XIX IMEKO World Congress
Fundamental and Applied Metrology
September 6−11, 2009, Lisbon, Portugal

TESTING FOR OUTLIERS BASED ON BAYES RULE

Giampaolo E. D'Errico [1]

[1] Istituto Nazionale di Ricerca Metrologica (INRIM), Torino, Italy, g.derrico@inrim.it

Abstract − A Bayesian treatment of conjectured outlying observations is developed, using the computational device of inverse probability. The device's performance is discussed in term of posterior probability of missed or false detections. The key role of prior probability is shown through a numerical example.

Keywords: outliers, Bayesian testing

1. INTODUCTION

The problem posed by likely occurrence of outliers in measurement experiments has attracted a great deal of research long since [1], [2]. Attention is raised by both theoretical and practical aspects: in fact, an unexpected experimental result might be announcing the discovery of an unforeseen phenomenon; on the other hand, an outlier might indicate a miscalibration or fault in instrumentation, or even a trivial reporting mistake.

In the framework of statistical methods, the use of a test to screen doubtful observations from a sample implies handling with care implicit assumptions that are likely subject to violations. Fundamental issues are involved too, for example [3]: how to maintain the hypothesis of randomness with application to a screened sample set; how to preserve conditions for observations to be mutually independent, though related through the property of having resisted the screening test.

The problem can be rather expressed in term of a treatment, rather than rejection, of outliers. For example, a weighted mean can be used to take into account presumptive outliers by assigning them weights close (or equal) to zero. However this points out that averaging is not always the best choice for estimating the expected value of a probability density function (pdf). The introduction of weighting factors supports the concept that all available information should be processed.

In probability calculus, the acknowledgement of this concept is conveyed by use of conditional probabilities and prior probabilities according to the so-called Bayes rule.

In the present paper, the problem of outliers treatment is approached from such a Bayesian point o view [4].

The paper is organised as follows.

In the next Section 2, the problem of outliers is set in the context of statistical hypothesis testing, based on Neyman-Pearson method (see, e. g., [5]). In Section 3, this method – challenged by a long-standing dispute about apparent fallacies and misuse (a recent criticism is detailed in [6]) – is contrasted with a Bayesian treatment. This treatment is here developed with application to a numerical example. Section 4 contains some concluding remarks.

2. POTENTIAL OUTLIERS AND HYPOTHESIS TESTING

The problem of outliers detection can be modelled in the framework of statistical hypothesis testing. The hypothesis under scrutiny is that some values observed by sampling from a given population can be considered as probable outliers – with respect to the population's pdf.

In order to allow formal definitions and related computations, a summary of basic concepts and technicalities is required.

According to Neyman-Pearson method, the rational for testing statistical hypotheses can be summarized as follows (see, e. g., [5]).

Suppose that the observation of a random variable Z gives rise to a set $d=\{d_1, \ldots, d_n\}$ of results representing an n-dimensional data point in the sample space Ω. If a region ω of Ω is selected, an hypothesis can be formulated concerning the probability P that d belongs to ω: $P(d \in \omega)$. This can be called the null hypothesis H_0, and ω the critical region.

In the simplest term, an alternative hypothesis H_1 is introduced, concerning the probability $P(d \in \varpi)$ that d belongs to the complementary region ϖ of ω in the sample space: ϖ is called the acceptance region. Here complementary region means that the set-union of ϖ and ω is the whole Ω, while their set-intersection is the empty set; thus, H_0 and H_1 are mutually exclusive and exhaustive over the hypotheses domain.

The probability value $P(d \in \omega \mid H_0) = \alpha$ is called the test size. It is also known as the level of significance (typically, α is pre-fixed at 5% or 1%, and the critical region is accordingly selected).

More, putting $P(d \in \varpi \mid H_1) = \beta$, the probability value $P(d \in \omega \mid H_1) = 1 - \beta$ is called the power of the test of the (null) hypothesis H_0 against the (alternative) hypothesis H_1.

In this framework, two types or errors may occur in testing a statistical hypothesis:

(I) it may be (with a probability α) wrongly rejected, when it is true;

(II) it may be (with a probability β) wrongly accepted, when it is false.

As regards outliers, in principle the approach is to evaluate a test statistic (e.g. Grubbs test [1]) using suspected observations. Then, this statistic is compared with the corresponding theoretical distribution generated under the null hypothesis (H_0) of unsuspected observations.

To decide, H_0 is rejected if the test result is excessively improbable (i.e., with a probability less than the significance level α).

The standard practice for dealing with outliers [7] remarks that rejection of aberrant observations should relay preferably upon physical – rather than statistical – grounds.

It is worthwhile noting that the significance level is related to the probability of the data to belong to the critical region, given the null hypothesis: $P(d \in \omega \mid H_0) = \alpha$ (and the test power is related to the probability of the data to belong to the critical region, given the alternative hypothesis).

As a matter of fact, it is also appropriate to evaluate the test performance using the criterion of the inverse probability, that is the posterior probability of the hypothesis after data d have been observed. The posterior being computed by use of Bayes rule – this is also known as a Bayesian approach.

3. A BAYESIAN TREATMENT

To develop a model of Bayesian testing for outliers, let the propositions H_0 and H_1 represent two mutually exclusive and exhaustive hypotheses under test. Let the proposition E state that the observed data belong to the critical region: E="$d \in \omega$". In this term, the test size $P(d \in \omega \mid H_0)$ translates into $P(E \mid H_0)$, and the power of the test $P(d \in \omega \mid H_1)$ into $P(E \mid H_1)$.

In the following, given two propositional variables X, Y the product XY will denote their logical conjunction ("X and Y"), the sum X+Y their logical disjunction ("X or Y"), and ¬X logical negation ("not X").

In term of propositional calculus, since $H_0 = \neg H_1$, E can be partitioned into $E = EH_0 + EH_1$.

Thus, in term of calculus of probability:

$$P(E) = P(EH_0) + P(EH_1) = P(H_0)P(E \mid H_0) + P(H_1)P(E \mid H_1). \quad (1)$$

Using Bayes rule:

$$P(H_0 \mid E) = \frac{P(H_0)P(E \mid H_0)}{P(E)} = \frac{P(H_0)P(E \mid H_0)}{P(H_0)P(E \mid H_0) + P(H_1)P(E \mid H_1)}. \quad (2)$$

It is clear that the posterior $P(H_0 \mid E)$ may be lower, equal or greater than $P(E \mid H_0)$. In particular, it depends also on the prior $P(H_0)$, that is the probability of the null hypothesis before the observation is performed. Table 1 shows some related figures.

Table 1. Posterior probabilities $P(H_0|E)$ vs. $P(E|H_0)$.

$P(H_0)$	$P(H_1)$	$P(E\|H_1)$	$P(E\|H_0)$	$P(H_0\|E)$
0,50	0,50	0,95	0,05	0,05
0,50	0,50	0,01	0,05	0,83
0,90	0,10	0,05	0,05	0,90
0,10	0,90	0,50	0,15	0,03

Since the test is aimed at detecting outliers, the involved propositions can be instantiated as follows:

- H_0="the observation is not an outlier", null hypothesis;
- H_1="the observation is an outlier", alternative hypothesis;
- E="the test result is positive for a suspected outlier", evidence.

Thus, given H_0 (respectively, H_1), E means the test gives a wrong (respectively, a correct) result. Therefore $P(E \mid H_0)$ may be conveniently rewritten as $P(\text{wrong\_test})$ and $P(E \mid H_1) = P(\text{correct\_test})$.

The posterior probability of an observation being not an outlier (H_0) given the test result is positive for a suspected outlier (E) can be computed by application of Bayes rule:

$$P(H_0 \mid E) = \frac{P(H_0)P(\text{wrong\_test})}{P(H_0)P(\text{wrong\_test}) + P(H_1)P(\text{correct\_test})}. \quad (3)$$

After that, the posterior probability of an observation being an outlier (H_1), given the test result is positive for a suspected outlier (E), is easily obtained by:

$$P(H_1 \mid E) = 1 - P(H_0 \mid E). \quad (4)$$

Beside $P(\text{wrong\_test})$ and $P(\text{correct\_test})$ – that are test's characteristics –, the prior probability of an outlier occurrence $P(H_1)$ is needed (since $H_0 = \neg H_1$, the other prior is just $P(H_0)=1–P(H_1)$) to compute Eq. (3).

If, for instance, the observation process is likely prone to a 1% rate of outlying values, then:

$$P(H_1)=0{,}01; \quad (5)$$

and:

$$P(H_0)=1–P(H_1)=0{,}99. \quad (6)$$

To retrieve a numerical example, let the test accuracy (i.e., the probability of correctly detecting a statistical outlier) be greater than 95%; this means, at least, a test power:

$$1 - \beta = P(\text{correct\_test}) = 0{,}95. \quad (7)$$

More, let the test specificity (i.e., the probability that a regular value will not be treated as an outlier) be around 98%; this translates into a significance level fixed at:

$$\alpha = P(\text{wrong\_test}) = 1 - 0{,}98 = 0{,}02\,. \qquad (8)$$

Putting these values in Eq. (3) yields:

$$P(\mathrm{H}_0 \mid \mathrm{E}) = \frac{0{,}99 \times 0{,}02}{0{,}99 \times 0{,}02 + 0{,}01 \times 0{,}95} \approx 0{,}676\,; \qquad (9)$$

and, using Eq. (4):

$$P(\mathrm{H}_1 \mid \mathrm{E}) = 1 - P(\mathrm{H}_0 \mid \mathrm{E}) \approx 0{,}324\,. \qquad (10)$$

Thus, the probability (≈32%) for a value to be an outlier – after the test has indicated it as a possible outlier – is very much lower than 95% of test power stated in Eq. (7).

Of course, this result stems from the relationship that – given two X, Y– generally (except particular cases, such as $P(\mathrm{H}_0|\mathrm{E})=P(\mathrm{E}|\mathrm{H}_0)=0{,}05$ reported in Tab. 1: first row) holds between relevant conditional probabilities:

$$P(\mathrm{X} \mid \mathrm{Y}) \neq P(\mathrm{Y} \mid \mathrm{X})\,. \qquad (11)$$

4. CONCLUSIONS

In the Bayesian perspective, a novel approach to test performance evaluation can be developed with application to the problem of screening probable outliers from a sample of observations.

To summarize, three propositions are statistically processed:

H_0="the observation is not an outlier" (null hypothesis);

H_1="the observation is an outlier", (alternative hypothesis);

E="the test result is positive for a suspected outlier" (evidence).

In Neyman-Pearson method, the test size, or significance $\alpha = P(\mathrm{E} \mid \mathrm{H}_0)$, and the power of the test, $1-\beta = P(\mathrm{E} \mid \mathrm{H}_1)$, are taken into account to estimate type I and type II errors likely affecting the test.

In the Bayesian approach, instead:

- posterior probabilities of hypotheses under test, $P(\mathrm{H}_0|\mathrm{E})$ and $P(\mathrm{H}_1|\mathrm{E})$, are used for test performance criteria;
- probabilities of type I and type II errors are processed in Bayes rule as intermediate factors $P(\mathrm{E}|\mathrm{H}_0)$ and $P(\mathrm{E}|\mathrm{H}_1)$;
- however, prior probabilities $P(\mathrm{H}_0)$ and $P(\mathrm{H}_1)$ are crucially needed – if unknown, they must be estimated – to compute posteriors.

REFERENCES

[1] F. Grubbs, "Procedures for Detecting Outlying Observations in Samples", *Technometrics*, vol. 11, nº. 1, pp. 1-21, February 1969.

[2] W. K. Kruskal, "Some Remarks on Wild Observations", *Technometrics*, vol. 2, nº. 1, pp. 1-3, February 1960.

[3] H. K. Ku, *Statistical Concepts in Metrology – With a Postscript on Statistical Graphics*, NBS Special Publication 747, Gaithersburg, MD, August 1988.

[4] E. T. Jaynes, *Probability – The Logic of Science*, Cambridge University Press, Cambridge, U. K., 2003.

[5] M. G. Kendall, and A. Stuart, *The Advanced Theory of Statistics*, (Vol. 2: Inference and Relationship, 2nd edition), Ch. Griffin and Co., London, U. K., 1967.

[6] X. R. Li, and X.-B. Li, "Common Fallacies in Applying Hypothesis Testing", *11th IEEE Conference on Information Fusion*, Cologne, Germany, 30 June-3 July 2008.

[7] ASTM Standard E 178-08, 1961, 2002 (2008) "Standard Practice for Dealing With Outlying Observations", ASTM International, West Conshohocken, PA, USA, 2008.