Robust Empirical Bayes Confidence Intervals

Michal Kolesár

May 27, 2020

The package `ebci` implements robust empirical Bayes confidence intervals (EBCIs) proposed by Armstrong et al. [2020] for inference in a normal means model $Y_i \sim N(\theta_i, \sigma_i^2)$, $i = 1, \ldots, n$.

Setup

Suppose we use an empirical Bayes estimator of θ_i that shrinks toward the predictor based on the regression of θ_i onto X_i (equivalently, regression of Y_i onto X_i),

$$\hat{\theta}_i = X_i'\delta + w_i(Y_i - X_i'\delta), \tag{1}$$

where $\delta = E[X_i X_i']^{-1} E[X_i\theta_i]$, $w_i = \frac{\mu_2}{\mu_2+\sigma_i^2}$, and

$$\mu_2 = E[(\theta_i - X_i'\delta)^2 \mid X_i, \sigma_i]. \tag{2}$$

We assume that μ_2 doesn't depend on σ_i. Morris [1983] proposes to use the *parametric EBCI*

$$\hat{\theta}_i \pm \frac{z_{1-\alpha/2}}{\sqrt{w_i}} w_i\sigma_i.$$

The critical value $z_{1-\alpha/2}/\sqrt{w_i}$ is larger than the usual critical value $z_{1-\alpha/2}$ =`qnorm(1-alpha/2)` if the estimator was unbiased conditional on θ_i. This CI is justified if we strengthen the assumption (2) by making the normality assumption $\theta_i \mid X_i, \sigma_i \sim N(X_i'\delta, \mu_2)$: to account for the bias $b_i = (1-w)(\theta_i - X_i'\delta)$ of the estimator (more precisely b_i is the bias conditional on $(X_i, \theta_i, \sigma_i)$), the parametric EBCI assumes that it is normally distributed.

A *robust EBCI* that is only uses (2) and not the normality assumption takes the form

$$X_i'\delta + w_i(Y_i - X_i'\delta) \pm cva_\alpha(m_2, \infty)w_i\sigma_i, \; m_2 = (1 - 1/w_i)^2\mu_2/\sigma_i^2, \tag{3}$$

where the critical value cva_α is derived in Armstrong et al. [2020]. Here m_2 is the second moment of the bias-standard deviation ratio $m_2 = b_i/(w_i\sigma_i)$, which we refer to as the normalized bias. This critical value imposes a constraint (2) on the second moment of θ_i, but no constraints on higher moments. We can make the critical value smaller by also imposing a constraint on the kurtosis of θ_i (or equivalently, the kurtosis of the normalized bias)

$$\kappa = E[(\theta_i - X_i'\delta)^4 \mid X_i, \sigma_i]/\mu_2^2 = E[b_i^4]/E[b_i^2]^2. \tag{4}$$

In analogy to (2), we assume here that the conditional fourth moment of $\theta_i - X_i'\delta$ doesn't depend on (X_i, σ_i). In this case, the robust EBCI takes the form

$$\hat{\theta}_i \pm cva_\alpha(m_2, \kappa)w_i\sigma_i, \; m_2 = (1 - 1/w_i)^2\mu_2\sigma_i^2,$$

These critical values are implemented in the package by the `cva` function:

```
library("ebci")
## If m_2=0, then we get the usual critical value
cva(m2 = 0, kappa = Inf, alpha = 0.05)$cv
#> [1] 1.959964
## Otherwise the critical value is larger:
cva(m2 = 4, kappa = Inf, alpha = 0.05)$cv
#> [1] 7.216351
## Imposing a constraint on kurtosis tightens it
cva(m2 = 4, kappa = 3, alpha = 0.05)$cv
#> [1] 4.619513
```

Example

Here we illustrate the use of the package using a dataset from Chetty and Hendren [2018] (CH hereafter). The dataset is included in the package as the list `cz`. Run `?cz` for a full description of the dataset. As in Chetty and Hendren [2018], we use precision weights proportional to the inverse of the squared standard error to compute (δ, μ_2, κ).

```
## As Y_i, use fixed effect estimate theta25 of causal
## effect of neighborhood for children with parents at
## the 25th percentile of income distribution. The
## standard error for this estimate is se25. As
## predictors use average outcome for permanent residents
## (stayers), stayer25. Let us use 90% CIs.
r <- ebci(formula = theta25 ~ stayer25, data = cz, se = se25,
    weights = 1/se25^2, alpha = 0.1, wopt = FALSE)
```

For shrinkage toward the grand mean, or toward zero, use the specification `theta25 ~ 1`, or `theta25 ~ 0` in the `formula` argument of `ebci`.

The return value contains (see `?ebci` for full description)

1. The least squares estimate of δ:

   ```
   r$delta
   #> (Intercept)    stayer25
   #> -1.44075193  0.03244676
   ```

2. Estimates of μ_2 and κ. The estimate used for EBCI calculations (`estimate`) is obtained by applying a finite-sample correction to an initial method of moments estimate (`uncorrected_estimate`). This correction ensures that we don't shrink all the way to zero (or past zero) if the method-of-moments estimate of μ_2 is negative (see Armstrong et al. [2020] for details):

```
c(r$mu2, r$kappa)
#>              estimate uncorrected_estimate             estimate 
#>         6.243867e-03         6.243867e-03         7.785337e+02 
#> uncorrected_estimate 
#>         3.453191e+02
```

3. A data frame with columns:

```
names(r$df)
#>  [1] "w_eb"    "w_opt"   "ncov_pa" "len_eb"  "len_op"  "len_pa"  "len_us" 
#>  [8] "th_us"   "th_eb"   "th_op"   "se"
```

The columns of the data frame refer to:

- `w_eb` Empirical Bayes shrinkage factor $w_i = \mu_2/(\mu_2 + \sigma_i^2)$.
- `th_eb` Empirical Bayes estimator $\hat{\theta}_i$ given in (1)
- `len_eb` Half-length $cva_\alpha(m_2, \kappa)w_i\sigma_i$ of the robust EBCI, so that the lower endpoint of the EBCIs are given by `th_eb-len_eb`, and the upper endpoint by `th_eb+len_eb`. For a given observation, this can be also computed directly using the `cva` function:

```
cva(m2 = ((1 - 1/r$df$w_eb[1])/r$df$se[1])^2 * r$mu2[1],
    r$kappa[1], alpha = 0.1)$cv * r$df$w_eb[1] * r$df$se[1]
#> [1] 0.1916245
r$df$len_eb[1]
#> [1] 0.1916245
```

- `len_pa` Half-length $z_{1-\alpha/2}\sqrt{w_i}\sigma_i$ of the parametric EBCI.
- `w_opt` Shrinkage factor that optimizes the length of the resulting confidence interval, that is, the value of w_i that minimizes (3) over w_i (missing here since we specified `wopt=FALSE`)
- `th_op` Estimator based on the length-optimal shrinkage factor `w_opt` (missing here since we specified `wopt=FALSE`)
- `len_op` Half-length $cva_\alpha(m_2, \kappa)w_i\sigma_i$ of the length-optimal EBCI (missing here since we specified `wopt=FALSE`).
- `th_us` The unshrunk estimate Y_i, as specified in the `formula` argument of the function `ebci`.
- `len_us` Half-length $z_{1-\alpha/2}\sigma_i$ of the CI based on the unshrunk estimate
- `se` The standard error σ_i, as specified by the argument `se` of the `ebci` function.
- `ncov_pa` Maximal non-coverage of the parametric EBCI.

Using the data frame, we can give a table summarizing the results. Let us show the results for the CZ in California:

```
df <- (cbind(cz[!is.na(cz$se25), ], r$df))
df <- df[df$state == "CA", ]
knitr::kable(data.frame(cz = df$czname, unshrunk_estimate = df$theta25,
    estimate = df$th_eb, lower_ci = df$th_eb - df$len_eb,
    upper_ci = df$th_eb + df$len_eb), digits = 3)
```

cz	unshrunk_estimate	estimate	lower_ci	upper_ci
Klamath Falls	0.372	0.112	-0.076	0.301
Redding	0.082	0.031	-0.144	0.206
Eureka	0.280	-0.097	-0.299	0.105
Crescent City	-0.078	-0.076	-0.302	0.149
Modesto	-0.004	0.015	-0.111	0.142
Bakersfield	0.028	0.084	-0.047	0.215
Fresno	-0.377	-0.163	-0.274	-0.052
Chico	0.072	0.034	-0.135	0.203
Sacramento	-0.012	0.006	-0.095	0.106
San Jose	-0.011	0.048	-0.073	0.169
Santa Rosa	-0.117	0.002	-0.167	0.171
San Francisco	0.017	0.029	-0.077	0.134
Mammoth Lakes	0.045	0.109	-0.120	0.338
San Diego	0.054	0.056	-0.035	0.148
Yuma	-0.216	0.031	-0.130	0.192
Santa Barbara	-0.040	0.068	-0.088	0.224
Los Angeles	-0.170	-0.129	-0.191	-0.067

Using shrinkage tightens the robust EBCIs relative to the unshrunk CI by a factor of

```
mean(r$df$len_us)/mean(r$df$len_eb)
#> [1] 4.024772
```

on average.

On the other hand, using the parametric EBCI yields CIs that violate the 90% coverage requirement, the worst-case coverage for the estimated value of (μ_2, κ) is given by

```
mean(r$df$ncov_pa)
#> [1] 0.2274703
```

References

Tim Armstrong, Michal Kolesár, and Mikkel Plagborg-Møller. Robust empirical Bayes confidence intervals. ArXiv: 2004.03448, April 2020. URL https://arxiv.org/abs/1606.01200.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228, August 2018. doi: 10.1093/qje/qjy006.

Carl N. Morris. Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55, March 1983. doi: 10.1080/01621459.1983.10477920.