International Journal of Approximate Reasoning 55 (2014) 654–665

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Probabilistic inference for multiple testing

Chuanhai Liu, Jun Xie *

Department of Statistics, Purdue University, 250 N. University St., West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:
Received 25 April 2013
Received in revised form 24 September 2013
Accepted 25 September 2013
Available online 3 October 2013

Keywords:
Belief functions
Fiducial inference
Inferential model
Many-normal-means
Predictive random sets

ABSTRACT

An inferential model is developed for large-scale simultaneous hypothesis testing. Starting with a simple hypothesis testing problem, the inferential model produces a probability triplet (p, q, r) on an assertion of the null or alternative hypothesis. The probabilities p and q are *for* and *against* the truth of the assertion, whereas $r = 1 - p - q$ is the remaining probability called the probability of "don't know". For a large set of hypotheses, a sequence of assertions concerning the total number of true alternative hypotheses are proposed. The inferential model provides levels of belief without a prior for the sequence of assertions and offers a new multiple comparison procedure (MCP). The proposed method is obtained by improving Fisher's fiducial and the Dempster–Shafer theory of belief functions so that it produces probabilistic inferential results with desirable frequency properties. The new multiple comparison procedure is shown to have a comparable performance with other existing MCPs and is favorable in terms of probabilistic interpretation. The proposed method is applied in identifying differentially expressed genes in microarray data analysis.

1. Introduction

There have been tremendous research efforts made in the last decade on solving large-scale simultaneous hypothesis testing, where one is concerned with a large number n of pairs of competing hypotheses: $H_0^{(i)}$ versus $H_a^{(i)}$ for $i = 1, \ldots, n$. The multiple testing problem is introduced by modern scientific techniques, for example, gene expression microarray in identifying differentially expressed genes from a large number of candidates or even the whole genome. Existing efforts have been made for a series of multiple comparison procedures (MCPs), for example, controlling generalized family-wise error rate (gFWER) or controlling false discovery rate [5,12,27,28,20]. Dudoit and Laan's book [10] provides good descriptions of the existing MCPs.

An alternative way of thinking about the multiple testing problem is to consider a sequence of assertions:

$$\mathscr{A}_k = \left\{\text{there are at least } k \; H_a^{(i)}\text{'s that are true}\right\}$$

for $k = 1, 2, \ldots, n$. We will develop a probabilistic inference for this type of assertions and will use the probabilistic inference to construct a multiple comparison procedure. We start with a single test for a null hypothesis H_0 versus an alternative hypothesis H_a.

The classic frequency theory of hypothesis testing developed by Neyman, Pearson, and Fisher has been known as the twentieth century's most influential piece of applied mathematics [6,11]. However, the p-value, computed from an observed test statistic assuming the truth of the null hypothesis, does not have a desirable probability interpretation of whether or not the null hypothesis is true. In fact, the interpretation of p-value is so confusing for nonstatisticians that many falsely think it

* Corresponding author.
E-mail addresses: chuanhai@purdue.edu (C. Liu), junxie@purdue.edu (J. Xie).

0888-613X/$ – see front matter
http://dx.doi.org/10.1016/j.ijar.2013.09.017

is a probability of H_0. In the context of Bayesian hypothesis testing, Bayes factors are often computed to measure evidence in favor one over the other hypothesis. However, like Fisher's p-value, Bayes factors do not have a desirable probability interpretation.

An attractive approach is to seek prior-free probabilistic inferential methods. Ronald R. Fisher, the founding father of the modern statistics, took a lead in this direction with his fiducial inference [16,17]. Intensive and stimulating investigations in 1940–1960s on fiducial had led to the nowadays commonly perceived conclusion that "fiducial inference stands as R.A. Fisher's one great failure" [32]. On the other hand, researchers have been inspired by the interesting ideas in fiducial inference. Perhaps due to the challenge in modern scientific inference, especially with high-dimensional problems such as multiple testing discussed in this article, it is exciting to see renewed efforts in generalizing and modifying Fisher's fiducial. These include the Dempster–Shafer theory [26,8], generalized fiducial [18,19], algorithmic inference [1–3], exact confidence interval methods [4], confidence distributions [30,31], belief functions with good frequency properties [9], and inferential models [21,14,23]. For a brief review of these methods, see Liu and Xie [25].

In this paper, we tackle the problem of multiple testing from a general perspective of producing uncertainty assessments and derive a probabilistic inferential model for it. To be more precise, we follow Dempster [8] to view that probabilistic inference for the truth of H_0 or H_a amounts to producing a probability p for the truth of H_0, a probability q for the truth of H_a, and a residual probability r, called the probability of "don't know", for neither H_0 nor H_a. That is, the triplet (p, q, r) is our uncertainty assessment of H_0 and H_a. As in [21], we modify Fisher's fiducial [17,18] to obtain a new probabilistic inferential model, so that the triplet (p, q, r) has desirable frequency properties. The modified inferential framework is called Inferential Model (IM). It is closely related to the Dempster–Shafer theory of belief functions [7,26]. Compared to the notions of *Belief* and *Plausibility* in belief functions [26,23], p equals *Belief* and $p+r$ equals *Plausibility* of H_0. IM provides direct statistical evidence for H_0 and H_a. Most importantly, the (p, q, r) triplet is calculated from the specification of our uncertainty on unknown model parameters but is not the conditional probability under either the truth of H_0 or the truth of H_a.

In Section 2, we introduce the framework of IM for single hypothesis testing. We specially discuss the interpretation of the probabilities used in IM, which are considered as degrees of belief for the null or alternative hypothesis. In Section 3, we study the many-normal-means problem, where the inferential model is used for multiple testing. We show that IM offers a new multiple comparison procedure and has comparable performances with other popular MCPs. In Section 4, we apply the inferential model of multiple testing in microarray data analysis, to identify differentially expressed genes. Finally, in Section 5 we conclude with a few remarks and point to a future work that an optimal IM can be derived for multiple testing.

2. The framework of IM

2.1. The basic framework

We start with a demonstration example. Assume that a set of observed data X is available and that model $f_\theta(X)$ for $X \in \mathcal{X}$ is specified, usually with unknown parameter $\theta \in \Theta$. We use the following example to explain the framework of IM. The key idea is to use an unobserved auxiliary random variable to represent $f_\theta(X)$ and to make uncertainty assessment by predicting this unobserved auxiliary random variable.

Example 1. The exponential distribution with a rate parameter θ (>0) has a pdf $f(x;\theta) = \theta e^{-\theta x}$ for $x \geqslant 0$ and $f(x;\theta) = 0$ for $x < 0$. It has a cdf $F(x;\theta) = 1 - e^{-\theta x}$ for $x \geqslant 0$. A random number generator of the exponential distribution, based on inverse of its cdf, is given by

$$X = -\theta^{-1} \ln U, \quad \text{where } U \sim \text{Unif}(0,1). \tag{1}$$

This sampling mechanism preserves the model for X given θ. Moreover, it specifies a relationship among the observed data X, the unknown parameter θ, and the unobserved variable U. The unobserved variable U is called auxiliary random variable and Eq. (1) is referred to as an association equation.

Our proposed IM inference about θ is based on predicting the unobserved U using what is called a predictive random set (PRS). An example of PRS is defined by a random interval in the following,

$$\mathcal{S} = \{u\colon |u - .5| \leqslant |V - .5|\} = [.5 - |V - .5|, .5 + |V - .5|], \quad \text{with } V \sim \text{Unif}(0,1).$$

A realization S of the PRS $\mathcal{S}$ is interpreted as the belief that the unobserved U lies in S. It follows from the association equation (1) that given the observed data X, the true value of θ belongs to

$$\{\theta\colon X = -\theta^{-1} \ln u \quad \text{for some } u \in S\},$$

which is denoted as Θ_X and more specifically

$$\Theta_X \equiv \left[-X^{-1} \ln(.5 + |V - .5|), -X^{-1} \ln(.5 - |V - .5|)\right], \quad V \sim \text{Unif}(0,1).$$

Now suppose that an assertion of interest about θ is given by a subset $\mathscr{A}$ in the space of θ, i.e., $\mathscr{A} \subset (0, \infty)$. The realization S supports the truth of $\mathscr{A}$ if and only if $\Theta_X \subseteq \mathscr{A}$, and supports the truth of the negation of $\mathscr{A}$, denoted by $\mathscr{A}^c$,