False Discovery Rates - A New Deal

Matthew Stephens

Department of Statistics and Department of Human Genetics, University of Chicago

The Set-up

For $j = 1, \ldots, J$, let

β_j denote the (unobserved) jth effect of interest.

$\hat{\beta}_j$ denote a noisy measurement of β_j.

$\hat{s}_j$ denote the (estimated) standard error of $\hat{\beta}_j$.

$z_j := \hat{\beta}_j / \hat{s}_j$, and p_j be the corresponding p value testing $H_j : \beta_j = 0$

A typical genomics pipeline analyses the p values with the goal of estimating or controlling the false discovery rate (FDR) [1].

Examples

`qvalue` [2] essentially attempts to decompose the distribution of p values into two components, one null (uniform distribution), and the other alternative.

`locfdr` [3] attempts to decompose the distribution of z scores into two components, one null (standard normal), and the other alternative.

`mixfdr` [4] attempts to decompose the distribution of z scores into two components, one null (standard normal), and the other alternative (mixture of normals).

For example, `mixfdr` assumes that

$$p(z_j) = \pi_0 N(z_j; 0, 1) + (1 - \pi_0) f_1(z_j) \quad (1)$$

where f_1 is assumed to follow a mixture of normal distributions. It then estimates π_0 and f_1 by (penalized) maximum likelihood, from which estimates of (local) FDR follow.

A New Deal

Similar to [3, 4] we take an Empirical Bayes (EB) approach to the problem, but with some key differences:

1. Decompose distribution of β_j (instead of p values or z score):
$$p(\beta_j | s_j) = \pi_0 \delta_0 + (1 - \pi_0) g_1(\beta_j)$$
2. **Unimodal Assumption (UA):** Assume g_1 to be unimodal about 0.
3. Connect the observations with the model via a normal likelihood assumption:
$$\hat{\beta}_j | \beta_j, \hat{s}_j \sim N(\beta_j, \hat{s}_j).$$

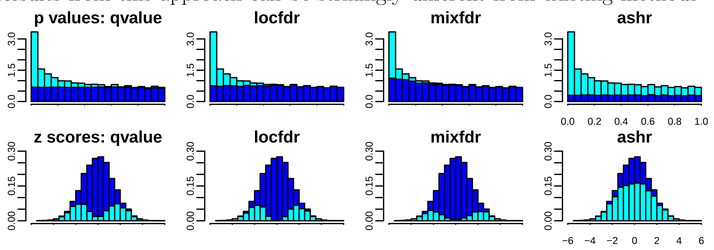
Details

1. Simple way to implement the UA is via a mixture of 0-centered normals: $g_1(\beta_j) = \sum_{k=1}^{K} \pi_k N(\beta_j; 0, \sigma_k^2)$ where K is large, and $\sigma_1, \ldots, \sigma_K$ are a fixed fine grid spanning values from very small to very large. Mixture proportions estimated by a simple EM algorithm.
2. The mixture of normals (above) also assumes g_1 is symmetric; can allow asymmetry using mixures of uniforms that end or start at 0.
3. Can generalize normal likelihood to t likelihood to help allow for estimation of $\hat{s}_j$.
4. Add penalty term to the likelihood to encourage π_0 to be as big as possible given data (avoids underestimating FDR).

The methods are implemented in an R package `ashr` ("Adaptive SHrinkage") available at `http://github.com/stephens999/ashr`. Code for this poster, draft paper in progress, and other info at `http://github.com/stephens999/ash`.

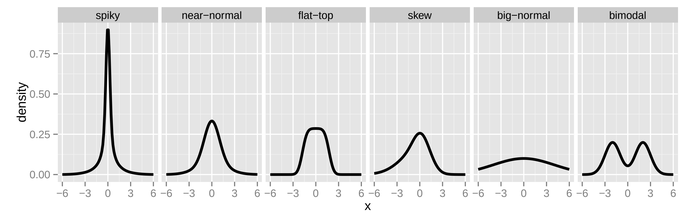
UA changes p value and z score decomposition (dark-blue=null; cyan = alt)

Results from this approach can be strikingly different from existing methods



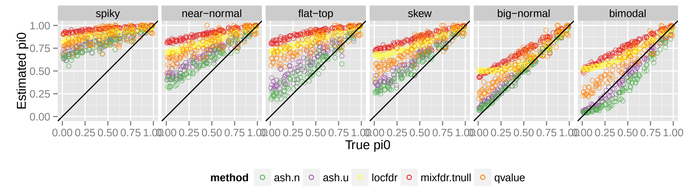
Illustrative Simulations

We conducted simulations using various effect distributions g_1



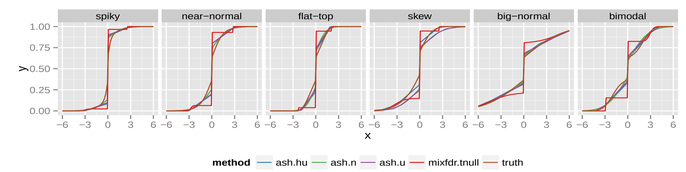
The UA (if it holds) allows less conservative estimates of π_0

All methods over-estimate π_0 (deliberately); `ashr` is least conservative.



The UA leads to stable estimates of g (vs `mixfdr`)

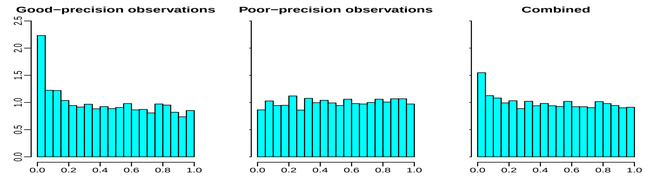
Figures compare true cdf with estimate from different methods.



See also [3] which discusses challenges of getting stable non-parametric estimates of g; the UA provides one simple solution to this problem.

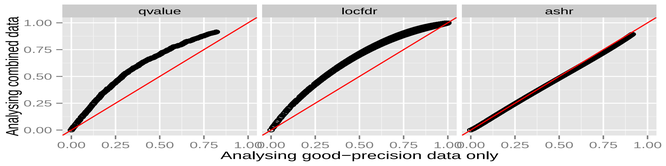
Problem: low precision measurements have less power, diluting signal

The p values from high-precision measurements (left) show more signal than those from low-precision measurements (center) due to higher power of former. Combining them (right) dilutes the signal.



Taking precision into account avoids dilution

Figure shows q values (FDR estimates) for good-precision measurements. For `qvalue` and `locfdr`, the q values are inflated by including low-precision measurements in the analysis. For `ashr` they are not, because the $N(\beta, \hat{s}_j)$ likelihood is essentially flat when $\hat{s}_j$ is large, so low-precision measurements effectively ignored.



Adaptive Shrinkage

Methods are applicable not only to FDR, but also shrinkage more generally. Eg as a competitor to EB approaches like [5], which make more restrictive parametric assumptions than the UA. Because `ashr` provides posterior distribution on each effect given observed data, can also tackle the "post-selection" problem of providing interval estimates for "significant" effects.

References

[1] Yoav Benjamini and Yosef Hochberg. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1995), pp. 289–300.

[2] J.D. Storey. "A direct approach to false discovery rates". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64.3 (2002), pp. 479–498.

[3] Bradley Efron. "Microarrays, empirical Bayes and the two-groups model". In: *Statistical Science* 23.1 (2008), pp. 1–22.

[4] Omkar Muralidharan. "An empirical Bayes mixture method for effect size and false discovery rate estimation". In: *The Annals of Applied Statistics* (2010), pp. 422–438.

[5] Iain M Johnstone and Bernard W Silverman. "Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences". In: *Annals of Statistics* (2004), pp. 1594–1649.