Exact Conditional Goodness-of-Fit Tests

Federico O’Reilly and Leticia Gracia-Medrano

IIMAS- UNAM, México

federico@sigma.iimas.unam.mx

lety@sigma.iimas.unam.mx

55th Session ISI, Sydney, Australia, 2005
Summary

The idea is to advocate the use of the **conditional distribution** of the goodness-of-fit test given the value of $T_n$ (min. suff. stat.). This, in testing fit of a distribution in presence of unknown $\theta$ and since not of interest, it is considered **nuisance** so conditioning seems appropriate.

Emphasis in no need for tables but rather for an **algorithm** based on simulation which produces “exact” conditional p-value. So it is an exact level $\alpha$, finite-$n$ procedure, in the continuous case.

It may be used in the **discrete case as well** but the level would be approximate due to discreteness of $T_n$.

Inverse Gaussian is discussed, comparing the results of the procedure with recent power studies, showing that for the alternatives considered, there is an **increase of power**.
1. Introduction

Tests of fit proposed with properties considered desirable, starting with the practical issue of availability of tables, to decide at a pre-assigned level $\alpha$, or by evaluating the significance, if there is evidence against the null hypothesis $H_0$.

Solutions proposed in the case where model completely specified, $H_0$ simple, were extended to case model is not completely specified due to presence of unknown $\theta$, employing the same test statistics replacing parameter by an estimate. A practice, referred as plug-in method.

Abundant literature exists for distribution of (EDF) tests and “correcting” formulas to employ the asymptotic critical points (at least for the right tail), see D’Agostino and Stephens (1986, Chapter 4).
Recently many articles suggest use of “in situ” fast algorithms that would “simulate” the distribution of the test statistic under $H_0$ by parametric bootstrap. See Henze and Klar (2002) for a continuous case and Gurtler and Henze (2000) for a discrete one. There is, in these articles, a limit ingredient which is two-fold. The first has to do with the number of bootstrap-samples and the second, has to do with the sample size, an attribute of the problem. These articles show, that the approximate critical value obtained by the bootstrap technique will yield a close approximation to the exact critical value of the test statistic when the sample size is “large”.

In the composite case, distribution of test varies with the null family specified by $H_0$; and in some, also the distribution depends on the parameter value. That is the case of the gamma and of the inverse-Gaussian (and others non location-scale).
Some history..., Srinivasan (1970), after Lil- Iefors (1967, 1969), advocated for the first time, use of the Rao-Blackwell estimate of the distribution function instead of the plug-in estimate in building Kolmogorov’s statistic. His paper had an error pointed out by Schafer et al. (1972) and gave rise to an interesting paper by Moore (1973), in which at least for the distributions concerned, was shown that the asymptotics for the empirical process estimating the parameter and the one by placing the Rao-Blackwell distribution, were the same.

If Rao-Blackwell leads to same asymptotics as the plug in, is there an **advantage in using it?**. Due to being a conditional distribution itself, there is something more important than providing an asymptotically equivalent estimate.
In a general setting, the Rao-Blackwell estimate allows generation of samples which would have exactly the same distributional properties as the observed initial sample, under $H_0$. This is discussed in Section 2, where these conditionally iid samples are referred to as “look alike” samples.

In Section 3 the inverse-Gaussian is used to illustrate this approach and power comparisons are made with the recent studies in Gracia-Medrano and O’Reilly (2004), showing an increase believed to be due to use of conditional distribution.

Finally in Section 4 some general comments are made regarding future research on the wide applicability of conditioning in goodness-of-fit.
2. Generation of look alike samples

In order to provide some notation, let $\tilde{F}_n(x)$ stand for the Rao-Blackwell estimate of $F(x, \theta)$, that is, $\tilde{F}_n(x) = P(X_i \leq x | T_n)$ where $i$ could be $1, \ldots, n$.

In many applications $T_n$ is doubly transitive, by which it is meant that knowledge of $T_n$ and $X_n$ is equivalent to knowledge of $T_{n-1}$ and $X_n$.

If statistic is doubly transitive and one needs to simulate $x_1^*, \ldots, x_n^*$ from the conditional distribution of $X_1, X_2, \ldots, X_n$ given $T_n$ one uses a procedure that only needs the use of the Rao-Blackwell distribution estimate for $F(x, \theta)$. The result is only the extension to higher dimensions and within a conditional setting, of the result that uses the inverse of any distribution function, to generate an outcome of a random variable with that distribution function; continuos or discrete.
Observe there is redundancy in the conditional distribution of the complete sample. In presence of parameter of dimension \((k)\), usually dimension of \(T_n\) coincides with \(k\) (as in the NEF), so even if it is correct to speak of conditional distribution of all sample, one really identifies the \textbf{conditional distribution of} \(n – k\) \textit{terms of the sample} observing that other \(k\) terms, follow a system of equations that defines them (up to a permutation).

Assume that the maximun \(i\) for which the conditional distribution of \(X_1, X_2, \ldots, X_i\), given \(T_n\), results in a distribution with no redundancies, is \(n – k\). Let \(x_1, x_2, \ldots, x_n\) be a realization of \(X_1, \ldots, X_n\), then a sample \(x_1^*, x_2^*, \ldots, x_n^*\) may be generated, which will be a conditionally independent realization given \(t_n = T_n(x_1, \ldots, x_n)\). And \(T_n\) applied to the new \(x^*\)–sample is also \(t_n\). For ease of notation, the last \(n – k\) terms of the \(x^*\)–sample, will be the first ones generated.
**Theorem.**

Under above conditions, the procedure for obtaining an $x^*$-sample is the following:

Let $\tilde{F}_n(x)$ be the RB estimate based on $t_n = T_n(x_1, \ldots, x_n)$ and denote by $\tilde{F}_n^{-1}(u)$ its inverse. If $u_n$ is a realization of a $U(0,1)$ r.v, define $x_n^* = \tilde{F}_n^{-1}(u_n)$ and recalculate $t_{n-1}$ from $t_n$ and $x_n^*$, denote it by $t_{n-1}^*$ (downdating ?).

With $\tilde{F}_{n-1}$ standing for the RB given $t_{n-1}^*$ and the obvious notation for its inverse, let $u_{n-1}$ be another independent realization of a $U(0,1)$ r.v. and define the number $x_{n-1}^* = \tilde{F}_{n-1}^{-1}(u_{n-1})$. Recalculate $t_{n-2}^*$ from $t_{n-1}^*$ and $x_{n-1}^*$, and keep going until $x_{k+1}^*$ is generated from $u_{k+1}$.

Next, find the $k$ remaining terms $x_1^*, \ldots, x_k^*$, solution to $T_n(x_1, \ldots, x_n) = T_n(x_1^*, \ldots, x_k^*)$. 

Note: for $T_n$ doubly transitive, the conditional distribution of $X_r$ given $T_n, X_n, X_{n-1}, \ldots, X_{r+1}$ is just the conditional of $X_r$ given $T_r$, so proof follows from the fact that the U r.v.’s are being mapped with inverses of CDF’s.

The result stated, allows the construction of $n$ random variables $X_1^*, \ldots, X_n^*$ conditioned to give the value $t_n$; out of $n - k$ iid uniforms, in some sense, a converse to an earlier result of O’Reilly and Quesenberry (1973) where for the (absolutely) continuous case, out of $n$ $X$–random variables, and conditionally also on $t_n$, a set of $n - k$ independent uniforms is obtained, bad in terms of power as method to simplify a composite hypothesis, Gracia-Medrano and O’Reilly (2004).
3. Inverse Gaussian Example

The inverse Gaussian distribution function may be written as

\[ F(x; \mu, \lambda) = \Phi(R) + \Phi(L) \exp\left\{ \frac{2\lambda}{\mu} \right\}, \]

where \( R = -\left( \frac{\lambda}{x} \right)^{1/2} + \left( \frac{\lambda x}{\mu^2} \right)^{1/2}, \)
\( L = -\left( \frac{\lambda}{x} \right)^{1/2} - \left( \frac{\lambda x}{\mu^2} \right)^{1/2} \)
and \( \Phi \) is the standard normal distribution function.

The Rao-Blackwell estimate, for \( x \) lying between \( l \) and \( u \), (0 below \( l \) and 1 above \( u \)) is

\[ \tilde{F}_n(x) = \]

\[ G_{n-2}(W) + \frac{n-2}{n} \left[ 1 + \frac{4(n-1)\hat{\lambda}_n}{n^2\hat{\mu}_n} \right] \frac{(n-3)}{2} G_{n-2}(-W') \]

where \( G_{n-2} \) is Student’s t distribution function with \( n - 2 \) degrees of freedom,

\[ W = \frac{1}{C} \sqrt{n(n-2)} \left( \frac{x}{\hat{\mu}_n} - 1 \right), \]
\[ W' = \frac{1}{C} \sqrt{n(n-2)(1 + \left( \frac{n-2}{n} \right) \frac{x}{\hat{\mu}_n})}, \]
\[ C = \sqrt{\frac{n}{\hat{\lambda}_n} (n - \frac{x}{\hat{\mu}_n}) x - n(1 - \frac{x}{\hat{\mu}_n})^2} \]

Observe that \( \tilde{F_n}(x) \) is well defined for \( n \) as little as 3 so \( k \) is 2 and one can take \( t_n \) to be the pair of MLE’s \( (\hat{\lambda}_n, \hat{\mu}_n) \).

These are all ingredients needed and there is no known explicit expression for the inverse of the Rao-Blackwell estimate, so computations of inverse done with a numerical iterative search.
To see that procedure that will identify $x^*$ samples (the look alike samples), yields in effect samples with the same conditional distribution given $T_n$, 1000 samples from an inverse-Gaussian were simulated and for each of them, first Anderson Darling’s statistic was computed and then 1000 look alike samples were generated, computing Anderson Darling’s statistic as well. With the 1000 values of Anderson Darling’s statistic from the look alike samples, it was decided to reject or not reject with an $\alpha = 0.05$; this, for each of the 1000 simulated samples. The number of rejections happened to be 42, which based on 1000 simulations is in agreement.
Table 1: Power Comparison.
Proportion of rejections out of 1000.
\[ \alpha = .05 \]

<table>
<thead>
<tr>
<th>Alternative</th>
<th>( A^2 ) look alike</th>
<th>( A^2 ) O&amp;R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>.643</td>
<td>.62</td>
</tr>
<tr>
<td>scale = 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lognormal</td>
<td>.131</td>
<td>.10</td>
</tr>
<tr>
<td>from ( n(1/2, 1) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uniform (0, 1)</td>
<td>.863</td>
<td>.85</td>
</tr>
<tr>
<td>Weibull</td>
<td>.406</td>
<td>.38</td>
</tr>
<tr>
<td>scale = 1 shape = 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I-G (( \mu = 1, \lambda = 8 ))</td>
<td>.042</td>
<td>.05</td>
</tr>
</tbody>
</table>

For \( A^2 \) “look-alike” in each replication the simulated critical value is found with 1000 “look-alike” samples. \( A^2 \) O&R uses the asymptotic critical value from O’Reilly & Rueda (1992), the entries are taken from Gracia-Medrano & O’Reilly (2004).

As seen in table, the **proposed procedure did well**. All differences in power of same sign (except in \( \alpha \)), and one quite significant.
4. General Comments

Result on Section 2 holds for **continuous and discrete** cases, but in latter there is a limitation regarding exactness of procedure respect to the desired \( \alpha \) level. This is due to discrete nature of the resulting test statistic. The discrete case is thus included in proposal of conditioning on \( T_n \) and discreteness certainly is no impediment in computing conditional p-values.

In location and scale cases, the conditional simulation may be done, but it is **unnecessary since the conditional distribution of the EDF test, is the same as the unconditional one**, so plain simulation from parent fixing parameter arbitrarily yields exact distribution (well known fact due to equivariant sufficient \( T_n \) and invariant empirical process).
There is something appealing regarding the use of the RB in the empirical process, and it is its similitude, with the empirical distribution function which is the conditional distribution given the vector of order statistics, an estimate of the true cdf whether the null hypothesis is true or false, and the Rao-Blackwell cdf is an estimate, expected to be “good”, only if the null is true. There is here, an implication of a property that would be very nice to hold, namely that if the null is not true, then the RB is not “good” estimate. True in many cases.

We have illustrated the use of the conditional distribution on a particular statistic, but may be used in simulating the exact conditional distribution of any statistic.
Finally, an example where applying any EDF test to the **Rao-Blackwellized empirical process** one gets back well known results, appears when testing the $U(0, \theta)$ distribution; and also when testing the $U(\theta_1, \theta_2)$. If one writes the empirical process with the Rao-Blackwell cdf, it is found that in the first case, the process is nonzero only if the argument $x$ of the process, varies in $(0, X_{(n)})$ and in the second case only if $x$ lies in $(X_{(1)}, X_{(n)})$ and these empirical processes, conditionally to their corresponding $T_n$ are proportional, both, to the empirical process associated to the case when sampling from the $U(0,1)$. The transformations implied for the first case, are the “new” ordered observations, $Y_{(i)} = X_{(i)}/X_{(n)}$ for $i = 1, \ldots, n-1$, and $Y_{(i)} = (X_{(i)} - X_{(1)})/(X_{(n)} - X_{(1)})$ for $i = 2, \ldots, n-1$. Both of which are well known, and frequently used transformations.
Acknowledgements

Part of this research was done while the first author was visiting Simon Fraser University, Canada, on a sabbatical leave. There, the fit of other non location-scale distributions was studied with Professors R.A. Lockhart and M.A. Stephens.

Bibliography


Thank you very much