsymbol{\mu})] &= \sum_i^n \mu_i(1-\mu_i) \\ &= n\bar{\mu}(1-\bar{\mu}) - n\text{Var}(\mu) \end{aligned}$$

where $\text{Var}(\mu) = \frac{1}{n} \sum_i^n (\mu_i - \bar{\mu})^2$ is the variance of the finite population. The conclusion is that there is underdispersion, as described by Box et al. ([8, pp. 135-137]), as compared to the variance of a sample of n identical elements having the same $E(\mu)$ parameter.

The phenomenon discussed in fact is a compensation effect: the first term of the total variance expression shows underdispersion. The second term gives the overdispersion, and this may compensate or over-compensate the underdispersion.

All items are checked k times

The picture is that a sample is taken from different (biased) coins and the sampled coins are tossed, k times each, with result y_{ij} .

The first term of the total variance expression is

$$\text{Var}(y|\boldsymbol{\mu}) = \sum_i^n \sum_j^k \text{Var}(y_{ij}|\mu_i) = k \sum_i^n \mu_i(1-\mu_i)$$

$$E[\text{Var}(y|\boldsymbol{\mu})] = kn[E(\mu)][1-E(\mu)] - kn\text{Var}(\mu)$$

The second term

$$E(y|\boldsymbol{\mu}) = E\left(\sum_i^n \sum_j^k y_{ij}|\mu_i\right) = k \sum_i^n \mu_i$$

$$\text{Var}[E(y|\boldsymbol{\mu})] = \text{Var}\left(k \sum_i^n \mu_i\right) = k^2 n \text{Var}(\mu)$$

The sum of the two terms

$$\text{Var}(y) = n[E(\mu)][1-E(\mu)] + k(k-1)n\text{Var}(\mu)$$

The result is that overdispersion occurs only when exposition of items is repeated ($k > 1$). This situation may not be typical in most fields of application (including biomedical experiments)

but it occurs in case of the attribute gauge R&R study. The typical experiment in that area consists of repeated ratings of different product items by several operators.

In the special case when both the population and the sample contain n elements (the binomial parameters are not random) the variance is given with Eq (3):

$$\begin{aligned} E[\text{Var}(y|\boldsymbol{\mu})] &= \sum_i^n \sum_j^k \mu_i (1 - \mu_i) = \\ &= kn\bar{\mu} (1 - \bar{\mu}) - kn \text{Var}(\boldsymbol{\mu}) \end{aligned} \quad (3)$$

In case of Poisson distribution similar situation can not occur, as there the *first term* of the total variance does not have a negative part. (λ denotes the Poisson parameter.)

$$\begin{aligned} \text{Var}(y|\boldsymbol{\lambda}) &= \sum_i^n \sum_j^k \text{Var}(y_{ij}|\lambda_i) = k \sum_i^n \lambda_i \\ E[\text{Var}(y|\boldsymbol{\lambda})] &= kn [E(\lambda)] \end{aligned}$$

The second term

$$\begin{aligned} E(y|\boldsymbol{\lambda}) &= E\left(\sum_i^n \sum_j^k y_{ij}|\lambda_i\right) = k \sum_i^n \lambda_i \\ \text{Var}[E(y|\boldsymbol{\lambda})] &= \text{Var}\left(k \sum_i^n \lambda_i\right) = k^2 n \text{Var}(\lambda) \end{aligned}$$

The sum of the two terms

$$\text{Var}(y) = kn ([E(\lambda)] + k \text{Var}(\lambda))$$

If the Poisson parameter is not constant the variance also exceeds that in the homogenous case.

4 Overdispersion at the multinomial case

In some other experimental situations there are more than two outcome categories.

An example of our interest is the attribute gauge R&R study where several (e.g. 50) product items are rated by several operators several times. The result of the study is the number of good items classified as good, the number of good items classified as bad etc, altogether 4 categories.

The result is analogous to the binomial case: there is no overdispersion if the rating is not repeated. Moreover, the pooled outcome probabilities are the population averages of the outcome probabilities of individual items, so the outcome frequencies follow multinomial distribution with these averaged probabilities. In the special case of binary outcomes the resulting distribution is binomial.

In the following we calculate the covariance matrix of the frequencies $\mathbf{Y}^T = (Y_1, \dots, Y_m)^T$ of an experiment repeated k times with m possible outcome categories conducted on N objects selected independently and with replacement. We can suppose without loss of generality that $n = 1$, for the objects are selected independently. In the general case the covariance matrix

is multiplied by n . Denote $\boldsymbol{\mu}^T = (\mu_1, \dots, \mu_m)^T$ the individual outcome probabilities of a randomly selected object. We have

$$\text{Cov}(\mathbf{Y}) = E(\text{Cov}(\mathbf{Y}|\boldsymbol{\mu})) + \text{Cov}(E(\mathbf{Y}|\boldsymbol{\mu})).$$

The conditional distribution of $\mathbf{Y}|\boldsymbol{\mu}$ is multinomial with order k and parameter $\boldsymbol{\mu}$, so $E(\mathbf{Y}|\boldsymbol{\mu}) = k\boldsymbol{\mu}$ and $\text{Cov}(\mathbf{Y}|\boldsymbol{\mu}) = k\{\text{diag}(\boldsymbol{\mu}) - \boldsymbol{\mu}\boldsymbol{\mu}^T\}$. Here $\text{diag}(\boldsymbol{\mu})$ is a diagonal matrix with $\boldsymbol{\mu}$ in its diagonal. From this we have

$$\begin{aligned} \text{Cov}(\mathbf{Y}) &= \\ &= k(\text{diag}(E(\boldsymbol{\mu})) - E(\boldsymbol{\mu}\boldsymbol{\mu}^T)) + k^2 \text{Cov}(\boldsymbol{\mu}). \end{aligned}$$

Here $\text{Cov}(\boldsymbol{\mu}) = E(\boldsymbol{\mu}\boldsymbol{\mu}^T) - E(\boldsymbol{\mu})E(\boldsymbol{\mu})^T$, thus we obtain

$$\begin{aligned} \text{Cov}(\mathbf{Y}) &= k(k-1) \text{Cov}(\boldsymbol{\mu}) \\ &+ k(\text{diag}(E(\boldsymbol{\mu})) - E(\boldsymbol{\mu})E(\boldsymbol{\mu})^T). \end{aligned}$$

In the general case when n can be greater than 1 we have

$$\begin{aligned} \text{Cov}(\mathbf{Y}) &= nk(k-1) \text{Cov}(\boldsymbol{\mu}) \\ &+ nk(\text{diag}(E(\boldsymbol{\mu})) - E(\boldsymbol{\mu})E(\boldsymbol{\mu})^T) \end{aligned}$$

The second term in the right hand side is the covariance matrix of a multinomial distribution with order nk and parameter $E(\boldsymbol{\mu})$, hence the first term represents the extra variation due to the variance of the individual outcome probabilities $\boldsymbol{\mu}$.

5 Conclusion

Both for binomial and multinomial cases it has been proven that overdispersion occurs only if repeated ratings are performed on the same items. In other words the marginal distribution is binomial or multinomial, if there is no repeated rating, in spite of the fact that the conditional distributions are different from item to item.

The same phenomenon does not occur at Poisson distribution, as here the expression of variance is of different nature (it does not have negative part).

References

- 1 **Agresti A.** *Categorical data analysis*, Wiley, New Jersey, 2002. 2nd ed.
- 2 **Hinde J, Demetrio C G B.** *Overdispersion: Models and estimation*, Computational Statistics and Data Analysis **27** (1998), no. 2, 151-170.
- 3 **Cox D R.** *Some remarks on overdispersion*, Biometrika **70** (1983), no. 1, 269-274.
- 4 **Ganio L M, Schaffer D W.** *Diagnostics for overdispersion*, Journal of the American Statistical Association **87** (1992), no. 419, 795-804.
- 5 *The Analysis of Count Data: Overdispersion and Autocorrelation*, Sociological Methodology **22** (1992), 179-220.
- 6 **Vágó E, Kemény S.** *A model based approach for attribute R&R Analysis*, Quality and Reliability Engineering International, posted on 2010, DOI 10.1002/qre.1154, (to appear in print).
- 7 *Measurement System Analysis; Reference Manual, 3rd ed.*, Detroit, MI: Automotive Industry Action Group, 2002.
- 8 **Box G E P, Hunter W. G, Hunter J. S.** *Statistics for Experimenters*, Wiley, 1978.