Lecture 23

05/14/2020 [MATH341]

• Bayes Factors (B)

$$B = \frac{P_{H_a}(x)}{P_{H_0}(x)} = \frac{\int_{\Theta_a} P(x|\theta)P(\theta)\,d\theta}{\int_{\Theta_0} P(x|\theta)P(\theta)\,d\theta} = \frac{\int_0^1 \binom{100}{61}\theta^{61}(1-\theta)^{39}(1)\,d\theta}{\int_{\theta\in\{0.5\}} \binom{100}{61}\theta^{61}(1-\theta)^{39}(1)\,d\theta}$$

want to prove

$H_0: \theta = 0.5$

$H_a: \theta \neq 0.5$

$\alpha = 5\%$

$P(\theta) = U(0,1)$

$$= \frac{\int_0^1 \theta^{61+1-1}(1-\theta)^{39+1-1}\,d\theta}{0.5^{61}(1-0.5)^{39} = 0.5^{100}}$$ ← Beta function.

$$= \frac{B(62, 40)}{0.5^{100}}$$

$$= 1.39 > 1$$

Jeffrey's (1961) scale for evidence of Ha:

B < 1 ⇒ no evidence for Ha.

B between 1, 3 ⇒ barely worth mentioning.

B between 3, 10 ⇒ substantial

B between 10, 30 ⇒ strong.

B between 30, 100 ⇒ very strong.

B > 100 ⇒ absolutely decisive

No concept of p-values and no concept of reject and fail to reject. Bayes factors AKA "ratio of evidences" are just a different way of thinking about this whole concept.

famous

ESP data $n = 104{,}490{,}000$ and $x = 52{,}263{,}471 \Rightarrow \hat{\theta} = \bar{x} = 0.500018$

$H_0: \theta = 0.5$ · Frequentist test rejects with pvalue $= 0.0003 < \alpha = 5\%$.

$H_a: \theta \neq 0.5$ · Bayes Factor calculation. Let $\theta \sim U(0,1)$

$\alpha = 5\%$

$$B = \frac{B(52{,}263{,}471,\ 52{,}226{,}529)}{0.5^{104{,}490{,}000}}$$

$$\Rightarrow \ln(B) = \log B(\cdot,\cdot) - 104{,}490{,}000\,\ln(0.5)$$

$$= -1.0899$$

$\Rightarrow B \approx e^{-1} \approx 0.37$

$\Rightarrow$ no evidence for Ha.

There is an idea of "statistical significance" assessed with the frequentist p-value, the Bayesian p-val and the Bayes Factor B. But the other important idea is "clinical significance" which measures how important the actual effect theta is. If n is very large, you always get statistical significance, because theta under Ho is never truly real.

The Bayes Factor kind of puts together both significance. Parenthetically, the Bayesian 2-sides hypothesis test with margin of equivalence does the same thing.

$H_0 : \theta \in [\theta_0 \pm \delta]$   $\delta$ forces you to think about clinical significance.

$H_a : \theta \notin [\theta_0 \pm \delta]$   · High delta $\Rightarrow$ you only care about large effects.

$\alpha$ is chosen

$P(\theta)$ is chosen

// Course over.

Gibbs sampler sometimes $P(\theta_j | \theta_{-j})$ is unknown and you only have its kernel. And (1) you can't solve for the constant c.

(2) Gibbs sampling too slow.

Idea: Draw $\theta_{t,j}$ from distribution $g(\theta_{t-1,j}, \phi)$ e.g. $N(\theta_{t-1,j}, 1^2)$

$\uparrow$ previous sample   $\nwarrow$ hyperparameter

g is called a "proportional distr."

We don't always accept $\theta_{t,j}$. We only accept it if:

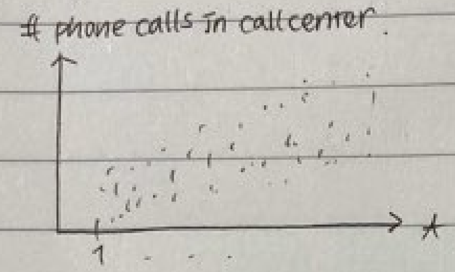
$$r = \frac{P(\theta_{1,t}, \ldots, \theta_{t,j}, \ldots \theta_p | X)}{P(\theta_{1,t}, \ldots, \theta_{t-1,j}, \ldots \theta_p | X)} \geq u,$$ where u is a realization

from $U(0,1) \Rightarrow$ Accept.

Assume $q(\theta_{t,j} \mid \theta_{t-1,j}, \theta) = q(\theta_{t-1,j} \mid \theta_{t,j}, \theta)$ (Metropolis et al., 1953)

If not symmetric, then it easy to fix by multiplying $r$ by the transition ratio ___ $\Rightarrow$ metropolis-H — Sampler (M.H)

within a Gibbs Sampler, it is called "Metropolis - within Gibbs".




Good Model $X_t \sim \text{Poisson}(a+bt)$

$\vec{\theta} = \begin{bmatrix} a \\ b \end{bmatrix}$, $P(a,b) = P(a)P(b) \propto 1$

$$P(a,b \mid X_1, \ldots, X_T) \propto P(X_1, \ldots, X_T \mid a,b) \cdot P(a,b).$$

$$= \left( \prod_{t=1}^{T} \frac{e^{-(a+bt)} (a+bt)^{x_t}}{x_t!} \right) P(a,b).$$

$$\propto e^{-n(a+b\bar{t})} \prod_{t=1}^{T} (a+bt)^{x_t}$$ obviously not a kernel we know.

Try to build Gibbs Sampler.

$P(a \mid -) \propto e^{-na} \prod_{t=1}^{T} (a+bt)^{x_t}$

$P(b \mid -) \propto e^{-n\bar{t}b} \prod_{t=1}^{T} (a+bt)^{x_t}$

Both these conditional kernels unknown. $\Rightarrow$ MH steps with

$q_a = N(a_{t-1}, 1^2)$, $q_b = N(b_{t-1}, 1^2)$.

since these are symmetric we employ the vanilla Metropolis algorithm at each sampling step.