Resenhas IME-USP 1995, Vol. 2, No. 1, 27 - 46.

BAYESIAN HYPOTHESIS TEST: Using Surface Integrals to Distribute Prior Information Among the Hypotheses

Telba Zalkind Irony and Carlos Alberto de Bragança Pereira

Abstract: A solution to the problem of transferring the prior preferences about the parameters of interest to the spaces defined by the hypotheses in test is proposed. We make use of surface (line) integrals to obtain this solution. The advantage is that the preferences assessed in the original space are mantained inside the subspaces defined by the hypotheses. The Bayesian test obtained is applied to standard situations in quality assurance. The first is the detection of shifts in production processes (test for comparison of two Poisson rates). The second is the comparison of the quality of two different manufacturers (homogeneity test). The last is the test for independence between defect types (independence test). The line integral solution can be applied not only to hypotheses defined by linear relationships (comparison of Poisson rates and homogeneity test) but also to hypotheses defined by non-linear relationships (independence test).

Key words: surface (line) integrals, odds ratio, homogeneity test, independence test, quality assurance.

AMS classification: 62A15, 62F15,62F03.

1. INTRODUCTION

The demand for hypotheses tests is enormous in all areas. For instance, hypotheses tests are often used to detect shifts in production processes, to compare the quality of different manufacturers or to check whether a specific type of defect has influence on the occurrence of other defects in a lot of manufactured items.

The methodology of testing hypotheses presented here is Bayesian. This means that, in the process of construction of the final decision rule, experts' opinions are considered and all costs involved are taken into account.

Under the Bayesian approach, the expert's prior opinion about the parameters of interest can be expressed by a system of preferences over the parameter space. This system of preferences should be formally represented by a prior probability (density) function defined on the parameter space (see De Groot, 1986). The scenario of testing hypotheses consists on a set of k (≥ 2) hypotheses about reality and the objective is to choose one among them. The hypotheses to be compared define a k-subset partition of the parameter space. Usually, some of these subsets are contained in subspaces whose dimensions are smaller than that of the original parameter space. The novelty of the methodology presented in this article is that the original system of preferences assessed by the experts to the whole parameter space is transferred, in a sensible and natural way, to those subsets of smaller dimensions that were generated by some of the hypotheses in test. The transference of the original system of preferences to a subset defined by a specific hypothesis, say H_0, is made through a straightforward computation of a line integral (or surface integral for higher dimensions) of the probability (density) function that was defined on the original parameter space. This line (surface) integral is computed over the line (surface) defined by the hypothesis H_0.

The test suggested in this article is exact in the sense that it does not require asymptotic approximations. Moreover, whenever the prior distributions assessed belong to conjugate families, the decision function will be a function exclusively of the parameters of the prior distribution and of the sufficient statistic. This means that, when dealing with prior distributions that belong to conjugate families, the analyst performing the hypothesis test does not have to reconstruct the decision rule for each member of the family or for each different data set under analysis. It suffices to compute the value of the sufficient statistic and to use it together with the parameters of the prior in the general decision rule.

Another important aspect of the Bayesian test introduced is that it uses directly the natural parameters defined by the hypotheses being tested. There is no need to look for alternative parametrizations in order to obtain simplifications.

We planned the paper according to the following:
In Section 2 the test procedure is developed. In Section 3 the test is applied to two cases in which the hypotheses being tested are linear. The first is the comparison of two Poisson distributions and the second, known as homogeneity test, is the comparison of two binomial distributions. Section 4 presents a more sofisticated application, known as test for independence, where the hypothesis being tested is nonlinear. We would like to point out that, in agreement with our intuition, this test of independence has a decision function that differs from the one obtained by the homogeneity test presented in Section 3.

2. THE BAYES TEST

Let d denote the data obtained as result of an experiment and $f(d|\omega)$ represent the associated probability (density) function. Here $\omega \in \Omega$ ($\subset \mathbb{R}^n$) is the unknown parameter that identifies the function. Ω is the parameter space.

Let $L(\omega|d)$ be the likelihood function; i.e., $L(\omega|d)$ is a function of ω such that, for each observed d, $L(\omega|d)$ is given by $f(d|\omega)$ evaluated at d. In many cases, the analyst in charge of the experiment will be able to express her preferences for the elements of Ω before the experiment is performed. In other words, her experience will suggest that some subset of values in Ω are, a priori, more likely than others and consequently she will be able to assess a prior probability (density) function $g(\cdot)$ on Ω. If there is no preference for any set of points in Ω, the analyst may say that all elements of Ω are equally likely and will assess a uniform distribution over Ω. In summary, the analyst preferences over Ω are translated into a prior probability distribution over Ω.

Let $\{\Omega_0, \Omega_1\}$ be a partition of Ω, i.e., $\Omega_0 \cup \Omega_1 = \Omega$ and $\Omega_0 \cap \Omega_1 = \emptyset$. The hypotheses to be considered are $H_0 : \omega \in \Omega_0$ and $H_1 : \omega \in \Omega_1$. The parameter θ will be the indicator of the hypotheses, that is, $\theta = i$ if $\omega \in \Omega_i$, and the analyst prior preferences for the hypotheses will be described by $P(\theta = i) = \xi_i$ (for $i = 0, 1$ and $\xi_0 + \xi_1 = 1$).

In order to decide in favor or against H_0 before observing the experimental results, the analyst must compare ξ_0 and ξ_1. The prior odds in favor of H_0 are $R_{01} = \frac{\xi_0}{\xi_1}$ and a Bayesian test will favor H_0 if $R_{01} > c$, where c is a positive constant related to the losses and gains associated to the decisions. Appendix I will clarify the meaning of the constant c. From a classical perspective the test may be viewed as a procedure that minimizes a linear combination of the error probabilities associated to the hypotheses in test. If the losses associated with the errors of first and second kind are the same, then $c = 1$.

Note that the values of ξ_0 and ξ_1 vary according to the analyst's experience. Thus, before observing d, her experience is described by ξ_0 and ξ_1 and once d is observed, her knowledge is updated and the new situation is described by $\xi_0(d) = P(\theta = 0|d)$ and $\xi_1(d) = P(\theta = 1|d)$, where $\xi_0(d)$ and $\xi_1(d)$ are respectively the posterior probabilities that $\theta = 0$ and $\theta = 1$, given d. The posterior odds in favor of H_0 are $R_{01}(d) = \frac{\xi_0(d)}{\xi_1(d)}$ and a Bayesian test will decide in favor of H_0 if $R_{01}(d) > c$.

Recall that the probability (density) function g was defined over Ω. In order to obtain the values of $\xi_i(d)$ $(i = 0, 1)$, the analyst needs to define densities over Ω_i and the novelty of this work is precisely the way in which we will define these densities.

Definition 1: The prior density of ω under H_i $(i = 0, 1)$ is defined by:

$$g(\omega|\theta = i) = g_i(\omega) = \frac{g(\omega) I_i(\omega)}{\int g(\omega)\, d\Omega_i} \quad (i = 0, 1)$$

where $I_i(\omega)$ is the indicator function[1] of $\omega \in \Omega_i$ $(\theta = i)$ and $\int g(\omega)\, d\Omega_i$ is the line integral[2] (in $\mathbb{R}^2$) or surface integral[2] (in $\mathbb{R}^n, n > 2$) of $g(\omega)$ in the subset Ω_i. Note that the denominator of the function g_i is a normalizing constant.

---

[1] The indicator function $I_i(\omega)$ is defined as: $I_i(\omega) = \begin{cases} 1 & \text{if } \omega \in \Omega_i \quad (\theta = i) \\ 0 & \text{if } \omega \notin \Omega_i \quad (\theta \neq i) \end{cases}$.

[2] See appendix II for the definition of such integrals.

Figure 1 presents an intuitive interpretation of Definition 1. Note that g describes the analyst's preferences both in Ω and in Ω_i $(i = 0, 1)$. The lateral view of the contour defined by g over the subset Ω_0 suggests a natural way to define the analyst's probability density over Ω_0. This density is presented in Figure 2.

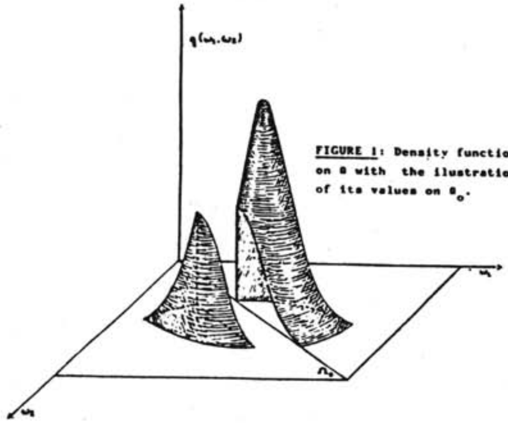

FIGURE 1: Density function on Ω with the illustration of its values on Ω_0.

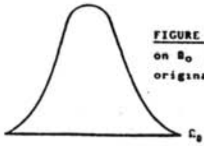

FIGURE 2: Density function on Ω_0 obtained from the original density on Ω.

Once d is observed, the posterior density of ω, which is given by Bayes' formula, is:

$$g(\omega|d) = \frac{L(\omega|d)g(\omega)}{\int_\Omega L(\omega|d)g(\omega)\, d\omega} \ .$$

If we apply Definition 1 to the posterior distribution of ω, the posterior distribution of ω under the hypothesis H_i is given by

$$g(\omega|d, \theta = i) = g_i(\omega|d) = \frac{g(\omega)L(\omega|d)I_i(\omega)}{\int g(\omega)L(\omega|d)\, d\Omega_i}$$

where again $I_i(\omega)$ is the indicator function of $\omega \in \Omega_i$ and $\int f(\omega)\, d\Omega_i$ is the line integral (in $\mathbb{R}^2$) or surface integral (in $\mathbb{R}^n, n > 2$) of $f(\omega)$ in the subset Ω_i.

It is important to note that the relationship between the parameter θ (the hypotheses in test) and the experimental result d is made through the parameter ω and the likelihood function associated to the experimental model, $L(\omega|d)$. Consequently, in order to obtain the "predictive" distribution, $f(d)$, we must integrate the parameter ω out as follows:

$$f(d) = \int_{\Omega} f(d|\omega) g(\omega)\, d\omega \ .$$

This suggests the following definition:

Definition 2: The predictive distribution (or predictive density) of the data under the hypothesis H_i, $f(d|\theta = i)$, which will be denoted by $f_i(d)$ is defined by:

$$f_i(d) = \frac{\int g(\omega) L(\omega|d)\, d\Omega_i}{\int g(\omega)\, d\Omega_i}$$

where the symbol $d\Omega_i$ denotes that we are taking the line integral (or surface integral) of the integrand constrained to the subset Ω_i.

Note also that $\int g(\omega)\, d\Omega_i$ is the normalizing constant that assures that $f_i(d)$ is a probability (density) function.

Since $f_i(d)$ is defined as the conditional probability of d given $\theta = i$, the posterior probability of the event $\{\theta = i\}$ given d is obtained by Bayes' formula:

$$P(\theta = i|d) = \xi_i(d) = \frac{\xi_i f_i(d)}{\sum_{i=0}^{1} \xi_i f_i(d)} \ .$$

Consequently, the posterior odds in favor of H_0 are given by $R_{01}(d) = \dfrac{\xi_0 f_0(d)}{\xi_1 f_1(d)}$ and the suggested test will favor H_0 if $R_{01}(d) > c$.

3. APPLICATIONS: POISSON AND BINOMIAL CASES

In this section we present typical applications in quality assurance that benefit enormously from the procedure suggested in the previous section.

Here, only for simplicity, we consider prior distributions that belong to conjugate classes of distributions (DeGroot, 1986). The advantage of using these distributions is that there is no difference between prior and posterior distributions except for the values of their parameters. This means that there is no need to recalculate the whole posterior distribution. It suffices to update its parameters.

¿From now on the symbol $\sim$ will stand for the expression "distributed as". For instance, $x \sim G(a, b)$ will mean that the random quantity x is distributed as a gamma distribution with parameters a and b. For a beta distribution with parameters a and b we write $x \sim \text{Be}\,(a, b)$.

To allow changes in the values of the parameters we write

$$x|(a,b) \sim G(a,b) \quad \text{or} \quad x|(a,b) \sim \mathrm{Be}\,(a,b)\ .$$

3.1 DETECTING SHIFTS IN PRODUCTION PROCESSES

In any production process there will always exist a certain amount of inherent variability that is the cumulative effect of many small, essentially unavoidable causes. When this background noise is relatively small, the process is acceptable and the system is a stable process with "common" causes of variation. Occasionally, however, special causes of variations will occur, resulting in a "shift" to an out-of-control state in which a larger proportion of the process output does not conform to engineering requirements. If this occurs, the process has to be adjusted and this inccurs in a certain cost. On the other hand, there is also a cost associated to allowing the process to operate out of control.

A key issue in quality assurance is to judge whether or not an audited production process is in statistical control and this is done through quality audits.

A quality audit is a system of inspections done periodically on a sampling basis. Samples are collected periodically, sampled product is inspected and defects are assessed whenever the product fails to meet the engineering requirements. At each rating period, the inspection results are combined and expressed by a quality index. See Irony, Pereira and Barlow (1992) for an example of how to model such inspection plans in order to estimate the quality index of a production line.

A sensible method to detect a shift in a production process is to compare the quality index of the current rating period, λ_T, with the quality index of the previous rating period, λ_{T-1}, using a hypothesis test.

The Bayes hipothesis test suggested in this paper seems to be the right tool to perform this comparison because it allows for the inclusion of: (a) the cost of unnecessary adjustments and (b) the costs of allowing the process to go out of control. It also allows for the incorporation of expert's opinion translated into prior probabilities for the quality indexes and prior probabilities for the hypotheses in test.

Suppose that two audit samples, one of size m and another of size n are collected at rating periods $T-1$ and T respectively. Let x represent the number of defects found in the first sample and y represent the number of defects found in the second sample. Notice that an item in a sample may have more than one defect.

Irony and Pereira (1994) show that most production lines have physical characteristics that justify the Poisson distribution for the number of defects in a sample. Consequently, we will adopt the Poisson model in our analysis; i.e., $x \sim \mathrm{Poisson}\,(m\lambda_{T-1})$ and $y \sim \mathrm{Poisson}\,(n\lambda_T)$. λ_{T-1} and λ_T are the quality indexes of the first and second rating periods respectively. They express the frequency of defects per unit. The probability distribution of x and y will be given by:

$$p(x|\lambda_{T-1}) = \frac{e^{-m\lambda_{T-1}}(m\lambda_{T-1})^x}{x!} \quad \text{and} \quad p(y|\lambda_T) = \frac{e^{-n\lambda_T}(n\lambda_T)^y}{y!}$$

Here, we are also assuming that for fixed values of λ_{T-1} and λ_T, x and y are two independent random quantities.

The hypotheses to be tested are:

H_0 : there was no shift in the process $(\lambda_{T-1} = \lambda_T)$

H_1 : there was a shift in the process $(\lambda_{T-1} \neq \lambda_T)$

These hypotheses define the following partition of the parametric space:

$$\begin{aligned}
\Omega_0 &= \{(\lambda_{T-1}, \lambda_T) : \lambda_{T-1} \geq 0\ , \quad \lambda_T \geq 0 \quad \text{and} \quad \lambda_{T-1} = \lambda_T\} \\
\Omega_1 &= \{(\lambda_{T-1}, \lambda_T) : \lambda_{T-1} \geq 0\ , \quad \lambda_T \geq 0\}
\end{aligned}$$

which are, respectively, the diagonal line of the positive quadrant and the whole positive quadrant.

Let the parameter θ be defined as:

$$\theta = \begin{cases} 0 & \text{if } H_0 \text{ is true, i.e., if } \omega = (\lambda_{T-1}, \lambda_T) \in \Omega_0 \\ 1 & \text{if } H_1 \text{ is true, i.e., if } \omega = (\lambda_{T-1}, \lambda_T) \in \Omega_1 \end{cases}$$

Suppose that the analyst in charge of the production process can express her preferences for the hypotheses by $P(\theta = 0) = \xi_0$ and $P(\theta = 1) = \xi_1$ $(\xi_0 + \xi_1 = 1)$. Suppose also that the preferences for the quality indexes, λ_{T-1} and λ_T, can be expressed by a product of two independent gamma distributions: $\lambda_{T-1}|(a,b) \sim G(a,b)$ and $\lambda_T|(c,d) \sim G(c,d)$. In other words:

$$g(\omega) = g(\lambda_{T-1}, \lambda_T) = \frac{b^a}{\Gamma(a)} \lambda_{T-1}^{a-1} e^{-b\lambda_{T-1}} \frac{d^c}{\Gamma(c)} \lambda_T^{c-1} e^{-d\lambda_T}\ ,$$

where $\lambda_{T-1} > 0$ *and* $\lambda_T > 0$ and $\Gamma(\alpha)$ denotes the gamma function that is defined as:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx \quad \text{for} \quad \alpha > 0\ .$$

Note that $\Gamma(n+1) = n!$ if n is an integer.

For fixed values of λ_{T-1} and λ_T, x and y are distributed as two independent Poisson random variables and the likelihood for this problem is given by:

$$L(\omega|d) = L(\lambda_{T-1}, \lambda_T|x,y) = \frac{(m\lambda_{T-1})^x}{x!} \frac{(n\lambda_T)^y}{y!} e^{-m\lambda_{T-1}} e^{-n\lambda_T}\ .$$

The class of conjugate distributions for $\omega = (\lambda_{T-1}, \lambda_T)$ is precisely the product of two independent gamma distributions (see DeGroot, 1986). Hence, the posterior distribution of $\omega = (\lambda_{T-1}, \lambda_T)$ given $d = (x, y)$ will also be the product of two independent gamma distributions, namely

$$
\begin{aligned}
(\lambda_{T-1}|a,b,c,d,x,y) &\sim \lambda_{T-1}|(a+x,b+m) \sim G(a+x,b+m) \quad \text{and} \\
(\lambda_T|a,b,c,d,x,y) &\sim \lambda_T|(c+y,d+n) \sim G(c+y,d+n) \ .
\end{aligned}
$$

The predictive distribution of the data under the hypothesis H_0 is given by:

$$
f_0(x,y) = \frac{m^x}{x!}\frac{n^y}{y!}\frac{\Gamma(A+C-1)(b+d)^{a+c-1}}{(B+D)^{A+C-1}\Gamma(a+c-1)}
$$

where $A = a+x$, $B = b+m$, $C = c+y$, $D = d+n$.

The predictive distribution of the data under H_1 is given by:

$$
f_1(x,y) = \frac{b^a}{\Gamma(a)}\frac{d^c}{\Gamma(c)}\frac{m^x}{x!}\frac{n^y}{y!}\frac{\Gamma(A)}{B^A}\frac{\Gamma(C)}{D^C} \ .
$$

The posterior odds in favor of H_0 are given by:

$$
\begin{aligned}
R_{01}(x,y) &= \frac{\xi_0 f_0(x,y)}{\xi_1 f_1(x,y)} \\
&= \frac{\xi_0}{\xi_1}\frac{\Gamma(A+C-1)}{\Gamma(A)\Gamma(C)}\frac{\Gamma(a)\Gamma(c)}{\Gamma(a+c-1)}\frac{B^A D^C}{(B+D)^{A+C-1}}\frac{(b+d)^{a+c-1}}{b^a d^c}
\end{aligned}
$$

If $a = b = c = d = 1$, corresponding to the prior assessment of two independent exponential distributions with means equal to 1 for λ_{T-1} and λ_T, we will have:

$$
R_{01}(x,y) = \frac{\xi_0}{\xi_1}\binom{x+y}{x}\left(\frac{m+1}{m+n+2}\right)^{x+1}\left(\frac{n+1}{m+n+2}\right)^{y+1}(m+n+2)2 \ .
$$

Hence, once x and y are obtained, the analyst will favor H_0 if $R_{01}(x,y) \geq c$, where c is a pre-specified constant. Finally for equal sample sizes, $m = n$ and $\xi_0 = \xi_1 = 1/2$ we would have

$$
R_{01}(x,y) = \binom{x+y}{x}\left(\frac{1}{2}\right)^{x+y}(n+1) \ .
$$

3.2 COMPARING THE QUALITY OF TWO DIFFERENT MANUFACTURERS (HOMOGENEITY TEST)

Suppose that the proportion of defective items produced by two different manufacturers is to be compared. Two audit samples of sizes m and n are collected from the first and second production lines respectively. Let x be the number of defective items found in the first sample and y be the number of defectives found in the second sample.

It is said that the manufacturer's quality is measured by the proportion of defectives that are produced. If p is the proportion of defectives turned out by the first manufacturer and q by the second manufacturer, the objective is to test:

H_0 : the quality of both manufacturers is equivalent ($p = q$)

H_1 : the manufacturers have different quality ($p \neq q$)

These hypotheses define the following partition of the parametric space:

$$\begin{aligned} \Omega_0 &= \{(p,q) : 0 \leq p \leq 1, \quad 0 \leq q \leq 1 \quad \text{and} \quad p = q\} \\ \Omega_1 &= \{(p,q) : 0 \leq p \leq 1, \quad 0 \leq q \leq 1\} \end{aligned}$$

which are, respectively, the increasing diagonal of the unit square and the whole unit square.

For fixed values of p and q, x and y are observations of two independent binomial random quantities with parameters $(m;p)$ and $(n;q)$ respectively.

Let the parameter θ be defined as:

$$\theta = \begin{cases} 0 & \text{if } H_0 \text{ is true; i.e., if } \omega = (p,q) \in \Omega_0 \\ 1 & \text{if } H_1 \text{ is true; i.e., if } \omega = (p,q) \in \Omega_1 \end{cases}$$

Suppose that the analyst in charge of the comparison can express her preferences for θ by $P(\theta = 0) = \xi_0$ and $P(\theta = 1) = \xi_1$. Suppose also that she is able to establish her system of preferences over Ω by a product of two independent beta distributions. In other words, in her opinion, $p \sim B(a,b)$ and $q \sim B(c,d)$, or

$$g(\omega) = g(p,q) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} p^{a-1}(1-p)^{b-1} q^{c-1}(1-q)^{d-1},$$

where $0 \leq p \leq 1$ and $0 \leq q \leq 1$ and $\Gamma(\cdot)$ is the gamma function as defined in Section 3.1.

Since the class of conjugate distributions for $\omega = (p,q)$ is precisely the product of two independent beta distributions (see DeGroot, 1986), the posterior distribution of $\omega = (p,q)$ given $d = (x,y)$ will also be the product of two independent beta distributions, namely:

$$p|(a,b,x,y) \sim p|(A,B) \sim \mathrm{Be}\,(A,B) \quad \text{and} \quad q|(c,d,x,y) \sim q|(C,D) \sim \mathrm{Be}\,(C,D)$$

where $A = a + x$, $B = b + m - x$, $C = c + y$ and $D = d + m - y$.

The likelihood of the problem is given by:

$$L(\omega|d) = L(p,q|xy) = \binom{m}{x}\binom{n}{y} p^x (1-p)^{m-x} q^y (1-q)^{n-y}$$

The arch that represents the hypothesis H_0 is given by the equations $p = p$ and $q = h(p) = p$. Hence, the predictive distribution of the data (x, y) under the hypothesis H_0 is given by:

$$f_0(x,y) = \binom{m}{x}\binom{n}{y}\frac{B(A+C-1, B+D-1)}{B(a+c-1, b+d-1)}$$

where $A = a + x$, $C = c + y$, $B = b + m - x$, $D = d + n - y$ and $B(u, v)$ is the beta function evaluated at the point (u, v), which is $B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$

Analogously, the predictive distribution of (x, y) under H_1 is:

$$f_1(x,y) = \binom{m}{x}\binom{n}{y}\frac{B(A,B)B(C,D)}{B(a,b)B(c,d)} \; .$$

Hence, the posterior odds in favor of H_0 are:

$$R_{01}(x,y) = \frac{\xi_0}{\xi_1}\frac{B(A+C-1, B+D-1)B(a,b)B(c,d)}{B(a+c-1, b+d-1)B(A,B)B(C,D)} \; .$$

If the assessed prior is uniform over Ω ($a = b = c = 1$) and $\xi_0 = \xi_1 = 1/2$, the posterior odds in favor of H_0 will be:

$$R_{01}(x,y) = \frac{\binom{m}{x}\binom{n}{y}}{\binom{m+n}{x+y}}\frac{(m+1)(n+1)}{(m+n+1)}$$

and the Bayes test will decide in favor of H_0 when $R_{01}(x, y) \geq c$. Finally, if we have equal sample sizes, $m = n$, we will have

$$R_{01}(x,y) = \frac{\binom{n}{x}\binom{n}{y}}{\binom{2n}{x+y}}\frac{(n+1)^2}{2n+1} \; .$$

Note that the first term in the product is a hypergeometric probability which is used in Fisher's exact test for 2×2 tables. Irony and Pereira (1986) compare the Bayes test for homogeneity with Fisher's exact test using a large number of simulated samples. The comparison favors the Bayes test in the sense that it minimizes the linear combination of the errors of first and second kind.

4. A TEST FOR INDEPENDENCE

The application presented in this section reveals the strength of the methodology introduced in Section 2. Here, the null hypothesis is represented by a set of non-linear equations which are usually avoided by standard techniques of hypothesis testing through reparametrization. The test developed by using the line integral does not involve ad hoc parametrizations and the final expression for the decision rule has an intuitive flavor.

The problem of testing independence between two events arises, for instance when items coming out from a production line are evaluated according to the presence or absence of two types of defects, say E and F. The objective is to check whether the presence of defect E (F) in a unit increases the chance of occurence of defect F (E). With this objective in mind, a sample of N units is selected from a lot. The sampled items are then classified exaustively and exclusively into the following categories: EF, $E\overline{F}$, $\overline{E}F$, and $\overline{EF}$, where $\overline{E}$ and $\overline{F}$ indicate the absence of defect E and F respectively. The data are displayed in Table 4.1. Analogously, the parameters of the multinomial distribution, the cell probabilities associated to Table 4.1, are displayed in Table 4.2.

Let the prior assessed to the parameters of the multinomial distribution be a Dirichlet distribution of order 4. Recall that the family of Dirichlet distributions is a conjugate family for multinomial models. The parameters of the prior (posterior) distributions are displayed in Table 4.3.

The prior density function and the likelihood function are, respectively, proportional to

$$p_{11}^{a_{11}-1}p_{12}^{a_{12}-1}p_{21}^{a_{21}-1}p_{22}^{a_{22}-1} \tag{4.1}$$

and

$$p_{11}^{x_{11}}p_{12}^{x_{12}}p_{21}^{x_{21}}p_{22}^{x_{22}} \ . \tag{4.2}$$

In order to obtain the posterior distribution, it suffices to change in (4.1) $A_{ij} = a_{ij} + x_{ij}$ for a_{ij} (for $i = 1,2$ and $j = 1,2$).

The objective here is to test the independence of defects E and F, that is, if the presence of defect E (F) does not change the chance of occurrence of defect F (E). Formally, this is equivalent to test the hypothesis H_0 defined by the following equations system:

$$H_0 \ : \ \begin{cases} p_{11} = pq \\ p_{12} = p(1-q) \\ p_{21} = (1-p)q \\ p_{22} = (1-p)(1-q) \end{cases} \tag{4.3}$$

The alternative hypothesis is represented by the simplex $\Omega_1 = \{(p_{11},\ldots,p_{22});\ p_{ij} \geq 0,\ \sum_{ij} p_{ij} = 1,\ i,j = 1,2\}$.

In order to compute the surface integral, associated to H_0, we need the matrix of derivatives of $(p_{11}, p_{12}, p_{21}, p_{22})$ relatively to (p,q) (see Courant & John, 1974, vol. 2) which is given by:

$$M = \begin{bmatrix} q & p \\ (1-q) & -p \\ -q & (1-p) \\ -(1-q) & -(1-p) \end{bmatrix} \tag{4.4}$$

Table 4.1: Category sample frequencies

	E	$\overline{E}$	Total
F	x_{11}	x_{12}	$x_{1\cdot}$
$\overline{F}$	x_{21}	x_{22}	$x_{2\cdot}$
Total	$x_{\cdot 1}$	$x_{\cdot 2}$	n

Table 4.2: Parameters of the multinomial model

	E	$\overline{E}$	Total
F	p_{11}	p_{12}	p
$\overline{F}$	p_{21}	p_{22}	$1-p$
Total	q	$1-q$	1

Table 4.3: Parameters of the prior (posterior) distribution

	E	$\overline{E}$	Total
F	a_{11} (A_{11})	a_{12} (A_{12})	$a_{1\cdot}$ $(A_{1\cdot})$
$\overline{F}$	a_{21} (A_{21})	a_{22} (A_{22})	$a_{2\cdot}$ $(A_{2\cdot})$
Total	$a_{\cdot 1}$ $(A_{\cdot 1})$	$a_{\cdot 2}$ $(A_{\cdot 2})$	a (A)

Note: $A_{ij} = a_{ij} + x_{ij}$

The six determinants of the squared matrices of rank 2 formed by the rows of M are: $D_1 = -p$, $D_2 = q$, $D_3 = p - q$, $D_4 = 1 - (p+q)$, $D_5 = -(1-q)$, and $D_6 = 1 - p$.

The factor used in the surface integral is then

$$\begin{aligned}\sqrt{\Delta} &= (D_1^2 + D_2^2 + D_3^2 + D_4^2 + D_5^5 + D_6^2)^{1/2} \\ &= 2[3/4 - p(1-p) - q(1-q)]^{1/2}\end{aligned}$$

The volume under the density function (4.1) over the surface defined by equations (4.3) is given by:

$$H(a_{11}, a_{12}, a_{21}, a_{22}) = \int_0^1 \int_0^1 \sqrt{\Delta} p^{a_{1.}-2}(1-p)^{a_{2.}-2} q^{a_{.1}-2}(1-q)^{a_{.2}-2}\, dp\, dq$$

The predictive probability function under the null hypothesis is given by:

$$f_0(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{H(A_{11}, A_{12}, A_{21}, A_{22})}{H(a_{11}, a_{12}, a_{21}, a_{22})} \frac{n!}{x_{11}!x_{12}!x_{21}!x_{22}!} .$$

The predictive probability function under the alternative hypothesis is given by:

$$f_1(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{B(A_{11}, A_{12}, A_{21}, A_{22})}{B(a_{11}, a_{12}, a_{21}, a_{22})} \frac{n!}{x_{11}!x_{12}!x_{21}!x_{22}!} ,$$

where

$$B(a_{11}, a_{12}, a_{21}, a_{22}) = \frac{(a_{11}-1)!(a_{12}-1)!(a_{21}-1)!(a_{22}-1)!}{(a-1)!} .$$

The function f_1 is a mixture of multinomial probability functions by a Dirichlet density function.

The posterior odds in favor of H_0 are:

$$R_{01}(d) = \frac{f_0(d)}{f_1(d)} \frac{\xi_0}{\xi_1} = \frac{\xi_0}{\xi_1} \frac{H(A_{11}, A_{12}, A_{21}, A_{22})}{B(A_{11}, A_{12}, A_{21}, A_{22})} \frac{B(a_{11}, a_{12}, a_{21}, a_{22})}{H(a_{11}, a_{12}, a_{21}, a_{22})} .$$

If uniform priors are assessed, that is, $a_{11} = a_{12} = a_{21} = a_{22} = 1$ and $\xi_0 = \xi_1 = 1/2$, we obtain:

$$R_{01}(d) = \frac{H(x_{11}+1, x_{12}+1, x_{21}+1, x_{22}+1)}{6H(1,1,1,1)} \frac{(n+3)!}{x_{11}!x_{12}!x_{21}!x_{22}!} .$$

Note that $H(1,1,1,1) = 2/3$.

The Taylor expansion of $[3/4 - p(1-p) - q(1-q)]^{1/2}$ about the point $(p,q) = (\frac{1}{2}, \frac{1}{2})$ is $\frac{1}{2} + (p - \frac{1}{2})^2 + (q - \frac{1}{2})^2$. Using this expansion together with some properties of the beta distribution we obtain:

$$\begin{aligned}H(a_{11}, a_{12}, a_{21}, a_{22}) &= 2B(a_{1.}-1; a_{2.}-1)B(a_{.1}-1; a_{.2}-1) \times \\ &\times \left\{1 - \frac{a-2}{a-1}[P(1-P) + Q(1-Q)]\right\}\end{aligned}$$

where $P = \dfrac{a_{1.} - 1}{a - 2}$ and $Q = \dfrac{a_{.1} - 1}{a - 2}$.

Finally, when assessing uniform priors, we have

$$R_{01}(d) = \frac{1}{4}\frac{(n+3)(n+2)}{n+1}\frac{\binom{x_{1.}}{x_{11}}\binom{x_{2.}}{x_{22}}}{\binom{n}{x_{.1}}}\left\{1 - \frac{n+2}{n+3}[\tilde{P}(1-\tilde{P}) + \tilde{Q}(1-\tilde{Q})]\right\},$$

where $\tilde{P} = \dfrac{x_{1.} + 1}{n+2}$ and $\tilde{Q} = \dfrac{x_{.1} + 1}{n+2}$. Note that $\tilde{P}$ and $\tilde{Q}$ are respectively the Bayes estimators of p and q.

The results in the following examples are obtained using Taylor expansion. Uniform priors and $\xi_0 = \xi_1 = 1/2$ were assessed. In order to evaluate the p-value presented, we ordered the sample space according to the value of $R_{01}(d)$ in each sample point. In other words, we consider a sample point d_1 to be more extreme than a point d_2 if $R_{01}(d_1) < R_{01}(d_2)$. Hence, the p-value associated to d_0 is the number $p_0 = \sum f_0(d)$ where the sum is over the set $\{d : R_{01}(d) \leq R_{01}(d_0)\}$.

Example 1. Suppose that in a sample of size 4, two items were found to have both defects E and F and the other two items had no defects at all. The objective is to test whether or not the presence of one type of defect increases the chance of occurence of the other type. The data is displayed in the following contingency table and an independence test is performed.

	E	$\overline{E}$	Total
F	2	0	2
$\overline{F}$	0	2	2
Total	2	2	4

The results are: $R_{01}(d) = 2/10$ (p-value $= 0.0057$)

As expected, the odds ratio indicates that independence must be rejected. In other words, the presence of one type of defect in an item should change the chance of occurence of the other type. The p-value also supports this conclusion.

Example 2. Now, suppose that in a sample of 4 items, all of them were found to have both defects. This is a critical case because the conclusion is not intuitive.

	E	$\overline{E}$	Total
F	4	0	4
$\overline{F}$	0	0	0
Total	4	0	4

The result of the independence test is: $R_{01}(d) = 8/5$ (p-value $= 1$)

This result favors independence, that is, the presence of one type of defect should not change the chance of occurence of the other type of defect.

Example 3. In this example, a sample of size 4 had an item with both types of defects, an item with no defects and two items with only one defect, one of them with defect E and the other with defect F. Here, at first glance, intuition would favor independence but, due to the small sample size, independence is not evident.

	E	$\overline{E}$	Total
F	1	1	2
$\overline{F}$	1	1	2
Total	2	2	4

The result of the independence test is: $R_{01}(d) = 4/5$ (p-value $= 0.2143$)

Note that the odds ratio seems not to support the intuition in this case. This is due to the small sample size that makes independence debatable. In fact, for a table (k,k,k,k) with sample size equal to 4k, the odds ratio will support the null hypothesis only for k>3

Example 4. Finally, in the last example, a sample of size 4 had 3 items with both defects and one item with no defect.

	E	$\overline{E}$	Total
F	3	0	3
$\overline{F}$	0	1	1
Total	3	1	4

The result of the independence test is: $R_{01}(d) = 13/40$ (p-value $= 0.0486$)

Although favoring dependence, the value of the odds ratio is not as small as the one obtained in Example 1.

Note that the expression of the odds ratio produced by the Taylor expansion highlights the intuitive flavor of the suggested test.

5. FINAL REMARKS

The tests introduced in this paper are exact, in the sense that they do not require asymptotic approximations and can be used even for small sample sizes. As shown in the previous section, one may need to use numerical integration, which sometimes can be difficult and tedious. Approximations like Taylor's expansions or Laplace methods (Tierney and Kadane, 1986) may be of some help to obtain closed analytical expressions. However, the approximations may be poor in cases where sample sizes are small. In addition, sophisticated numerical methods may be needed in situations where multivariate models are required.

Acknowledgements: This article was written during the visit of the first author to the department of Statistics at the University of São Paulo. It was supported by the National Science Foundation under Grant N. DDM-9209344 and by FAPESP under Grant N. 91/3527-9.

APPENDIX I - The meaning of the constant c

Suppose we want to test $H_0 : \theta = 0$ vs $H_1 : \theta = 1$.
A hypothesis test is a binary function $\delta(\cdot)$ from the data d such that

$$\begin{array}{lll} \text{if } \delta(d) = 1 & \text{then,} & H_0 \text{ is rejected and} \\ \text{if } \delta(d) = 0 & \text{then,} & H_0 \text{ is not rejected.} \end{array}$$

Let $\alpha(\delta)$ = probability of rejecting H_0 given that H_0 is true and $\beta(\delta)$ = probability of not rejecting H_0 given that H_0 is false. They are the probabilities of the errors of first and second order respectively, associated to the test δ.

Let $\ell(\theta, \delta)$ be the loss function associated to the test. It can be presented in the following table

	$\delta = 0$	$\delta = 1$
$\theta = 0$	0	a
$\theta = 1$	b	0

If the prior probabilities of the hypotheses are given by $P(\theta = 0) = \xi_0$ and $P(\theta = 1) = \xi_1$, the expected loss (risk) of the test, $r(\delta)$, is given by

$$r(\delta) = E[\ell(\theta, \delta)] = \xi_0 E[L(\theta, \delta)|\theta = 0] + \xi_1 E[\ell(\theta, \delta)|\theta = 1]$$

where the symbol $E[X]$ represents the expected value of X and $E(X|Y = y)$ represents the conditional expectation of X given $Y = y$.
Note that

$$E[\ell(\theta,\delta)|\theta = 0] = a.P(\delta = 1|\theta = 0) = a\alpha(\delta) \quad \text{and}$$
$$E[\ell(\theta,\delta)|\theta = 1] = b.P(\delta = 0|\theta = 1) = b\beta(\delta)$$

Consequently, $r(\delta) = \xi_0 a\alpha(\delta) + \xi_1 b\beta(\delta)$. A test that minimizes the risk $r(\delta)$ is a Bayes test.

DeGroot (1986) proves that a test δ defined on d, such that

$\delta(d) = 0$ (H_0 is not rejected) if $R_{01}(d) \geq \frac{b}{a} = c$,
$\delta(d) = 1$ (H_0 is rejected) if $R_{01}(d) \leq \frac{b}{a} = c$

minimizes the risk $r(\delta)$. Note that the well known Neyman-Pearson Lemma is a corollary of this result.

The values of a and b determine the relative importance (cost) of the errors of first and second order. If $\xi_0 = \xi_1 = 1/2$ and $b/a = c = 1$, both errors have the same cost whereas if $b/a = c \neq 1$, the error of second order costs c times more than the error of first order.

APPENDIX II - Line integral (surface integral)[3]

Let C be a smooth curve defined on the plane $p \times q$. A smooth curve is a curve represented by the following parametric equations:

$$\begin{aligned} p &= \phi(t) \\ q &= \psi(t) \quad \text{where } k \leq t \leq \ell \end{aligned}$$

and where ϕ and ψ are continuous functions with continuous derivatives on the interval $k \leq t \leq \ell$. If A is the point $(\phi(k), \psi(k))$ and B is the point $(\phi(\ell), \psi(\ell))$, C could be seen as the pathway traveled by a point that is moving continuously from A to B. If C is a smooth curve, the arch S is well defined. It is the length of the line that starts at the point $t = k$ ending at the generic t:

$$S = \int_k^t \sqrt{\left(\frac{dp}{dq}\right)^2 + \left(\frac{dq}{dt}\right)^2}\, dt \,.$$

If C is oriented in the direction of increasing t, then S will also increase in the direction of the motion and its value will vary from 0 up to the length L of C.

[3] See Courant and John (1974) for further information about this topic.

Now, we divide C according to Figure II-1:

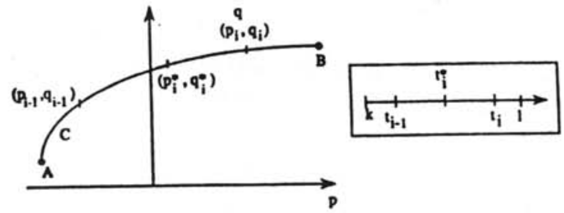

FIGURE II-1: Pathway for the line integral

Let Δ_{iS} denote the increment in S from t_{i-1} up to t_i, i.e., the pathway traveled in this interval. The line integral of f over the curve C is defined by:

$$\int_C f(p,q)\,dS = \lim_{\substack{n\to 0 \\ \max \Delta_{iS}\to 0}} \sum_{i=1}^{n} f(p_i^*, q_i^*)\Delta_{iS} \ .$$

If f is continuous in C, this integral exists and is given by:

$$\int_C f(p,q)\,dS = \int_k^{\ell} f[\phi(t),\psi(t)]\sqrt{\phi'^2(t)+\psi'^2(t)}\,dt \ .$$

If C can be represented by $q = h(p)$, $a \le p \le b$, C will be given by the equations:

$$\begin{aligned} &p = q \\ &q = h(p) \quad a \le p \le b \quad \text{and} \\ &\int_C f(p,q)\,dS = \int_a^b f(p,h(p))\sqrt{1+h'^2(p)}\,dp \end{aligned}$$

For instance, if the curve C is a curve whose density varies along C, the wire mass will be given by $m = \int_C f(p,q)\,dS$ where $f(p,q)$ is the density at the point (p,q).

Surface integral. Surface integrals are an extension of line integrals to higher dimensions.

Now, let C be a piece of smooth surface defined on a space of dimension n. A piece of smooth surface of dimension m $(< n)$ is a surface represented by the following parametric equations:

$$\begin{aligned} p_1 &= \phi_1(t_1,\ldots,t_m) = \phi_1(\mathbf{t}) \\ p_2 &= \phi_2(t_1,\ldots,t_m) = \phi_2(\mathbf{t}) \end{aligned}$$

$$\vdots$$
$$p_n = \phi_n(t_1, \ldots, t_m) = \phi_n(\mathbf{t})$$

where $\phi_1, \phi_2, \ldots, \phi_n$ are continuous functions with continuous derivatives in all points of the piece of the surface.

Analogously, as in the line integral case, the volume of the piece of smooth surface is given by

$$\iint \cdots \int \sqrt{D_1^2 + D_2^2 + \cdots + D_j^2}\, dt_1, dt_2, \ldots, dt_m = \int \cdots \int \sqrt{\Delta}\, d\mathbf{t}$$

where:

- the limits of the integrals are obtained in order to cover the piece of smooth surface;
- $j = \binom{n}{m}$; and
- $D_1, D_2, \ldots, D_j$ are the j determinants of all squared submatrices of order m obtained from the rows of the following matrix of partial derivatives:

$$\begin{bmatrix} \frac{d\phi_1}{dt_1}, & \cdots & \frac{\partial\phi_1}{\partial t_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial\phi_n}{\partial t_1}, & \cdots & \frac{\partial\phi_n}{\partial t_m} \end{bmatrix}$$

If f is a continuous function on C, the surface integral of f over the piece of smooth surface C is given by

$$\int_C f(p_1, p_2, \ldots, p_n)\, dS = \int \cdots \int f(\phi_1(\vec{t}) \cdots \phi_n(\vec{t}))\sqrt{\Delta}\, d\vec{t}\,.$$

where the limits of the multiple integral are the ones obtained in order to cover C.

REFERENCES

Courant, R.& John, F. (1974) *Introduction to calculus and analysis.* Vol. II. John Wiley. New York.

DeGroot, M. (1986) *Probability and statistics.* 2.ed. Addison-Wesley. Menlo Park.

Irony, T.Z. & Pereira, C.A. de B. (1986) Exact tests for equality of two proportions; Fisher vs. Bayes. *Journal of Statistical Computation and Simulation*, **25**; 93-114.

Irony, T.Z. & Pereira, C.A. de B. (1994) Motivation for the use of discrete distributions in quality assurance. *Test*, **3** (2) : 181-193.

Irony, T.Z; Pereira, C.A. de B. & Barlow, R.E. (1992) Bayesian models for quality assurance. In: ***Bayesian statistics*** **4**, (Bernardo, J.M.; Berger, J.O.; Dawid, A.P. and Smith, A.F.M. editors). Oxford Science Publications.

Tierney, L. & Kadane, J.B. (1986) Accurate approximations for posterior moments and marginal densities. *JASA*, **81** (393) : 82-112.

Telba Zalkind Irony
The George Washington University
Department of Operations Research, Washington, DC 20052.
e-mail: irony@seas.gwu.edu
USA

Carlos Alberto de Bragança Pereira
Universidade de São Paulo
Instituto de Matemática e Estatística
C.Postal 20570, CEP 01498, SP.
e-mail: cpereira@ime.usp.br
BRASIL