Short benchmark of Fleiss' kappa, Cohen's kappa, and Brennan-Prediger

Benchmark for the non-aggregrated functions

Jonas Moss

4/25/23

The main benefit of `quadagree` versus `irrCAC` for data on long form is its support for missing data and continuous data, but bootstrapping and transformations can come in handy as well. The extra machinery makes it slightly slower, however, in typical workloads. If n is very large, `irrCAC.fleiss.raw` severely outperforms `fleissci`. The following benchmark is being run on $n = 50$ ratings and $R = 4$ raters.

```
library("quadagree")
x = dat.zapf2016
irr_conger <- \(x) irrCAC::conger.kappa.raw(x, weights = "quadratic")
irr_fleiss <- \(x) irrCAC::fleiss.kappa.raw(x, weights = "quadratic")
irr_bp <- \(x) irrCAC::bp.coeff.raw(x, weights = "quadratic")
microbenchmark::microbenchmark(
  irr_conger(x),
  congerci(x),
  irr_fleiss(x),
  fleissci(x),
  irr_bp(x),
  bpci(x),
  times = 1000
)
```

```
Unit: microseconds
          expr   min      lq      mean  median      uq      max neval
 irr_conger(x) 752.0  820.55  943.0781  846.85  896.40   6050.4  1000
   congerci(x) 955.4 1026.35 1666.3746 1061.00 1131.95 474518.5  1000
 irr_fleiss(x) 375.2  436.15  525.5756  456.20  486.95   5404.5  1000
   fleissci(x) 961.0 1030.40 1215.0220 1063.15 1124.80  25896.8  1000
```

```
    irr_bp(x) 376.9  427.00  516.1516  449.30  479.20   5765.9  1000
      bpci(x) 904.8  990.60 1188.0605 1021.45 1079.65  86680.7  1000
```

We see that `irrCAC` is more than twice as fast for Fleiss' kappa and the Brennan-Prediger coefficient, which is due to the binning used by `irrCAC`. On the other hand, binning is not as simple for Cohen's kappa, hence the speed differential is smaller. Let's try $n = 500$.

```
y = rbind(x, x, x, x, x, x, x, x, x, x)

microbenchmark::microbenchmark(
  irr_conger(y),
  congerci(y),
  irr_fleiss(y),
  fleissci(y),
  irr_bp(y),
  bpci(y),
  times = 1000
)
```

```
Unit: microseconds
          expr    min      lq      mean  median      uq     max neval
 irr_conger(y) 1204.8 1277.45 1601.0278 1310.00 1395.60 71697.7  1000
   congerci(y) 1347.1 1425.30 1773.6871 1463.80 1554.30 71617.1  1000
 irr_fleiss(y)  478.6  526.60  598.2008  547.70  577.55  3565.4  1000
   fleissci(y) 1348.3 1432.70 1721.1952 1466.15 1564.65  5312.1  1000
     irr_bp(y)  468.5  515.30  591.2828  536.90  561.70  3699.2  1000
       bpci(y) 1322.1 1410.60 1689.4510 1448.15 1540.55  7156.5  1000
```

The speed differential shrinks for Conger's kappa but increases for Fleiss' kappa / Brennan-Prediger as n becomes larger. This is due to the calculation of the asymptotic variance, which is slow and unoptimized in `quadagree`.

For $n = 3000$, `irrCAC` is roughly 5 times faster than `quadagree` for Fleiss' kapa.

```
z = rbind(y, y, y, y, y, y, y, y, y, y)
microbenchmark::microbenchmark(
  irr_conger(z),
  congerci(z),
  irr_fleiss(z),
  fleissci(z),
  irr_bp(z),
  bpci(z)
```

```
)
```

```
Unit: milliseconds
          expr    min      lq      mean   median       uq     max neval
 irr_conger(z) 6.7168 8.96060 11.111147 10.23575 13.09320 19.2711   100
   congerci(z) 5.8056 8.81820 11.776415 10.27405 13.51165 80.5201   100
 irr_fleiss(z) 1.4095 1.60325  2.345702  1.93615  2.59200  7.3114   100
   fleissci(z) 6.3114 9.07585 11.356691 11.10265 13.32645 18.5786   100
     irr_bp(z) 1.3890 1.54490  2.348898  1.83980  2.48755  6.4055   100
       bpci(z) 6.0253 8.94175 13.209353 11.28685 15.16765 85.9913   100
```