Robust Empirical Bayes Confidence Intervals

Michal Kolesár

April 09, 2020

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

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.

```
## For illustration, only use 20 largest CZ
df <- cz[sort(cz$pop, index.return = TRUE, decreasing = TRUE)$ix[1:20],
    ]
## 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, and let us
## set kappa=Inf (so we only don't use a constraint on
## kurtosos) in order to decrease the computational time.
## Given this, we do not need to parallelize the
## computation, and can set cores=1:
r <- ebci(formula = theta25 ~ stayer25, data = df, se = se25,
    alpha = 0.1, cores = 1, kappa = Inf)
```

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:

1. The least squares estimate of δ:

   ```
   r$delta
   #> (Intercept)    stayer25
   #> -0.75375536  0.01707074
   ```

2. Estimates of $\sqrt{\mu_2}$ and κ, obtained using method of moments. We set `kappa=Inf`, so the kurtosis value is taken to be infinity:

```
c(r$sqrt_mu2, r$kappa)
#> [1] 0.06085232        Inf
```

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$sqrt_mu2/r$df$se[1])^2,
    Inf, alpha = 0.1)$cv * r$df$w_eb[1] * r$df$se[1]
#> [1] 0.05766286
r$df$len_eb[1]
#> [1] 0.05766286
```

- `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
- `th_op` Estimator based on the length-optimal shrinkage factor `w_opt`
- `len_op` Half-length $cva_\alpha(m_2, \kappa)w_i\sigma_i$ of the length-optimal EBCI.
- `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:

```
knitr::kable(data.frame(name = paste0(df$czname, ", ", df$state),
    estimate = r$df$th_eb, lower_ci = r$df$th_eb - r$df$len_eb,
    upper_ci = r$df$th_eb + r$df$len_eb), digits = 3)
```

name	estimate	lower_ci	upper_ci
Los Angeles, CA	-0.111	-0.168	-0.053
New York, NY	-0.098	-0.158	-0.039
Chicago, IL	-0.117	-0.190	-0.043
Newark, NJ	0.011	-0.066	0.088

name	estimate	lower_ci	upper_ci
Philadelphia, PA	-0.015	-0.106	0.075
Detroit, MI	-0.093	-0.176	-0.011
Boston, MA	0.039	-0.059	0.138
San Francisco, CA	0.022	-0.076	0.120
Washington DC, DC	0.083	0.005	0.161
Houston, TX	-0.017	-0.094	0.059
Miami, FL	-0.018	-0.085	0.049
Atlanta, GA	-0.096	-0.161	-0.032
Seattle, WA	0.100	0.005	0.196
Dallas, TX	-0.024	-0.109	0.062
Bridgeport, CT	-0.030	-0.125	0.064
Phoenix, AZ	0.005	-0.070	0.079
Minneapolis, MN	0.068	-0.045	0.182
San Diego, CA	0.043	-0.041	0.127
Cleveland, OH	-0.020	-0.124	0.084
Sacramento, CA	0.006	-0.087	0.099

An analogous table, but constructed using length-optimal shrinkage:

```
knitr::kable(data.frame(name = paste0(df$czname, ", ", df$state),
    estimate = r$df$th_op, lower_ci = r$df$th_op - r$df$len_op,
    upper_ci = r$df$th_op + r$df$len_op), digits = 3)
```

name	estimate	lower_ci	upper_ci
Los Angeles, CA	-0.114	-0.172	-0.057
New York, NY	-0.101	-0.161	-0.042
Chicago, IL	-0.124	-0.196	-0.052
Newark, NJ	0.010	-0.065	0.085
Philadelphia, PA	-0.010	-0.094	0.074
Detroit, MI	-0.093	-0.172	-0.014
Boston, MA	0.039	-0.049	0.128
San Francisco, CA	0.021	-0.067	0.109
Washington DC, DC	0.093	0.017	0.169
Houston, TX	-0.020	-0.095	0.055
Miami, FL	-0.018	-0.085	0.048
Atlanta, GA	-0.096	-0.161	-0.032
Seattle, WA	0.120	0.033	0.207
Dallas, TX	-0.024	-0.105	0.057
Bridgeport, CT	-0.043	-0.130	0.043
Phoenix, AZ	0.004	-0.069	0.077
Minneapolis, MN	0.076	-0.020	0.172
San Diego, CA	0.044	-0.036	0.124
Cleveland, OH	0.003	-0.088	0.094
Sacramento, CA	0.003	-0.082	0.089

Using length-optimal shrinkage tightens the robust EBCIs by about 6 percentage points on average

```
mean(r$df$len_op)/mean(r$df$len_eb)
#> [1] 0.9387541
```

On the other hand, using the parametric EBCI yields CIs that mildly violate the 90% coverage requirement:

```
mean(r$df$ncov_pa)
#> [1] 0.1200214
```

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.