Statistical Inference

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA
Spring, 2001

April 19, 2001

1. Cross Classification

Data from many sources including clinical trials, survey analysis, etc. include subjects who are classified in two (or more) different ways, with the inferential aim of discovering whether or not the two classifications are independent and, if not, of quantifying the departure from independence. Let us illustrate by examining two common kinds of medical trials

1.1. 2×2 Tables: RCT and CCS

For example, the subjects in a prospective clinical trial or Randomized Controlled Trial (RCT) might initially be classified into *treated* and *untreated* (more often called "control") groups, then later be classified into *successes* and *failures*. The hypothesis that these two classifications are independent would be tantamount to the hypothesis that the treated and control groups do not differ in their success rates.

For rare conditions or diseases it is common to do the analysis the other way around— to identify individuals with the condition or disease (usually called "Cases," but to help us consider the two halves of this example together, we'll call these *failures*) and to match these individuals with otherwise-similar non-cases (usually called "Controls" but, for us, *successes*). Subjects are then classified into Treated (or sometimes "exposed") and Untreated (unexposed) in such a Case-Control Study (CCS).

In each of these cases the twice classified subjects fall into one of four groups, with counts n_{ij} as follows:

	S	F	
U	n_{00}	n_{01}	n_{0+}
T	n_{10}	n_{11}	n_{1+}
	n_{+0}	n_{+1}	n

and so the classifications for subjects drawn independently (with replacement) at random from all those included in such a trial would follow a multinomial distribution $\vec{n} \sim \mathsf{MN}(n, \vec{p})$ with pmf

$$f(\vec{n} \mid n, \vec{p}) = \binom{n}{n_{00}, n_{01}, n_{10}, n_{11}} p_{00}{}^{n_{00}}\, p_{01}{}^{n_{01}}\, p_{10}{}^{n_{10}}\, p_{11}{}^{n_{11}}$$

For RCT's the natural hypothesis to test is

$$H_0 : p_{\mathsf{S|T}} = p_{\mathsf{S|C}},$$

where $p_{\mathsf{S|T}} = p_{11}/p_{1+}$ and $p_{\mathsf{S|U}} = p_{01}/p_{0+}$ are the conditional success probabilities for treated and untreated subjects, respectively; for CCS's the natural hypothesis to test is

$$H_0 : p_{\mathsf{T|S}} = p_{\mathsf{T|F}},$$

where $p_{\mathsf{T|S}} = p_{11}/p_{+1}$ and $p_{\mathsf{T|F}} = p_{10}/p_{+0}$ are the conditional treatment probabilities for non-cases and cases, respectively. Actually both of these are the *same* hypothesis, mathematically equivalent to

$$H_0 : p_{00}\, p_{11} - p_{01} p_{10} = 0$$

or

$$H_0 : \log\left(\frac{p_{00}\, p_{11}}{p_{01}\, p_{10}}\right) = 0.$$

There are a variety of ways of testing this hypothesis (Breslow and Day, 1980, Chap. IV) or, alternatively, of estimating the "log odds ratio" $\theta = \log(p_{00}\, p_{11}/p_{01}\, p_{10})$. Studies of CCS or RCS design differ in the constraints they impose on the $\{p_{ij}\}$; typically the row sums $p_{\mathsf{i}+}$ are fixed by the experimentor in RCT's, while the column sums $p_{+\mathsf{j}}$ are fixed in CCS's, leaving in each case one unspecified "nuisance" parameter. The Likelihood Ratio test for RCS's would reject H_0 for large values of the statistic

$$
\begin{aligned}
\lambda &= \frac{\sup_{p,q} \binom{n_{1+}}{n_{11}} p^{n_{11}} (1-p)^{n_{10}} \quad \binom{n_{0+}}{n_{01}} q^{n_{01}} (1-q)^{n_{00}}}{\sup_p \binom{n_{1+}}{n_{11}} p^{n_{11}} (1-p)^{n_{10}} \quad \binom{n_{0+}}{n_{01}} p^{n_{01}} (1-p)^{n_{00}}} \\
&= \frac{\left(\frac{n_{00}}{n_{0+}}\right)^{n_{00}} \left(\frac{n_{01}}{n_{0+}}\right)^{n_{01}} \left(\frac{n_{10}}{n_{1+}}\right)^{n_{10}} \left(\frac{n_{11}}{n_{1+}}\right)^{n_{11}}}{\left(\frac{n_{+0}}{n}\right)^{n_{+0}} \quad \left(\frac{n_{+1}}{n}\right)^{n_{+1}}} \\
&= \left(\frac{n_{00}}{n_{0+}n_{+0}/n}\right)^{n_{00}} \left(\frac{n_{01}}{n_{0+}n_{+1}/n}\right)^{n_{01}} \left(\frac{n_{10}}{n_{1+}n_{+0}/n}\right)^{n_{10}} \left(\frac{n_{11}}{n_{1+}n_{+1}/n}\right)^{n_{11}} \\
&= \prod_{ij} (n_{ij}/e_{ij})^{n_{ij}}
\end{aligned}
$$

where $e_{ij} = (n_{\mathsf{i}+}n_{+\mathsf{j}}/n)$ is the "expected" count if H_0 were true using data-based maximum likelihood estimates $p = p_{\mathsf{S|T}} \approx n_{11}/n_{1+}$ and $q = p_{\mathsf{S|U}} \approx n_{01}/n_{0+}$ for the success probabilities. The likelihood ratio statistic for CCS's turns out to be *exactly* the same.

In 1900 Carl Pearson showed that the logarithm of λ is well approximated by $\log\lambda \approx Q/2$ for the test statistic

$$Q \equiv \sum \frac{(n_{ij} - e_{ij})^2}{e_{ij}}$$

and that, if H_0 is true, then Q has approximately a χ^2 distributon with one degree of freedom (the analogous statement is true more generally for cross-classified data, with $(r-1)(c-1)$ degrees of freedom for tables with r rows and c columns). Thus H_0 may be tested classically by rejecting for large values of Q, with a p-value of

$$p = 2\Phi(-Q) \approx 2\Phi(-2\log\lambda),$$

or (perhaps better) by finding an interval estimate for the log odd-ratio

$$\theta \equiv \log \frac{p_{00}\, p_{11}}{p_{01}\, p_{10}} = \log \frac{p_{\mathsf{S|T}}\,(1 - p_{\mathsf{S|U}})}{(1 - p_{\mathsf{S|T}})\, p_{\mathsf{S|U}}} = \log \frac{p_{\mathsf{T|S}}\,(1 - p_{\mathsf{T|F}})}{(1 - p_{\mathsf{T|S}})\, p_{\mathsf{T|F}}}$$

that will quantify any departure from H_0, i.e., will measure any possible effect of treatment. The maximum likelihood estimate $\hat\theta = \log(n_{00}n_{11}/n_{01}n_{10})$ is infinite of any of the cell counts vanish, so many investigators instead use the modified MLE (Breslow and Day, 1980, p. 139)

$$\tilde\theta = \log \frac{(n_{00} + \frac{1}{2})\ (n_{11} + \frac{1}{2})}{(n_{01} + \frac{1}{2})\ (n_{10} + \frac{1}{2})}$$

whose sampling distribution is approximately $\tilde{\theta} \sim \mathsf{No}(\theta, \sigma^2)$ with variance $\sigma^2 \approx \Sigma \frac{1}{n_{ij}+1/2}$. Obviously Bayesian methods and more careful Classical estimation techniques are available as well. For much more consult Bishop, Fienberg, and Holland (1977).

1.2. 2×2 Tables: Estimating θ

The Dirichlet distribution $\vec{p} \sim \mathsf{Di}(\vec{\alpha})$ with pdf

$$\pi(\vec{p} \mid \vec{\alpha}) = \frac{\Gamma(\sum_{ij} \alpha_{ij})}{\prod_{ij} \Gamma(\alpha_{ij})} \prod_{ij} {p_{ij}}^{\alpha_{ij}-1}$$

on the simplex $\{\vec{p} : 0 \le p_{ij}, \sum_{ij} p_{ij} = 1\}$ is conjugate for multinomial data $\vec{n} \sim \mathsf{MN}(n, \vec{p})$ with pmf

$$f(\vec{n} \mid n, \vec{p}) = \binom{n}{n_{00}, n_{01}, n_{10}, n_{11}} {p_{00}}^{n_{00}} \, {p_{01}}^{n_{01}} \, {p_{10}}^{n_{10}} \, {p_{11}}^{n_{11}};$$

the posterior distribution is evidently $\vec{p}|\vec{n} \sim \mathsf{Di}(\vec{\alpha} + \vec{n})$. Thus we can learn about Bayesian posterior distribution of $\theta \equiv \log \frac{p_{00}\, p_{11}}{p_{01}\, p_{10}}$ by considering the distribution of θ for Dirichlet $\vec{p} \sim \mathsf{Di}(\vec{\alpha})$.

If $p \sim \mathsf{Di}(\vec{\alpha})$ is independent of $\lambda \sim \Gamma(\sum_{ij} \alpha_{ij}, \beta)$ then the random variables $\lambda_{ij} \equiv p_{ij}\lambda$ have independent $\Gamma(\alpha_{ij}, \beta)$ distributions; a common way to generate Dirichlet random vectors is to begin with independent $\lambda_{ij} \sim \Gamma(\alpha_{ij}, \beta)$ and set $p_{ij} = \lambda_{ij} / \sum_{ij} \lambda_{ij}$. Notice that λ and β cancel in the expression

$$\begin{aligned} \theta &= \log \frac{p_{00}\, p_{11}}{p_{01}\, p_{10}} \\ &= \log \frac{\lambda_{00}\, \lambda_{11}}{\lambda_{01}\, \lambda_{10}} \\ &= \log \lambda_{00} - \log \lambda_{01} - \log \lambda_{10} + \log \lambda_{11}, \end{aligned}$$

leading us to consider θ as a linear combination of four independent "log gamma" random variables. Let's look at that distribution in more detail.

First recall the definitions of the Gamma function, its logarithm, and its derivatives the "digamma" (or "psi") and "trigamma" functions:

$$
\begin{array}{lllllll}
\Gamma(\alpha) & \equiv & \int_0^\infty t^{\alpha-1}\, e^{-t}\, dt & \Gamma(n+1) & = & \Gamma(n) * n & = & n! \\
\gamma(\alpha) & \equiv & \log\Gamma(\alpha) & \gamma(n+1) & = & \gamma(n) + \log n & = & \sum_1^n \log k \\
\psi(\alpha) & \equiv & \gamma'(\alpha) & \psi(n+1) & = & \psi(n) + \frac{1}{n} & = & -\gamma + \sum_1^n \frac{1}{k} \\
\psi'(\alpha) & \equiv & \gamma''(\alpha) & \psi'(n+1) & = & \psi'(n) - \frac{1}{n^2} & = & \frac{\pi^2}{6} - \sum_1^n \frac{1}{k^2}
\end{array}
$$

Their properties are expounded in (Abramowitz and Stegun, 1964, §9); we'll want the recursion relations above and the initial and asymptotic results

$$
\begin{array}{lllllllll}
\gamma(1) & = & 0 & \gamma(\frac{1}{2}) & = & \frac{1}{2}\log\pi & \gamma(z+\frac{1}{2}) & \approx & z\log z - z + \frac{\log 2\pi}{2} \\
\psi(1) & = & -\gamma & \psi(\frac{1}{2}) & = & -\gamma - \log 4 & \psi(z+\frac{1}{2}) & \approx & \log(z) \\
\psi'(1) & = & \pi^2/6 & \psi'(\frac{1}{2}) & = & \pi^2/2 & \psi'(z+\frac{1}{2}) & \approx & 1/z
\end{array}
$$

with errors of order $O(z^{-1})$, $O(z^{-2})$, and $O(z^{-3})$, respectively, as $z \to \infty$. Note that the initial values and recursion relations above give each function exactly at the integers and half-integers.

Let $X \sim \mathsf{Ga}(\alpha, \beta)$ and $Y \equiv \log X$; then the log moment generating function for Y is

$$\phi(\omega) \equiv \log \mathsf{E}[e^{\omega Y}] = \log \mathsf{E}[X^\omega] = \log \frac{\Gamma(\alpha+\omega)}{\Gamma(\alpha)\ \beta^\omega} = \gamma(\alpha+\omega) - \gamma(\alpha) - \omega\log\beta,$$

so the mean and variance of Y are given exactly and approximately as

$$
\begin{aligned}
\mathsf{E}[Y] &= \phi'(0) = \psi(\alpha) - \log\beta \approx \log(\alpha - \tfrac{1}{2}) - \log\beta \\
\mathsf{V}[Y] &= \phi''(0) = \psi'(\alpha) \approx \tfrac{1}{\alpha - 1/2} \\
\mu = \mathsf{E}[\theta] &= \psi(\alpha_{00}) - \psi(\alpha_{01}) - \psi(\alpha_{10}) + \psi(\alpha_{11}) \\
&\approx \log \tfrac{(\alpha_{00}-1/2)\ (\alpha_{11}-1/2)}{(\alpha_{01}-1/2)\ (\alpha_{10}-1/2)} \\
\sigma^2 = \mathsf{V}[\theta] &= \psi'(\alpha_{00}) + \psi'(\alpha_{01}) + \psi'(\alpha_{10}) + \psi'(\alpha_{11}) \\
&\approx \tfrac{1}{\alpha_{00}-1/2} + \tfrac{1}{\alpha_{01}-1/2} + \tfrac{1}{\alpha_{10}-1/2} + \tfrac{1}{\alpha_{11}-1/2}
\end{aligned}
$$

For the Jeffreys prior $\pi_J(\vec{p}) \sim \mathsf{Di}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ and the Uniform prior $\pi_U(\vec{p}) \sim \mathsf{Di}(1, 1, 1, 1)$ the induced posteriors for θ have exact and approximate means

$$
\begin{aligned}
\pi_J:\ \mu &= \psi(n_{00} + \tfrac{1}{2}) - \psi(n_{01} + \tfrac{1}{2}) - \psi(n_{10} + \tfrac{1}{2}) + \psi(n_{11} + \tfrac{1}{2}) \\
&\approx \log \tfrac{n_{00}\ n_{11}}{n_{01}\ n_{10}} \\
\sigma^2 &= \psi'(n_{00} + \tfrac{1}{2}) + \psi'(n_{01} + \tfrac{1}{2}) + \psi'(n_{10} + \tfrac{1}{2}) + \psi'(n_{11} + \tfrac{1}{2}) \\
&\approx 1/n_{00} + 1/n_{11} + 1/n_{01} + 1/n_{10} \\
\pi_U:\ \mu &= \psi(n_{00} + 1) - \psi(n_{01} + 1) - \psi(n_{10} + 1) + \psi(n_{11} + 1) \\
&\approx \log \tfrac{(n_{00}+\frac{1}{2})\ (n_{11}+\frac{1}{2})}{(n_{01}+\frac{1}{2})\ (n_{10}+\frac{1}{2})} \\
\sigma^2 &= \psi'(n_{00} + 1) + \psi'(n_{01} + 1) + \psi'(n_{10} + 1) + \psi(n_{11} + 1) \\
&\approx \tfrac{1}{n_{00}+1/2} + \tfrac{1}{n_{01}+1/2} + \tfrac{1}{n_{10}+1/2} + \tfrac{1}{n_{11}+1/2}
\end{aligned}
$$

Note that the approximations (valid whenever all four cell counts n_{ij} are at least five or so) are the same as the MLE $\hat{\theta}$ and modified MLE $\tilde{\theta}$ and their standard errors introduced before. This leads to approximate reference Bayesian credible intervals and two-sided hypothesis tests of $H_0 : [\theta = 0]$ of the same form as the asymptotic confidence intervals and p-values above,

$$
\begin{aligned}
\theta &\approx \log \frac{n_{00}\, n_{11}}{n_{01}\, n_{10}} \pm Z_{\alpha/2} \sqrt{1/n_{00} + 1/n_{01} + 1/n_{10} + 1/n_{11}} \\
p &\approx 2\Phi\big(- \big| \log \frac{n_{00}\ n_{11}}{n_{01}\ n_{10}} \big| / \sqrt{1/n_{00} + 1/n_{01} + 1/n_{10} + 1/n_{11}} \big),
\end{aligned}
$$

but now we can simulate from the Bayesian posterior exactly (setting $\theta = \log \frac{\lambda_{00}\ \lambda_{11}}{\lambda_{01}\ \lambda_{10}}$ for $\lambda_{ij} \sim \mathsf{Ga}(n_{ij} + \frac{1}{2}, 1)$ or $\lambda_{ij} \sim \mathsf{Ga}(n_{ij} + 1, 1)$ for Jeffreys or Uniform priors) or can give the posterior mean and variance exactly, using μ and σ^2 given above. The difference $|\psi(n + \frac{1}{2}) - \log(n)|$ is bounded by $1/24n^2$, so only exceeds about 0.01 for $n = 0, 1$; still, experiments with one or more of the n_{ij} this small do occur.

1.3. Example: ECMO

One of the most widely studied and controversial experiments was the 1985 trial of Extra-Corporial Membrane Oxygenation, or *ECMO*, a treatment for newborns intended to prevent potentially lethal respiratory disease. It turns out that ECMO is *much* better than what was then the conventional treatment, and that untreated babies experience much higher risk of death; nonethesless the original small trial we study below was regarded as "inconclusive" at the time, leading to a much larger trial in which scores of babies in the "control" arm were denied the benefit of ECMO and died.

This trial had an unusual "play the winner" sequential patient allocation protocol in which subjects were preferentiably assigned (through an urn scheme) to whichever treatment is doing better. Nine subjects were given ECMO, all of whom survived; ten subjects were given the conventional treatment, of whom six survived and four did not. Any analysis begins with the four cell counts $n_{00} = 6$, $n_{01} = 4$, $n_{10} = 9$, and $n_{11} = 0$.

The conventional asymptotic analysis is unavailable because of the zero cell count; $\hat{\theta} = \log \frac{6*0}{4*9}$ is negative infinity.

Under Jeffreys and uniform prior distributions our Bayesian analysis above would find posterior means and variances for θ of

$$
\begin{aligned}
\pi_J: \ \mu &= \psi(6.5) - \psi(4.5) - \psi(9.5) + \psi(0.5) &&\approx -3.7572 \\
\sigma^2 &= \psi'(6.5) + \psi'(4.5) + \psi'(9.5) + \psi'(0.5) &&\approx 5.4608 \\
\pi_U: \ \mu &= \psi(7) - \psi(5) - \psi(10) + \psi(1) &&\approx -2.4623 \\
\sigma^2 &= \psi'(7) + \psi'(5) + \psi'(10) + \psi(1) &&\approx 2.1250
\end{aligned}
$$

leading to approximate 90% credible intervals of

$$
\begin{aligned}
\theta &= \mu \pm z_{.05} \times \sigma \\
&= -3.7572 \pm 1.645 \times 2.3368 \\
&= [-7.601, \ 0.087], \ \text{for Jeffreys prior, or} \\
&= -2.4623 \pm 1.645 \times 1.4577 \\
&= [-4.860, \ -0.065], \ \text{for Uniform prior.}
\end{aligned}
$$

Zero is just on the boundary of both 90% intervals, suggesting that even this small trial has enough evidence to suggest that ECMO is superior to the conventional treatment at approximately the 5% level; indeed the one-sided probabilities that ECMO is *not* superior are (continuing with the normal approximation)

$$\begin{aligned}\pi_J[\theta > 0 \mid \vec{n} = (6,4,9,0)] &\approx \texttt{pnorm(mu/sd);} \approx 0.054\\ \pi_U[\theta > 0 \mid \vec{n} = (6,4,9,0)] &\approx \texttt{pnorm(mu/sd);} \approx 0.046\end{aligned}$$

suggesting that there is only a five percent chance or so that ECMO is ineffective. The normal approximation isn't really necessary; the exact probability is easily approximated through simulation by the S-Plus function

```
prob <- function (n00, n01, n10, n11, prior=0.50, mc=500000) {
  lam00 <- rgamma(mc, n00+prior);  lam01 <- rgamma(mc, n01+prior);
  lam10 <- rgamma(mc, n10+prior);  lam11 <- rgamma(mc, n11+prior);
  return (mean( (lam00*lam11)/(lam01*lam10) > 1) );
}
```

giving

$$\begin{aligned}\pi_J[\theta > 0 \mid \vec{n} = (6,4,9,0)] &\approx \texttt{prob(6,4,9,0,0.50)}\\ &\approx 0.0108\\ \pi_U[\theta > 0 \mid \vec{n} = (6,4,9,0)] &\approx \texttt{prob(6,4,9,0,1.00)}\\ &\approx 0.0226\end{aligned}$$

and suggesting even stronger evidence against $H_0 : [\theta \geq 0]$.

1.4. Exact Computations

If $\theta = \log \frac{\lambda_{00}\ \lambda_{11}}{\lambda_{01}\ \lambda_{10}}$ for independent $\lambda_{ij} \sim \mathsf{Ga}(\alpha_{ij}, 1)$ then we could change variables to $p = \lambda_{01}/\lambda_{0+}$ and $q = \lambda_{11}/\lambda_{1+}$ to write $\theta = \log \frac{(1-p)q}{p(1-q)}$ for independent $p \sim \mathsf{Be}(\alpha_{01}, \alpha_{00})$ and $q \sim \mathsf{Be}(\alpha_{11}, \alpha_{10})$, and can change variables from (p, q) to (θ, q), using the relations

$$\frac{p}{1-p} = \frac{q}{1-q} e^{-\theta} \Longrightarrow p = \frac{q}{q + (1-q)e^{\theta}}, \qquad \frac{\partial \theta}{\partial p} = \frac{-1}{p\,(1-p)},$$

to find the marginal density function for θ to be

$$\begin{aligned}\pi(\theta|\vec{\alpha}) &= \int_0^1 c_1 \left(\frac{q}{q+(1-q)e^{\theta}}\right)^{\alpha_{01}} \left(\frac{(1-q)e^{\theta}}{q+(1-q)e^{\theta}}\right)^{\alpha_{00}} q^{\alpha_{11}-1}(1-q)^{\alpha_{10}-1}\, dq\\ &= c_1\, e^{\theta\alpha_{00}} \int_0^1 \frac{q^{\alpha_{+1}-1}\,(1-q)^{\alpha_{+0}-1}}{\left(q+(1-q)e^{\theta}\right)^{\alpha_{0+}}}\, dq\\ &= c_2\, e^{-\theta\alpha_{01}}\,{}_2F_1\left(\alpha_{+1}, \alpha_{0+}, \alpha_{++}; 1-e^{-\theta}\right)\end{aligned}$$

where $c_1 = \frac{\Gamma(\alpha_{0+})\Gamma(\alpha_{1+})}{\Gamma(\alpha_{00})\Gamma(\alpha_{01})\Gamma(\alpha_{10})\Gamma(\alpha_{11})}$ and $c_2 = \frac{\Gamma(\alpha_{0+})\Gamma(\alpha_{1+})\Gamma(\alpha_{+0})\Gamma(\alpha_{+1})}{\Gamma(\alpha_{00})\Gamma(\alpha_{01})\Gamma(\alpha_{10})\Gamma(\alpha_{11})\Gamma(\alpha_{++})}$ are constants and where ${}_2F_1(a,b,c;z)$ is the confluent hypergeometric function (Abramowitz and Stegun, 1964, pg. 558). Thus the exact posterior distribution for θ in the ECMO example is available for both Jeffreys and Uniform priors as

$$\begin{aligned}
\pi_J(\theta|6,4,9,0) &= \frac{\Gamma(11)\Gamma(10)\Gamma(16)\Gamma(5)\, e^{-4.5\theta}}{\Gamma(6.5)\Gamma(4.5)\Gamma(9.5)\Gamma(0.5)\Gamma(21)}\, {}_2F_1\big(5,11,21;1-e^{-\theta}\big)\\
\pi_U(\theta|6,4,9,0) &= \frac{\Gamma(12)\Gamma(11)\Gamma(17)\Gamma(6)\, e^{-5\theta}}{\Gamma(7)\Gamma(5)\Gamma(10)\Gamma(1)\Gamma(23)}\, {}_2F_1\big(6,12,23;1-e^{-\theta}\big)
\end{aligned}$$

Each of these may be plotted in Mathematica or Maple, or integrated numerically to give posterior probabilities of ECMO ineffectiveness of

$$\pi_J(\theta \geq 0|6,4,9,0) = 0.0107338 \qquad \pi_U(\theta \geq 0|6,4,9,0) = 0.0227038$$

so we confirm that the probability of ECMO ineffectiveness is no more than one or two percent. You can read more about ECMO and earlier analyses in (Ware, 1989), particularly the discussion by Kass and Greenhouse in which dozens of prior distributions are compared, and (Lavine et al., 1991) where the assumptions of prior independence are explored.

1.5. Fisher's Classical Approach

Classical testing of $H_0 : [\theta = 0]$ is complicated by the fact that the hypothesis does not completely specify the distribution of the cell counts $\vec{n} = (n_{00}, n_{01}, n_{10}, n_{11})$ or of derived quantities like the likelihood ratio statistic. Fisher proposed computing the *conditional* probability distribution of such quantities, given the observed row and column sums (say, given n_{0+} and n_{+0}). Conditional on these quantities, and on $H_0 : [\theta = 0]$, Fisher showed that n_{ij} has a hypergeometric $\mathsf{HG}(n_{+\mathrm{j}}, n - n_{+\mathrm{j}}, n_{\mathrm{i}+})$ distribution, so

$$\mathsf{P}[N_{ij} = n_{ij} | n_{\mathrm{i}+}, n_{+\mathrm{j}}, \theta = 0] = \frac{\binom{n_{\mathrm{i}+}}{n_{ij}}\binom{n_{\bar{\mathrm{i}}+}}{n_{\bar{\mathrm{i}}\bar{\mathrm{j}}}}}{\binom{n}{n_{+\mathrm{j}}}} = \frac{n_{0+}!\; n_{1+}!\; n_{+0}!\; n_{+1}!}{n_{00}!\; n_{01}!\; n_{10}!\; n_{11}!\; n!}$$

fnwhere $\bar{\mathrm{i}} \equiv 1 - i$ and in particular the Fisher Exact p-value for the ECMO experiment would be

$$p = \frac{\binom{9}{0}\binom{10}{4}}{\binom{19}{15}} = \texttt{phyper(0,9,10,4)} = 0.05417957,$$

again close to 1/20; concerns have arisen earlier about the low power of Fisher's test, a consequence of its conditioning on the non-ancillary row and column sums.

References

Abramowitz, M. and Stegun, I. A., eds. (1964), *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, volume 55 of *Applied Mathematics Series*, Washington, D.C.: National Bureau of Standards.

Bishop, Y., Fienberg, S. E., and Holland, P. W. (1977), *Discrete Multivariate Analysis: Theory and Practice*, Cambridge: MIT Press.

Breslow, N. E. and Day, N. E. (1980), *Statistical methods in cancer research. Volume 1 - The analysis of case-control studies*, number 32 in IARC Scientific Publications, Lyon: International Agency for Research on Cancer.

Kass, R. E. and Greenhouse, J. B. (1989), "Comment: A Bayesian Perspective," *Statistical Science*, 4, 310–317, discussion of *Investigating therapies of potentially great benefit: ECMO*.

Lavine, M. L., Wasserman, L., and Wolpert, R. L. (1991), "Bayesian inference with specified prior marginals," *Journal of the American Statistical Association*, 86, 964–971.

Ware, J. H. (1989), "Investigating therapies of potentially great benefit: ECMO," *Statistical Science*, 4, 298–340.