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

Lebesgue Integral

$H_0: \theta = 0.5$

$H_a: \theta \neq 0.5$

$\alpha = 5\%$

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

$$= \frac{\int_0^1 \theta^{(61+1)-1}(1-\theta)^{(39+1)-1}\,d\theta}{\underbrace{0.5^{61}(1-0.5)^{39}}_{0.5^{100}}} = \frac{B(62,40)}{0.5^{100}} = 1.39 > 1$$

Jeffrey's (1961) scale for evidence of H_a:
B < 1 => no evidence for H_a
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 $x = 52{,}263{,}471 \Rightarrow \hat{\theta} = \bar{x} = 0.500018 \approx 0.5$

$H_0: \theta = 0.5$

$H_a: \theta \neq 0.5$

$\alpha = 5\%$

Frequentist test rejects with $p_{val} = 0.0003 < \alpha = 5\%$.

Bayes Factor calculation. Let $\theta \sim U(0,1)$.

$$B = \frac{B(52{,}263{,}472,\ 52{,}226{,}031)}{0.5^{104{,}490{,}000}} \Rightarrow \ln(B) = \log B(\cdot,\cdot) - 104{,}490{,}000\ln(0.5) = -1.0897$$

$\Rightarrow B \approx e^{-1} \approx 0.33 \Rightarrow$ No evidence for H_a.

There is an idea of "statistical significance" assessed with the frequentist p-val, 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 H_0 is never truly real.

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

$H_0: \theta \in [\theta_0 \pm \delta]$

$H_a: \theta \notin [\theta_0 \pm \delta]$

δ forces you to think about clinical significance. High delta => you only care about large effects.

α 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) Grid sampling too slow. Idea:

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

$\theta_{t-1,j}$: previous sample; ϕ: hyperparameters

q is called a "proposal 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.

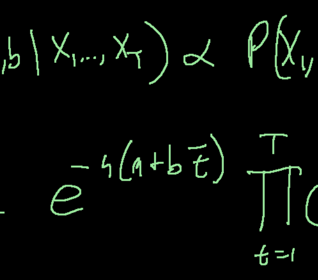
assume $q(\theta_{t,j} | \theta_{t-1,j}, \phi) = q(\theta_{t-1,j} | \theta_{t,j}, \phi)$

i.e. symmetrical transitions (Metropolis et al., 1953)

If not symmetric, then it's easy to fix by multiplying r by the transition ratio (Hastings, 1974). $\Rightarrow$ Metropolis-Hastings Sampler. (MH)

Within a Gibbs sampler, it is called "Metropolis-within Gibbs".

\# phone calls in call center



Good model $X_t \overset{ind}{\sim} \text{Poisson}(a + bt)$

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

$$P(a,b | X_1, \ldots, X_T) \propto P(X_1, \ldots, X_T | a,b)\,P(a,b) = \left(\prod_{t=1}^{T} \frac{e^{-(a+bt)}(a+bt)^{x_t}}{x_t!}\right) P(a,b)$$

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

Try to build Gibbs sampler

$$P(a | \text{—}) \propto e^{-Ta} \prod_{t=1}^{T} (a+bt)^{x_t} \quad \Big| \quad P(b | \text{—}) \propto e^{-T\bar{t}b} \prod_{t=1}^{T} (a+bt)^{x_t}$$

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

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

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