Recall from Equation (8) that the EDF of the P values $\widetilde{F}_m(t)$ has expected value $E[\widetilde{F}_m(t)]=F_m(t)$ for every t; that is, $\widetilde{F}_m(\cdot)$ is an unbiased estimator of the P value ensemble cdf $F_m(\cdot)$. Cheng *et al.* (2004) **[24]** observe that if the tests $\hat{\theta}_i$ $(i = 1, \ldots, m)$ are not too much correlated asymptotically in the sense $\sum_{i=j} Cov(\hat{\theta}_i, \hat{\theta}_j) = o(m^2)$ as $m \to \infty$, $\widetilde{F}_m(\cdot)$ is "asymptotically consistent" for $F_m(\cdot)$ in the sense $\left|\widetilde{F}_m(t) - F_m(t)\right| \to_p 0$ for every $t \in$ IR. These results provide heuristics for the estimation of π_0, the estimation of FDR, and data-adaptive determination of α for the $HT(\alpha)$ procedure. Estimation of π_0 is reviewed in this section.

As noted in the previous sections, the proportion of the true null hypotheses π_0 is an important parameter in FDR-related procedures. Consider first the P value ensemble cdf $F_m(\cdot)$. Because for any $t \in (0, 1)$ $\pi_0 = [H_m(t) - F_m(t)] / [H_m(t) - t]$, a plausible estimator of π_0 is $\hat{\pi}_0 = \frac{\Lambda - \widetilde{F}_m(t_0)}{\Lambda - t_0}$ for properly chosen Λ and t_0. The inverse function of $F_m(\cdot)$, defined as $Q_m(u) := F_m^{-1}(u) := \inf\{t : F_m(t) \geq u\}$, is the *P value ensemble quantile function*. The sample version is the *empirical quantile function* (EQF) defined as $Q_m(u) := \widetilde{F}_m^{-1}(u) := \inf\{x : \widetilde{F}_m(x) \geq u\}$. Then
$\pi_0 = [H_m(Q_m(u)) - u] / [H_m(Q_m(u)) - Q_m(u)]$, for $u \in (0, 1)$, and with Λ_1 and u_0 properly chosen, $\hat{\pi}_0 = \frac{\Lambda_1 - u_0}{\Lambda_1 - Q_m(u_0)}$ is a plausible estimator. Many of the estimators take either of the above two basic representation with some modifications.

Clearly it is necessary to have $\Lambda_1 \geq u_0$ in order to have a meaningful estimator. Because $Q_m(u_0) \leq u_0$ by the stochastic order assumption [cf. (1)], choosing Λ_1 too close to $u0$ will produce an estimator much biased downward. A heuristic is that if u_0 is so chosen that all P values corresponding to the alternative hypotheses concentrate in the interval $[0, Q_m(u_0)]$ then $H_m(Q_m(u_0)) = 1$; thus setting $\Lambda_1 = 1$. A similar heuristic leads to setting $\Lambda = 1$.

4.1 Slope estimator

Taking a graphical approach Schweder and Spjøtvoll (1982) **[25]** consider the slope from the point $(\lambda, \widetilde{F}_m(\lambda))$ to the point (1,1), and an estimator of m_0 as $\hat{m}_0 = m(1 - \widetilde{F}_m(\lambda)) / (1 - \lambda)$ for a properly chosen λ; hence a corresponding estimator of π_0 is $\hat{\pi}_0(\lambda) = \hat{m}_0 / m = (1 - \widetilde{F}_m(\lambda)) / (1 - \lambda)$. Storey's (2002) **[7]** estimator is exactly this one. Additionally, Storey (2002) **[7]** observes that λ is a tuning parameter that dictates the bias and variance of the estimator, and proposes computing $\hat{\pi}_0$ on a grid of λ values, smoothing them by a spline function, and taking the smoothed $\hat{\pi}_0$ at a λ close to 1, (e.g. 0.95) as the final estimator. Storey *et al.* (2003) **[8]** propose a bootstrap procedure to estimate the mean-squared error (MSE) and pick the λ that gives the minimal estimated MSE; a simulation study in Cheng (2006) **[15]** and investigation in Langaas *et al.* (2005) **[26]** show that this estimator tends to be biased downward.

4.2 Quantile slope estimator

Approaching to the problem from the quantile perspective Benjamini and Hochberg (2000) **[10]** propose $\hat{m}_0 = \min\{1 + m + 1(m + 1 - j)/(1 - P_{j:m}), m\}$ for a properly chosen j; hence $\hat{\pi}_0 = \hat{m}_0 / m$. The index j is determined by examining the slopes $S_i = (1 - P_{i:m})/(m + 1 - i)$, $i = 1, \ldots, m$, and is taken to be the smallest index such that
$S_j < S_{j-1}$. Then $\hat{m}_0 = \min\{1 + 1 / S_j, m\}$. Cheng (2006) **[15]** shows that as m gets large the event $\{S_j < S_{j-1}\}$ tends to occur early (i.e., at small j) with high probability; therefore the estimator tends to be increasingly conservative (i.e., biased upward) as the number of tests m increases. The conservativeness is also demonstrated by the simulation study in Cheng (2006). **[15]**

4.3 Quantile slope estimator by quantile modeling

Cheng (2006) **[15]** develops an improvement of Benjamini and Hochberg's' (2000) **[10]** estimator by considering a shape requirement on the P value ensemble quantile function $Q_m(\cdot)$. Heuristically, the stochastic order requirement in model (1) implies that $F_m(\cdot)$ is approximately concave and hence $Q_m(\cdot)$ is approximately convex. When there is a substantial proportion of true null and true alternative hypotheses, there is a "bend point" $\tau_m \in (0, 1)$ such that $Q_m(\cdot)$ assumes roughly a nonlinear shape on the interval $[0, \tau_m]$, primarily dictated by the distributions of the P values corresponding to the true alternative hypotheses, and $Q_m(\cdot)$ is essentially linear on the interval $[\tau_m ,1]$, primarily dictated by the $U(0, 1)$ distribution of the null P values. The estimation of π_0 can benefit from properly capturing this shape characteristic using a model. Cheng (2006) **[15]** considers a two-piece function approximation (model) for $Q_m(\cdot)$. In an interval $[0, \tau_m]$ $Q_m(u)$ is approximated by a polynomial of the form $\eta u^\gamma + \delta u$ with $\gamma \geq 1$, $\eta \geq 1$, and $0 \leq \delta \leq 1$; on the interval $[\tau_m ,1]$ it is approximated by a linear function $\beta_0 + \beta_1 u$ with $\beta_0 \leq 0$ and $\beta_1 \geq 1$. The two pieces are joint smoothly at τ_m by the constraints $\eta\tau_m^\gamma + \delta\tau_m = \beta_0 + \beta_1\tau_m$ (continuity) and $\eta\gamma\tau_m^{\gamma-1} + \delta = \beta_1$ (differentiability). For identifiability it is further required that $\gamma = \eta = 1$ and $\delta = 0$ if and only if $\tau_m = 0$. These parameters are determined by minimizing the integrated absolute difference (L^1 distance) between $Q_m(u)$ and $Q_m^*(u) := I(0 \leq u \leq \tau_m)(\eta u^\gamma + \delta u) + I(\tau_m \leq u \leq 1)(\beta_0 + \beta_1 u)$, subject to the above constraints. Cheng (2006) develops a procedure to estimate these parameters from the P value EQF $\widetilde{Q}_m(\cdot)$. The estimator of π_0 is the reciprocal of the estimator of β_1: $\hat{\pi}_0 := 1/\hat{\beta}_1$.

A simulation study by Cheng (2006) **[15]** indicate that in a reasonably wide range of scenarios this estimator is slightly biased upward (i.e., conservative); the upward bias is usually less than the downward bias of the bootstrap estimator of Storey *et al.* (2003), **[8]** and is much less than the upward bias of Benjamini and Hochberg (2000) **[10]** estimator. In this regard this quantile slope estimator outperforms the other two estimators, as well as in terms of the mean square error.

4.4 Monotone convex and smooth density estimators

Note that under model (1) the probability density function (pdf) of $F_m(\cdot)$, the P value ensemble pdf, is $f_m(t) := \frac{d}{dt}F_m(t) = \pi_0 + (1-\pi_0)h_m(t),\ t \in [0, 1]$,

where $h_m(t) := \frac{d}{dt}H_m(t),$ the P value ensemble alternative pdf. Note $\pi_0 \approx f_m(1)$ if $h_m(1) \approx 0$; this is achievable under the heuristic that essentially all the P values corresponding to the true alternative hypotheses concentrate in an interval away from 1. Langaas *et al.* (2005) **[26]** consider estimating π_0 by requiring $F_m(\cdot)$ be strictly concave and thus $f_m(\cdot)$ be monotone and convex. They propose to estimate $f_m(\cdot)$ by the nonparametric maximum likelihood estimator $\hat{f}_m^*(\cdot)$ under the constraint of monotonicity and convexity, and to estimate π_0 by $\hat{\pi}_0 := \hat{f}_m^*(1)$. The simulation study therein indicates this estimator performs very well in a range of scenarios.

Cheng *et al.* (2004) **[24]** consider a spline function estimator $\hat{F}_m(\cdot)$ of $F_m(\cdot)$. $\hat{F}_m(\cdot)$ is a B-spline function constructed by smoothing the P value EDF $\widetilde{F}_m(\cdot)$. The spline knots are placed in a way that gives little smoothing in the vicinity of 0 but a large amount of smoothing in the right tail. An estimator of $f_m(\cdot)$ is the derivative function $\hat{f}_m(1) := \frac{d}{dt}\hat{F}_m(t), t \in [0,1]\cdot$ Then an estimator of π_0 is given by $\hat{\pi}_0 := \hat{f}_m(1)$. The simulation study in Cheng *et al.* (2004) **[24]** indicate that this estimator is slightly upward biased (conservative) in a range of scenarios as long as the true π_0 is not too close to 1.

4.5 Mixture model estimators

Allison *et al.* (2002) **[27]** and Pounds and Morris (2003) **[28]** describe methods that estimate the FDR via P value modeling. These methods also estimate π_0. Allison *et al.* (2002) **[27]** describe a method that models the P values as arising from a mixture distribution with one $U(0, 1)$ component and potentially several beta components. The model is fit by maximum likelihood estimation and the bootstrap is used to determine the number of beta components that are used in the model. Allison *et al.* (2002) **[27]** note that it is often unnecessary in practice to include more than one beta component in the model. Pounds and Morris (2003) **[28]** give a detailed description of the use of a specific model with one beta component. Assuming null p-values follow a U(0,1) distribution, Pounds and Morris (2003) **[28]** show that π_0 must be less than or equal to the minimum of the ensemble P value pdf. Thus, they propose to estimate π_0 by the minimum of the pdf of the mixture model fit to the p-values. Allison *et al.* (2002) **[27]** estimate π_0 by the mixing weight for the uniform component of the fitted model. It is theoretically possible that the mixing weight of the uniform component could be substantially smaller than the minimum of the fitted pdf. In this case, the mixing weight estimator understates the proportion of the fitted density that could be attributed to a uniform (0,1) distribution.

4.6 Moment estimator

Pounds and Cheng (2006) **[29]** describe a simple moment-based estimator of π_0. Let $\overline{P} = m^{-1}\sum_{i=1}^{m} P_i$. Assuming that $E[P_i] \geq 1/2$ if $\theta_i = 0$ (i.e., H_{0i} is true), it follows that $E\left[\overline{P}\right] \geq 2\pi_0$. This observation motivates $\hat{\pi}_0 = \min(1, 2\overline{P})$ as an estimator of π_0. This estimator has several advantages over those described above. It is very simple to compute, and it does not rely on continuity or model assumptions for the P values. However, it is considerably more conservative than the other estimators when the assumptions of those estimators hold.