Supplementary Material

Main paper: UW Statistics Technical Report 1175 (v7)
Making the cut: improved ranking and selection for large-scale inference
Authors: Nicholas C. Henderson and Michael A. Newton
Version: June 22, 2015
Contacts: Nicholas C. Henderson, `nhenders@stat.wisc.edu`
Michael A. Newton, `newton@stat.wisc.edu`

- Section 1: On data analyses
- Section 2: On threshold functions, Table 1
- Section 3: On r-value computation
- Section 4: Theory (supporting proof of Theorem 4, main)
- Section 5: Models and assumptions
- Section 6: Connection to hypothesis testing
- Section 7: Simulation figures

1 On data analyses

1.1 T2D example

Data were obtained as file `DIAGRAM.metabochip.txt` from
`http://http://diagram-consortium.org/downloads.html`
on September 5, 2013. These refer to the stage 2 metabochip summary statistics, and provide summary data on 127903 SNPs. Point estimates $\{X_i\}$ were computed as logarithms of the reported odds ratios; standard errors $\{\sigma_i\}$ were computed as 1/4 of the length of the associated 95% confidence intervals. Recognizing that a large number of the SNPs may be null (true log odds of zero), we filtered the data by first fitting a mixture model:

$$X_i \sim \begin{cases} \text{Normal}(0, \sigma_i^2) & \text{w.p. } \pi_0 \\ \text{Normal}(\mu, \tau^2 + \sigma_i^2) & \text{w.p. } 1 - \pi_0 \end{cases}$$

By EM algorithm we estimated $\hat{\pi}_0 = 0.801$, $\hat{\mu} = 0.007$, $\hat{\tau}^2 = 0.0014$. The 80/20 mixing of null/non-null SNPs is comparable to the mixing rate estimated by other means in Morris *et al.* (2012). The EM algorithm generated SNP-specific posterior probabilities of non-null effect. We experimented with several selection schemes for producing a reduced set of SNPs that were probably non-null. Data illustrated in Figures S1 and 1 (main paper) correspond to a single instance of sampling from these posterior non-null probabilities to obtain a set of 25,558 SNPs that are probably non-null, and thus associated with T2D.

1.2 RNAi example

We started with the 984 genes (unique Entrez gene ID's) identified in the Hao *et al.* (2013) meta-analysis as having been detected in at least one genome-wide RNAi screen for influenza-virus association, as these data are available in the R package `metaflu` (version 1.0) associated with Hao *et al.* (2013). We utilized version 2.8.0 of Bioconductor package `org.Hs.eg.db` to associate the 984 influenza-associated genes with gene ontology (GO) gene sets (GO terms). Fully 17959 human genes were annotated to at least one GO terms, and fully 16572 GO terms were available, though we restricted attention to 5719 terms, all those annotating between 10 and 1000 human genes. The median set size in this class is 32 genes; the mean rate of influenza-association is 0.07. Maximum likelihood was used to estimate the shape parameters of the Beta distribution presumed to govern the fluctuation over sets in their underlying enrichment levels.

1.3 Pyeon example (Fig S8)

We started with a 54675×84 array of gene expression values $\{Y_{ij}\}$ corresponding to measurements on 33 case subjects and 51 control subjects, as reported in Pyeon *et al.* (2007). The effect size estimates $\{X_i\}$ were computed by using the differences between the sample means in the case group (HPV+) and the sample means in the control group (HPV-) (i.e., $X_i = \bar{Y}_{i1} - \bar{Y}_{i2}$ for each i). The associated standard errors $\{s_i\}$ were computed by

$$s_i^2 = \frac{1}{82}\Big(\frac{1}{33} + \frac{1}{51}\Big)\Big(\sum_{j \in \text{case}} (Y_{ij} - \bar{Y}_{1\cdot})^2 + \sum_{j \in \text{control}} (Y_{ij} - \bar{Y}_{2\cdot})^2\Big). \quad (1)$$

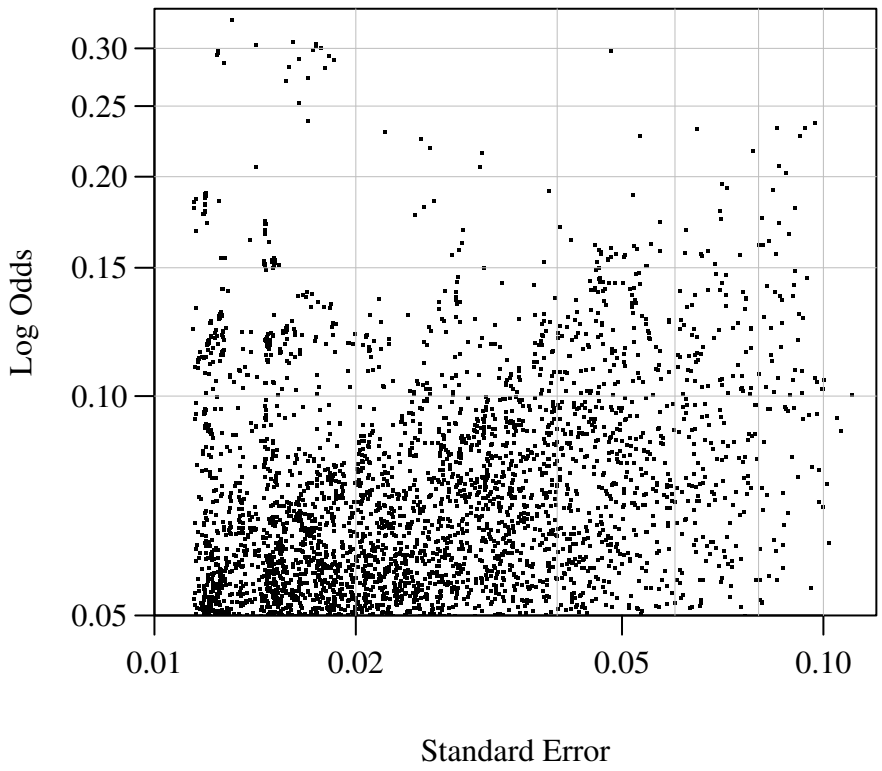

Figure S1: Type-2 diabetes example: From the full complement of 127,903 SNPs used in the second stage meta-analysis from Morris *et al.* 2012, we filtered to a reduced set of 25,558 SNPs that are probably associated with T2D, and plot 3371 of those having highest observed association (log odds exceeding 0.05). These estimates are based on genotype data from 22,669 T2D cases and 58,119 control subjects.

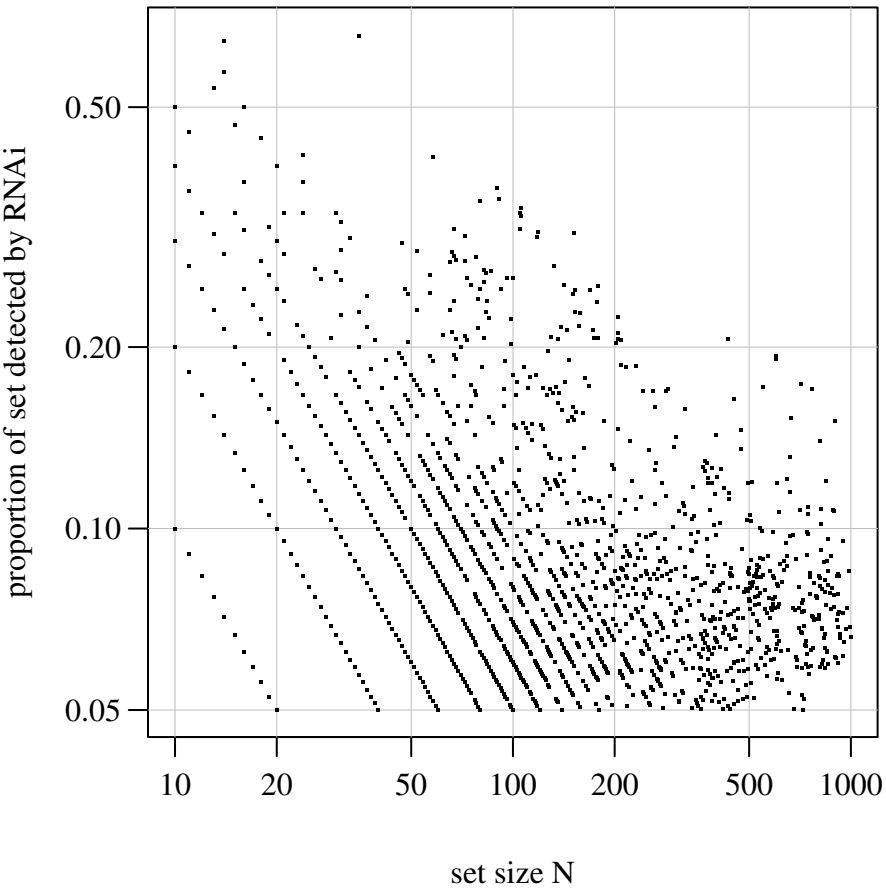

Figure S2: RNAi example: From a recent version of Gene Ontology, 5719 terms (gene sets) annotate between 10 and 1000 human genes. Shown is a summary of the integration of these terms with the list of 984 genes detected by RNAi as being involved in influenza virus replication (from Hao *et al.* 2013). The x-axis shows set size and the y-axis shows the proportion of the set that was detected by RNAi. The plot is restricted to 3626 sets for which the observed proportion exceeds 5%.

The parameters of interest $(\theta_1, \ldots, \theta_n)$ were modeled as a mixture of two t-distributions with 3.5 degrees of freedom (i.e., given a mixture label $Z_i = z$, $\theta_i \sim t_{3.5}(\mu_z, \sigma_z^2)$ with $z \in \{1, 2\}$ and $P(Z_i = z) = \pi_z$). The notation $t_{df}(\mu, \sigma^2)$ is used to denote the distribution of a random variable W such that $W = \mu + \sigma T$ where T follows a standard t-distribution with 3.5 degrees of freedom. Using an EM algorithm, we estimated the hyperparameters as

$$(\hat{\mu}_1, \hat{\mu}_2) = (0.337, -0.128), \qquad (\hat{\sigma}_1, \hat{\sigma}_2) = (0.189, 0.098), \qquad (\hat{\pi}_1, \hat{\pi}_2) = (0.24, 0.76). \quad (2)$$

To generate samples from the posterior distributions $\theta_i | D_i$, we used a Gibbs sampling approach employing the following full-conditionals

$$\begin{aligned} \theta_i | X_i, u_i^2, Z_i = z &\sim N\Big(\frac{X_i}{1 + u_i^2 s_i^2} + \frac{\mu_z u_i^2 s_i^2}{1 + u_i^2 s_i^2}, \frac{s_i^2}{1 + u_i^2 s_i^2}\Big) \\ u_i^2 | X_i, \theta_i, Z_i = z &\sim \text{Gamma}\Big(\frac{df + 1}{2}, \frac{(\theta_i - \mu_z)^2 + df\sigma_z^2}{2}\Big) \\ Z_i | X_i, \theta_i, u_i^2 &\sim p(\theta_i | u_i^2, Z_i = z) p(u_i^2 | Z_i = z) \pi_z, \end{aligned} \quad (3)$$

where $df = 3.5$ and $\text{Gamma}(\alpha, \beta)$ denotes a random variable with density $f(x) = \beta^\alpha x^{\alpha-1} e^{-\beta x} / \Gamma(\alpha)$. Using the above Gibbs procedure (with a burn-in of 100 iterations), computation of r-values was then based on MCMC output which, for each unit, contained 2000 draws from the targeted posterior.

Results are summarized in Figure S9.

1.4 NBA example, mid-season analysis

Table S1 reports the leading free-throw shooters part way into the 2013-2014 regular season (at the end of December, 2013). Columns as in Table 2, main paper, which shows leaders after the entire season. Figure S3 used the mid-season rankings as well as true $\{\theta_i\}$ vectors simulated from the end-of-season posterior to validate the r-value ranking.

2 On threshold functions and Table 1 (main)

2.1 Posterior expected rank

In the notation of the main paper, the (relative) rank of unit i's parameter θ_i (from the top) is

$$\rho_i = \frac{1}{n} \sum_{j=1}^{n} 1[\theta_i \leq \theta_j].$$

Various authors have recommended to use posterior expected rank $\hat{\rho}_i = E(\rho_i | \text{data})$ as the basis of a ranking method:

$$\hat{\rho}_i = \frac{1}{n} \sum_{j=1}^{n} P(\theta_i \leq \theta_j | D_i, D_j)$$

where D_i and D_j are the data available on units i and j, respectively. For any fixed D_i, and in the limit of a large system $n \longrightarrow \infty$,

$$\hat{\rho}_i \longrightarrow_{a.s.} P(\theta_i \leq \theta | D_i) \quad (4)$$

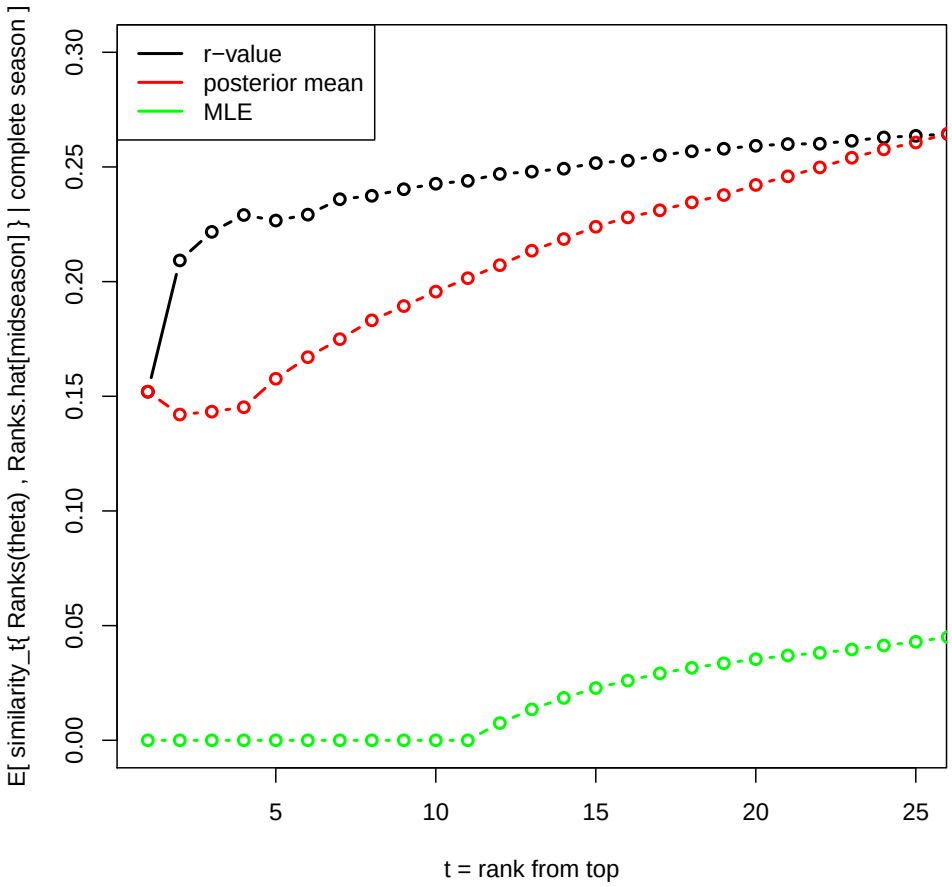

Figure S3: Validation experiment, NBA example, shows that r-values computed mid-season produce a ranking that is in better agreement with the true ranking of top players, on the average conditional upon end-of-season data, than rankings produced by maximum likelihood (MLE), posterior mean (PM) or posterior expected rank (PER).

Table S1: Leading free-throw shooters, by end of December 2013, in 2013-2014 regular NBA season.

player i	y_i	n_i	FTP	PM	RV	MLE.R	PM.R	RV.R
Ryan Anderson, PF	54	56	0.964	0.904	0.002	23	1	1
Brian Roberts, PG	18	18	1	0.862	0.003	11.5	12	2
Mike Harris, SF	26	27	0.963	0.867	0.006	24	11	3
Dirk Nowitzki, PF	124	135	0.919	0.895	0.008	30	2	4
Manu Ginobili, SG	63	68	0.926	0.883	0.009	29	4	5
Kevin Martin, SG	144	158	0.911	0.892	0.013	35	3	6
J.J. Redick, SG	51	55	0.927	0.877	0.015	28	5	7
Zaza Pachulia, C	42	45	0.933	0.873	0.017	26	7	8
Trey Burke, PG	41	44	0.932	0.871	0.019	27	8	9
O.J. Mayo, SG	61	67	0.91	0.871	0.023	36	9	10
Damian Lillard, PG	155	174	0.891	0.875	0.028	44	6	11
Reggie Jackson, PG	54	60	0.9	0.859	0.028	39	13	12
Jeremy Lamb, SG	18	19	0.947	0.84	0.029	25	25	13
Greivis Vasquez, PG	44	49	0.898	0.852	0.033	41	15	14
Ray Allen, SG	42	47	0.894	0.847	0.038	43	19	15
Travis Outlaw, SF	22	24	0.917	0.836	0.039	31.5	27	16
Kevin Durant, SF	255	290	0.879	0.87	0.04	52	10	17
Marc Gasol, C	54	61	0.885	0.849	0.043	49	17	18
Jimmer Fredette, PG	9	9	1	0.82	0.045	11.5	44	19
Khris Middleton, PF	47	53	0.887	0.846	0.045	48	21	20
Shaun Livingston, PG	60	68	0.882	0.849	0.046	51	16	21
D.J. Augustin, PG	21	23	0.913	0.832	0.048	33.5	32.5	22.5
Derek Fisher, PG	21	23	0.913	0.832	0.048	33.5	32.5	22.5
Jordan Crawford, SG	76	87	0.874	0.848	0.055	54	18	24
E'Twaun Moore, PG	19	21	0.905	0.824	0.06	37	39	25

where θ is a random draw from F. To see why, consider the object $U_j = P(\theta_i \leq \theta_j | D_i, D_j)$ as a random variable induced by D_j, and considering D_i as fixed. Then $\hat{\rho}_i$ is nothing but an average of these U_j's, and so by the strong law of large numbers we get convergence of $\hat{\rho}_i$ to $E(U_j|D_i)$, which, upon reflection, is seen to equal the right hand side of (4).

A substantive point of interest is the notion that while ranking is essentially a comparative exercise (comparing each θ_i with all other θ_j), in large systems the comparison amounts to checking θ_i against a distribution estimated from the data.

In the canonical normal/normal model of Section 2.1, we get:

$$P(\theta_i \leq \theta | X_i, \sigma_i^2) = P\left(Z_i \leq \frac{\theta - X_i/(\sigma_i^2+1)}{\sqrt{\sigma_i^2/(\sigma_i^2+1)}} \middle| X_i, \sigma_i^2 \right)$$

where Z_i is standard normal. By differencing, this becomes a probability for a single normal,

and thus:

$$P(\theta_i \leq \theta | X_i, \sigma_i^2) = \Phi\left(\frac{-X_i}{\sqrt{(\sigma_i^2+1)(2\sigma_i^2+1)}}\right). \quad (5)$$

The left hand side of (5) is the ranking variable for the posterior-expected-rank method; to obtain the corresponding threshold function, we fix the left side at some constant (depending on α) and replace X_i on the right side with $t_\alpha(\sigma_i^2)$, from which we obtain $t_\alpha(\sigma^2) = u_\alpha\sqrt{(2\sigma^2+1)(\sigma^2+1)}$. We reiterate that this threshold is quite similar to $u_\alpha(\sigma^2+1)$ corresponding to ranking by posterior mean $X_i/(\sigma_i^2+1)$.

2.2 Bayes factor

The BF threshold reported in Table 1 is admittedly an approximation to the most suitable one for this model (the ideal one is not analytically tractable). We imagine ranking units by evidence against a null as measured in the Bayes factor:

$$\text{BF} = \frac{P(\text{data}|\theta_i \neq 0)}{P(\text{data}|\theta_i = 0)}.$$

Since we are focusing on large θ_i, a more appropriate alternative would be $\theta_i > 0$, however for units with $X_i > 0$ and large BF the ideal and approximate values will be similar. The denominator above is evaluated using the fact that $X_i \sim \text{Normal}(0, \sigma_i^2)$ on the null. On the unordered alternative, $X_i \sim \text{Normal}(0, \sigma_i^2+1)$, marginal to θ_i. Thus the BF is

$$\text{BF} = \sqrt{\frac{\sigma_i^2}{\sigma_i^2+1}}\frac{\phi\left(X_i/\sqrt{\sigma_i^2+1}\right)}{\phi(X_i/\sigma_i)},$$

where ϕ is the standard normal density function. The threshold function is computed by setting the left side (ranking variable) to a constant (depending on α) and replacing X_i on the right side by the function $t_\alpha(\sigma_i^2)$, from which we obtain:

$$[t_\alpha(\sigma^2)]^2 = \sigma^2(\sigma^2+1)u_\alpha + \log\frac{\sigma^2+1}{\sigma^2}.$$

We include the indicator $X_i > 0$ in Table 1 to emphasize our focus on the large positive parameter values, recognizing that this prohibits ranking when $X_i < 0$.

3 On r-value computation

3.1 Pseudo-code

1. Input:
   - Data on n inference units: D[1], D[2], ... , D[n]
   - J grid points: 0 < A[1] < A[2], ... , < A[J] < 1; Avec = (A[j]).
   - Function $K(D,t) = P(\theta \geq t|D)$
   - Function $Q(a)$ such that $P\{\theta \geq Q(a)\} = a$
2. Construct θ quantiles:
   - Set T[j] = Q(A[j]) for all j
3. Construct conditional tail probabilities:
   - Set V[i,j] = K(D[i], T[j]) for all i and j
4. Construct marginal quantile function for conditional tail probability:
   - Find empirical quantile for each j:
     Lambda0[j] = quantile(ecdf(V[,j]), prob=(1-A[j]))
   - Smooth: Lambda[j] = $\sum_{k=1}^{J} w_{k,j}$Lambda0[k] for smoothing weights $w_{k,j}$.
     LambdaVec = (Lambda[j])
   - Construct an interpolating function:
     LambdaFun(a) = approxfun(Avec, LambdaVec)
5. Compute r-values:
   - Construct the function: DeltaFun(a, D) = LambdaFun(a) - $K(D, Q(a))$
   - For all i, solve DeltaFun(a, D[i]) = 0 for a = R[i]; take smallest root if multiple roots
6. Output: R=(R[i])

There is the matter of how to define the smallest r-value. In many cases there will be a smallest α value below which none of the n curves $V_\alpha(D_i)$ will cross one another as they converge towards zero. For example, in the normal/normal model one can show that curves either don't cross [equal variance] or they cross exactly once. Thus the `quantile` function will identify the same unit for all small α; the infimum of α values where $V_\alpha(D_i)$ is greater than or equal to the quantile $\hat{\lambda}_{\alpha,n}$ then is zero. Recognizing that λ_α is poorly estimated as α tends to zero, we could instead put the smallest value at some preset positive number equal to the minimum α on the grid. In our experiments we have used the lower bound $1/n$ for this purpose.

3.2 R package

The R package `rvalues` enables computation of r-values in a range of model structures. Some elements from version 0.2 are included here. See `http://cran.r-project.org/web/packages/rvalues/index.html` for the current version.

```
> ?rvalues

R-values

Description

Given data on a collection of units, this function computes r-values which are percentiles constructed
to maximize the agreement between the reported percentiles and the percentiles of the effect of interest.
Additional details about r-values are provided below and can also be found in the listed references.

Usage

rvalues(data, family = gaussian, hypers = "estimate", prior = "conjugate",
       alpha.grid = NULL, ngrid = NULL, smooth = 0, control = list())
Arguments

data
A data frame or a matrix with the number of rows equal to the number of sampling units. The first column
should contain the main estimates, and the second column should contain the nuisance terms.

family
An argument which determines the sampling distribution; this could be either family = gaussian,
family = tdist, family = binomial, family = poisson

hypers
values of the hyperparameters; only meaningful when the conjugate prior is used; if set to "estimate",
the hyperparameters are found through maximum likelihood; if not set to "estimate" the user should
supply a vector of length two.

prior
the form of the prior; either prior="conjugate" or prior="nonparametric".

alpha.grid
a numeric vector of points in (0,1); this grid is used in the discrete approximation of r-values

ngrid
number of grid points for alpha.grid; only relevant when alpha.grid=NULL

smooth
either smooth="none" or smooth takes a value between 0 and 10; this determines the level of smoothing
applied to the estimate of lambda(alpha) (see below for the definition of lambda(alpha)); if smooth
is given a number, the number is used as the bass argument in supsmu.

control
a list of control parameters for estimation of the prior; only used when the prior is nonparametric

Details

The r-value computation assumes the following two-level sampling model
\eqn{ X_i|\theta_i} ~ \eqn{p(x|\theta_i,\eta_i)}
and \eqn{\theta_i} ~ \eqn{F}, for \eqn{i = 1,...,n},
with parameters of interest \eqn{\theta_i}, effect size estimates \eqn{X_i},
and nuisance terms \eqn{\eta_i}. The form of \eqn{p(x|\theta_i,\eta_i)} is determined
by the \code{family} argument. When \code{family = gaussian}, it is assumed that
\eqn{X_i|\theta_i,\eta_i} ~ N(\eqn{\theta_i,\eta_i^{2})}.
When \code{family = binomial}, the \eqn{(X_i,\eta_i)} represent the number of successes
and number of trials respectively, and it is assumed that \eqn{X_i|\theta_i,\eta_i} ~
Binomial\eqn{(\theta_i,\eta_i)}. When \code{family = poisson}, the \eqn{{X_i}} should be
counts, and it is assumed that \eqn{X_i|\theta_i,\eta_i} ~ Poisson(\eqn{\theta_i * \eta_i)}.

The distribution of the effect sizes \eqn{F} may be a parametric distribution
that is conjugate to the corresponding \code{family} argument,
```

```
or it may be estimated nonparametrically. When it is desired that \eqn{F} be
parametric (i.e., \code{prior = "conjugate"}), the default is to estimate the
hyperparameters (i.e., \code{hypers = "estimate"}), but these may be supplied by the
user as a vector of length two. To estimate \eqn{F} nonparametrically, one
should use \code{prior = "nonparametric"} (see \code{\link{npmle}} for
further details about nonparametric estimation of \eqn{F}).

The \emph{r-value}, \eqn{r_i}, assigned to the ith case of interest is determined by
\eqn{  r_i = } inf[ \eqn{0 < \alpha < 1: V_\alpha(X_i,\eta_i) \ge \lambda(\alpha) } ]
where \eqn{V_\alpha(X_i,\eta_i) = P( \theta_i \ge \theta_\alpha|X_i,\eta_i) }
is the posterior probability that \eqn{\theta_i} exceeds the threshold \eqn{\theta_\alpha},
and \eqn{\lambda(\alpha)} is the upper-\eqn{\alpha}th quantile associated
with the marginal distribution of \eqn{V_\alpha(X_i,\eta_i)} (i.e.,
\eqn{ P(V_\alpha(X_i,\eta_i) \ge \lambda(\alpha)) = \alpha). } Similarly,
the threshold \eqn{\theta_\alpha} is the upper-\eqn{\alpha}th quantile of
\eqn{F} (i.e., \eqn{P(\theta_i \ge \theta_\alpha) = \alpha} ).

Value

An object of class "rvals" which is a list containing at least the following components:

main
a data frame containing the r-values, the r-value rankings along with the rankings from
several other common procedures

aux
a list containing other extraneous information

rvalues
a vector of r-values

Author(s)

Nicholas Henderson and Michael Newton

References

Henderson, N.C. and Newton, M.A. (2014) Making the Cut: Improved Ranking and Selection for
Large-Scale Inference. http://arxiv.org/abs/1312.5776

See Also

rvaluesMCMC, PostSummaries, Valpha

Examples

### Binomial example with Beta prior:
data(fluEnrich)
flu.rvals <- rvalues(fluEnrich, family = binomial)
hist(flu.rvals$rvalues)

### look at the r-values for indices 10 and 2484
fig_indices  <- c(10,2484)
fluEnrich[fig_indices,]

flu.rvals$rvalues[fig_indices]

### Gaussian sampling distribution with nonparametric prior
### Use a maximum of 5 iterations for the nonparam. estimate
data(hiv)
hiv.rvals <- rvalues(hiv, prior = "nonparametric", control = list(maxiter=5))
```

For example, the top gene sets by r-value are:

```
> head( flu.rvals$main )
                 RValue RV.rank MLE.rank PM.rank xx  nn  PostMean
GO:0022627 0.0001220703     1.0      1.0     1.0 23  35 0.2786319
```

```
GO:0022626 0.0001851183     2.0     23.0     2.0 33  90 0.2441963
GO:0006413 0.0003715832     3.0     45.0     3.0 47 152 0.2389331
GO:0045047 0.0005835630     4.5     27.5     4.5 36 106 0.2388799
GO:0072599 0.0005835630     4.5     27.5     4.5 36 106 0.2388799
GO:0006415 0.0009772328     6.0     25.0     6.0 32  91 0.2362017
```

Here's the internal R function that deploys the grid-based algorithm for a general case, where preprocessing has computed a matrix V of tail probabilities as well as a function for further evaluation:

```
> rvalues:::rvalueGuts

rvalueGuts <- function(dat,alpha.grid,V,vfun,hypers,smooth)
{
  ###########################################################
  # dat           nunits x 2 [specific to family]
  # alpha.grid    ngrid
  # V             nunits x ngrid
  # vfun          function tail prob
  # hypers        length 2; hyper-parameters [specific to familiy in vfun]
  ###########################################################

  nunits <- nrow(dat)
  ngrid <- length(alpha.grid)
  cc <- numeric(ngrid)
  for( j in 1:ngrid )
  {
      cc[j] <- quantile(V[,j], prob= 1 - alpha.grid[j], names = FALSE, type = 1)
  }

  ## smooth and functionalize
  if(smooth=="none") {
      ccfun <- approxfun( alpha.grid, cc, yleft = 1, yright = 0 )
  }
  else {
      cc2 <- supsmu( alpha.grid, cc, bass=smooth )
      ccfun <- approxfun( c(0,cc2$x,1), c(1,cc2$y,0))
  }
  ### Think of ccfun as the lambda_{\alpha} function

  dfun <- function( alpha, unitdata, hypers )
  {
    dd <- ccfun( alpha ) - vfun( alpha, unitdata, hypers )
    dd
  }
  ## march through units finding rvalue by uniroot
  rvals <- numeric(nunits)
  for( i in 1:nunits ) {
     rvals[i] <- uniroot(dfun,interval=c(0,1),unitdata=dat[i,],hypers=hypers)$root
  }

  ans <- list()
  ans$rvals <- rvals
  ans$lamfun <- cc
  ans$smoothlamfun <- ccfun
  return(ans)
}
<environment: namespace:rvalues>
```

4 Theory, supporting proof of Theorem 4 (main)

4.1 Properties of $H_\alpha(v)$ and $H_\alpha^{-1}(p)$

The following calculations support Lemma 5 of the main paper. Recall that $H_\alpha(v) = P[V_\alpha(X_i, \sigma_i^2) \leq v]$. Let $B = (0,1) \times (0,1)$ denote the open unit square, and $H_\alpha^{-1}(p) = \inf\{v : H_\alpha(v) \geq p\}$.

Lemma S1. *Under assumptions A1-A3, the cdf $H_\alpha(v)$ is continuous and strictly increasing on the range of $V_\alpha(D_i)$. Further, $H_\alpha^{-1}(p)$ is continuous for $(\alpha, p) \in B$ and is nondecreasing in each argument.*

Proof. Firstly, continuity of $D \mapsto V_\alpha(D)$ (Lemma 4, main) together with A1 on the positive density for D, and together with strict monotonicity of V_α in x, imply that the cdf $H_\alpha(v)$ is strictly increasing and continuous in v for each α.

On the nondecreasing monotonicity of $H_\alpha^{-1}(p)$ in both arguments: note that for fixed α the monotonicity follows from the function being a quantile function. For fixed p, the monotonicity follows since the random element V_α is increasing in α.

On the continuity of $H_\alpha^{-1}(p)$ in both arguments: Consider any $\alpha \in (0,1)$ and $p \in (0,1)$ and let (α_n, p_n) be a sequence converging to (α, p). To show that $H_{\alpha_n}^{-1}(p_n)$ converges to $H_\alpha^{-1}(p)$, we proceed in three steps.

(i) $H_\alpha^{-1}(p_n) \longrightarrow H_\alpha^{-1}(p)$.

This follows from the continuity of $p \mapsto H_\alpha^{-1}(p)$, using continuity and monotonicity of $H_\alpha(v)$ in v.

(ii) $H_{\alpha_n}^{-1}(p) \longrightarrow H_\alpha^{-1}(p)$.

Note that $V_{\alpha_n}(D_i) = V_\alpha(D_i) + u_{\alpha_n}(D_i)$, where

$$u_{\alpha_n}(D_i) = \begin{cases} P\{\theta_i \in [\theta_{\alpha_n}, \theta_\alpha) | D_i\}, & \text{if } \alpha_n \geq \alpha \\ -P\{\theta_i \in [\theta_\alpha, \theta_{\alpha_n}) | D_i\}, & \text{if } \alpha_n < \alpha. \end{cases}$$

Because $E(|u_{\alpha_n}(D_i)|) = |\alpha - \alpha_n|$, $u_{\alpha_n}(D_i) \longrightarrow_P 0$, and therefore, $V_{\alpha_n}(D_i) \longrightarrow_d V_\alpha(D_i)$. Thus, due to the continuity of $H_\alpha(v)$, $H_{\alpha_n}(v) \longrightarrow H_\alpha(v)$. From this, Lemma 21.2 of Van der Vaart implies that $H_{\alpha_n}^{-1}(u)$ converges to $H_{\alpha_n}^{-1}(u)$ at all continuity points u of $H_\alpha^{-1}(\cdot)$, which in particular, implies that $H_{\alpha_n}^{-1}(p) \longrightarrow H_\alpha^{-1}(p)$.

(iii) $H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n) \longrightarrow 0$.

Because $p \mapsto H_\alpha^{-1}(p)$ is nondecreasing, we have that

$$\begin{aligned} H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n) &= \Big(H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n)\Big)\mathbf{1}\{|p_n - p| \leq \varepsilon\} + \Big(H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n)\Big)\mathbf{1}\{|p_n - p| > \varepsilon\} \\ &\leq \Big(H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n)\Big)\mathbf{1}\{|p_n - p| \leq \varepsilon\} + \mathbf{1}\{|p_n - p| > \varepsilon\} \\ &\leq H_{\alpha_n}^{-1}(p + \varepsilon) - H_\alpha^{-1}(p - \varepsilon). + \mathbf{1}\{|p_n - p| > \varepsilon\}. \end{aligned} \tag{6}$$

Because $H_{\alpha_n}^{-1}(p) \longrightarrow H_\alpha^{-1}(p)$ [from (ii)] and $p_n \longrightarrow p$, we have

$$\limsup_{n \longrightarrow \infty} H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n) \leq H_\alpha^{-1}(p + \varepsilon) - H_\alpha^{-1}(p - \varepsilon). \tag{7}$$

Likewise

$$\liminf_{n \longrightarrow \infty} H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n) \geq H_\alpha^{-1}(p - \varepsilon) - H_\alpha^{-1}(p + \varepsilon). \tag{8}$$

Due to the continuity of H_α^{-1}, $H_\alpha^{-1}(p-\varepsilon) - H_\alpha^{-1}(p+\varepsilon) \longrightarrow 0$ as $\varepsilon \longrightarrow 0$, which from (7) and (8) means that $H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n) \longrightarrow 0$.

To conclude the proof, we just need to observe that

$$H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p) = H_\alpha^{-1}(p_n) - H_\alpha^{-1}(p) + H_{\alpha_n}^{-1}(p_n) - H_\alpha^{-1}(p_n), \tag{9}$$

and note that (i) and (iii) imply that $H_{\alpha_n}^{-1}(p_n) \longrightarrow H_\alpha^{-1}(p)$. □

4.2 Convergence of Sample Quantiles

Lemma S2. *Let F_n be a sequence of distribution functions such that for each t, we have $F_n(t) \xrightarrow{P} F(t)$, where F is a continuous strictly increasing distribution function. Then, for any $p \in (0,1)$, $\hat{F}_n^{-1}(p) \xrightarrow{P} F^{-1}(p)$, where $\hat{F}_n^{-1}(p) = \inf\{x : F_n(x) \geq p\}$.*

Proof. By noting the relation $F_n^{-1}(p) \leq x$ if and only if $F_n(x) \geq p$, we have $P\big\{F_n^{-1}(p) \leq x\big\} = P\big\{F_n(x) \geq p\big\}$. Hence, for any $p \in (0,1)$

$$\lim_{n\longrightarrow\infty} P\big\{F_n^{-1}(p) \leq x\big\} = \lim_{n\longrightarrow\infty} P\big\{F_n(x) \geq p\big\} = \begin{cases} 0 & \text{if } F(x) < p \\ 1 & \text{if } F(x) > p. \end{cases} \tag{10}$$

Because F has a proper inverse F^{-1}, (10) is equivalent to

$$\lim_{n\longrightarrow\infty} P\big\{F_n^{-1}(p) \leq x\big\} = \begin{cases} 0 & \text{if } F^{-1}(p) > x \\ 1 & \text{if } F^{-1}(p) < x. \end{cases} \tag{11}$$

The above means $F_n^{-1}(p)$ converges in distribution to $F^{-1}(p)$ which, since $F^{-1}(p)$ is a constant, implies that $\hat{F}_n^{-1}(p) \xrightarrow{P} F^{-1}(p)$. □

4.3 A two-dimensional Polya theorem

Lemma S3. *Consider $G : (0,1)\times(0,1) \longrightarrow [0,1]$ and the sequence of functions $\widehat{G}_n : (0,1)\times(0,1) \longrightarrow [0,1]$. Suppose that $\widehat{G}_n(t,s)$ converges to $G(t,s)$ pointwise in probability on the rectangle $A_\delta = [\delta, 1-\delta]\times[\delta,1-\delta]$ where $\delta \in (0,1/2)$. Further suppose that G and each $\hat{G}_n$ is coordinate-wise monotone on A_δ in the sense that for each fixed s, $t \mapsto G(t,s)$ (and $t \mapsto \widehat{G}_n(t,s)$) is nondecreasing and for each fixed t, $s \mapsto G(t,s)$ (and $s \mapsto \widehat{G}_n(t,s)$) is nondecreasing. Then, if $G(t,s)$ is continuous on A_δ, $\widehat{G}_n$ converges uniformly over A_δ to G in probability in the sense that*

$$\sup_{(t,s)\in A_\delta} |\widehat{G}_n(t,s) - G(t,s)| \longrightarrow_P 0. \tag{12}$$

Proof. Let $\varepsilon > 0$. Because A_δ is compact, G is uniformly continuous on A_δ. So, there is a δ_ϵ such that $d\big((x_1,y_1),(x_2,y_2)\big) < \delta_\epsilon \implies G(x_1,x_2) < \varepsilon$, where $d\big((x_1,y_1),(x_2,y_2)\big) = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$.

Choose grid points $\delta = t_1 < t_2 < \ldots < t_J = 1-\delta$ and $\delta = s_1 < s_2 < \ldots < s_K = 1-\delta$ such that for any $1 \leq j \leq J-1$ and $1 \leq k \leq K-1$ we have $d\big((t_j,s_k),(t_{j+1},s_{k+1})\big) < \delta_\epsilon$. Consider

any $(t,s) \in A_\delta$ and choose t_j, s_k such that $t_j \le t \le t_{j+1}$ and $s_k \le s \le s_{k+1}$. Then,

$$\begin{aligned}
\widehat{G}_n(t,s) - G(t,s) &\le \widehat{G}_n(t_{j+1},s) - G(t_j,s) \\
&\le \widehat{G}_n(t_{j+1},s_{k+1}) - G(t_j,s_k) \\
&= \widehat{G}_n(t_{j+1},s_{k+1}) - G(t_{j+1},s_{k+1}) + G(t_{j+1},s_{k+1}) - G(t_j,s_k) \\
&\le \widehat{G}_n(t_{j+1},s_{k+1}) - G(t_{j+1},s_{k+1}) + \varepsilon.
\end{aligned}$$

Likewise,

$$\begin{aligned}
\widehat{G}_n(t,s) - G(t,s) &\ge \widehat{G}_n(t_j,s) - G(t_{j+1},s) \\
&\ge \widehat{G}_n(t_j,s_k) - G(t_{j+1},s_{k+1}) \\
&= \widehat{G}_n(t_j,s_k) - G(t_j,s_k) + G(t_j,s_k) - G(t_{j+1},s_{k+1}) \\
&\ge \widehat{G}_n(t_j,s_k) - G(t_j,s_k) - \varepsilon. \qquad (13)
\end{aligned}$$

Hence,

$$\sup_{(t,s)\in A_\delta} |\widehat{G}_n(t,s) - G(t,s)| \le \max_{1\le j\le J, 1\le k\le K} |\widehat{G}_n(t_j,s_k) - G(t_j,s_k)| + \varepsilon. \qquad (14)$$

Define $E_n^{2\epsilon}$ as the event

$$E_n^{2\epsilon} = \Big\{ \sup_{(t,s)\in A_\delta} \Big|\widehat{G}_n(t,s) - G(t,s)\Big| > 2\varepsilon \Big\}. \qquad (15)$$

If for each $1 \le j \le J, 1 \le k \le K$, we define $B_{jk,n}^\epsilon = \big\{|\widehat{G}_n(t_j,s_k) - G(t_j,s_k)| > \varepsilon\big\}$, then $E_n^{2\epsilon} \subseteq \cup_{j=1}^J \cup_{k=1}^K B_{jk,n}^\epsilon$. Hence,

$$P(E_n^{2\epsilon}) \le P\Big(\bigcup_{j=1}^J \bigcup_{k=1}^K B_{jk,n}^\epsilon\Big) \le \sum_{j=1}^J \sum_{k=1}^K P(B_{jk,n}^\epsilon). \qquad (16)$$

Because $\widehat{G}_n$ converges pointwise to G in probability, $\lim_{n\longrightarrow\infty} P(B_{jk,n}^\epsilon) = 0$, for each (j,k). Moreover, because the number of terms in the summation in (16) is finite and does not depend on n, we have that $\lim_{n\longrightarrow\infty} P(E_n^{2\epsilon}) = 0$ which concludes the proof. □

4.4 Truncated r-values

Lemma 7 of the main paper is concerned with r-values truncated away from endpoints 0 and 1. The essential elements are in the proof in the main document, but here we work through all cases in detail.

Lemma S4. *Suppose that $\alpha \mapsto g_\alpha(D_i)$ is continuous with a unique root $r(D_i)$ in $[0,1]$. For any fixed $\delta \in (0, 1/2)$, if we define the truncated r-values $r^\delta(D_i)$ and estimated truncated r-values $\hat{r}_n^\delta(D_i)$ as*

$$\begin{aligned}
r^\delta(D_i) &= \min\Big\{ \inf\{\alpha \in [\delta,1] : g_\alpha(D_i) \ge 0\}, 1-\delta \Big\} \\
\hat{r}_n^\delta(D_i) &= \min\Big\{ \inf\{\alpha \in [\delta,1] : \hat{g}_{\alpha,n}(D_i) \ge 0\}, 1-\delta \Big\},
\end{aligned}$$

then $\hat{r}_n^\delta(D_i) \longrightarrow_P r^\delta(D_i)$.

Proof. Consider any $0 < \varepsilon < \frac{1-2\delta}{2}$. By the extreme value theorem, $|g_\alpha(D_i)|$ attains a minimum value $\delta_\epsilon(D_i)$ on the set $[0,1] \setminus (r(D_i)-\varepsilon, r(D_i)+\varepsilon)$. From the assumption that $g_\alpha(D_i)$ has a unique root $\delta_\epsilon(D_i) > 0$, and from the assumed continuity of $\alpha \mapsto g_\alpha(D_i)$, we have $\sup_{0\le\alpha\le r(D_i)-\varepsilon} g_\alpha(D_i) \le -\delta_\epsilon(D_i)$ (when $r(D_i) \ge \varepsilon$), and $\inf_{r(D_i)+\varepsilon\le\alpha\le 1} g_\alpha(D_i) \ge \delta_\epsilon(D_i)$ (when $r(D_i)+\varepsilon \le 1$).

Now, define E_n^ϵ as the event

$$E_n^\epsilon = \Big\{ ||\hat{g}_n(D_i) - g(D_i)||_\infty \le \frac{\delta_\epsilon(D_i)}{2} \Big\}, \tag{17}$$

and define

$$||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} = \sup_{\delta\le\alpha\le 1-\delta} |\hat{g}_{\alpha,n}(D_i) - g_\alpha(D_i)|. \tag{18}$$

We consider three cases separately: (i) $r(D_i) < \delta+\varepsilon$ (ii) $r(D_i) > 1-\delta-\varepsilon$ and (iii) $\delta+\varepsilon \le r(D_i) \le 1-\delta-\varepsilon$.

(i) $r(D_i) < \delta+\varepsilon$

When $r(D_i) < \delta+\varepsilon$, $g_{\delta+2\epsilon}(D_i) \ge \delta_\epsilon(D_i)$. This along with $||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} \le \frac{\delta_\epsilon(D_i)}{2}$ implies that $\hat{g}_{\delta+2\epsilon,n}(D_i) \ge \frac{\delta_\epsilon(D_i)}{2}$, and thus $\hat{r}_n(D_i) \le \delta+2\varepsilon$. Hence,

$$\begin{aligned} P\Big(E_n^\epsilon \cap \{r(D_i) < \delta+\varepsilon\}\Big) &\le P\Big(\delta \le \hat{r}_n^\delta(D_i) \le \delta+2\varepsilon, r(D_i) < \delta+\varepsilon\Big) \\ &\le P\Big(r^\delta(D_i) - 2\varepsilon \le \hat{r}_n^\delta(D_i) \le r^\delta(D_i)+2\varepsilon, r(D_i) < \delta+\varepsilon\Big). \end{aligned} \tag{19}$$

(ii) $r(D_i) > 1-\delta-\varepsilon$.

In this case, $\sup_{0\le\alpha\le 1-\delta-2\epsilon} g_\alpha(D_i) \le -\delta_\epsilon(D_i)$. This, along with $||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} \le \frac{\delta_\epsilon(D_i)}{2}$ implies $\sup_{\delta\le\alpha\le 1-\delta-2\epsilon} \hat{g}_{\alpha,n}(D_i) \le -\frac{\delta_\epsilon(D_i)}{2}$, and thus $\hat{r}_n^\delta(D_i) \ge 1-\delta-2\varepsilon$. Hence,

$$\begin{aligned} P\Big(E_n^\epsilon \cap \{r(D_i) > 1-\delta-\varepsilon\}\Big) &\le P\Big(1-\delta-2\varepsilon \le \hat{r}_n^\delta(D_i) \le 1-\delta, r(D_i) > 1-\delta-\varepsilon\Big) \\ &\le P\Big(r^\delta(D_i) - 2\varepsilon \le \hat{r}_n^\delta(D_i) \le r^\delta(D_i)+2\varepsilon, r(D_i) > 1-\delta-\varepsilon\Big) \end{aligned} \tag{20}$$

(iii) $r(D_i) \in [\delta+\varepsilon, 1-\delta-\varepsilon]$

In this case, $\sup_{0\le\alpha\le r(D_i)-\epsilon} g_\alpha(D_i) \le -\delta_\epsilon(D_i)$ and $\inf_{r(D_i)+\epsilon\le\alpha\le 1} g_\alpha(D_i) \ge \delta_\epsilon(D_i)$. These, along with $||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} \le \frac{\delta_\epsilon(D_i)}{2}$ imply that $\sup_{\delta\le\alpha\le r(D_i)-\epsilon} \hat{g}_{\alpha,n}(D_i) \le -\frac{\delta_\epsilon(D_i)}{2}$ and $\inf_{r(D_i)+\epsilon\le\alpha\le 1-\delta} \hat{g}_{\alpha,n}(D_i) \ge \frac{\delta_\epsilon(D_i)}{2}$, and thus $\hat{r}_n^\delta(D_i) \in [r(D_i)-\varepsilon, r(D_i)+\varepsilon]$. Hence,

$$\begin{aligned} &P\Big(E_n^\epsilon \cap \{\delta+\varepsilon \le r(D_i) \le 1-\delta-\varepsilon\}\Big) \\ &\le P\Big(r(D_i)-\varepsilon \le \hat{r}_n^\delta(D_i) \le r(D_i)+\varepsilon, \delta+\varepsilon \le r(D_i) \le 1-\delta-\varepsilon\Big) \\ &= P\Big(r^\delta(D_i)-\varepsilon \le \hat{r}_n^\delta(D_i) \le r^\delta(D_i)+\varepsilon, \delta+\varepsilon \le r(D_i) \le 1-\delta-\varepsilon\Big). \end{aligned} \tag{21}$$

By combining (19), (20), and (21) we have that

$$P\Big(r^\delta(D_i) - 2\varepsilon \le \hat{r}_n^\delta(D_i) \le r^\delta(D_i)+2\varepsilon\Big) \ge P\Big(E_n^\epsilon\Big). \tag{22}$$

Thus, to complete the proof, we just need to show that $\lim_{n\longrightarrow\infty} P(E_n^\epsilon) = 1$. To this end, if we choose K_ϵ such that $P\{\delta_\epsilon(D_i) \leq K_\epsilon\} \leq \varepsilon$ and look at the complement event $(E_n^\epsilon)^c$, then

$$\begin{aligned} P\{(E_n^\epsilon)^c\} &= P\Big(||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} > \frac{\delta_\epsilon(D_i)}{2}\Big) \\ &\leq P\Big(||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} > \frac{\delta_\epsilon(D_i)}{2}, \delta_\epsilon(D_i) \leq K_\epsilon\Big) + P\Big(\delta_\epsilon(D_i) \leq K_\epsilon\Big) \\ &\leq P\Big(||\hat{g}_n(D_i) - g(D_i)||_{\delta,\infty} > \frac{K_\epsilon}{2}\Big) + \varepsilon. \end{aligned} \tag{23}$$

It then follows from the uniform convergence result, Lemma 6 (main), that $\lim_{n\longrightarrow\infty} P\{(E_n^\epsilon)^c\} = 0$, and hence, $\lim_{n\longrightarrow\infty} P(E_n^\epsilon) = 1$. □

5 Models and assumptions

5.1 Normal/Normal

As indicated in the main document the normal/normal model under the standard parameterization is:

$$\begin{aligned} p(x|\theta,\sigma^2) &= \frac{1}{\sigma}\phi\left(\frac{x-\theta}{\sigma}\right) \qquad x \in \mathbb{R} \\ f(\theta) &= \phi(\theta) \qquad \theta \in \mathbb{R} \\ g(\sigma^2) &= \text{arbitrary density on } \mathbb{R}^+ \ . \end{aligned}$$

From this we derive that $\theta_i|x_i,\sigma_i^2$ is normal with mean $x_i/(\sigma_i^2+1)$ and variance $\sigma_i^2/(\sigma_i^2+1)$, and further that $x_i|\sigma_i^2$ is normal with mean 0 and variance $1+\sigma_i^2$. The upper α quantile $\theta_\alpha = \Phi^{-1}(1-\alpha)$, and the posterior tail probability is

$$V_\alpha(x,\sigma^2) = 1 - \Phi\left[\sqrt{\frac{\sigma^2+1}{\sigma^2}}\left(\theta_\alpha - \frac{x}{\sigma^2+1}\right)\right].$$

The sampling component $p(x|\theta,\sigma^2)$ satisfies A3 by inspection. Any permutation invariant consistent estimator $\hat{F}_n$ of the normal F with satisfy A2, and Theorem 3 assures no crossing (A4). Further, $V_\alpha(x,\sigma^2)$ is strictly increasing and continuous in x (for Theorem 2), and the optimal thresholds do not cross (Theorem 3).

5.2 t/Normal

Adopt the model structure from Section 2.1 (main paper), but rather than having $\sigma_i^2 = \text{var}(X_i|\theta_i,\sigma_i^2)$, suppose that σ_i^2 is an sample-based estimate of the true variance, say ξ_i^2, such that $\nu\sigma_i^2/\xi_i^2$ has a chi-square distribution on ν degrees of freedom given ξ_i, as in a standard normal sampling model with unknown variance. Assume further that ξ_i^2 has an inverse-Gamma marginal distribution, with shape $a/2$ and rate $b/2$, say. Marginally, it follows that the estimated variance σ_i^2 has a compound Gamma distribution. But more to the point, the sampling model:

$$p(x_i|\sigma_i^2,\theta_i) = \int_0^\infty p(x_i|\xi_i,\theta_i)\, p(\xi_i|\sigma_i^2)\, d\xi_i.$$

Conjugacy gives ξ_i^2 given σ_i^2 as another inverse Gamma, having shape $(a+\nu)/2$ and rate $(b+\nu\sigma_i^2)/2$. This allows integration above to a t conditional of x_i given σ_i^2 and θ:

$$p(x_i|\sigma_i^2,\theta) \propto \left[(x_i-\theta_i)^2 + b + \nu\sigma_i^2\right]^{-(a+\nu+1)/2}.$$

Assumptions A1 and A2 follow from the basic model structure given a consistent estimate of the normal mixing distribution F. The t sampling model satisfies $A3(i)$ and $A3(ii)$ by inspection. Although the monotonicity condition A3(iii) does not hold for the t, $V_\alpha(x,\sigma^2)$ is continuous and strictly increasing in x. Monotonicity of V follows since the posterior mean $E(\theta|x,\sigma^2)$ is increasing in x, because the posterior variance does not involve x, and by inspecting the posterior density (e.g. Nadarajah and Pogany, 2012). Note, A3(iii) is used only in Lemma 4 to establish monotonicity in x of the posterior tail probabilities. Finally, we conjecture that A4 is true owing to unimodality and symmetry of the posterior densities, however a proof remains to be found.

5.3 Gamma/Inverse Gamma

Suppose that $X_i|\theta_i,\gamma_i \sim \text{Gamma}(1, 1/\theta_i\gamma_i)$ and $\theta_i \sim \text{Inv-Gamma}(1,1)$ so that

$$p(x|\theta,\gamma) = \frac{1}{\theta\gamma}\exp(-x/\theta\gamma) \qquad \text{and} \qquad p(\theta) = \theta^{-2}e^{-1/\theta}.$$

Because $p(x|\theta,\gamma)p(\theta) \propto \theta^{-3}\exp\{-\theta^{-1}(x/\gamma+1)\}$, we have $\theta_i|X_i,\gamma_i \sim \text{Inv-Gamma}(2, 1+X_i/\gamma_i)$ and hence $\frac{\theta_i}{(1+X_i/\gamma_i)}|X_i,\gamma_i \sim \text{Inv-Gamma}(2,1)$. The tail probabilities $V_\alpha = P\{\theta \geq \theta_\alpha|x,\gamma\}$, are then given by

$$V_\alpha(x,\gamma) = P\Big\{\frac{\theta}{1+x/\gamma} \geq \frac{\theta_\alpha}{1+x/\gamma}\Big|x,\gamma\Big\} = 1 - F_{2,1}\Big(\frac{\theta_\alpha}{1+x/\gamma}\Big),$$

where $F_{2,1}(t)$ is the cdf of an Inv-Gamma(2,1) random variable. By looking at the relation $V_\alpha(t_\alpha(\gamma),\gamma) = \lambda_\alpha$, we have that

$$1 - F_{2,1}\Big(\frac{\theta_\alpha}{1+t_\alpha(\gamma)/\gamma}\Big) = \lambda_\alpha,$$

which upon simplification gives

$$\frac{\theta_\alpha}{1+t_\alpha(\gamma)/\gamma} = F_{2,1}^{-1}\big(1-\lambda_\alpha\big). \tag{24}$$

By simplifying again, we have

$$t_\alpha(\gamma) = \gamma\Big(\frac{\theta_\alpha}{F_{2,1}^{-1}\big(1-\lambda_\alpha\big)} - 1\Big). \tag{25}$$

Now, to determine the form of λ_α, we need the marginal distribution of X_i given γ_i which is given by

$$p(x|\gamma) = \int_0^\infty p(x|\theta,\gamma)p(\theta|\gamma)d\theta = \frac{1}{\gamma}\int_0^\infty \theta^{-3}e^{-\theta^{-1}(1+x/\gamma)}d\theta = \frac{1}{\gamma(1+x/\gamma)^2},$$

and the marginal distribution function is

$$F_{x|\gamma}(t) = \int_0^t \frac{1}{\gamma(1+x/\gamma)^2}dx = \frac{-1}{1+x/\gamma}\Big|_0^t = 1 - \frac{1}{1+t/\gamma}. \tag{26}$$

Using (26) and letting $h(\gamma)$ denote the density of γ, the size constraint equation can be written as

$$\int_0^\infty P\big\{X \geq t_\alpha(\gamma)|\gamma\big\}h(\gamma)d\gamma = \int_0^\infty \frac{1}{1+t_\alpha(\gamma)/\gamma}h(\gamma)d\gamma = \alpha. \tag{27}$$

Looking back at (24), we know that $1 + t_\alpha(\gamma)/\gamma = \theta_\alpha/F_{2,1}^{-1}\big(1-\lambda_\alpha\big)$ and hence

$$\int_0^\infty \frac{1}{1+t_\alpha(\gamma)/\gamma}h(\gamma)d\gamma = \int_0^\infty \frac{F_{2,1}^{-1}\big(1-\lambda_\alpha\big)}{\theta_\alpha}h(\gamma)d\gamma = \frac{F_{2,1}^{-1}\big(1-\lambda_\alpha\big)}{\theta_\alpha} = \alpha. \tag{28}$$

So, $F_{2,1}^{-1}\big(1-\lambda_\alpha\big)/\theta_\alpha = \alpha$, regardless of the choice of h. Looking back at (25), the threshold function can now be written as

$$t_\alpha(\gamma) = \gamma\Big(\frac{\theta_\alpha}{F_{2,1}^{-1}\big(1-\lambda_\alpha\big)} - 1\Big) = \gamma\Big(\frac{1}{\alpha} - 1\Big), \tag{29}$$

from which it is clear that any two threshold functions $t_{\alpha_1}(\gamma)$ and $t_{\alpha_2}(\gamma)$ do not cross for any $\gamma > 0$.

Because $F_{2,1}^{-1}\big(1-\lambda_\alpha\big)/\theta_\alpha = \alpha$ and $\theta_\alpha = -1/\log(1-\alpha)$, we have

$$\lambda_\alpha = 1 - F_{2,1}\big(\theta_\alpha\alpha\big) = 1 - F_{2,1}\big(-\alpha/\log(1-\alpha)\big).$$

It is interesting to note that λ_α is an increasing function which may be seen by differentiating λ_α with respect to α

$$\frac{d\lambda_\alpha}{d\alpha} = \Big(\frac{\log(1-\alpha) + \frac{\alpha}{1-\alpha}}{\log^2(1-\alpha)}\Big)f_{2,1}\big(-\alpha/\log(1-\alpha)\big) > 0.$$

It is also interesting to note that

$$\lim_{\alpha\longrightarrow 0}\lambda_\alpha = 1 - F_{2,1}\Big(\lim_{\alpha\longrightarrow 0}\frac{-\alpha}{\log(1-\alpha)}\Big) = 1 - F_{2,1}(1).$$

6 Connection to hypothesis testing

Consider testing the null hypothesis $H_0 : \theta_i < \theta_\alpha$ against the alternative $H_A : \theta_i \geq \theta_\alpha$, for some $\alpha \in (0,1)$ in the measurement model of Section 2.1 (main document). The conditional probability $P\{\theta_i \geq \theta_\alpha | X_i \geq t_\alpha^*(\sigma_i^2)\}$ is a false discovery rate (FDR) for this test, and so the agreement in (equation 2, main paper) satisfies

$$\text{agreement} = (1 - \text{FDR}) \cdot \alpha.$$

That t^* maximizes agreement means that it minimizes FDR for this test. The procedure also maximizes an average power, since

$$\text{agreement} = \text{average power} \cdot \alpha$$

if we flip the conditioning around and notice that $P\left\{X_i \geq t_\alpha^*(\sigma_i^2)|\theta_i \geq \theta_\alpha\right\}$ is an average power, averaging over H_A. In both cases, the association is slightly contrived, since the significance level of the test is not controlled. Nonetheless the observation may provide further context for the approach.

7 Simulation figures

Figures S4-S8 show results of a simulation study. In this study, r-values were computed through a fitted normal/normal model, and using the empirical estimate of the quantile function $\hat{\lambda}_{\alpha,n}$.

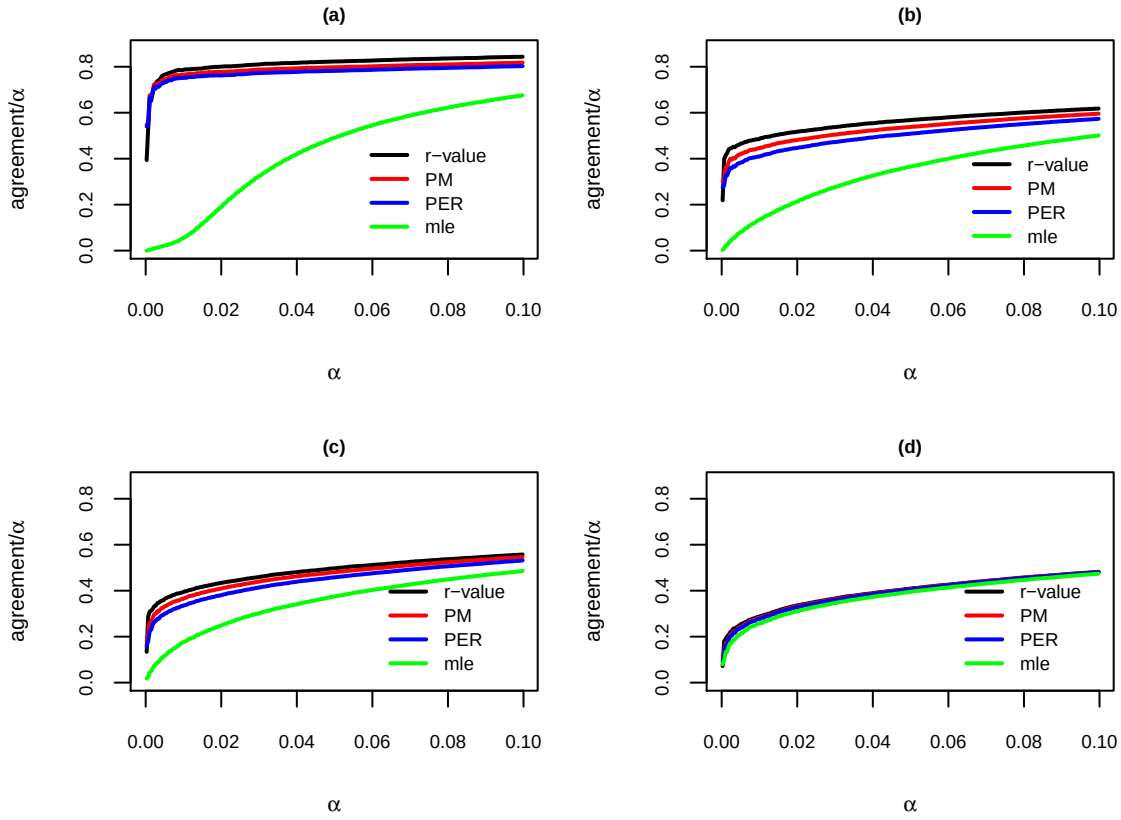

Figure S4: Finite-sample performance of r-value, posterior mean (PM), posterior expected rank (PER), and maximum likelihood estimate (MLE) in the normal/normal model for varying distributions of σ_i^2. The simulation-based agreement compares the true top-α list with the estimated top-α list for various methods and for $1/n \leq \alpha \leq 0.1$ (common horizontal axis), when the marginal distribution of θ_i and the quantile λ_α are both estimated from available data (no smoothing). The distribution of σ_i^2 in each of the four panels is as follows: (a) $\sigma_i^2 \sim \text{Gamma}(1/16, 1/16)$ (b) $\sigma_i^2 \sim \text{Gamma}(1/2, 1/2)$ (c) $\sigma_i^2 \sim \text{Gamma}(1, 1)$ (d) $\sigma_i^2 \sim \text{Gamma}(10, 10)$. The common vertical axis is agreement/α; and results from 1000 simulated data sets were averaged for each panel.

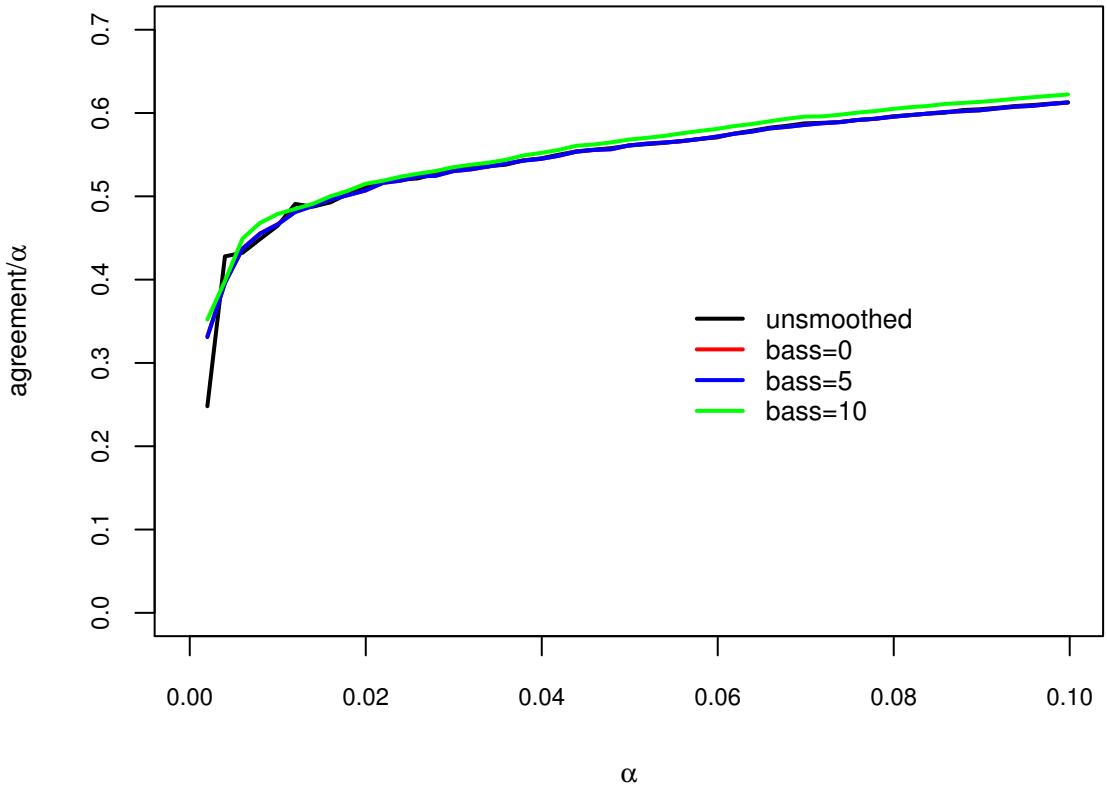

Figure S5: Finite-sample performance of r-value in the normal/normal model for varying levels of smoothness in the estimate $\hat{\lambda}_{\alpha,n}$ of λ_α. The simulation-based agreement compares the true top-α list with the estimated top-α list for various methods and for $1/n \leq \alpha \leq 0.1$ (common horizontal axis), when the marginal distribution of θ_i and the quantile λ_α are both estimated from available data (no smoothing). Smoothing is performed using Friedman's "Super-smoother" as implemented by the **R** function **supsmu** where the level of smoothing is determined by the **bass** argument (values up to 10 indicate increasing smoothness). The smoothing in each of the four panels is as follows: (a) no smoothing (b) bass=0 (c) bass=5 (d) bass=10. The vertical axis is agreement/α; $\sigma_i^2 \sim$ Gamma$(1/2, 1/2)$, and results from 1000 simulated data sets were averaged.

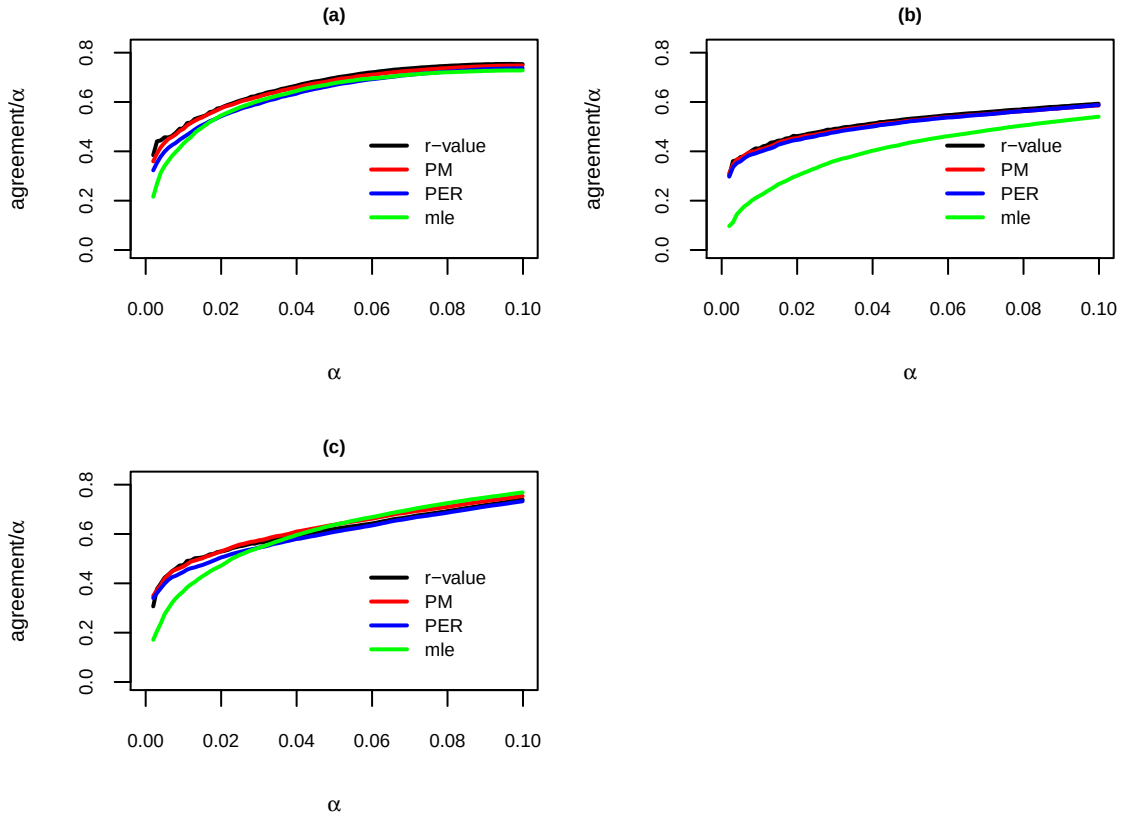

Figure S6: Effects on agreement of model misspecification: true distribution of θ_i is a Normal mixture with three components. The k^{th} component of the mixture has parameters (μ_k, τ_k^2), and the mixture proportions are denoted with (π_1, π_2, π_3). The three panels are as follows: (a) $(\pi_1, \pi_2, \pi_3) = (0.2, 0.7, 0.1)$ and $(\mu_1, \mu_2, \mu_3) = (-2, 0, 3)$ (b) $(\pi_1, \pi_2, \pi_3) = (0.2, 0.5, 0.3)$ and $(\mu_1, \mu_2, \mu_3) = (-2, 0, -1)$ (c) $(\pi_1, \pi_2, \pi_3) = (0.2, 0.6, 0.2)$ and $(\mu_1, \mu_2, \mu_3) = (-2, 0, 4)$. In (a), (b), and (c), $(\tau_1^2, \tau_2^2, \tau_3^2) = (1, 1, 1)$. The common vertical axis is agreement/α; $\sigma_i^2 \sim$ Gamma$(1, 1)$, and results from 1000 simulated data sets were averaged for each panel. R-values are computed assuming the normal/normal model.

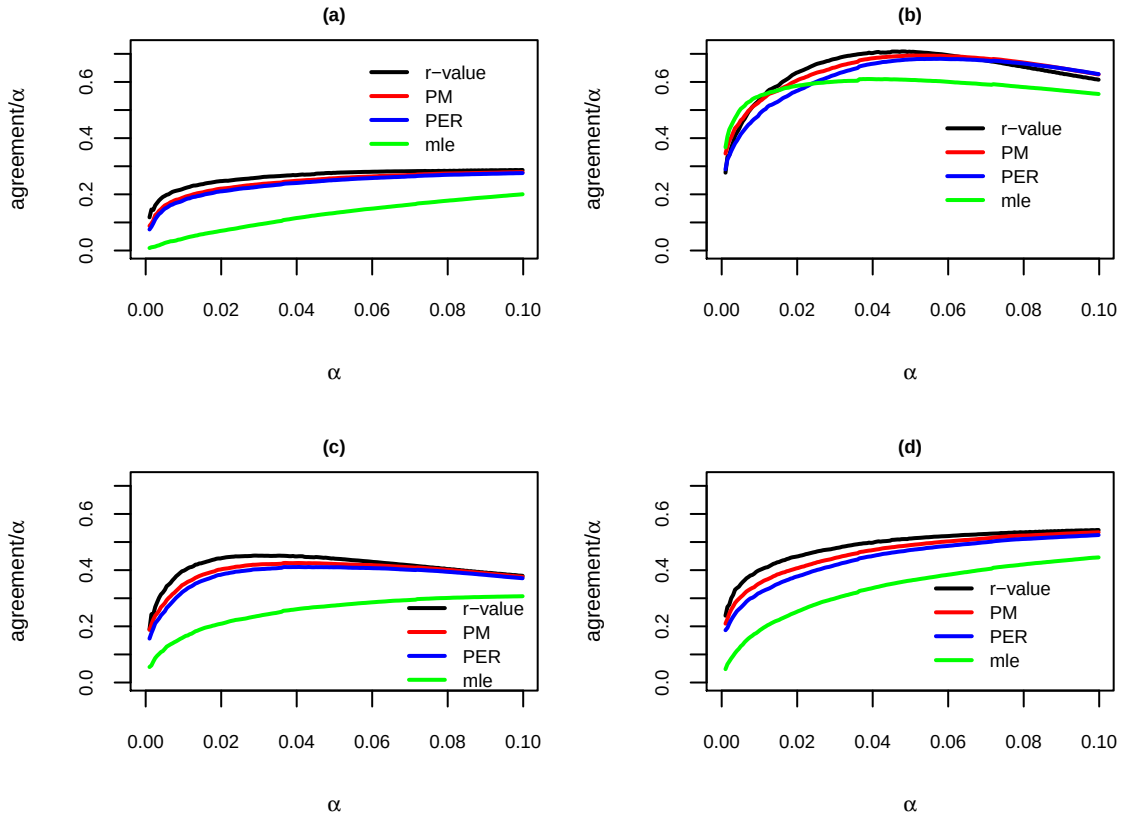

Figure S7: Effects on agreement of model misspecification: true distribution of θ_i is a mixture of a Normal distribution $N(0, \tau^2)$ (with probability $1 - \pi$) and a point mass at zero (with probability π). The four panels are as follows: (a) $\pi = .75$ and $\tau = 1/2$ (b) $\pi = .75$ $\tau = 2$ (c) $\pi = .8$ and $\tau = 1$ (d) $\pi = .5$ and $\tau = 1$. The common vertical axis is agreement/α; $\sigma_i^2 \sim$ Gamma$(1, 1)$, and results from 1000 simulated data sets were averaged for each panel. R-values are computed assuming the normal/normal model.

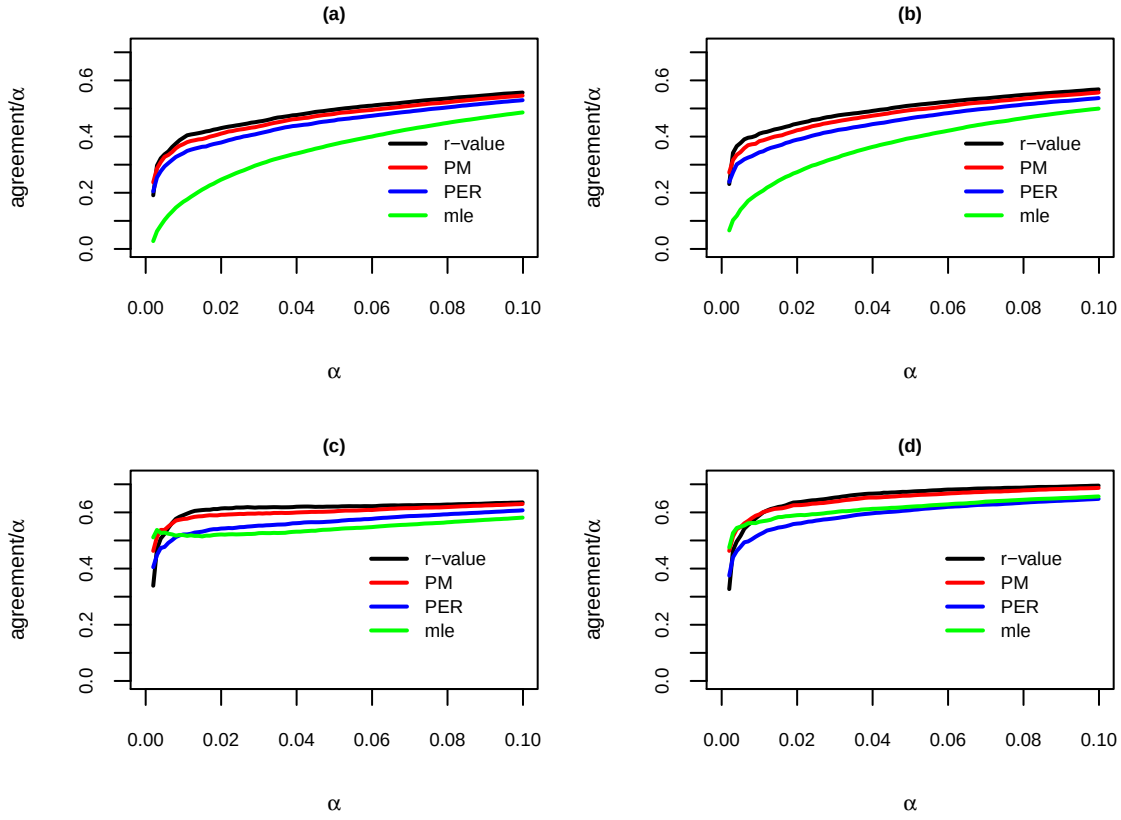

Figure S8: Effects on agreement of model misspecification: θ_i and σ_i^2 are correlated. The four panels are as follows: (a) $\text{Corr}(\theta_i, \sigma_i^2) = \sqrt{1/29}$ (b) $\text{Corr}(\theta_i, \sigma_i^2) = \sqrt{1/17}$ (c) $\text{Corr}(\theta_i, \sigma_i^2) = \sqrt{4/13}$ (d) $\text{Corr}(\theta_i, \sigma_i^2) = \sqrt{1/2}$. The common vertical axis is agreement/α; $\sigma_i^2 \sim \text{Gamma}(1, 1)$, and results from 1000 simulated data sets were averaged for each panel. R-values are computed assuming the normal/normal model.

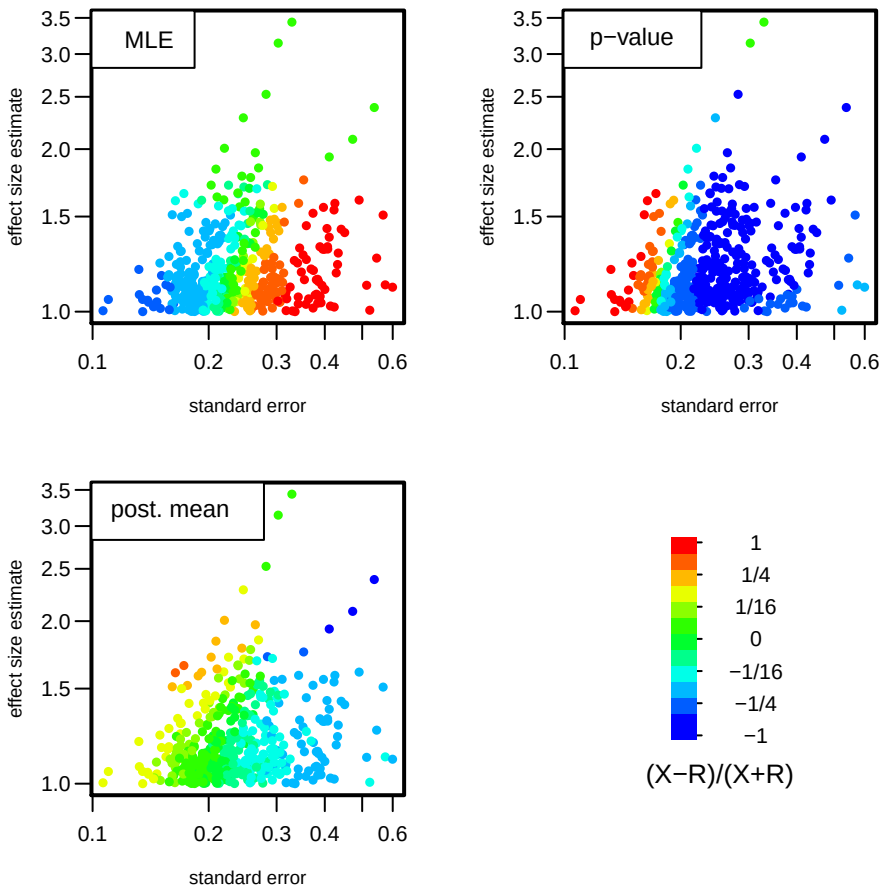

Figure S9: Ranking via various methods compared to r-value ranking, Pyeon example. Coloring format as in Fig. 3 (main paper). From 54675 transcripts assayed on a microarray, shown are data from 458 having the highest mean difference between the two groups ($\log_2$ fold change exceeding 1).

Other References

1. Nadarajah, S and Pogany, TK (2012). On the convolution of normal and t random variables. *Statistics: A journal of theoretical and applied statistics*. DOI:10.1080/02331888.2012.694447.