Fisher, Neyman and Bayes: Part II

Philosophical excursion

André Meichtry

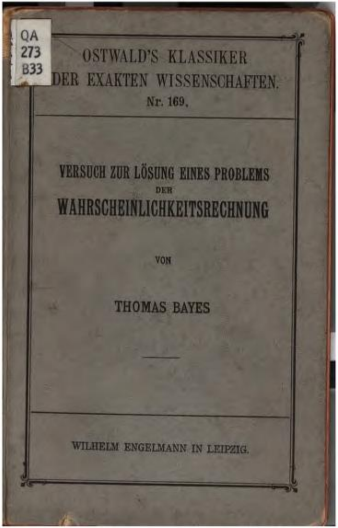
Departement of Health Professions
Bern University of Applied Sciences

May 2, 2023

Uncertainty

- Thinking about Uncertainty, Probability (17th century)
- Birth of probability as a mathematical discipline in 1654 (Pascal, Fermat)
- Gambling (calculus for long-run frequencies)
- Existence of God (calculus of beliefs)

Thomas Bayes, Pierre Simon Laplace



- Named after Reverend Thomas Bayes, an English part-time mathematician (1702-1761): "An essay in towards solving a problem in the doctrine of chances" (1763).
- With Thomas Bayes, Pierre Simon Laplace was the first to invert the probability statement and obtain probability statements about unknowns quantities, given observed quantities.
- Bayesian Statistics: Standard in the 18th/19th century
- 20th century: Classical/Frequentist Statistics (Fisher, NP)

Revival of Bayesian Statistics

- Foundations in the 1950s. Savage, Lindley and others...
- Bayesian statistics was percieved as
  - Foundationally/philosophically sound
  - But impractical due to computational limitations
- Modern Bayesian Statistics
  - Computational tools
  - Powerful simulation algorithms

Hume's problem of induction

- Can we learn about the future from the past?
- Can we learn from incomplete information?
- Conditional probability statements about unknown given known.
  - Unknown, not directly observable, parameters of a data-generating model.
  - Unknown and potentially observable, unobserved data, missing data, future data.
  - Known, observed data.

Parameters and data

Assume θ is some unknown quantity of interest, for example the true success rate of a new therapy.

Prior, data and posterior

- Let $p(\theta)$ denote the prior probability (density) distribution of θ, **Your** judgment about θ.
- Assume we have some evidence y, for example the results of a clinical trial, whose probability of occurrence depends on θ. This dependence is formalized by the likelihood $L(\theta) = p(y \mid \theta)$.
- We would like to obtain the posterior probability (density) distribution of θ, given the evidence, $p(\theta \mid y)$.

A theorem about probabilities

Theorem

Bayes' Theorem

$$p(\theta \mid y) = \frac{p(y \mid \theta) \times p(\theta)}{p(y)},$$

where $p(y)$ is the total probability of the data.[1]

- Follows directly from the axioms.[2]
- Is the basis for the whole apparatus of Bayesian statistics.
- How not fall in love with this, either from a philosophical or statistical perspective...

[1] normalizing factor, to ensure that the posterior integrates to 1.

[2] probabilities are numerical positive quantities, defined on a set of "outcomes" that are additive over mutually exclusive outcomes, and sum to 1 over all possible mutually exclusive outcomes.

Prior distribution

- The prior must be justified to a skeptical audience.
- Different priors can be used.
- Priors are explicitly and epistemologically relevant.
- Inappropriate to not use a prior (e.g. diagnostics, predictive values versus sensitivity/specificity).
- Often, if we have "no information", we use uninformative priors $\rightarrow$ automatic Bayes.

Inference and Decisions

- We use the posterior distribution of the quantity of interest to answer questions with unambiguous probability statements.
- Standard posterior summaries: mean, median, mode, standard deviation, quantile (e.g, for 95% credibility intervals, the $Q_{0.025}$ and $Q_{0.975}$).

Example questions about θ, given data

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

Table: Posterior distribution of θ has it all

Recap: Posterior distribution under different priors

			x=1	x=2	x=3	x=4
	θ		Likelihood: $p(x \mid \theta)$			
	θ_0		.980	.005	.005	.010
	θ_1		.098	.001	.001	.900
Prior odds	θ	Prior prob: $p(\theta)$	Posterior: $p(\theta \mid x)$			
1:1	θ_0	1/2	.91	.83	.83	.01
	θ_1	1/2	.09	.17	.17	.99
1:5	θ_0	1/6	.67	.50	.50	.002
	θ_1	5/6	.33	.50	.50	.998

Table: Posterior probabilities with different prior odds. Decision based on higher posterior probability. As example: $\Pr(\theta = \theta_0 \mid X = 1)$ with $\Pr(\theta_0) = \Pr(\theta_1) = 0.5$ is $\frac{\Pr(X=1|\theta=\theta_0)\Pr(\theta_0)}{\Pr(X=1)} = \frac{\Pr(X=1|\theta=\theta_0)\Pr(\theta_0)}{\Pr(X=1|\theta=\theta_0)\Pr(\theta_0)+\Pr(X=1|\theta=\theta_1)\Pr(\theta_1)} = \frac{0.98\times 0.5}{0.98\times 0.5+0.098\times 0.5} = 0.91.$

NHST Strawman Test: *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
```

```r
psych::describe(cbind(y1, y2, diff = y1 - y2))
```

```
##      vars n  mean     sd median trimmed    mad  min max range  skew kurtosis    se
## y1      1 8 130.0 116.00    120   130.0 140.85  -50 300   350 -0.13    -1.39 41.01
## y2      2 8   7.5 107.94      0     7.5 118.61 -120 190   310  0.29    -1.38 38.16
## diff    3 8 122.5  28.16    120   122.5  14.83   70 170   100 -0.19    -0.45  9.96
```

```r
t.test(y1, y2)
```

```
##
##  Welch Two Sample t-test
##
## data:  y1 and y2
## t = 2.19, df = 13.9, p-value = 0.046
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##    2.2877 242.7123
## sample estimates:
## mean of x mean of y
##     130.0       7.5
```

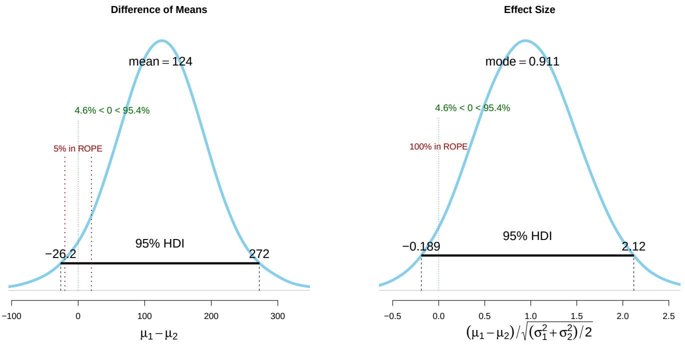
Bayesian estimation with Markov Chain Monte Carlo (MCMC) with uninformative priors

```
library(BEST)  ## 'Bayesian Estimation Supersedes the t Test'
mod <- BESTmcmc(y1, y2)
summary(mod, ROPEm = c(-20, 20))

##              mean  median    mode HDI%    HDIlo  HDIup compVal %>compVal ROPElow ROPEhigh %InROPE
## mu1        130.39 130.565 128.000   95   21.193 237.07
## mu2          6.36   6.314   9.350   95  -94.307 106.65
## muDiff     124.03 124.488 125.369   95  -25.175 269.40       0      95.4     -20       20    4.51
## sigma1     139.94 128.384 113.126   95   62.562 243.63
## sigma2     129.87 119.617 104.779   95   59.048 223.70
## sigmaDiff   10.06   8.329   7.993   95 -133.286 155.28       0      56.1
## nu          34.46  25.621  10.374   95    1.148  94.10
## log10nu      1.38   1.409   1.487   95    0.585   2.09
## effSz        0.96   0.952   0.911   95   -0.170   2.10       0      95.4
```

Bayesian estimation with Markov Chain Monte Carlo (MCMC) with uninformative priors

```
plot(mod, compVal = 0, ROPE = c(-20, +20), showCurve = TRUE)
plot(mod, compVal = 0, ROPE = c(-20, +20), which = "effect", showCurve = TRUE)
```



Binomial data: estimation of a success rate

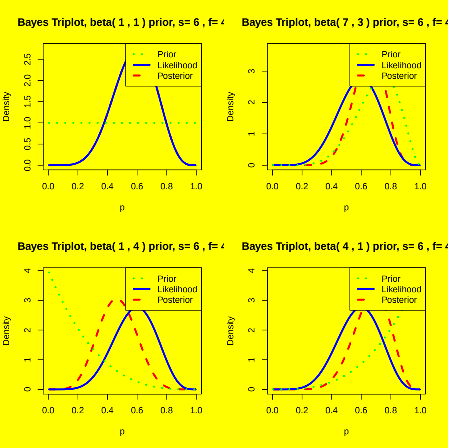


Figure: Prior (green), likelihood (blue) and posterior distributions (red) for success probability π for 6 successes and 4 failures, with different priors.

Moving from NHST to Bayesian estimation

An open letter from [Kru10]

- 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.
- 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.

Moving from NHST to Bayesian estimation

- Give arguments to researchers, reviewers, colleagues, editors, etc. still wanting *p*-values.
- The less we use the word “significant”, the better.
- If you test, do not attack the Strawman.
- If possible, estimate the quantity of interest.
- If possible, consider Bayesian estimation.

Perspective

At any rate – even if you are a Frequentist – try to view the world with Bayesian eyeglasses, as most people out there – probably – do.

Thank you

"There's no theorem like Bayes' theorem
Like no theorem we know
Everything about it is appealing
Everything about it is a wow
Let out all that a priori feeling
You've been concealing right up to now!"

— *George Box; Music: Irving Berlin*

Software

All analyses were performed using the R statistical software R version 4.3.0 (2023-04-21) [R C22].
Session Info

- R version 4.3.0 (2023-04-21), `x86_64-pc-linux-gnu`
- Running under: `Ubuntu 22.04.2 LTS`
- Matrix products: default
- BLAS: `/usr/lib/x86_64-linux-gnu/blas/libblas.so.3.10.0`
- LAPACK: `/usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0`
- Base packages: base, datasets, graphics, grDevices, methods, stats, utils
- Other packages: BayesCourse 0.7, BEST 0.5.4, coda 0.19-4, HDInterval 0.2.2, knitr 1.37, LearnBayes 2.15.1, MASS 7.3-53.1, MCMCpack 1.5-0, psych 2.2.9, xtable 1.8-4
- Loaded via a namespace (and not attached): compiler 4.3.0, conquer 1.0.2, evaluate 0.20, formatR 1.9, grid 4.3.0, highr 0.9, lattice 0.20-44, magrittr 2.0.3, Matrix 1.3-2, MatrixModels 0.5-0, matrixStats 0.63.0, mcmc 0.9-7, mnormt 2.1.1, nlme 3.1-152, parallel 4.3.0, quantreg 5.85, Rcpp 1.0.10, rjags 4-13, SparseM 1.81, stringi 1.7.8, stringr 1.4.0, tools 4.3.0, xfun 0.30

Bibliographie

[Chr05] Ronald Christensen. Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, 59(2):121–126, 2005.

[GCS+13] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis, Third Edition (Chapman and Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, 3 edition, 11 2013.

[KM15] John K. Kruschke and Mike Meredith. *BEST: Bayesian Estimation Supersedes the t-Test*, 2015. R package version 0.3.0.

[Kru10] John Kruschke. An open letter. `http://www.indiana.edu/~kruschke/AnOpenLetter.htm`, 2010. Accessed: 2015-09-30.

[Kru14] John Kruschke. *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*. Academic Press, 2 edition, 11 2014.

[R C22] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022.

[SAM04] David J. Spiegelhalter, Keith R. Abrams, and Jonathan P. Myles. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, 1 edition, 1 2004.

[ZM08] Stephen T. Ziliak and Deirdre N. McCloskey. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives (Economics, Cognition, and Society)*. University of Michigan Press, 1st edition, 2 2008.