Design of the `ump` Package

Charles J. Geyer Glen D. Meeden

March 26, 2017

1 Introduction

1.1 The UMP and UMPU Problems

The following is mostly taken from the first draft of the fuzzy confidence intervals and P-values paper.

1.1.1 UMP

Lehmann (1959, pp. 68–69) says for a one-parameter model with likelihood ratio monotone in the statistic $T(X)$ there exists a UMP test having null hypothesis $H_0 = \{ \vartheta : \vartheta \le \theta \}$, alternative hypothesis $H_1 = \{ \vartheta : \vartheta > \theta \}$, significance level α, and critical function ϕ defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) > C \\ \gamma, & T(x) = C \\ 0, & T(x) < C \end{cases} \tag{1}$$

where the constants γ and C are determined by

$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha.$$

The description of the analogous lower-tailed test is the same except that all inequalities are reversed.

The constant C is clearly any $(1 - \alpha)$-th quantile of the distribution of $T(X)$ for the parameter value θ. If C is not an atom of this distribution, then the test is effectively not randomized and the value of γ is irrelevant. Otherwise

$$\gamma = \frac{\alpha - \Pr_\theta\{T(X) > C\}}{\Pr_\theta\{T(X) = C\}}. \tag{2}$$

1.1.2 UMPU

Lehmann (1959, pp. 126–127) says for a one-parameter exponential family model with canonical statistic $T(X)$ and canonical parameter θ there exists a

UMPU test having null hypothesis $H_0 = \{ \vartheta : \vartheta = \theta \}$, alternative hypothesis $H_1 = \{ \vartheta : \vartheta \neq \theta \}$, significance level α, and critical function ϕ defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) < C_1 \\ \gamma_1, & T(x) = C_1 \\ 0, & C_1 < T(x) < C_2 \\ \gamma_2, & T(x) = C_2 \\ 1, & C_2 < T(x) \end{cases} \tag{3}$$

where $C_1 \leq C_2$ and the constants γ_1, γ_2, C_1, and C_2 are determined by

$$E_\theta\{\phi(X, \alpha, \theta)\} = \alpha \tag{4a}$$
$$E_\theta\{T(X)\phi(X, \alpha, \theta)\} = \alpha E_\theta\{T(X)\} \tag{4b}$$

If $C_1 = C_2 = C$ in (3) then $\gamma_1 = \gamma_2 = \gamma$ also. This occurs only in a very special case. Define

$$p = \Pr_\theta\{T(X) = C\} \tag{5a}$$
$$\mu = E_\theta\{T(X)\} \tag{5b}$$

Then in order to satisfy (4a) and (4b) we must have

$$1 - (1 - \gamma)p = \alpha$$
$$\mu - C(1 - \gamma)p = \alpha\mu$$

which solved for γ and C gives

$$\gamma = 1 - \frac{1 - \alpha}{p} \tag{6a}$$
$$C = \mu \tag{6b}$$

Thus this special case occurs only when μ an atom of the distribution of $T(X)$ for the parameter value θ, and then only for very large significance levels: $\alpha > 1 - p$. Hence this special case is of no practical importance, although it is of some computational importance to get every case right, no weird bogus results or crashes in unusual special cases.

Returning to the general case, assume for a second that we have particular C_1 and C_2 that work for some x, α, and θ (we will see how to determine C_1 and C_2 presently). With μ still defined by (5b) and with the definitions

$$p_i = \Pr_\theta\{T(X) = C_i\}, \qquad i = 1, 2 \tag{7a}$$
$$p_{12} = \Pr_\theta\{C_1 < T(X) < C_2\} \tag{7b}$$
$$m_{12} = E_\theta\{T(X) I_{(C_1, C_2)}[T(X)]\} \tag{7c}$$

(4a) and (4b) become

$$1 - (1 - \gamma_1)p_1 - (1 - \gamma_2)p_2 - p_{12} = \alpha \tag{8a}$$
$$\mu - C_1(1 - \gamma_1)p_1 - C_2(1 - \gamma_2)p_2 - m_{12} = \alpha\mu \tag{8b}$$

which solved for γ_1 and γ_2 give

$$\gamma_1 = 1 - \frac{(1-\alpha)(C_2 - \mu) + m_{12} - C_2 p_{12}}{p_1(C_2 - C_1)} \tag{9a}$$

$$\gamma_2 = 1 - \frac{(1-\alpha)(\mu - C_1) - m_{12} + C_1 p_{12}}{p_2(C_2 - C_1)} \tag{9b}$$

Note that (9a) and (9b) are linear in α. They are valid over the range of α (if any) such that both equations give values between zero and one.

Now we turn to the determination of C_1 and C_2. We present an algorithm that determines $\phi(x, \alpha, \theta)$ for any discrete one-parameter exponential family with canonical statistic $T(X)$ for all values of x and α for one fixed value of θ.

1. Start with $\alpha = 1$.

   (a) If μ given by (5b) is an atom, then $\phi(x, \alpha, \theta)$ is given by (3) with $C_1 = C_2 = \mu$ and $\gamma_1 = \gamma_2 = \gamma$ given by (5a), (6a), and (6b) over the range of α such that (6a) is between zero and one.

   (b) If μ given by (5b) is not an atom, then choose C_1 and C_2 to be adjacent atoms such that $C_1 < \mu < C_2$ and $\phi(x, \alpha, \theta)$ is given by (3) with γ_1 and γ_2 given by (7a), (7b), (7c), (9a), and (9b) over the range of α such that both (9a) and (9b) are between zero and one. [Because C_1 and C_2 are adjacent atoms, $p_{12} = m_{12} = 0$, and $\alpha = 1$ gives $\gamma_1 = \gamma_2 = 1$, a valid solution.]

2. Start with the smallest α for which $\phi(x, \alpha, \theta)$ was determined in step 1 or a previous iteration of step 2. At this point, either γ_1 or γ_2 is zero (or both are).

   (a) If γ_1 is zero, then decrease C_1 to the adjacent lower atom and set $\gamma_1 = 1$ [which does not change the value of $\phi(x, \alpha, \theta)$ for any x].

   (b) If γ_2 is zero, then increase C_2 to the adjacent higher atom and set $\gamma_2 = 1$ [which does not change the value of $\phi(x, \alpha, \theta)$ for any x].

   (c) Now $\phi(x, \alpha, \theta)$ is given by (6a) with γ_1 and γ_2 given by (7a), (7b), (7c), (9a), and (9b) over the range of α such that both (9a) and (9b) are between zero and one [because of steps (a) and (b), both γ_i are now greater than zero, so α can be decreased].

3. Repeat step 2 until the whole range $0 \le \alpha \le 1$ is covered.

This algorithm is certainly unwieldy, but it does make clear that $\alpha \mapsto \phi(x, \alpha, \theta)$ is (1) continuous, (2) piecewise linear, (3) nondecreasing, and (4) onto $[0, 1]$. Hence it is the distribution function of a continuous random variable (an abstract randomized P-value). Clearly, it is differentiable on each linear piece and the derivative is piecewise constant (a step function).

1.2 Endpoint Behavior

The UMPU test is not well defined when the null hypothesis is on the boundary of the parameter space. But equations (3), (4a), and (4b) still make sense and define a test. Since the probability and the expectation in those equations are continuous in θ this also characterizes the behavior as θ converges to a boundary point (which we need to know to calculate fuzzy confidence intervals, which involve all θ in the parameter space).

Theorem 1. *Suppose the setup for a UMPU test described in the first paragraph of Section 1.1.2. In addition suppose that $T(X)$ has a topologically discrete distribution (concentrated on a countable set of atoms and the atoms topologically isolated) not concentrated at one point. If the support S of $T(X)$ has a lower bound and L is the two-point set consisting of the two smallest atoms of the support, then*

$$1 - \phi(x, \alpha, \theta) \to (1 - \alpha) I_L(x), \qquad \text{as } \theta \to -\infty. \tag{10}$$

Similarly, if the support has an upper bound and L consists of the two largest atoms, then (10) *holds with $-\infty$ replaced by $+\infty$.*

Proof. We do the case where the support has a lower bound. The upper bound case is entirely analogous.

The densities in the family of distributions of $T(X)$ have the form

$$f_\theta(s) = \frac{1}{c(\theta)} e^{s\theta} \lambda(s), \qquad s \in S, \tag{11}$$

where λ is some strictly positive function, and

$$c(\theta) = \sum_{s \in S} e^{s\theta} \lambda(s). \tag{12}$$

The natural parameter space of the family is the set Θ of θ such that (12) is finite. Because S is bounded below, if $c(\psi) < \infty$, then $c(\theta) < \infty$ for all $\theta < \psi$. Thus Θ is either the whole real line or a semi-infinite interval extending to $-\infty$.

For every θ in the interior of Θ, the distribution with density f_θ has a moment generating function M_θ defined by

$$M_\theta(t) = E_\theta\{e^{tT(X)}\} = \frac{c(\theta + t)}{c(\theta)},$$

and hence this distribution has moments of all orders, the mean and variance being given by derivatives at zero of the cumulant generating function $\log M_\theta$

$$\mu(\theta) = E_\theta\{T(X)\} = \frac{d}{d\theta} \log c(\theta)$$
$$\sigma^2(\theta) = \operatorname{var}_\theta\{T(X)\} = \frac{d^2}{d\theta^2} \log c(\theta)$$

Because of our assumption that $T(X)$ is not concentrated at one point, $\sigma^2(\theta) = d\mu(\theta)/d\theta$ can never be zero. Hence μ is a strictly increasing continuous function that maps the interior of Θ to some open interval of the real line.

Write $L = \{s_0, s_1\}$ with $s_0 < s_1$. Since

$$\frac{f_\theta(s)}{f_\theta(s_0)} = e^{(s-s_0)\theta} \frac{\lambda(s)}{\lambda(s_0)}$$

the distribution clearly converges to the distribution concentrated at s_0 as $\theta \to -\infty$.

Now

$$\frac{\mu(\theta) - s_0}{f_\theta(s_1)} = \sum_{s \in S \setminus \{s_0\}} (s - s_0) e^{(s-s_1)\theta} \frac{\lambda(s)}{\lambda(s_1)}$$

goes to $s_1 - s_0$ by monotone convergence as $\theta \to -\infty$. Hence $\mu(\theta) \to s_0$ as $\theta \to -\infty$. Thus μ is a diffeomorphism from the interior of Θ to some open interval of the real line, the lower endpoint of which is s_0.

Now

$$\frac{\Pr_\theta\{T(X) > s_1\}}{f_\theta(s_1)} = \sum_{s \in S \setminus L} e^{(s-s_1)\theta} \frac{\lambda(s)}{\lambda(s_1)}$$

goes to zero by monotone convergence as $\theta \to -\infty$.

And these facts together imply

$$\begin{aligned} \Pr_\mu\{T(X) = s_0\} &= 1 - \frac{\mu - s_0}{s_1 - s_0} + o(\mu - s_0) \\ \Pr_\mu\{T(X) = s_1\} &= \frac{\mu - s_0}{s_1 - s_0} + o(\mu - s_0) \\ \Pr_\mu\{T(X) > s_1\} &= o(\mu - s_0) \end{aligned} \tag{13}$$

where $\mu = \mu(\theta)$ is the mean value parameter.

Now we claim that for small enough values of θ or μ we have $C_1 = s_0$ and $C_2 = s_1$ and the UMPU test is given by equations (3.10a) and (3.10b) in the paper with p_1 and p_2 given by (3.8a) in the paper and $p_{12} = m_{12} = 0$. Let's check. These equations give now

$$\begin{aligned} \gamma_1 &= 1 - \frac{(1-\alpha)(s_1 - \mu)}{s_1 - \mu + o(\mu - s_0)} \\ \gamma_2 &= 1 - \frac{(1-\alpha)(\mu - s_0)}{\mu - s_0 + o(\mu - s_0)} \end{aligned}$$

and clearly both converge to α as $\mu \to s_0$ hence both are between zero and one for small enough θ or μ and hence define the UMPU test. □

This explains the behavior of the fuzzy confidence intervals for the binomial distribution for the two x values nearest each boundary in Figure 2 of the paper. As $\theta \to 0$, the fuzzy confidence interval $1 - \phi(x, \alpha, \theta)$ converges to $1 - \alpha$ for $x = 0$ or $x = 1$ and converges to zero for all other x. And as $\theta \to 1$, the fuzzy confidence interval converges to $1 - \alpha$ for $x = n - 1$ or $x = n$ and converges to zero for all other x.

1.3 Models With Nuisance Parameters

UMP and UMPU theory extends to multiparameter exponential families when the parameter of interest θ is one of the canonical parameters (Lehmann, TSH, 1st ed., pp. 134–136).

Suppose the family has densities of the form

$$\frac{1}{c(\theta, \boldsymbol{\eta})} \exp\left(\theta T(x) + \sum_{i=1}^{k} \eta_i U_i(x)\right)$$

with respect to some measure on the sample space. Then the situation is exactly the same as described above except that the reference distribution of the test is the conditional distribution of $T(X)$ given $\mathbf{U}(X)$, which (a standard fact about exponential families) depends only on θ and not on the nuisance parameter $\boldsymbol{\eta}$.

1.3.1 UMP Tests With Nuisance Parameters

Now there exists a UMP test having null hypothesis $H_0 = \{\, \vartheta : \vartheta \le \theta \,\}$, alternative hypothesis $H_1 = \{\, \vartheta : \vartheta > \theta \,\}$, and significance level α, and its critical function ϕ is defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) > C[\mathbf{U}(x)] \\ \gamma[\mathbf{U}(x)], & T(x) = C[\mathbf{U}(x)] \\ 0, & T(x) < C[\mathbf{U}(x)] \end{cases} \tag{14}$$

where the functions γ and C are determined by

$$E_\theta\{\phi(X, \alpha, \theta) \mid \mathbf{U}(X)\} = \alpha. \tag{15}$$

Everything is exactly the same as for the one-parameter case except for the conditioning on $\mathbf{U}(x)$. The only point of the discussion is that the test is UMP whether considered conditionally or unconditionally.

As before, the UMP upper-tailed test is obtained by reversing all the inequalities above.

1.3.2 UMPU Tests With Nuisance Parameters

Now there exists a UMPU test having null hypothesis $H_0 = \{\, \vartheta : \vartheta = \theta \,\}$, alternative hypothesis $H_1 = \{\, \vartheta : \vartheta \ne \theta \,\}$, and significance level α, and its critical function ϕ is defined by

$$\phi(x, \alpha, \theta) = \begin{cases} 1, & T(x) < C_1[\mathbf{U}(x)] \\ \gamma_1[\mathbf{U}(x)], & T(x) = C_1[\mathbf{U}(x)] \\ 0, & C_1[\mathbf{U}(x)] < T(x) < C_2[\mathbf{U}(x)] \\ \gamma_2[\mathbf{U}(x)], & T(x) = C_2[\mathbf{U}(x)] \\ 1, & C_2[\mathbf{U}(x)] < T(x) \end{cases} \tag{16}$$

where the functions γ_1, γ_2, C_1, and C_2 are determined by

$$E_\theta\{\phi(X, \alpha, \theta) \mid \mathbf{U}(X)\} = \alpha \tag{17a}$$
$$E_\theta\{T(X)\phi(X, \alpha, \theta) \mid \mathbf{U}(X)\} = \alpha E_\theta\{T(X) \mid \mathbf{U}(X)\} \tag{17b}$$

Again, the point is that the test is UMPU whether considered conditionally or unconditionally.

2 Models

2.1 Binomial

Let $X \sim \text{Bin}(n, p)$ with $0 < p < 1$.

All of the quantities in (9a) and (9b) are easily calculated (in R) except possibly m_{12}. Actually, as Lehmann points out (TSH, 1st, ed., pp. 128–129), this quantity is also easy to calculate

$$\sum_{x=C_1+1}^{C_2-1} x \binom{n}{x} p^x (1-p)^{n-x} = np \sum_{x=C_1+1}^{C_2-1} \binom{n-1}{x-1} p^{x-1} (1-p)^{(n-1)-(x-1)}$$

and the sum on the right is just a binomial probability for the $\text{Bin}(n-1, p)$ distribution, that is, the right side can be calculated in R (with the obvious definitions of the variables) by

```
n * p * (pbinom(c2 - 2, n - 1, p) - pbinom(c1 - 1, n - 1, p))
```

assuming `pbinom(c1 - 1, n - 1, p)` does not crash and produces zero when `c1` is zero (which it does in recent versions I can check).

2.2 Poisson

Let $X \sim \text{Pois}(\mu)$ with $0 < \mu$.

Again, all of the quantities in (9a) and (9b) are easily calculated except possibly m_{12}. Does it work like the binomial case? Yes!

$$\sum_{x=C_1+1}^{C_2-1} x \frac{\mu^x}{x!} e^{-\mu} = \mu \sum_{x=C_1+1}^{C_2-1} \frac{\mu^{x-1}}{(x-1)!} e^{-\mu}$$

and the sum on the right is just another Poisson probability, that is, the right side can be calculated in R (with the obvious definitions of the variables) by

```
mu * (ppois(c2 - 2, mu) - ppois(c1 - 1, mu))
```

2.3 Negative Binomial

Let $X \sim \text{NegBin}(r, p)$ with $0 < p < 1$. Like R we consider the sample space to start at zero rather than r. This also allows for non-integer r. The densities of the family have the form

$$f(x) = \frac{\Gamma(x + r)}{\Gamma(r)x!} p^r (1 - p)^x$$

Note that if we are to have an exponential family r cannot be an unknown parameter! The only unknown parameter is p.

Again, all of the quantities in (9a) and (9b) are easily calculated except possibly m_{12}. Does it work like the binomial case? Yes!

$$\sum_{x=C_1+1}^{C_2-1} x \frac{\Gamma(x + r)}{\Gamma(r)x!} p^r (1 - p)^x = \frac{1 - p}{p} \sum_{x=C_1+1}^{C_2-1} \frac{\Gamma(x - 1 + r + 1)}{\Gamma(r)(x - 1)!} p^{r+1} (1 - p)^{x-1}$$

and the sum on the right is just another negative binomial probability, that is, the right side can be calculated in R (with the obvious definitions of the variables) by

```
(1 - p) / p * (pnbinom(c2 - 2, r + 1, p)
    - pnbinom(c1 - 1, r + 1, p))
```

2.4 Two Independent Poisson Random Variables

Let $X_i \sim \text{Pois}(\mu_i)$ with $0 < \mu_i$, for $i = 1, 2$ be independent random variables. We wish to compare the means μ_1 and μ_2. We cannot just test or produce fuzzy confidence intervals for a function pulled out of the air, such as $\mu_1 - \mu_2$. The parameter we test must be canonical.

The canonical statistics of this exponential family are X_1 and X_2 and the corresponding canonical parameters are $\psi_i = \log(\mu_i)$. Linear functions of canonical parameters are again canonical so we can test or produce fuzzy confidence intervals for $\psi_1 - \psi_2 = \log(\mu_1/\mu_2)$.

Introduce new parameters

$$\begin{aligned} \psi_1 &= \eta + \theta \\ \psi_2 &= \eta \end{aligned}$$

Then

$$X_1 \psi_1 + X_2 \psi_2 = X_1 \theta + (X_1 + X_2) \eta = T(\mathbf{X}) \theta + U(\mathbf{X}) \eta$$

Where

$$\begin{aligned} T(\mathbf{X}) &= X_1 \\ U(\mathbf{X}) &= X_1 + X_2 \end{aligned} \tag{18}$$

It is a standard result that the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$ is

$$X_1 \mid X_1 + X_2 \sim \text{Bin}\left(X_1 + X_2, \frac{\mu_1}{\mu_1 + \mu_2}\right)$$

So the theory says we do the UMP or UMPU test based on this distribution with $\mu_1/(\mu_1 + \mu_2)$ as the parameter of interest (Lehmann, TSH, 1st ed., pp. 140–142, gives further details).

2.5 Two Independent Binomial Proportions

Let $X_i \sim \text{Bin}(n_i, p_i)$ with $0 < p_i < 1$, for $i = 1, 2$ be independent random variables. We wish to compare the proportions p_1 and p_2. We cannot just test or produce fuzzy confidence intervals for a function pulled out of the air, such as $p_1 - p_2$. The parameter we test must be canonical.

The canonical statistics of this exponential family are X_1 and X_2 and the corresponding canonical parameters are $\psi_i = \text{logit}(p_i)$. Linear functions of canonical parameters are again canonical so we can test or produce fuzzy confidence intervals for $\psi_1 - \psi_2$.

As in the Poisson case we see that we can base the test on the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$, where these variables are defined by (18). This distribution is known although is is not "nice" and is not a "brand name" distribution. It is (Lehmann, TSH, 1st ed., pp. 142–143) the exponential family generated by the hypergeometric distribution.

$$\Pr\{X_1 = t \mid X_1 + X_2 = u\} = \frac{1}{c(\rho)} \binom{n_1}{t} \binom{n_2}{u - t} \rho^t, \qquad t = 0, \ldots, u, \tag{19}$$

where

$$\rho = \frac{p_1(1 - p_2)}{(1 - p_1)p_2}$$

and

$$c(\rho) = \sum_{t=0}^{u} \binom{n_1}{t} \binom{n_2}{u - t} \rho^t.$$

So the theory says we do the UMP or UMPU test based on this distribution with ρ as the parameter of interest

This distribution is very hairy. For starters the actual range of the random variable $X_1 \mid X_1 + X_2$ is not 0 to $X_1 + X_2$ as (19) seems to indicate. The binomial coefficients can evaluate to zero.

We will leave the complexity of this model and move on to other models. In principle it is just a one-parameter exponential family and hence "nice" in certain respects. In practice, there is no readily available software to calculate distribution and density functions, so this model requires more effort in computer implementation.

2.6 Two Independent Negative Binomial Variables

Let $X_i \sim \text{NegBin}(r_i, p_i)$ with $0 < r_i$ and $0 < p_i < 1$, for $i = 1, 2$ be independent random variables. As in Section 2.3 we are using the convention that the sample space starts at zero. We wish to compare the proportions p_1 and p_2. We cannot just test or produce fuzzy confidence intervals for a function pulled out of the air, such as $p_1 - p_2$. The parameter we test must be canonical.

The canonical statistics of this exponential family are X_1 and X_2 and the corresponding canonical parameters are $\psi_i = \log(1 - p_i)$. Linear functions of canonical parameters are again canonical so we can test or produce fuzzy confidence intervals for $\psi_1 - \psi_2$.

As in the Poisson case we see that we can base the test on the conditional distribution of $T(\mathbf{X})$ given $U(\mathbf{X})$, where these variables are defined by (18). This distribution is not fully explained in Lehmann, although the $r_1 = r_2 = 1$ case is the subject of a homework problem in the second edition.

Let's see what happens. The joint distribution of the X's is

$$\begin{aligned} f(x_1, x_2) &= \prod_{i=1}^{2} \frac{\Gamma(x_i + r_i)}{\Gamma(r_i) x_i!} p_i^{r_i} (1 - p_i)^{x_i} \\ &= \exp(x_1 \psi_1 + x_2 \psi_2) \prod_{i=1}^{2} \frac{\Gamma(x_i + r_i)}{\Gamma(r_i) x_i!} p_i^{r_i} \\ &= \exp(t\theta + u\eta) \frac{\Gamma(t + r_1)}{\Gamma(r_1) t!} \frac{\Gamma(u - t + r_2)}{\Gamma(r_2)(u - t)!} p_1^{r_1} p_2^{r_2} \end{aligned} \tag{20}$$

where

$$\begin{aligned} t &= x_1 \\ u &= x_1 + x_2 \\ \theta &= \psi_1 - \psi_2 \\ \eta &= \psi_2 \end{aligned}$$

It matters not that we have not specified p_1 and p_2 as a function of θ and η. We want to consider the conditional distribution anyway. Thought of as a function of t for fixed u and dropping all terms that do not contain both parameters and t we get

$$f_\theta(t \mid u) = \frac{1}{c(\theta)} \exp(t\theta) \frac{\Gamma(t + r_1)\Gamma(u - t + r_2)}{t!(u - t)!}, \qquad t = 0, \ldots, u, \tag{21}$$

where

$$c(\theta) = \sum_{t=0}^{u} \exp(t\theta) \frac{\Gamma(t + r_1)\Gamma(u - t + r_2)}{t!(u - t)!}.$$

Hmmmm. Also not a brand name family and not the exponential family generated by the hypergeometric distribution we saw in the case of two independent

binomial proportions. Still just another one-parameter exponential family. No big deal.

Actually, Mathematica says this has some sort of relation to hypergeometric functions, for example

```
In[4]:= Sum[ rho^t Gamma[t + r] Gamma[u - t + s] /
        (Gamma[t + 1] Gamma[u - t + 1]), {t, 0, u} ]

        Gamma[r] Gamma[s + u] Hypergeometric2F1[r, -u, 1 - s - u, rho]
Out[4]= ------------------------------------------------------------
                                  Gamma[1 + u]
```

So Mathematica knows how to calculate the normalizing function and it involves a "hypergeometric ${}_2F_1$ function. Maybe we should call this the exponential family generated by the negative hypergeometric distribution.

According to

```
http://planetmath.org/encyclopedia/NegativeHypergeometricDistribution.html
```

there is a *negative hypergeometric distribution* having density

$$f(x) = \frac{\binom{x+b-1}{x}\binom{W+B-b-x}{W-x}}{\binom{W+B}{W}}, \qquad x = 0, 1, \ldots, W,$$

where W, B, and b are positive integers. The web page explains that this is the distribution of the number of "special items" X (from W special items in the population) present before the bth object from a population with B items altogether.

Comparing with (21) we see that the base measure of the family does indeed have this form when r_1 and r_2 are integers.

2.7 Testing Independence in a Two-by-two Table

This is the UMP/UMPU competitor for Fisher's exact test. The data consist of a matrix X_{ij}, $i = 1, 2$, $j = 1, 2$, that has a multinomial distribution with sample size n and cell probability matrix p_{ij}, $i = 1, 2$, $j = 1, 2$. This is also called a two-by-two contingency table.

The canonical statistics are the X_{ij}, but the canonical parameters are not uniquely defined in terms of the p_{ij} because the model is really only three dimensional, not four, because the X_{ij} sum to n.

As is well known, this is a three-dimensional exponential family, the canonical statistics being any three of the four X_{ij}, the fourth being determined from the other three by the requirement that the X_{ij} sum to n.

In this problem (Lehmann, TSH, 1st ed., pp. 143–146) the marginals are the statistics for the nuisance parameters, and we can consider any other statistic linearly independent of the marginals and the sum of all cells as the statistic of

interest. Lehmann chooses

$$\begin{aligned} T(\mathbf{X}) &= X_{11} \\ U_1(\mathbf{X}) &= X_{11} + X_{12} \\ U_2(\mathbf{X}) &= X_{11} + X_{21} \end{aligned} \tag{22}$$

The conditional distribution of $T(\mathbf{X})$ given the marginals $U_1(\mathbf{X})$ and $U_2(\mathbf{X})$ is well known. It is the hypergeometric distribution involved in Fisher's exact test under the null hypothesis of independence and under general null hypotheses is the exponential family generated by the hypergeometric we encountered in Section 2.5.

$$\begin{aligned} &Pr\{X_{11} = t \mid X_{11} + X_{12} = u_1, X_{11} + X_{21} = u_2\} \\ &\qquad = \frac{1}{c(\rho)} \binom{u_1}{t} \binom{n}{u_2 - t} \rho^t, \qquad t = 0, \ldots, n \end{aligned} \tag{23}$$

where

$$\rho = \frac{p_{11} p_{22}}{p_{12} p_{21}}$$

and

$$c(\rho) = \sum_{t=0}^{n} \binom{u_1}{t} \binom{n}{u_2 - t} \rho^t.$$

As in Section 2.5 we note that this distribution given by (23) often does not have the full range 0 to n because the binomial coefficients may be zero.

2.8 Different Answers to the Same Question in a Poll

This section and the next give the UMP/UMPU/fuzzy competitors to the analysis of correlated binomial proportions in Wild and Seber (pp. 343–350). The first considers different answers to the same question on a poll. This is a multinomial problem. Say the categories of interest have counts X_1 and X_2, then we know

$$X_1 \mid X_1 + X_2 \sim \text{Bin}\left(X_1 + X_2, \frac{p_1}{p_1 + p_2}\right)$$

and so the UMP/UMPU/fuzzy procedures are based on this distribution.

2.9 Answers to Different Questions in a Poll

Here again, like in Section 2.7, we have a two-by-two table with data X_{ij} and cell probabilities p_{ij} but the question of interest is different. Now we are interested in just the opposite question, whether the marginals differ. This in a sense (a rather vague sense) interchanges the role of interest and nuisance parameters, what were interest parameters in Section 2.7 are now nuisance parameters and vice versa.

Ordinarily, this would be nonsense. There is exactly one interest parameter. The rest (in this case two) must be nuisance parameters. So, strictly speaking,

they cannot be interchanged. But a two-by-two table has a redundant canonical statistic: there are four X_{ij} but they sum to n so only three are linearly independent. So if we add the redundant statistic to the statistics corresponding to parameters of interest we two sets of two that can be interchanged.

It is clear that (22) could have been written with subscripts 1 and 2 interchanged, which would make X_{22} the statistic of interest. This tells us that here we should condition on X_{11} and X_{22} leaving either X_{12} or X_{21} as the statistic of interest. Thus in this case the UMP/UMPU/fuzzy procedure is based on the distribution

$$X_{12} \mid X_{11}, X_{22} \sim \operatorname{Bin}\left(n - X_{11} - X_{22}, \frac{p_{12}}{p_{12} + p_{21}}\right)$$

And in hindsight we see that we have invented the conditional, exact, uniformly most whatever (UMW) competitor of McNemar's test (Lindgren, *Statistical Theory*, pp. 381–383).

3 Algorithms

3.1 Version 0.1

After a great struggle, a very simple algorithm was decided on for calculating $\phi(x, \alpha, \theta)$. Given α and θ, calculate the appropriate c_1, c_2, γ_1, and γ_2 as follows.

First handle the special cases where α is zero or one and θ is on the boundary of the parameter space, using Theorem 1 above and the obvious fact that ϕ is identically equal to one when $\alpha = 1$ and identically equal to zero when $\alpha = 0$.

When in the general case $0 < \alpha < 1$ and θ not on the boundary choose some c_1 and c_2 such that $c_1 \le E_\theta\{T(X)\} \le c_2$. We pick the c_1 and c_2 that give, with randomization, an equal tailed test, in the hope that this is close to correct.

Then we go into an infinite loop that does the following.

- Calculate γ_1 and γ_2 using the current guesses for c_1 and c_2 and equations (9a) and (9b) above. If the results satisfy $0 \le \gamma_1 \le 1$ and $0 \le \gamma_2 \le 1$, then we are done and stop the loop.
- Otherwise, we change c_1 or c_2, the one corresponding to the γ_i that violates the constraints worst. If this γ_i is negative, we move the c_i out by one (i. e. decrease c_1 or increase c_2) and if this γ_i is greater than one, we move the c_i in by one.

Actually we don't do an infinite loop, because we have no theorem that says this algorithm converges, so we have a maximum iteration count (default 10) and just give up when it is reached. In the examples we have done, there has been no need to increase the iteration count.

See `ump/src/umpubinom.c` for an example of this algorithm. See `ump/tests/umpub.R` for the tests it passed.

3.2 Version 0.3

An attempt to implement the density of abstract randomized P-values and test the implementation shows that equations (9a) and (9b) are no good. The exhibit catastrophic cancellation for small alpha.

So we back up to (7a), (7b), and (7c) and keep (7a) but ignore the other two, replacing them with

$$P_1 = \Pr_\theta\{T(X) < C_1\} \tag{24a}$$
$$P_2 = \Pr_\theta\{T(X) > C_2\} \tag{24b}$$
$$M_1 = E_\theta\{T(X) I_{(-\infty, C_1)}[T(X)]\} \tag{24c}$$
$$M_2 = E_\theta\{T(X) I_{(C_2, \infty)}[T(X)]\} \tag{24d}$$

Then we replace (8a) and (8b) by

$$P_1 + \gamma_1 p_1 + \gamma_2 p_2 + P_2 = \alpha \tag{25a}$$
$$M_1 + \gamma_1 C_1 p_1 + \gamma_2 C_2 p_2 + M_2 = \alpha\mu \tag{25b}$$

which solved for γ_1 and γ_2 give

$$\gamma_1 = \frac{\alpha(C_2 - \mu) + (M_1 - C_2 P_1) + (M_2 - C_2 P_2)}{p_1(C_2 - C_1)} \tag{26a}$$
$$\gamma_2 = \frac{\alpha(\mu - C_1) - (M_1 - C_1 P_1) - (M_2 - C_1 P_2)}{p_2(C_2 - C_1)} \tag{26b}$$

This seems to work better, but honesty compels us to admit that this formula also is subject to catastrophic cancellation. Perusal of the source code reveals several ad hoc bits of code that deal with special cases in which the code without the adhockery fails due to catastrophic cancellation or other problems with the inexactitude of floating point arithmetic.

It is fair to say that our code is far from an elegant and provably correct solution to this problem. We think the algorithm presented on page 3 would actually be better than the one we used in all respects except that it takes time proportional to the sample size, which was deemed unacceptable (perhaps wrongly).

References

LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York. (2nd ed., Wiley, 1986; Springer, 1997).