Bayesian robust t-Test

Statistik-Kolloquium, André

September 4, 2024

The problem of induction

- Can we learn about the future from the past?
- Can we learn from incomplete information?
- Bayesian inference is fully probabilistic
- Unambiguous (conditional) probability statements about unknown given known!
  - Unknown are model parameters and unobserved data
  - Known are observed data

Parameters and data

- Assume a quantity of interest θ
- or a vector θ of quantities of interest (e.g., the regression coefficients of multiple regression)
- Assume we have data x

Bayes theorem

- The posterior distribution of θ given data x, $p(\theta \mid x)$, can be calculated from the
- likelihood (statistical model) $L(\theta) = p(x \mid \theta)$
- and the prior distribution $p(\theta)$,

Bayes-Theorem

$$\boxed{p(\theta \mid x) = \frac{p(x \mid \theta) \cdot p(\theta)}{p(x)}},$$

with $p(x)$ as the marginal likelihood or “evidence”[1].

---

[1] $p(x) = \int_\theta f(x \mid \theta) \cdot p(\theta) \, \mathrm{d}\theta$

Inference and decisions

- Standard posterior summaries
  - Mean, median, mode, standard deviation, quantile (e.g, for 95% credibility intervals, the $Q_{0.025}$ and $Q_{0.975}$)

We use the posterior distribution to answer questions

- Clinically relevant effect (δ)? $\Pr(\theta > \delta \mid Y)$
- Effect in some range (equivalence)? $\Pr(\delta_1 < \theta < \delta_2 \mid Y)$
- Treatment comparisons? $\Pr(\theta_1 > \theta_2 \mid Y)$ or $\Pr(\theta_1 > \theta_2 - \delta \mid Y)$
- Effect on transformed scale? $\Pr(g(\theta) \mid Y)$
- Combine posterior with utilities (if available)

Example: Bayesian Estimation versus t-Test

- Bayesian approach results in probability statements about the values of interest, rather than p-values and significance levels.
- In addition, the procedure accounts for departures from normality by using a t-distribution to model the variable of interest and estimating a measure of normality.

Example: Bayesian Estimation versus t-Test

```r
y1 <- c(-0.5, 0, 1.2, 1.2, 1.2, 1.9, 2.4, 3) * 100
y2 <- c(-1.2, -1.2, -0.5, 0, 0, 0.5, 1.1, 1.9) * 100
data <- data.frame(y1, y2)
psych::describe(data)
```

```
   vars n  mean    sd median trimmed   mad  min max range  skew kurtosis    se
y1    1 8 130.0 116.0    120   130.0 140.8  -50 300   350 -0.13    -1.39 41.01
y2    2 8   7.5 107.9      0     7.5 118.6 -120 190   310  0.29    -1.38 38.16
```

```r
print(t.test(y1, y2, mu = 0, var.equal = FALSE), digits = 6)
```

```

    Welch Two Sample t-test

data:  y1 and y2
t = 2.187, df = 13.93, p-value = 0.0463
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
   2.28771 242.71229
sample estimates:
mean of x mean of y
    130.0       7.5
```

Example: Robust Bayesian estimation

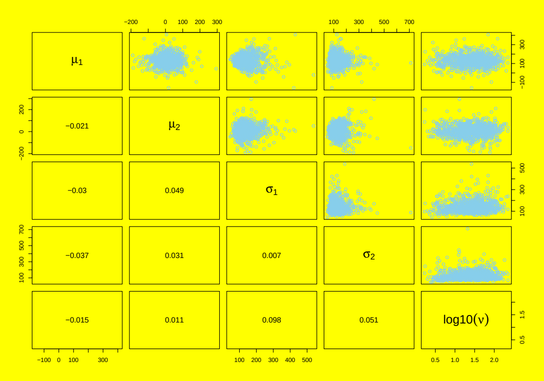
- We use uninformative priors for all parameters.
- the posterior distributions of μ_1, μ_2, $\mu_1 - \mu_2$, σ_1, σ_2, $\sigma_1 - \sigma_2$, ν (degrees of freedom)
- We can not only test $H_0 : \mu_1 - \mu_2 = \delta_0$,
- but we can test a region of practical equivalence (ROPE), as example, $H_0 : |(\mu_1 - \mu_2) - \delta_0| \leq 20$.

Implementation

```r
library(BEST)
BESTout <- BESTmcmc(y1, y2, verbose = FALSE)
```

Posterior distribution

```
par(bg = "yellow")
pairs(BESTout)
```



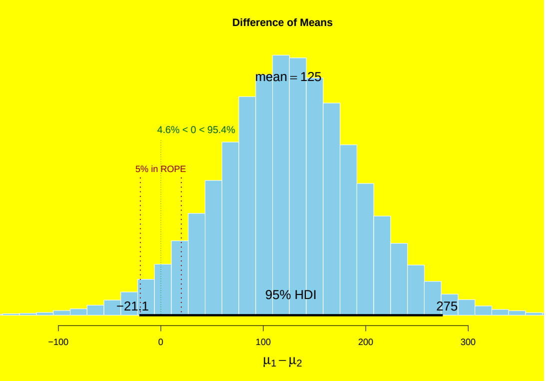
Posterior distribution of derived quantities

```
a <- 20  #ROPE
c <- 0  #delta_0
summary(BESTout, ROPEm = c(-a + c, c + a), compValm = c, digits = 5)
```

```
             mean  median    mode HDI%    HDIlo  HDIup compVal %>compVal ROPElow ROPEhigh %InROPE
mu1       130.803 130.990 133.189   95   18.898 236.48
mu2         6.295   5.948   7.520   95  -94.215 107.84
muDiff    124.508 124.514 129.924   95  -21.064 275.20       0      95.4     -20       20    4.54
sigma1    139.838 128.414 111.368   95   63.894 244.22
sigma2    130.036 119.381 106.326   95   58.757 225.01
sigmaDiff   9.802   8.560  10.105   95 -133.739 155.11       0      56.4
nu         34.579  25.816   9.564   95    1.038  94.11
log10nu     1.382   1.412   1.561   95    0.605   2.11
effSz       0.965   0.955   0.965   95   -0.158   2.12       0      95.4
```

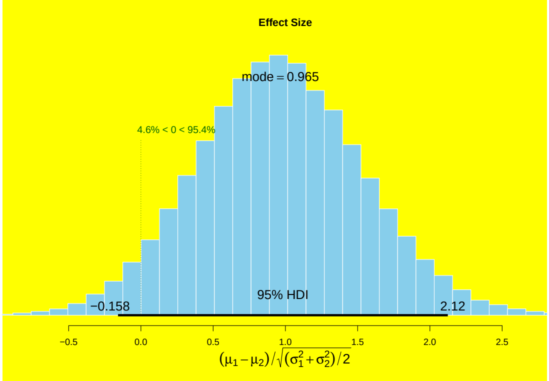
Posterior distribution of mean difference

```
par(bg = "yellow")
plot(BESTout, compVal = c, ROPE = c(-a + c, c + a), showCurve = FALSE)
```



Posterior distribution of effect size

```r
par(bg = "yellow")
plot(BESTout, which = "effect", showCurve = FALSE)
```



Robust Bayesian estimation versus t-test

- The t-test would reject $H_0 : \mu_1 - \mu_2 = 0$
- the Bayesian analysis would not reject the hypothesis, nor that the effect is in ROPE.
- The two treatments are practically or clinically equivalent.

Moving from NHST to Bayes

- Scientific disciplines from astronomy to zoology are moving to Bayesian data analysis. We should be leaders of the move, not followers.
- Modern Bayesian methods provide richer information, with greater flexibility and broader applicability than 20th century methods. Bayesian methods are intellectually coherent and intuitive. Bayesian analyses are readily computed with modern software and hardware.
- Null-hypothesis significance testing (NHST), with its reliance on p values, has many problems. There is little reason to persist with NHST now that Bayesian methods are accessible to everyone.

Software

We used R version 4.4.1 (R Core Team 2024) and the following R packages: BEST v. 0.5.4 (Kruschke and Meredith 2021), HDInterval v. 0.2.4 (Meredith and Kruschke 2022).

```
     Package Version    Citation
1       base   4.4.1       @base
2       BEST   0.5.4       @BEST
3 HDInterval   0.2.4 @HDInterval
```

Bibliography

Kruschke, John, and Mike Meredith. 2021. *BEST: Bayesian Estimation Supersedes the t-Test*. https://CRAN.R-project.org/package=BEST.

Meredith, Mike, and John Kruschke. 2022. *HDInterval: Highest (Posterior) Density Intervals*. https://CRAN.R-project.org/package=HDInterval.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.