Storey-Tibshirani method for "Qvalues": conditional false discovery rate

http://www.pnas.org/content/100/16/9440.abstract

[TODO: fix links for qvalue-demo.R Publication bias demo.R]

Define pFDR = E(V/R | R > 0). (If R = 0, why would we care about FDR?)
(In this paper, V/R is written as F/S.)

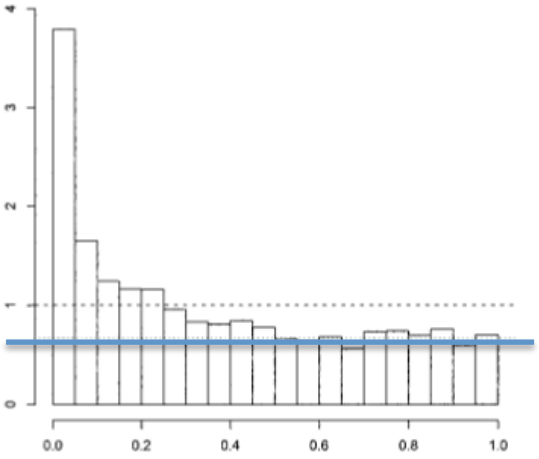

Fig. 1. A density histogram of the 3,170 *p* values from the Hedenfalk *et al.* (14) data. The dashed line is the density histogram we would expect if all genes were null (not differentially expressed). The dotted line is at the height of our estimate of the proportion of null *p* values.

A Bayes/frequentist hybrid:

Pr(H0) is an *estimated* prior probability,
and FDR is from Bayes theorem,
but we use P values.

Write FDR as a posterior probability:

$$\mathrm{FDR} = \mathrm{E}\left[\frac{m_0 \cdot [1 - \text{specificity}]}{m_0 \cdot [1 - \text{specificity}] + m_1 \cdot \text{sensitivity}}\right].$$

where m_0 = #true null hypotheses, m_1 = #false null hypotheses,.
We ***estimate*** m_0 from the observed distribution of all the P-values, assuming a mixture of a uniform multiplied by $\boldsymbol{\pi_0} = m_0 / (m_0 + m_1)$ and a portion near zero multiplied by $1 - \boldsymbol{\pi_0}$.
Different methods can estimate $\boldsymbol{\pi_0}$. In the Figure, the LOWER dotted line corresponds to $\hat{\pi}_0$. See the R package **qvalue** by Dabney and Storey. For each threshold ***t***, define

$$\widehat{\mathrm{FDR}}(t) = \frac{\hat{\pi}_0 m \cdot t}{S(t)} = \frac{\hat{\pi}_0 m \cdot t}{\#\{p_i \le t\}}$$

Finally, for each null hypothesis ***i*** , qvalue_i =

$$\hat{q}(p_i) = \min_{t \geq p_i} \widehat{\mathrm{FDR}}(t).$$

The interpretation of q is the minimum FDR you get if p_i is called “significant”.