Recall for **Section 3.3** that Storey (2002) **[7]** considers estimating the FDR for a fixed P value cut-off α by $\widehat{FDR}(\alpha) := \frac{\hat{\pi}_0(\lambda)\alpha}{\max\{R(\alpha),1\}/m}$, where $\hat{\pi}_0(\lambda)$ is an estimator of π_0 (See Section 4.1) and $R(\alpha)$ is the number of P values less than or equal to α. In term of the P value EDF,

$\widehat{FDR}(\alpha) := \frac{\hat{\pi}_0(\lambda)\alpha}{\max\{\tilde{F}_m(\alpha),1/m\}}$. Storey *et al.* (2003) **[8]** show that this estimator is biased upward and asymptotically conservative in the sense that with probability 1 $\lim_{m\to\infty} \inf_{t\geq\alpha} \{\widehat{FDR}(t) - FDR(t)\} \geq 0$ for each $\alpha > 0$. Storey (2002) considers an estimator of the pFDR given by

$p\widehat{FDR}(\alpha) := \frac{\hat{\pi}_0(\lambda)\alpha}{\max\{R(\alpha),1\}/m][1-(1-\alpha)^m]}$. **[7]** In term of the P value EDF,

$p\widehat{FDR}(\alpha) = \frac{\hat{\pi}_0(\lambda)\alpha}{\max\{\tilde{F}_m(\alpha),1/m\}[1-(1-\alpha)^m]}$.

Hence $\lim_{\alpha\to 0} p\widehat{FDR}(\alpha) = \lim_{\alpha\to 1} p\widehat{FDR}(\alpha) = \hat{\pi}_0(\lambda)$ for any fixed $m > 1$, and in general $p\widehat{FDR}(\alpha)$ is not monotone in α. Storey (2002) **[7]** establishes mean-squared error properties of this estimator and its asymptotic conservativeness that with probability 1, $\lim_{m\to\infty} p\widehat{FDR}(\alpha) \geq pFDR(\alpha)$. It not difficult to see that with the multiplier $1-(1-\alpha)^m$ in its denominator this estimator may tend to have large variance (thus be unstable) for small α.

The "empirical" *q*-values are defined as $\hat{q}_i := \hat{q}(P_{i:m}) := \min_{j\geq i}\{p\widehat{FDR}(P_{j:m})\}$, $i = 1, \ldots, m$. **[7]** Clearly $\hat{q}_1 \leq \cdots \leq \hat{q}_m$. Storey *et al.* (2003) **[8]** consider the more general *q*-value estimator $\hat{q}(\alpha) := \inf_{s\geq\alpha}\{p\widehat{FDR}(s)\}$ for $q(\alpha)$ defined in **Section 2.2,** and show its conservativeness that $\lim_{m\to\infty} \inf_{t\geq\alpha}\{\hat{q}(t) - q(t)\} \geq 0$ with probability 1 for each $\alpha > 0$ under a specific Bayesian model (see section 2.2) and certain ergodicity conditions.

5.2 Smooth ensemble cdf and pdf estimator

Cheng *et al.* (2004) **[24]** consider an estimator of the FDR of the $HT(\alpha)$ procedure (3) by $\widehat{FDR}(\alpha) = \frac{\hat{\pi}_0\alpha}{\hat{F}_m(\alpha)}$, where $\hat{\pi}_0$ and $\hat{F}_m(\cdot)$ are respectively the estimators of π_0 and the P value ensemble cdf $F_m(\cdot)$, derived from a spline smoothing of the P value EDF $\tilde{F}_m(\cdot)$; see **Section 4.4.** Cheng *et al.* (2004) **[24]** consider using this estimator to provide an FDR estimate at a P value cut-off threshold $\hat{\alpha}$ generated by a data-driven significance criterion **(see Section 6)**. Simulation results therein indicate that the estimator is able to provide a reasonably conservative (upward biased) FDR estimate at the data-driven significance threshold in a wide range of scenarios.

Pounds and Cheng (2004) **[30]** propose an estimator of the P value ensemble pdf $f_m(\cdot)$ by properly transforming and smoothing a histogram constructed from the spacings defined by the ordered P values. An estimator $\hat{f}_m(\cdot)$ is constructed by back-transforming and normalizing the smooth function, an estimator of π_0 by $\hat{\pi}_0 := \min\{\hat{f}_m(P_i), i = 1, \ldots, m\}$, an estimator $\hat{F}_m(\cdot)$ of $F_m(\cdot)$ by the trapezoid rule of integration applied to $\hat{f}_m(\cdot)$, and an FDR estimator by plugging the estimators into the above formula. Simulation results therein indicate that this estimator performs well in estimating the cFDR (or pFDR). For estimating pFDR it is much more stable (i.e., having less variance) than Storey's (2002) **[7]** estimator at the α values close to zero, which are often used in microarray applications.

5.3 Mixture model estimator

For the mixture models discussed in section 4.5, the FDR estimate is determined by substituting the fitted model's π_0 estimate and cdf into $\pi_0\alpha / F_m(\alpha)$. For the specific model of Pounds and Morris (2003), **[28]** the FDR estimate monotonically increases as α increases.

5.4 Robust estimator

As previously described, most of the available FDR estimation methods assume that $G_i(t) = t$ when $\theta_i = 0$ (i.e., H_{0i} is true). Pounds and Cheng (2006) **[29]** noted that this critical assumption is violated by discrete P values and P values from testing one-sided hypotheses. In particular, any test $\hat{\theta}_i$ that is one-sided or based on a discrete test statistic may have $G_i(t) < t$ for some t when $\theta_i = 0$. This violation can have severe and undesirable consequences for methods that estimate π_0 as part of their calculations. Pounds and Cheng (2005, 2006) **[31, 29]** describe these consequences in greater detail. Thus, Pounds and Cheng (2006) **[29]** develop a robust FDR estimator. The robust FDR estimator is conservative provided that $\Pr\left(\overline{P} \le 1/2\right) \approx 1$ and $G_i(t) \le t$ for $\theta_i = 0$, even when applied to one-sided tests or discrete P values. The method borrows ideas from least trimmed squares **[32]** and rank regression **[33]** to smooth raw FDR estimates obtained from the P value EDF. For one-sided tests, a folding transformation is used to make p-values essentially two-sided for purposes of estimating π_0 and then other calculations are performed on the original one-sided p-values.

5.5 Estimation of local FDR or empirical Bayes posterior

With estimators $\hat{\pi}_0$ and $\hat{f}_m(\cdot)$, an estimator of the empirical Bayes posterior probability (EBP or local FDR) **[16, 34]** of the null hypothesis H_{0i} conditional on $P_i = p_i$ is given by $\hat{\pi}_0 / \hat{f}_m(p_i)$. Efron (2004) **[34]** advocates to estimate the null ensemble density function of the test statistics from the empirical distribution and cautions against the use of random permutations. In the P value domain, this means to estimate the P value ensemble distribution under the "grand null" $H_0^* = \cap_{i=1}^m H_{0i}$ in lieu of assuming the $U(0, 1)$ distribution as in model (1). In a similar spirit, Datta and Datta (2005) **[35]** proposed an empirical Bayes method that first transforms p-values using the quantile-function of the standard normal distribution and then apply kernel density estimation methods to the transformed P values to obtain an EBP.

ISSN 0973-2063

Bioinformation 1(10): 436-446 (2007)

www.bioinformation.net