4 # COMBINING p-VALUES: A DEFINITIVE PROCESS

5 D.A.S. Fraser

6 *Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3*

7 *Email: dfraser@utstat.toronto.edu*

8 A.K.Md.Ehsanes Saleh

9 *Department of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6*

10 *Email: esaleh@math.carleton.ca*

11 K. Ji

12 *Department of Statistics, University of Toronto, Toronto, Canada M5S 3G3*

13 *Email: kexin.ji@utoronto.ca*

SUMMARY

In the Bayes approach the model-data summary is the observed likelihood function; in the frequentest approach it is arguably a p-value function assessing a least squares or maximum-likelihood departure; and in the higher-order likelihood approach it is the observed likelihood together with a canonical reparameterization. For likelihood the obvious method of combining is to add log-likelihoods from independent sources: this is in the nature of likelihood itself and is also a Bayes imperative as only likelihood is used in the Bayesian argument. For the familiar frequentest approach the combining of p-values is often ad hoc: we discuss first a Fisher proposal and then offer a likelihood based alternative. For the higher order likelihood approach the combining begins with the standard summary, which is likelihood plus a canonical reparameterization: we develop the appropriate higher order combining procedure. For the p-value summary, Fisher (1948) proposed a quick and easy method for combining p-values from independent investigations: multiply them together and use chi-square tables. The proposal received criticism that it did not address power and other conventional criteria, but Fisher had assumed quite clearly that such related information was unavailable. We use first order likelihood theory to derive a simple modification: the p-values are converted to likelihood values and the likelihood values to observed likelihood functions and these in turn to a new composite p-value. Higher order likelihood offers further refinement: use the standard summary involving the log-likelihood $\ell(\theta)$ and the canonical reparameterization $\varphi(\theta)$; the combining from independent investigations then amounts to adding the observed log-likelihoods $\ell_i(\theta)$ and weighting and adding the reparameterizations $\varphi_i(\theta)$. We develop this information combining procedure: add log-likelihoods and suitably weight and add canonical parameters. Some examples follow.

1 Introduction

(i) Fisher on combining p-values

Two independent investigations lead to p-values .145 and .087 and it is sometimes desired
"...to obtain a single (test) ... based on the product of the (individual p-values) observed."
Thus Fisher (1948) introduced a quick and easy method for combining p-values "taking
account only of (the p-values) and not of the detailed composition of the (initial) data
...." Fisher used three p-values but two values will suffice for illustration here. Various
criticisms emerged that his proposal did not address conventional optimality criteria. He of
course ignored the criticisms for he had been upfront in mentioning the absence of "detailed
composition" concerning the original data. He thus presented a quick and dirty method, a
useful method like a mean-and-standard deviation assessment of data.

Consider a single p-value p, which under a hypothesis being examined would be Uniform$(0, 1)$
in distribution. The transformation $\chi^2 = -2 \log p$ is a decreasing transformation and it con-
verts the Uniform$(0, 1)$ distribution for p to a chi-square (2 df) for χ^2; and it has the
attractive simplicity that the right tail distribution function of the new variable is just
$\exp\{-\chi^2/2\}$. Also if small values of p indicate significance then large value of χ^2 corre-
spondingly represent significance.

Fisher then made use of the additivity of chi-square variables. The two p-values 14.5%
and 8.7% give the chi-square values 3.86 and 4.88 and thus give the composite $\chi^2 = 3.86 +
4.88 = 8.74$; the corresponding right tail p-value for chi-square (4 df) is 6.8%, and represents
some blend of the original p-values. In Figure 1a we plot the observed p-value vector
$(p_1, p_2) = (.145, .087)$; then in Figure 1b we plot the observed chi-square vector $(\chi_1^2, \chi_2^2) =
(3.86, 4.88)$ together with the contour of points having the same $\chi^2 = 8.74$ value and thus
having the same composite p-value $p = 6.8\%$; the image of this observed contour is then
recorded in Figure 1a. To give some feel for this we calculate the median value of chi-square
(4 df)): $\chi^2 = 3.36$; then plot in Fig 1b the contour of points (χ_1^2, χ_2^2) having χ^2 at that
median value, and then the corresponding contour of (p_1, p_2) in Figure 1a.

Fisher described the procedure as a "simple test of the significance of the aggregate",
certainly a simple quick and easy way of combining p-values without "detailed (information
concerning) the (original) data." Clearly the procedure is driven by the simplicity of the
$\chi^2 = -2 \log p$ mapping. And Figure 1 indicates that it treats p-values in a balanced way,
while often p-values can have quite different importance in the evaluation of a parameter,
and we will see this in Section 4.

(ii) First order likelihood combining

Now consider how first order likelihood analysis can inform the combining of p-values from
independent investigations. We use results that say log-likelihood functions should be added
and that observed information provides evidence concerning the strength of the p-value
information. The manner in which p-values are calculated is clearly relevant. For this
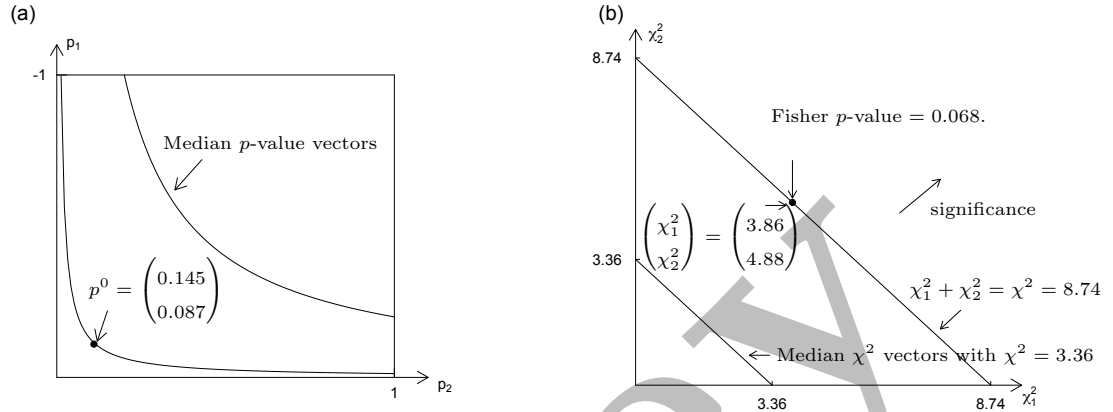suppose the p-values are providing a one sided assessment of a scalar parameter $\theta = \theta_0$,

Figure 1: (a) The observed p-vector $(p_1^0, p_2^0) = (0.145, 0.087)$, the corresponding contour, and the median contour. (b) The observed χ^2 vector with $\chi^2 = 8.74$ and p-value 0.068 with corresponding contour, and the median contour with $\chi^2 = 3.36$.

1 such as $p = \Phi^{-1}\{(\bar{y} - \mu_0)/(\hat{\sigma}/\sqrt{n})\}$ might be giving a Central Limit Theorem assessment of
2 $\mu = \mu_0$, with a concern that μ might be large relative to μ_0. Thus we are considering basic
3 one-sided p-values with small indicating significance, and not the common 2-sided values
4 that are often proposed.

The extension of Central Limit Theorem analysis into likelihood theory says that the signed likelihood root (SLR) $r(\theta)$ and the Wald statistic $q(\theta)$ are first order Normal $(0, 1)$:

$$r(\theta) = \mathrm{sgn}(\hat{\theta} - \hat{\theta}_0)[2\{\ell(\hat{\theta}) - \ell(\theta)\}]^{1/2}, \qquad q(\theta) = j_{\theta\theta}^{1/2}(\hat{\theta} - \theta),$$

5 where $\hat{\theta}$ maximizes the observed log-likelihood $\ell(\theta) = \ell(\theta; y^0)$ and $j_{\theta\theta}$ is the observed in-
6 formation $-\partial^2/\partial\theta^2 \ell(\theta; y)|_{\hat{\theta}}$ being the second derivative of likelihood at the maximum. The
7 SLR is typically better behaved than the Wald statistic but the latter overtly includes ob-
8 served information, which in turn gives an evaluation of the effectiveness of p-values. We
9 assume now that the observed information $\hat{j}_{\theta\theta}$ is available from each investigation and that
10 regularity is present for the individual models used.

11 Thus suppose that we have $\hat{j}_1 = 4$ and $\hat{j}_2 = 1$ suggesting that the first investigation
12 is providing more sensitive information. Then directly using the equivalence of $r_i(\theta)$ and
13 $q_i = \hat{j}_i^{1/2}(\hat{\theta}_i - \theta_0)$ we can start with a p-value p_i, solve for the normal score r_i, calculate the
14 log-likelihood drop ℓ_i and then solve for the maximum likelihood departure $\hat{\theta} - \theta_0$:

We might be tempted to directly combine the likelihood drops ℓ_i but they are referring to different maximum values. For ease of discussion we take the θ-scale to be centered at

Investigation	$\hat{j}_i$	p_i	$r_i = \Phi^{-1}(p_i)$	$\ell_i = -r_i^2/2$	$\hat{\theta}_i - \theta_0 = r_i/\hat{j}_i^{1/2}$
1	4	.145	-1.06	-.562	-.53
2	1	.087	-1.36	-.925	-1.36

θ_0, so in effect $\theta_0 = 0$. The first order individual likelihood functions are then

$$\ell_1(\theta) = -\frac{4}{2}(-.53 - \theta)^2, \quad \ell_2(\theta) = -\frac{1}{2}(-1.36 - \theta)^2,$$

giving the combined likelihood

$$\ell(\theta) = -\frac{4}{2}(-.53 - \theta)^2 - \frac{1}{2}(-1.36 - \theta)^2;$$

this sum of quadratics can then be rewritten relative to its maximum as

$$\ell(\theta) = -\frac{5}{2}(-.696 - \theta)^2,$$

perhaps most easily by using weight-by-information calculations giving

$$\hat{\theta} = \frac{4(-.53) + 1(-1.36)}{5} = -.696, \quad \hat{j} = 4 + 1 = 5.$$

Now from the rewritten log-likelihood for the combined data we obtain the likelihood drop $\ell = -(5/2)(-.696)^2 = -1.21$; we then obtain the composite signed likelihood root r and corresponding composite p value:

$$r = -\sqrt{(2 \times 1.21)} = -1.56, \quad p = \Phi(-1.56) = .059$$

This composite p-value .059 is smaller than the Fisher p-value, seemingly influenced by the more informative first investigation. The individual log-likelihoods and the combined likelihood based on first order theory are plotted in Figure 2.

(iii) An overview

We have used first order theory to indicate how observed information can be used in the combining of inferences from independent investigations. And observed information in turn depends directly on how the parameter is scaled or functionally calibrated. In some parallel sense the familiar Bayesian approach uses only observed likelihood and with independent data adds the log-likelihoods as the modus operandi, but provides no direct role for the use of available p-value reliability. The frequentist approach as developing in current likelihood theory also adds log-likelihoods as being clearly a routine procedure, but in addition addresses the need to calibrate the scale on which the parameter is presented; a partial indication of this need arises in the Bayesian search for an appropriate prior for analysis which
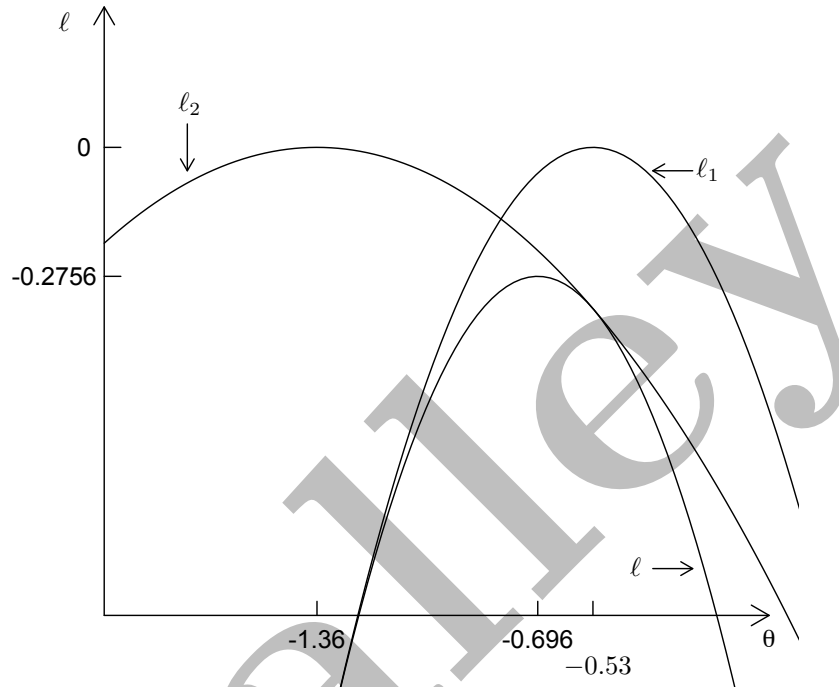
Figure 2: The imputed likelihoods $\ell_1(\theta)$, $\ell_2(\theta)$ from *p*-values 0.145 and 0.087; and the combined likelihood which in turn gives the *p*-value 0.068.

provides in turn a partial calibration of the parameter. In Section 3 we report on the third order theory that leads to the addition of log-likelihood from independent investigations,

$$\ell(\theta) = \ell_1(\theta) + \ell_2(\theta),$$

as one might reasonably expect. However the third order theory also leads to the use of a weighted combination of the canonical parameterizations that are usually critical in higher order statistical inference,

$$\varphi(\theta) = v_1 \varphi_1(\theta) + v_2 \varphi_2(\theta),$$

1 but the weights v_1 and v_2 are modifications of the observed informations usually found in
2 estimation theory.
3   But first in Section 2, we give a brief outline of the derivation of the canonical parame-
4 terizations and also an indication of the procedure for calculating the third order *p*-values.
5 Section 4 then contains examples.

2   The approximate model: Third order

The Normal distribution is the distribution of choice for most approximations in statistical inference: A typical departure from expectation is a sum of independent or approximately independent terms; This is then statistically standardized and a plug-in estimate used to eliminate nuisance parameters; The Central Limit Theorem gives approximate Normality for the sum and Slutzky's Lemma enables the transfer of Normality to the standardized quantity; And rarely is there any suggestion that higher order distributional methods might contribute to the relevance or accuracy.

Higher order approximations in statistics appeared first (Daniels, 1954) in a mathematically focussed journal and then after substantial delay (Barndorff-Nielsen & Cox, 1979) in a more statistically oriented medium, thus gaining needed recognition. As saddlepoint approximations they were highly accurate but little professional interest arose for going beyond the familiar Normal approximations to the perhaps less transparent Fourier approximations. A subsequent extension from density approximation to distribution function approximation (Lugannani & Rice, 1980) and then from cumulant generating function contexts to asymptotic contexts (Barndorff-Nielsen, 1986) provided a major increase in the range of applicability. Approximate conditioning then became available and enabled the route to very general contexts; for a recent overview see Fraser et al (2010).

The higher-order approximations do not focus on particular conventional statistics, but rather on approximating the full statistical model, that is, approximating the model at essentially all sample points, although in an application only the observed data point seems relevant or needs the calculations. The approximation is achieved using an exponential model pattern with many more free parameters to increase the fit. In addition, the exponential pattern allows highly accurate density and distribution function approximations and leads to the statistics needed for statistical inference calculations. Together this provides an incredibly flexible approach.

At any data point of interest, say the observed data, the most immediate part of a statistical model is given by the density function at that data point, that is, by the observed likelihood function, usually recorded in logarithmic form as $\ell(\theta) = \log f(y^0; \theta)$. Of course this is often ignored in frequentest statistics but in the Bayes approach is focal and exclusive.

After the likelihood function, however, what next would be important? Well, seemingly model form near the observed data point, or even just the first derivative form at the data point, or even, as it turns out, just model form in a few critical directions, with other directions of no additional interest. Such critical directions are in fact immediately available from continuity, how a change in the parameter affects the distribution at the data point.

For convenience, now, consider the case of independent coordinates, $y_1, \ldots, y_n$ and let $F_i(y_i; \theta)$ be the ith coordinate distribution function and $y_i(u_i; \theta)$ be its inverse, the quantile function, where u_i is the coordinate p-value or some equivalent scoring variable; the critical or sensitivity directions are then obtained by seeing how a change in $\theta = (\theta_1, \ldots, \theta_p)'$ affects

the response variable at the observed data. For notation, let

$$V = (v_1, \ldots, v_p) = \frac{dy}{d\theta}\big|_{y^0, \hat{\theta}^0}$$

be the derivative of the quantile vector with respect to the p coordinates of the parameter, as evaluated at the observed data y^0 and corresponding maximum likelihood value $\hat{\theta}^0$. This gives the direction of probability "flow" at the observed data when the parameter is changed at the observed maximum likelihood value; the p vectors in V record sensitivity directions, how the different coordinates of θ move the data point; for some background detail and discussion, see Fraser & Reid (1995, 2001). This gives the needed gradient of the log-model at the data:

$$\varphi(\theta) = \frac{d \log f(y; \theta)}{dV}\big|_{y^0},$$

where the derivative is calculated in the p directions $v_1, ..., v_p$. The approximating model is then the exponential model with the same likelihood and the same log-model gradient $\varphi(\theta)$ as the original model; the approximating exponential model is then

$$g(s; \theta) = \exp\{\ell(\theta) + s'\varphi(\theta)\}h(s)$$

with corresponding observed data $y^0 = 0$. For convenience we designate the preceding exponential model as $\{\ell(\theta), \varphi(\theta)\}$. Calculation with this exponential model is straightforward using routine saddlepoint methods and the model provides third order inference for arbitrary scalar parameters. The theory for determining the sensitivity directions is reviewed in Fraser et al (2010), the use of the tangent exponential model is discussed in Reid & Fraser (2010) and Davison et al (2006), and some general discussion may be found in Fraser et al (2009).

We thus treat $\{\ell(\theta), \varphi(\theta)\}$ as a definitive summary of statistical information from a statistical investigation, a summary that defines the exponential model from which full third order accurate inference is available for scalar parameters using current third order analysis, as summarized for example in Fraser et al (1999), Davison et al (2006) and Bédard et al (2008).

3   The combining of information summaries

(i) Two investigations

Now consider the third order combining of information from independent investigations and for convenience here restrict our attention to the scalar full parameter case; the vector full parameter case has some intriguing additional features that will be discussed separately. Let

$$\{\ell_1(\theta), \varphi_1(\theta)\}, \quad \{\ell_2(\theta), \varphi_2(\theta)\}$$

be the representative exponential models with observed data $s_1 = 0, s_2 = 0$ from the independent investigations with a common unknown parameter θ. Background theory for such

combining may be found for example in Fraser & Reid (1978, 2001), and Fraser et al (2010). The log-likelihoods for the combined model is of course the sum of the log-likelihoods for the components; thus

$$\ell(\theta) = \ell_1(\theta) + \ell_2(\theta).$$

For likelihood gradient a second order ancillary direction is needed in order to retain third order accuracy for the combined model. For this let $\hat{\theta}^0$ be the maximum likelihood value for the combined model log-likelihood $\ell(\theta)$ just recorded. We need to determine how θ change at $\hat{\theta}^0$ affects or moves the variables s_1, s_2 at their observed values s_1^0, s_2^0.

(ii) Sensitivity with an exponential model

First consider a scalar exponential model and assume the smoothness and regularity that supports the usual asymptotic analysis. For this we use the location relationship implicit in Welch & Peers (1963). For the exponential model $\{\ell(\varphi), \varphi\}$ the somewhat unusual quantity

$$z = \int^{\hat{\varphi}} j_{\varphi\varphi}^{1/2}(\tilde{\varphi})d\tilde{\varphi} - \int^{\varphi} j_{\varphi\varphi}^{1/2}(\tilde{\varphi})d\tilde{\varphi}$$

has a fixed distribution free of the parameter φ to the second order; each integral represents the constant information reparameterization and the difference gives the related observed minus expected. The first integral can be rewritten by changing the variable from the maximum likelihood variable $\hat{\varphi}$ to the score variable s thus producing the following alternative form for the standardized departure z:

$$z = \int^s j_{\varphi\varphi}^{-1/2}\{\hat{\varphi}(s)\}ds - \int^{\varphi} j_{\varphi\varphi}^{1/2}(\tilde{\varphi})d\tilde{\varphi}.$$

This departure relates the score variable to the canonical parameter, and in an exponential model allows us to calculate the special derivative $ds/d\varphi$ for fixed p-value. This derivative of variable with respect to parameter then evaluated at observed data determines a probability flow and thus determines the conditioning and related sensitivity directions that underlie the development of the canonical reparameterization. Accordingly by taking the total derivative in the preceding equation we obtain

$$\frac{ds}{d\theta}\Big|_{s^0} = j_{\varphi\varphi}^{1/2}\{\hat{\varphi}(s^0)\}j_{\varphi\varphi}^{1/2}(\varphi)$$

at the observed s^0. In particular, at the maximum likelihood value, this derivative of score with respect to parameter or maximum likelihood value of the parameter is given by a full information. These results are of particular interest in second order asymptotics, as the model can be represented equally as a location model or as an exponential model, and the obvious derivative in the location model case becomes the product of two root informations in the exponential model case.

1 **(iii) Sensitivity with two investigations**

Now consider two investigations and how parameter change at the overall maximum likelihood value $\hat{\theta}^0$ affects the i-th investigation, in particular how it moves the variable s_i at its observed value s_i^0. We apply the above formula to each investigation and thus obtain the change in s_i under change in φ_i, and then in turn the change $d\varphi_i = \varphi_i'(\hat{\theta}^0)d\theta$ in terms of change $d\theta$ at that overall maximum likelihood value; thus

$$ds_i = j_{\varphi_i\varphi_i}^{1/2}(\hat{\varphi}_i^0)j_{\varphi_i\varphi_i}^{1/2}\{\varphi_i(\hat{\theta}^0)\}\varphi_i'(\hat{\theta}^0)d\theta = v_id\theta;$$

this gives us the rate

$$v_i = j_{\varphi_i\varphi_i}^{1/2}(\hat{\varphi}_i^0)j_{\varphi_i\varphi_i}^{1/2}\{\varphi_i(\hat{\theta}^0)\}\varphi_i'(\hat{\theta}^0)$$

for the parameter effect on the individual exponential model variables s_i. And in turn it gives us the sensitivity matrix $V = (v_1, v_2)$, which is just a 2-vector as a row with a coordinate v_1 for the first investigation and a coordinate v_2 for the second investigation. The likelihood theory then gives the composite canonical parameter for the combined exponential models as

$$\varphi(\theta) = v_1\varphi_1(\theta) + v_2\varphi_2(\theta),$$

or equivalently

$$\varphi(\theta) = j_{\varphi_1\varphi_1}^{1/2}(\hat{\varphi}_1^0)j_{\varphi_1\varphi_1}^{1/2}\{\varphi_1(\hat{\theta}^0)\}\varphi_1'(\hat{\theta}^0)\varphi_1(\theta) + j_{\varphi_2\varphi_2}^{1/2}(\hat{\varphi}_2^0)j_{\varphi_2\varphi_2}^{1/2}\{\varphi_2(\hat{\theta}^0)\}\varphi_2'(\hat{\theta}^0)\varphi_2(\theta),$$

2 which is just a weighted linear combination of the component canonical parameters $\varphi_1(\theta)$
3 and $\varphi_1(\theta)$ as mentioned in the Introduction.

4 # 4 Some examples.

5 **Example 1: Two exponential models**

Consider two scalar-parameter scalar-variable exponential models with observed data. These models can be put in the modified form:

$$g_1(s_1; \theta) = \exp\{\ell_1(\theta) + s_1\varphi_1(\theta)\}h_1(s_1), \quad g_2(s_2; \theta) = \exp\{\ell_2(\theta) + s_2\varphi_1(\theta)\}h_2(s_1)$$

6 with observed data $(s_1^0, s_2^0) = (0, 0)$. If $\varphi_1(\theta)$ and $\varphi_2(\theta)$ are affinely related then we just
7 add the likelihoods and use either of the φ in the composite model. More generally, the
8 composite model has $\ell(\theta)$ as the sum of the component likelihoods and has φ as the weighted
9 sum of the component $\varphi(\theta)$ with weights as recorded in the preceding paragraph.

10 **Example 2: Two location models**

Consider two independent location-model investigations yielding the following tangent models

$$\{\ell_1(\theta), \varphi_1(\theta)\}, \quad \{\ell_2(\theta), \varphi_2(\theta)\}$$

where $y_1 = a_1\theta + \sigma_1 z_1$ and $y_2 = a_2\theta + \sigma_2 z_2$ are the quantile functions and the z_i are say standard Normal with σ's known. The combined likelihood is

$$\ell = -\frac{1}{2}\left(\frac{y_1 - a_1\theta}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{y_2 - a_2\theta}{\sigma_2}\right)^2.$$

The gradients of the individual likelihoods are

$$\ell_1 = \frac{a_1\theta - y_1}{\sigma_1^2}, \quad \ell_2 = \frac{a_2\theta - y_2}{\sigma_2^2}.$$

For the sensitivities we directly use the quantile expressions rather than again verify the use of the Welch-Peers second-order approximation; we obtain $v_1 = a_1$ and $v_2 = a_2$. We then have

$$\varphi = \frac{a_1\theta - y_1}{\sigma_1^2}a_1 + \frac{a_2\theta - y_2}{\sigma_2^2}a_2$$

1   which is just an affine function of θ. This gives the tangent full model as $\{\ell(\theta), \theta\}$ using
2   the above combined likelihood and reparameterization. This is just the combined location
3   model and the simplicity is based on the canonical parameter in the location Normal being
4   just the location parameter of the Normal.

5   **Example 3: Two location models with nonlinear location structures**

Consider two independent location-model investigations as above but now with nonlinear location structure: $y_1 = a_1(\theta) + \sigma_1 z_1$ and $y_2 = a_2(\theta) + \sigma_2 z_2$. The combined likelihood is

$$\ell = -\frac{1}{2}\left(\frac{y_1 - a_1(\theta)}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{y_2 - a_2(\theta)}{\sigma_2}\right)^2.$$

For the combined $\varphi(\theta)$ we are again able to bypass the Welch-Peers sensitivity formula and directly use the sensitivity available from the quantile expression in the model description; we then obtain the derivatives $v_1 = a'(\hat{\theta}^0)$ and $v_2 = a_2'(\hat{\theta}^0)$. This gives the composite reparameterization

$$\varphi(\theta) = \frac{a_1(\theta) - y_1}{\sigma_1^2}a_1'(\hat{\theta}_0) + \frac{a_2(\theta) - y_2}{\sigma_2^2}a_2'(\hat{\theta}_0)$$

which is affinely equivalent to

$$\varphi(\theta) = \frac{a_1'(\hat{\theta}^0)}{\sigma_1^2}a_1(\theta) + \frac{a_2'(\hat{\theta}^0)}{\sigma_2^2}a_2(\theta),$$

6   and thus blends the two location parameterizations as based on scaling at the overall max-
7   imum likelihood value.

1 **Example 4: Weather modeling and different sources of information**

2 Different weather models can sometimes give radically different predictions (Stainforth et
3 al, 2007). Various possible causes for this have been implicated, in particular the use of
4 flat priors for parameters in the models. In some cases an input parameter would be based
5 on how it was measured in the physical environment; and in another model the related
6 input parameter could in turn come from a quite different measurement process in the
7 environment. In such cases of course a flat prior for the parameter in the first model need
8 not correspond to a flat prior for the corresponding parameter in the second model, thus
9 giving a potential for significant differences. We do not here address this major concern
10 with default priors for model simplification, but rather focus on the more specific issue of
11 combining *p*-values when the sources of information concerning a parameter are from quite
12 different measurement contexts.

13   We consider a simple context where the arrival of particles is viewed as a Poisson process
14 with θ particles expected per unit time interval, and we examine whether θ is less than or
15 equal .1 which represents some background rate for particle arrival or whether the rate θ
16 has increased above that threshold due say to the presence of a new particle, which would
17 be the objective of the investigation. This is of interest in High Energy Physics with the
18 search for a new particle at the Large Hadron Collider in Geneva; for some related statistical
19 issues see Fraser, Reid & Wong, (2004) and Reid & Fraser (2003). We consider two very
20 oversimplified investigations that yield different inference summaries.

(i) Investigation I Let $x_1, \ldots, x_{10}$ record the number of particles arriving in 10 consec-
utive unit time intervals and suppose the observed value of the sum Σx_i is 3. The observed
log-likelihood function is

$$\ell_1(\theta) = 3 \log \theta - 10\theta = 3\varphi_1 - 10 \exp \varphi_1$$

where $\varphi_1(\theta) = \log(\theta)$ is the canonical parameter of the model. Twice differentiating the
log-model with respect to the canonical parameter φ_1 gives the information function $\jmath_{\varphi_1\varphi_1} =
10 \exp \phi_1 = 10\theta$. The observed maximum likelihood value is $\hat{\theta}_1 = .3$ and $\hat{\varphi}_1 = -1.2040$. A
standard *p*-value is obtained from the Poisson(1) distribution for the sum of 10 Poisson(.1)
values and is recorded as a mid-*p*-value for assessing $\theta = .1$

$$p_1(.1) = .3679 + .3679 + .1840 + (1/2).0613 = .9504;$$

21 the data is in the upper portion of the null distribution, indicating a higher occurrence rate
22 than would be expected under the hypothesis being assessed.

(ii) Investigation II As an alternative suppose we measure the time t to first occurrence
from some initial time point and that the observed value is $t^0 = 2.232$. The distribution is
a standard exponential life and the observed log-likelihood is

$$\ell_2(\theta) = -2.23\theta + \log \theta = -2.23\varphi_2 + \log \varphi_2$$

where $\varphi_2 = \theta$ is the canonical parameter of this second model. Twice differentiating the log-model with respect to the canonical parameter φ_2 gives the information function $\jmath_{\varphi_2 \varphi_2} = 1/\varphi_2^2 = 1/\theta^2$. The observed maximum likelihood value is $\hat{\theta}_2 = .448$ and correspondingly $\hat{\theta}_2 = \hat{\varphi}_2 = .448$. A standard p-value is obtained as a right tail sample space value acknowledging the different role of θ in this second investigation:

$$p_2(.1) = \exp\{-2.23(.1)\} = .800.$$

(iii) The Combined Investigations Of course we add the log-likelihoods from the component investigations as has long been Bayes and frequentest common procedure:

$$\ell(\theta) = \ell_1(\theta) + \ell_2(\theta) = 4\log(\theta) - 12.23\theta.$$

The component log-likelihoods both involve θ and $\log(\theta)$ but in somewhat reversed roles. The form of the likelihood does however still make the maximum likelihood value immediately accessible: $\hat{\theta} = 4/12.23 = .3271$. Now to find the weights v_1 and v_2 for combing the component φ_1 and φ_2. For the first, we have

$$v_1 = j_{\varphi_1 \varphi_1}^{1/2}\{\varphi(\hat{\theta}_1^0)\} j_{\varphi_1 \varphi_1}^{1/2}\{\varphi_1(\hat{\theta}^0)\}\varphi_1'(\hat{\theta}^0) = 1.732 \times 1.809 \times 3.057 = 9.578$$

where the first investigation root information is evaluated at the individual and then overall maximum likelihood values and then the derivative $\partial\varphi_1/\partial\theta = 1/\theta$ is evaluated at the overall maximum likelihood value. And for the second we have

$$v_2 = j_{\varphi_2 \varphi_2}^{1/2}\{\varphi(\hat{\theta}_2^0)\} j_{\varphi_2 \varphi_2}^{1/2}\{\varphi_2(\hat{\theta}^0)\}\varphi_2'(\hat{\theta}^0) = 2.23 \times 3.057 \times 1 = 6.8172$$

where the second investigation root informations are evaluated at the individual and the overall maximum likelihood values and the derivative $\partial\varphi_2/\partial\theta = 1$ is unity. This gives the combined canonical parameter

$$\varphi(\theta) = 9.578\varphi_1 + 6.8172\varphi_2 = 9.578\log(\theta) + 6.8172\theta;$$

and then by substituting $\hat{\theta} = .3271$ we obtain $\hat{\varphi} = -8.473$.

The preceding gives us the third order accurate ℓ and φ for the combined investigation. The consequent p-value is then available from the standard saddlepoint approximation for the imputed exponential model $\{\ell(\theta), \varphi(\theta)\}$. The signed likelihood root is available immediately from the combined likelihood:

$$r = +[2\{\ell(.3271) - \ell(.1)\}]^{1/2} = 1.981.$$

And the standardized maximum likelihood departure $q(.1)$ is calculated in the canonical parameterization φ. For this we need the departure

$$\varphi(\hat{\theta}) - \varphi(.1) = -8.473 - (-21.372) = 12.90,$$

as obtained by substituting the observed and expected maximum likelihood values .3271 and .1 in the expression for the combined canonical parameter $\varphi(\theta)$; and we need the root observed information as obtained from the combined likelihood:

$$\jmath_{\varphi\varphi}^{1/2}(\hat{\theta}) = \left\{ \frac{4}{(.3271)^2} \cdot \left(\frac{9.576}{.3271} + 6.817 \right)^{-2} \right\}^{1/2} = .1694,$$

where an indicated first factor is a second derivative with respect to θ and the second factor gives rescaling to the φ parameterization. Then combining the preceding gives

$$q = .1694 \times 12.90 = 2.185.$$

We can now calculate the third order standardized departure

$$r^* = 1.981 - 1.981^{-1} \log \left(\frac{1.981}{2.185} \right) = 2.031$$

and then the third order combined p-value

$$p(.1) = \Phi(2.031) = .979.$$

The individual highly accurate p-values 95.01% and 80.00% lead to the highly accurate combined value 97.9%. By contrast the simple and easy Fisher procedure works from the reverse values .0499 and .2000 giving .056 and thus produces the p-value 94.4%.

   In this instance the third order accurate p-value is substantially more significant in the convention senses. While simulations to evaluate the present third order value might easily be suggested we refer to very extensive validations of third order as in Bédard et al (2008) and Fraser et al (2009) and the references therein.

5   Discussion

We have developed general theory for combining p-values from statistically independent investigations. Foremost, of course, is the addition of log-likelihoods from component investigations. This aspect is the credo of the Bayesian approach where it is often presented as if such likelihood combining was not a routine standard of central statistics, although often neglected. And then to go beyond the first order accuracy of likelihood-only analysis, we target the weighted combination of the canonical parameters which provides the central feature of higher order likelihood theory. The weighting is not by reciprocal variance or information as perhaps might be expected, but by the product of two root informations, one at the component maximum likelihood value and the other at the overall maximum likelihood value; and the informations need to be calculated in the scale of the individual canonical parameters.

   As future work we anticipate the development of third order tangent models for interest parameters and the development of appropriate combining procedures for the corresponding information summaries.

1 Acknowledgements

9 References

[1] Barndorff-Nielsen, O.E. (1986). Inference on full or partial parameters based on the standardized log likelihood ratio. *Biometrika*, **73**, 307–322.

[2] Barndorff-Nielsen, O.E. and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications. *J.R. Statist. Soc. B*, **41**, 279–312.

[3] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Higher accuracy for Bayesian and frequentest inference: Large sample theory for small sample likelihood. *Statistical Science* **22**, 301-321.

[4] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Cakmak, S., Fraser, D.A.S., McDunnough, P., Reid, N., and Yuan, X. (1998). Likelihood centered asymptotic model: exponential and location model versions. *J. Statist. Planning and Inference* **66**, 211-222.

[5] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Annals Math. Statist.* **25**, 631-650.

[6] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Davison, A.C., Fraser, D.A.S. and Reid, N. (2006). Improved likelihood inference for discrete data. *J. Roy. Statist. Soc.***B 68**, 495-508.

[7] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Fraser, A.M., Fraser, D.A.S. and Staicu, A.-M. (2010). Second order ancillary: A differential view from continuity. *Bernoulli*, to appear

[8] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Fraser, D.A.S., and Reid, N. (1995). Ancillaries and third order significance. *Utilitas Mathematica* **47**, 33-53.

[9] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Fraser, D.A.S., and Reid, N. (2001). Ancillary information for statistical inference. In S.E. Ahmed and N. Reid (Eds), *Empirical Bayes and Likelihood Inference*, 185-209. New York: Springer-Verlag.

[10] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Fraser, D.A.S., Reid, N. and Wong, A. (2004). Inference for bounded parameters *Physics Review D*, **69**, 033002.

1 [11] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Fraser, D.A.S., Reid, N., & Wu,
2   J.(1999). A simple general formula for tail probabilities for Bayes and frequentist infer-
3   ence. *Biometrika* **86**, 249-264.

4 [12] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Fraser, D.A.S., Wong, A. and Sun,
5   Ye (2009). Three enigmatic examples and inference from likelihood. *Canadian Journal*
6   *of Statistics*, **37**, 1-21.

7 [13] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Lugannani, R. and Rice, S. (1980),
8   Saddlepoint approximation for the distribution of the sum of independent random vari-
9   ables, *Advances in Applied Probability* **12**, 475–490.

10 [14] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Reid, N, and Fraser, D.A.S. (2010).
11   Mean likelihood and higher order inference. *Biometrika* **97**, to appear.

12 [15] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Reid, N. and Fraser, D.A.S. (2003).
13   Likelihood inference in the presence of nuisance parameters. In *Proceedings of PHYS-*
14   *TAT2003*, L. Lyons, R. Mount, R. Reitmeyer, eds. SLAC e-Conf C030908, 265–271.

15 [16] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Stainforth, D. A., Allen, M. R.,
16   Tredger, E. R. and Smith, L. A. (2007). Confidence, uncertainty and decision-support
17   relevance in climate predictions. *Phil. Trans. Roy. Soc.* A, **365**, 2145-2162.

18 [17] Bédard, M., Fraser, D.A.S. and Wong, A. (2008). Welch, B.L. and Peers, H.W. (1963).
19   On formulae for confidence points based in intervals of weighted likelihoods. *J. Roy.*
20   *Statist. Soc.* B **25**, 318–329.