Web Appendix C: Alternative loss functions

In some cases, it may be interesting to penalize either (a) the number of variables in the Bayes estimator or (b) the number of atoms in the Bayes estimator. These penalties may be useful to constrain the size of the resulting estimates. Here, we introduce some variations on the loss function corresponding to the $0-1$ discrepancy measure which was introduced earlier, by looking at what happens when linear constraints are introduced. In order to penalize for a large number of variables in the Bayes estimator, we consider the following constrained optimization problem:

Minimize $L(\tau, U)$ with the constraint that $|U| < \rho$, for some $\rho > 0$. This is equivalent to minimizing the loss function:

$$L_f^{\lambda}(\tau, U) \quad = \quad (1-w) * |U \setminus \tau| + w * |\tau \setminus U| + \lambda |U| \text{ for some } \lambda > 0$$

(see for instance Gill et al., 1981). Similarly, one may be interested in penalizing for a large number of atoms in the Bayes estimator. In such a case, we would consider the following constrained optimization problem:

Minimize $L(\tau, U)$ with the constraint that $J < \eta$, for some $\eta > 0$, where J is the number of atoms in U. This is equivalent to minimizing the loss function:

$$L_a^{\xi}(\tau, U) \quad = \quad (1-w) * |U \setminus \tau| + w * |\tau \setminus U| + \xi J \text{ for some } \xi > 0.$$

The penalties can also be applied atoms with a small number of genes or atoms, respectively, by simply changing the signs of λ and ξ. For these two alternative minimization problems, it is also possible to analytically calculate the resulting Bayes estimators as in Theorem 2. Lemma C1 extends the results of Theorem 2 to the penalized case.

Lemma C1: *For a fixed value of $w \in [0, 1]$, the Bayes estimator for the posterior expected losses L_f^{λ} and L_a^{ξ} are defined by the indicators δ_l of whether atom A_l is in the Bayes estimator:*

$$
\begin{aligned}
\delta_l &= 1\{afdr_l + \lambda \leqslant w\}, \\
\delta_l &= 1\{afdr_l + \frac{\xi}{n_l} \leqslant w\}.
\end{aligned}
$$

Alternatively, loss functions which employ the ratio of missed discoveries and false discoveries, instead of the number of missed discoveries and false discoveries, may also be used. Thus, would lead to the estimator

$$
\begin{aligned}
L_r(\tau, U) &= (1-w) * \frac{|U \setminus \tau|}{|U|} + w * \frac{|\tau \setminus U|}{M - |U|} \\
&= (1-w) * \text{Ratio of false discoveries } + w* \text{ Ratio of missed discoveries.}
\end{aligned}
$$

Theorem C1 shows how this loss can be written in terms of the expected numbers of false discoveries for each atom, similar to Lemma A3 in Web Appendix A.

Theorem C1: *The posterior expected loss $\mathcal{L}_r(U)$ which results from the loss function L_r may be rewritten as:*

$$
\boldsymbol{\delta}' \left\{ \frac{(1-w)}{\boldsymbol{\delta}'\mathbf{n}} \mathbf{EFD}_A - \frac{w}{(M - \boldsymbol{\delta}'\mathbf{n})} (\mathbf{n} - \mathbf{EFD}_A) \right\} + \frac{w}{M - \boldsymbol{\delta}'\mathbf{n}} (\mathbf{n} - \mathbf{EFD}_A)
$$

For this more complicated loss function, an analytic solution is not available. However, an approximate algorithmic solution is available. We first consider the solution to the problem where we constrain the size of the Bayes estimator $|U| = \boldsymbol{\delta}'\mathbf{n}$ to some size ρ. In this case, we get the following constrained linear binary problem:

$$
\begin{aligned}
\min_{\boldsymbol{\delta}} \quad & \boldsymbol{\delta}' \left\{ \frac{(1-w)}{\rho} \mathbf{EFD}_A - \frac{w}{(M-\rho)} (n - \mathbf{EFD}_A) \right\} \\
\text{s.t.} \quad & \boldsymbol{\delta}'\mathbf{n} = \rho
\end{aligned} \tag{1}
$$

This is an instance of the well-known 0-1 knapsack problem (Garey and Johnson, 1979), which can be solved approximately by Dantzig's greedy algorithm. This uses a sorting strategy where atoms are sorted increasingly by the quantity

$$
\frac{(1-w)}{\rho} afdr_l - \frac{w}{(M-\rho)} (1 - afdr_l)
$$

and t_l is set to 1, in order, until $\boldsymbol{\delta}'\mathbf{n} = \rho$. Note that when $\rho = M/2$, atoms are sorted according to $afdr_l$ as in Theorem 2.

In principle, to solve the fractional problem, one can solve the 0-1 knapsack problem for each possible value of $\rho = |U|$ and select the best solution. Since ρ can range over a large number of possible values, we use a strategy based on the projected gradient of the fractional function at a given point to find a small number of estimator sizes $|U|$ to test. The steps are presented in Algorithm C1.

Algorithm C1: Algorithm to obtain the Bayes estimator for the loss function L_r.

$\rho \leftarrow \min\{\mathbf{n}\}$;

$I_{iter} \leftarrow 0$;

while $I_{iter} \leqslant I_{iter}^{max}$ **do**

  Find $\boldsymbol{\delta}_\rho = \arg\min_{\boldsymbol{\delta}} \boldsymbol{\delta}' \left\{ \frac{(1-w)}{\rho} \mathbf{EFD}_A - \frac{w}{(M-\rho)}(n - \mathbf{EFD}_A) \right\}$, along $\boldsymbol{\delta}'\mathbf{n} = \rho$;

  Find $\mathbf{s} = \min_{\boldsymbol{\delta}} \mathcal{L}_r(\boldsymbol{\delta})$, along $\boldsymbol{\delta} = \boldsymbol{\delta}_\rho - \alpha(\nabla_{\boldsymbol{\delta}} L_r(\boldsymbol{\delta}_\rho))_+$;

  if $\mathbf{s}'\mathbf{n} = \rho$ **then**

    stop;

  else

    $\rho \leftarrow \mathbf{s}'\mathbf{n}$;

  end

  $I_{iter} \leftarrow I_{iter} + 1$;

end

We compared the Bayes estimators resulting from the various loss functions considered here based on the results of 100 simulation runs with 2250 features, 10% of which were from the alternative distribution. We considered 8 atoms, 4 of size 50 and 4 of size 100. For each of the two sizes considered, atoms had fractions of alternatives of 0, 0.1, 0.5, or 0.9. The results are presented in Table C1. We used $\lambda = \xi = 0.2$. The simulations illustrate that our estimation methods generally produce conservative results for the loss functions considered, as in the vast majority of cases the atoms which were chosen as part of the Bayes estimator should in fact have been in it.

[Table 1 about here.]

References

Garey, M. R. and Johnson, D. S. (1979). *Computers and intractability: a guide to NP-completeness.* WH Freeman and Company, San Francisco.

Gill, P. E., Murray, W., and Wright, M. H. (1981). *Practical optimization.* London: Academic Press.

Table C1

Atoms present in the Bayes estimators obtained from different loss functions over 100 runs. We took $\lambda = \xi = 0.2$. Eight atoms were considered, having the fractions of alternatives 0, 0.1, 0.5, and 0.9 and the sizes 50 and 100. The number of times each atom appeared in the Bayes estimator is given for each loss function for three values of w; in parentheses is stated whether or not that specific atom should have been in the Bayes estimator (given the ideal scenario which gives posterior probabilities of 1 and 0 to the variables from the alternative, respectively from the null distribution.)

Fraction of alternatives	Size	w	L	L_f^λ	L_a^ξ	L_r
0	50	0.25	0 (No)	0 (No)	0 (No)	0 (No)
		0.50	0 (No)	0 (No)	0 (No)	0 (No)
		0.67	0 (No)	0 (No)	0 (No)	1 (No)
0	100	0.25	0 (No)	0 (No)	0 (No)	0 (No)
		0.50	0 (No)	0 (No)	0 (No)	0 (No)
		0.67	0 (No)	0 (No)	0 (No)	4 (No)
0.1	50	0.25	0 (No)	0 (No)	0 (No)	0 (No)
		0.50	0 (No)	0 (No)	0 (No)	0 (No)
		0.67	0 (No)	0 (No)	0 (No)	53 (Yes)
0.1	100	0.25	0 (No)	0 (No)	0 (No)	0 (No)
		0.50	0 (No)	0 (No)	0 (No)	0 (No)
		0.67	0 (No)	0 (No)	0 (No)	53 (Yes)
0.5	50	0.25	0 (No)	0 (No)	0 (No)	0 (No)
		0.50	0 (Yes)	0 (No)	0 (No)	0 (No)
		0.67	100 (Yes)	0 (No)	0 (No)	55 (Yes)
0.5	100	0.25	0 (No)	0 (No)	0 (No)	0 (No)
		0.50	0 (Yes)	0 (No)	0 (No)	0 (No)
		0.67	100 (Yes)	0 (No)	0 (No)	55 (Yes)
0.9	50	0.25	94 (Yes)	0 (No)	0 (No)	99 (Yes)
		0.50	100 (Yes)	97 (Yes)	97 (Yes)	100 (Yes)
		0.67	100 (Yes)	100 (Yes)	100 (Yes)	100 (Yes)
0.9	100	0.25	99 (Yes)	0 (No)	0 (No)	68 (Yes)
		0.50	100 (Yes)	100 (Yes)	100 (Yes)	92 (Yes)
		0.67	100 (Yes)	100 (Yes)	100 (Yes)	100 (Yes)