

Handout 13: Hypothesis tests

Lecturer & author: Georgios P. Karagiannis

georgios.karagiannis@durham.ac.uk

Aim: To explain, design, and use hypothesis tests in the Bayesian framework

References:

- Berger, J. O. (2013; Section 4.3.3). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.
- Robert, C. (2007; Section 5.2(exclude 5.2.6)). The Bayesian choice: from decision-theoretic foundations to computational implementation. Springer Science & Business Media.
- DeGroot, M. H. (2005, Sections 11.5-11.13). Optimal statistical decisions (Vol. 82). John Wiley & Sons

1 Set-up of a hypothesis test

Aim: Let $y = (y_1, \dots, y_n)$ generated from the real unknown data-generating process $y \sim R(y)$. Statistician approximates/parametrizes $R(\cdot)$ by a statistical model $F(y|\theta)$ with unknown $\theta \in \Theta$. Then you wish to find useful statements about θ : E.g. is there a smaller $\Theta_* \subseteq \Theta$ where You can restrict the possible values of unknown θ ?

Notation 1. Let $y = (y_1, \dots, y_n)$ be a sequence of observables modeled to have been generated from the sampling distribution $F(y|\theta)$ labeled by an unknown parameter $\theta \in \Theta$ following a priori distribution $\Pi(\theta)$; namely

$$\begin{cases} y|\theta & \sim F(y|\theta) \\ \theta & \sim \Pi(\theta) \end{cases} \quad (1)$$

Assume there is interest to test/compare the following hypotheses/statements

$$H_0 : \theta \in \Theta_0; \text{ vs. } H_1 : \theta \in \Theta_1 \quad (2)$$

where $\Theta = \Theta_0 \cup \Theta_1$, under the Bayesian model (1).

Note 2. The pair of hypotheses (2) partitions the overall prior $\Pi(\theta)$ (representing overall prior believes about θ) as

$$d\Pi(\theta) = \pi_0 \times d\Pi_0(\theta) + \pi_1 \times d\Pi_1(\theta) \quad (3)$$

where π_0 , and π_1 describe the prior probabilities of hypotheses H_0 and H_1

$$\pi_0 = \underbrace{P_{\Pi}(\theta \in \Theta_0)}_{=P_{\Pi}(H_0)} = \int 1(\theta \in \Theta_0) d\Pi(\theta), \quad \pi_1 = \underbrace{P_{\Pi}(\theta \in \Theta_1)}_{=P_{\Pi}(H_1)} = \int 1(\theta \in \Theta_1) d\Pi(\theta),$$

respectively while $\Pi_0(\theta) := \Pi(\theta|\theta \in \Theta_0)$ and $\Pi_1(\theta) := \Pi(\theta|\theta \in \Theta_1)$ are prior distributions with pdf/pmf

$$\pi_0(\theta) := \underbrace{\pi(\theta|\theta \in \Theta_0)}_{=\pi(\theta|H_0)} = \frac{\pi(\theta)1(\theta \in \Theta_0)}{\int 1(\theta \in \Theta_0) d\Pi(\theta)}; \quad \text{and} \quad \pi_1(\theta) := \underbrace{\pi(\theta|\theta \in \Theta_1)}_{=\pi(\theta|H_1)} = \frac{\pi(\theta)1(\theta \in \Theta_1)}{\int 1(\theta \in \Theta_1) d\Pi(\theta)},$$

describing how the prior mass of θ is spread out over the hypotheses H_0 and H_1 respectively. Then the Bayesian hypothesis test can also be expressed as

$$H_0 : \begin{cases} y|\theta & \sim F(y|\theta) \\ \theta & \sim \Pi_0(\theta), \theta \in \Theta_0 \end{cases} \quad \text{vs} \quad H_1 : \begin{cases} y|\theta & \sim F(y|\theta) \\ \theta & \sim \Pi_1(\theta), \theta \in \Theta_1 \end{cases} \quad (4)$$

with prior $\pi_0 = P_{\Pi}(\theta \in \Theta_0)$ and $\pi_1 = P_{\Pi}(\theta \in \Theta_1)$.

Question 3. Which Bayesian model (H_0 or H_1) describes ‘better’ the real data generating process?

Note 4. In Bayesian framework, hypothesis testing is rather straightforward. All You need to do is to calculate the corresponding posterior probabilities $P_{\Pi}(\theta \in \Theta_0|y)$, and $P_{\Pi}(\theta \in \Theta_1|y)$, and decide between H_0 and H_1 .

2 Decision theory perspective

Note 5. Bayes hypothesis test (4) can be addressed as a Bayesian statistical decision problem with decision space $\mathcal{D} = \{\text{accept } H_0, \text{accept } H_1\}$ or simpler $\mathcal{D} = \{0, 1\}$, and under Bayesian model (1). It can be seen as a parametric point estimation about the indicator function

$$1_{\Theta_1}(\theta) = \begin{cases} 0 & , \theta \in \Theta_0 \\ 1 & , \theta \in \Theta_1 \end{cases} \quad (5)$$

under Bayesian model (1), prior (3), and a loss function $\ell(\theta, \delta)$, with $\theta \in \Theta$, $\delta \in \mathcal{D}$; E.g., the 0 – 1 loss function.

Theorem 6. The Bayes estimator of $1_{\Theta_1}(\theta)$ in (5), under the prior $\Pi(\theta)$ in (3) and the $c_I - c_{II}$ loss function

$$\ell(\theta, \delta) = \begin{cases} 0 & , \text{if } \theta \in \Theta_0, \delta = 0 \\ 0 & , \text{if } \theta \notin \Theta_0, \delta = 1 \\ c_{II} & , \text{if } \theta \notin \Theta_0, \delta = 0 \\ c_I & , \text{if } \theta \in \Theta_0, \delta = 1 \end{cases} \quad (6)$$

where $c_I > 0$ and $c_{II} > 0$ are specified by the researcher is

$$\delta(y) = \begin{cases} 0 & , P_{\Pi}(\theta \in \Theta_0|y) > \frac{c_{II}}{c_{II}+c_I} \\ 1 & , \text{otherwise} \end{cases} \quad (7)$$

where $\{\Theta_0, \Theta_1\}$ constitute a partition for Θ , and $P_{\Pi}(\theta \in \Theta_0|y) = \int 1(\theta \in \Theta_0) d\Pi(\theta|y)$.

Proof. The posterior expected loss is¹

$$\begin{aligned} \varrho(\pi, \delta|y) &= E_{\Pi}(\ell(\theta, \delta)|y) = \int \ell(\theta, \delta) d\Pi(\theta|y) = \int_{\Theta_0} \ell(\theta, \delta) d\Pi(\theta|y) + \int_{\Theta_1} \ell(\theta, \delta) d\Pi(\theta|y) \\ &= \begin{cases} \int_{\Theta_0} 0 d\Pi(\theta|y) + \int_{\Theta_1} c_{II} d\Pi(\theta|y) & , \text{if } \delta = 0 \\ \int_{\Theta_0} c_I d\Pi(\theta|y) + \int_{\Theta_1} 0 d\Pi(\theta|y) & , \text{if } \delta = 1 \end{cases} = \begin{cases} c_{II} \int 1(\theta \in \Theta_1) d\Pi(\theta|y) & , \text{if } \delta = 0 \\ c_I \int 1(\theta \in \Theta_0) d\Pi(\theta|y) & , \text{if } \delta = 1 \end{cases} \\ &= c_{II} P_{\Pi}(\theta \notin \Theta_0|y) + c_I P_{\Pi}(\theta \in \Theta_0|y) \end{aligned}$$

¹Notation: $\int_{\Theta_j} d\Pi(\theta|y) = \int 1(\theta \in \Theta_j) d\Pi(\theta|y)$

The Bayes rule (estimator) of (5) is $\delta(y) = 0$ when

$$\varrho(\pi, \delta = 0|y) < \varrho(\pi, \delta = 1|y) \iff c_{\Pi} P_{\Pi}(\theta \notin \Theta_0|y) < c_I P_{\Pi}(\theta \in \Theta_0|y) \iff P_{\Pi}(\theta \in \Theta_0|y) > \frac{c_{\Pi}}{c_{\Pi} + c_I}$$

The Bayes rule (estimator) of (5) is $\delta(y) = 1$ when

$$\varrho(\pi, \delta = 0|y) > \varrho(\pi, \delta = 1|y) \iff c_{\Pi} P_{\Pi}(\theta \notin \Theta_0|y) > c_I P_{\Pi}(\theta \in \Theta_0|y) \iff P_{\Pi}(\theta \in \Theta_0|y) < \frac{c_{\Pi}}{c_{\Pi} + c_I}$$

So $\varrho(\pi, \delta|y)$ is minimised for (7). \square

3 Bayes factors perspective

Note 7. Hypothesis tests in Bayesian statistics can be addressed by using Bayes factors.

Definition 8. The Bayes factor $B_{01}(y)$ is the ratio of the posterior probabilities of H_0 and H_1 over the ratio of the prior probabilities of H_0 and H_1 .

$$B_{01}(y) = \frac{P_{\Pi}(\theta \in \Theta_0|y) / P_{\Pi}(\theta \in \Theta_0)}{P_{\Pi}(\theta \in \Theta_1|y) / P_{\Pi}(\theta \in \Theta_1)} \quad (8)$$

where

$$P_{\Pi}(\theta \in \Theta_j) = \int 1(\theta \in \Theta_j) d\Pi(\theta); \quad \text{and} \quad P_{\Pi}(\theta \in \Theta_j|y) = \int 1(\theta \in \Theta_j) d\Pi(\theta|y); \quad \text{for } j = 0, 1.$$

Proposition 9. For Hypothesis pair (2) and Bayes model (1), where the prior is formed as in (3), the Bayes factor in (8) can be written as

$$B_{01}(y) = \frac{\int f(y|\theta) d\Pi_0(\theta)}{\int f(y|\theta) d\Pi_1(\theta)} = \frac{f_0(y)}{f_1(y)}$$

where $f_j(y) = \int_{\Theta_j} f(y|\theta) d\Pi_j(\theta)$ is the conditional marginal likelihood (or prior predictive pdf/pmf) given H_j , for $j = 0, 1$.

Proof. It results by showing that for $j = 0, 1$, it is

$$\begin{aligned} P_{\Pi}(\theta \in \Theta_j|y) &= \int 1(\theta \in \Theta_j) d\Pi(\theta|y) = \int 1(\theta \in \Theta_j) \frac{f(y|\theta) d\Pi(\theta)}{\int_{\Theta} f(y|\theta) d\Pi(\theta)} \\ &= \int 1(\theta \in \Theta_j) \frac{f(y|\theta) (\pi_0 \times d\Pi_0(\theta) + \pi_1 \times d\Pi_1(\theta))}{\underbrace{\int_{\Theta} f(y|\theta) d\Pi(\theta)}_{=f(y)}} \\ &= \frac{\pi_0}{f(y)} \int 1(\theta \in \Theta_j) f(y|\theta) d\Pi_0(\theta) + \frac{\pi_1}{f(y)} \int 1(\theta \in \Theta_j) f(y|\theta) d\Pi_1(\theta) = \begin{cases} \frac{\pi_0}{f(y)} \int_{\Theta_0} f(y|\theta) d\Pi_0(\theta) & , \text{if } j = 0 \\ \frac{\pi_1}{f(y)} \int_{\Theta_1} f(y|\theta) d\Pi_1(\theta) & , \text{if } j = 1 \end{cases} \end{aligned}$$

\square

Remark 10. Obviously, $B_{10}(y) = 1/B_{01}(y)$

Remark 11. Bayes factor $B_{01}(y)$:

- ...is the ‘odds in favour of H_0 against H_1 that are given by the data’ y .
- ...evaluate the modification of the odds of Θ_0 against Θ_1 due to the observations y .
- ...is the ratio of the likelihoods, weighted by the conditional priors $d\Pi_0(\theta)$ and $d\Pi_1(\theta)$.

Proposition 12. One can write

$$P_{\Pi}(\theta \in \Theta_0|y) = \left[1 + \frac{1 - P_{\Pi}(\theta \in \Theta_0)}{P_{\Pi}(\theta \in \Theta_0)} B_{01}(y)^{-1} \right]^{-1} = \left[1 + \frac{\pi_1}{\pi_0} B_{01}(y)^{-1} \right]^{-1}$$

where $\pi_j = P_{\Pi}(\theta \in \Theta_j)$, for $j = 0, 1$, by rearranging (8) (please check).

Criterion 13. Consider a hypothesis test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$ as described in (4) with loss function (6), and given a Bayesian model (1). The hypothesis H_0 is accepted when

$$B_{01}(y) > \frac{c_{II} \pi_1}{c_I \pi_0} \quad (9)$$

where $\pi_j = P_{\Pi}(\theta \in \Theta_j)$, for $j = 0, 1$.

Proof. Straightforward result from Definition 8 and Theorem 6. □

Remark 14. Eq. 9 shows the duality between loss function and the prior distribution. Different combinations of priors and loss functions may lead to the same result. For instance, for $c'_{II} = c'_I = 1$, $\pi'_0 = \frac{c_I \pi_0}{c_I \pi_0 + c_{II} \pi_1}$, and $\pi'_1 = \frac{c_{II} \pi_1}{c_I \pi_0 + c_{II} \pi_1}$, we get again (9) !!!

Criterion 15. Jeffreys developed a scale to judge the strength of evidence in favor of H_0 or against H_0 brought by the data, outside a true decision-theoretic setting (aka; without the need to specify c_I and c_{II} in (9)).

| B_{01} | $\log_{10}(B_{01})$ | Strength of evidence |
|------------------------|---------------------|----------------------------------------------------------------------|
| $(1, +\infty)$ | $(0, +\infty)$ | H_0 is supported |
| $(10^{-1/2}, 1)$ | $(-1/2, 0)$ | Evidence against H_0: not worth more than a bare |
| $(10^{-1}, 10^{-1/2})$ | $(-1, -1/2)$ | Evidence against H_0: substantial |
| $(10^{-3/2}, 10^{-1})$ | $(-3/2, -1)$ | Evidence against H_0: strong |
| $(10^{-2}, 10^{-3/2})$ | $(-2, -3/2)$ | Evidence against H_0: very strong |
| $(0, 10^{-2})$ | $(-\infty, -2)$ | Evidence against H_0: decisive |

The precise bounds separating one strength from another are a matter of convention. Note that similar criticism exists in frequentist hypothesis tests with the choice of the significance level $\alpha = \{0.01, 0.05, 0.1, \dots\}$.

4 Special cases in hypotheses tests

Definition 16. Traditionally, hypotheses, H_j , are categorized as:

- **Single (or point) hypothesis** for θ is called the hypothesis $H_j : \theta \in \Theta_j$ where $\Theta_j = \{\theta_j\}$ contains a single element, namely when $\Pi_j(\theta)$ assigns probability one to a specific value for θ .
- **Composite hypothesis** for θ is called the hypothesis $H_j : \theta \in \Theta_j$ where $\Theta_j \subseteq \Theta$ contains many elements. Namely when $\Pi_j(\theta)$ defines a non-degenerate density $\pi_j(\theta)$ over $\Theta_j \subseteq \Theta$.
- **General alternative hypothesis** for θ is called the composite hypothesis $H_1 : \theta \in \Theta_1$ where $\Theta_1 = \Theta - \{\theta_0\}$ and θ_0 a single value. It is often denoted as $H_1 : \theta \neq \theta_0$ and compared against a single null hypothesis $H_0 : \theta = \theta_0$.

We present some special cases of hypothesis tests.

Case 1. **Composite vs Composite** is the hypothesis test:

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

where both $\Theta_0 \subseteq \Theta$ and $\Theta_1 \subseteq \Theta$ contain more than one elements. Overall prior can be partitioned as

$$d\Pi(\theta) = \pi_0 \times d\Pi_0(\theta) + \pi_1 \times d\Pi_1(\theta)$$

Then the conditional marginal likelihoods are

$$f_0(y) = \int f(y|\theta) d\Pi_0(\theta); \quad f_1(y) = \int f(y|\theta) d\Pi_1(\theta).$$

Example. A composite vs. composite hypothesis is:

$$H_0 : \begin{cases} y_i | \mu, \sigma^2 & \overset{\text{iid}}{\sim} N(\mu, \sigma^2), i = 1, \dots, n \\ \mu | \sigma^2 & \sim N(\mu_0, \sigma^2 \frac{1}{\lambda_0}) \\ \sigma^2 & \sim \text{Ga}(a_0, b_0) \end{cases} \quad \text{vs} \quad H_1 : \begin{cases} y_i | \mu & \overset{\text{iid}}{\sim} T(\mu, 1, k_0), i = 1, \dots, n \\ \mu & \sim N(\xi_0, v_0) \end{cases}$$

In H_0 : I consider a sampling model $y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with prior $(\mu, \sigma^2) \sim N(\mu_0, \sigma^2/\lambda_0) \text{IG}(a_0, b_0)$, and $\Theta_0 = \{N\} \cup \mathbb{R} \cup (0, \infty)$. Here $\mu_0, \lambda_0, a_0, b_0$ are fixed.

In H_1 : I consider a sampling model $y_i \overset{\text{iid}}{\sim} T(\mu, 1, k_0)$ with prior $(\mu, \sigma^2) \sim N(\mu_0, \sigma^2/\lambda_0) \text{IG}(a_0, b_0)$, and $\Theta_1 = \{T\} \cup \mathbb{R}$. Here μ_0, k_0, ξ_0, v_0 are fixed.

Case 2. **Single vs. General alternative** is the pair of hypotheses

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0.$$

If θ is continuous, the difficulty is that we cannot use a continuous prior for $\Pi_0(\theta)$ to conduct a test with point null hypothesis $H_0 : \theta = \theta_0$ because it would give a prior probability zero for $\theta = \theta_0$. To overcome this, we specify the conditional distribution $\Pi_0(\theta)$ as a Dirac² prior distribution with concentration point at θ_0 ; namely $d\Pi_0(\theta) = d\delta_{\theta}(\theta_0)$. The conditional distribution $\Pi_1(\theta)$ can be any reasonable distribution $d\Pi_1(\theta) = d\Pi(\theta|\theta \in \Theta_1)$. Then the overall prior

$$d\Pi(\theta) = \pi_0 \times d\delta_{\theta}(\theta_0) + \pi_1 \times d\Pi_1(\theta) \quad (10)$$

and it is called spike-and-slab. Then the conditional marginal likelihoods are

$$f_0(y) = \int f(y|\theta) d\Pi_0(\theta) = \int f(y|\theta) d\delta_{\theta_0}(\theta) = f(y|\theta_0)$$

$$f_1(y) = \int f(y|\theta) d\Pi_1(\theta) = \begin{cases} \int f(y|\theta) \pi(\theta) d\theta & \text{cont.} \\ \sum_{\forall \theta \neq \theta_0} f(y|\theta) \pi_1(\theta) & \text{discr.} \end{cases}$$

Example. The standard two side test $H_0 : \mu = \theta_0$ vs. $H_1 : \mu \neq \theta_0$, where the sampling distribution is assumed to be $y_i \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$ with known variance σ^2 for $i = 1, \dots, n$, is a simple vs. general alternative hypothesis test and can also be formulated as:

$$H_0 : y_i | \theta_0, \sigma_0^2 \overset{\text{iid}}{\sim} N(\theta_0, \sigma_0^2), i = 1, \dots, n \quad \text{vs} \quad H_1 : \begin{cases} y_i | \mu, \sigma^2 \overset{\text{iid}}{\sim} N(\mu, \sigma_0^2), i = 1, \dots, n \\ \mu \sim N(\mu_0, \sigma_0^2) \end{cases}$$

Here it is $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta \in \mathbb{R} : \theta \neq \theta_0\}$, while $\theta_0, \mu_0, \sigma_0^2$ are fixed values.

²Dirac measure $d\delta_x(A)$ of random variable x at A concentrates all the mass at a set A such as $\delta_x(A) = \begin{cases} 1 & , x \in A \\ 0 & , x \notin A \end{cases}$. If

$A = \{a\}$, we just denote the measure as $d\delta_x(a)$ with mass $\delta_x(a) = \begin{cases} 1 & , x = a \\ 0 & , x \neq a \end{cases}$.

Case 3. **Single vs. Single** is the pair of hypothesis

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta = \theta_1$$

where $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$ for some values of θ_0 and θ_1 . The hypotheses H_0 and H_1 are single, and hence the corresponding priors can be considered as having a point mass around θ_0 and θ_1 . Mathematically, we assign Dirac prior distributions $d\Pi_0(\theta) = d\delta_\theta(\theta_0)$ and $d\Pi_1(\theta) = d\delta_\theta(\theta_1)$, which imply

$$d\Pi(\theta) = \pi_0 \times d\delta_\theta(\theta_0) + \pi_1 \times d\delta_\theta(\theta_1)$$

Then the conditional marginal likelihoods are

$$\begin{aligned} f_0(y) &= \int f(y|\theta) d\Pi_0(\theta) = \int f(y|\theta) d\delta_\theta(\theta_0) = f(y|\theta_0) \\ f_1(y) &= \int f(y|\theta) d\Pi_1(\theta) = \int f(y|\theta) d\delta_\theta(\theta_1) = f(y|\theta_1) \end{aligned}$$

Note 17. In this case, Bayes factor is the likelihood ratio of H_0 against H_1 which most statisticians (whether Bayesian or not) view as the odds in favor of H_0 against H_1 that are given by the data.

Example. Given the statistical model $y_i \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ the comparison $H_0 : \mu = \theta_0$ vs. $H_1 : \mu = \theta_1$, is a simple vs. simple hypothesis, where $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$ are sets with a single elements $\theta_0 \neq \theta_1$.

Example. The model comparison

$$H_0 : y_i|\phi_0 \stackrel{\text{iid}}{\sim} \text{Nb}(\phi_0, 1) \text{ vs. } H_1 : y_i|\lambda_0 \stackrel{\text{iid}}{\sim} \text{Pn}(\lambda_0)$$

where $\phi_0 > 0$, $\lambda_0 > 0$ are known, is a simple vs. simple hypothesis. Here it is $\Theta_0 = \{\text{Nb}\}$, $\Theta_1 = \{\text{Pn}\}$.

Example 18. Let $y = (y_1, \dots, y_n)$ a sequence of observables, and assume that $n = 5$, and $y_* = \sum_{i=1}^5 y_i = 3$. Assume a sampling distribution $y_i|\theta \stackrel{\text{iid}}{\sim} \text{Br}(\theta)$, with unknown parameter $\theta \in [0, 1]$, a priori following a uniform distribution.

1. By using Jeffreys' scaling rule, perform the following hypothesis test for $\theta_0 = 1/2$

$$H_0 : \theta = \theta_0 \quad \text{vs} \quad H_1 : \theta \neq \theta_0$$

2. Compute the posterior probability of the NULL hypothesis.

Solution. This is a simple vs. general alternative hypothesis. I specify the overall prior with pdf

$$\pi(\theta) = \pi_0 1(\theta = \theta_0) + (1 - \pi_0) U(\theta|0, 1)$$

for some $\pi_0 > 0$. I leave π_0 abstract, however the usual choice (but maybe not the best) is $\pi_0 = 1/2$.

1. The Bayes factor is

$$B_{01}(y) = \frac{\prod_{i=1}^n \text{Br}(y_i|\theta_0)}{\int_{(0,1)} \prod_{i=1}^n \text{Br}(y_i|\theta) U(\theta|0, 1) d\theta} = \frac{\theta_0^{y_*} (1 - \theta_0)^{n-y_*}}{\int_{(0,1)} \theta^{y_*} (1 - \theta)^{n-y_*} d\theta} = \frac{\theta_0^{y_*} (1 - \theta_0)^{n-y_*}}{B(y_* + 1, n - y_* + 1)} = \frac{(1/2)^5}{B(4, 3)} = \frac{15}{8}$$

Then $B_{01}(y) = \frac{15}{8} \approx 2$, and $\log_{10}(B_{01}(y)) \approx 0.27$. According to Jeffreys' scaling rule, H_0 is supported. We can accept the null hypothesis.

2. The posterior probability of H_0 is

$$P_\Pi(\theta = \theta_0|y) = P_\Pi(\theta \in \Theta_0|y) = \left[1 + \frac{1 - \pi_0}{\pi_0} B_{01}(y)^{-1} \right]^{-1} = \left[1 + \frac{1/2}{1 - 1/2} \left(\frac{15}{8} \right)^{-1} \right]^{-1} = \frac{15}{23} \approx 0.65$$

and hence the posterior distribution tends to support H_0 .

Example 19. Let $y = (y_1, \dots, y_n)$ a sequence of observables. There is interest in performing the following hypothesis test

$$H_0 : \begin{cases} y_i | \phi \sim \text{Nb}(1, \phi); & \phi > 0 \\ \phi \sim \text{Be}(a_0, b_0); & a_0 = 2, b_0 = 2 \end{cases} \quad \text{vs} \quad H_1 : \begin{cases} y_i | \lambda \sim \text{Pn}(\lambda); & \lambda > 0 \\ \lambda \sim \text{Ga}(a_1, b_1); & a_1 = 2, b_1 = 1 \end{cases}$$

1. Perform the test for $n = 2$, and $y_1 = y_2 = 0$, by using Jeffrey's scaling.
2. Perform the test for $n = 2$, and $y_1 = y_2 = 2$, by using Jeffrey's scaling.

Hint-1 Poisson distribution $x \sim \text{Pn}(\lambda)$ has PMF: $\text{Pn}(x|\lambda) = \frac{1}{x!} \lambda^x \exp(-\lambda) 1_{\mathbb{N}}(x)$, where $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\lambda > 0$.

Hint-2 Negative Binomial distribution $x \sim \text{Nb}(r, \theta)$ has PMF: $\text{Nb}(x|r, \theta) = \binom{r+x-1}{r-1} \theta^r (1-\theta)^x 1_{\mathbb{N}}(x)$ with $\theta \in (0, 1)$, $r \in \mathbb{N} - \{0\}$, and $\mathbb{N} = \{0, 1, 2, \dots\}$.

Hint-3 Gamma distribution $x \sim \text{Ga}(a, b)$ has PDF: $\text{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) 1_{(0, \infty)}(x)$, with $a > 0$ and $b > 0$.

Hint-4 Beta distribution $x \sim \text{Be}(a, b)$ has PDF: $\text{Be}(x|a, b) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} 1_{(0, 1)}(x)$, with $a > 0$ and $b > 0$.

Solution. This is a Composite vs composite hypotheses. The overall a priori distribution $d\Pi(\theta)$ with $\theta \in \Theta$ and $\Theta = \{\text{Nb}\} \times (0, 1) \cup \{\text{Pn}\} \times (0, \infty)$ has density

$$\pi(\theta) = \pi_0 \text{Be}(\phi|a_0, b_0) + \pi_1 \text{Ga}(\lambda|a_1, b_1);$$

where $\pi_0 = \pi_1 = 0.5$. Let's compute the Bayes factor

$$\begin{aligned} f_0(y) &= \int \prod_{i=1}^n \text{Nb}(y_i|\phi, 1) \text{Be}(\phi|a_0, b_0) d\phi = \frac{1}{B(a_0, b_0)} \int_0^1 \phi^{n+a_0-1} (1-\phi)^{n\bar{y}+b_0-1} d\phi = \frac{B(n+a_0, n\bar{y}+b_0)}{B(a_0, b_0)} \\ f_1(y) &= \int \prod_{i=1}^n \text{Pn}(y_i|\lambda) \text{Ga}(\lambda|a_1, b_1) d\lambda = \frac{1}{\prod_{i=1}^n y_i!} \frac{b_1^{a_1}}{\Gamma(a_1)} \int_0^\infty \lambda^{n\bar{y}+a_1-1} \exp(-(n+b_1)\lambda) d\lambda \\ &= \frac{\Gamma(n\bar{y}+a_1)}{\Gamma(a_1)} \frac{b_1^{a_1}}{(n+b_1)^{n\bar{y}+a_1}} \frac{1}{\prod_{i=1}^n y_i!} \end{aligned}$$

So the Bayes Factor is

$$B_{01}(y) = \frac{B(n+a_0, n\bar{y}+b_0)}{B(a_0, b_0)} \frac{\Gamma(a_1)}{\Gamma(n\bar{y}+a_1)} \frac{(n+b_1)^{n\bar{y}+a_1}}{b_1^{a_1}} \prod_{i=1}^n y_i!$$

1. Then $B_{01}(y) = 2.70$, and $\log_{10}(B_{01}(y)) \approx 0.43$. According to Jeffrey's scaling rule, H_0 is supported.
2. Then $B_{01}(y) = 0.29$, and $\log_{10}(B_{01}(y)) \approx -0.53$. According to Jeffrey's scaling rule, the evidence against H_0 is substantial.

Example 20. Let $y = (y_1, \dots, y_n)$ a sequence of observables. There is interest in performing the following hypothesis test

$$H_0 : y_i | \phi \sim \text{Nb}(\phi, 1); \text{ with } \phi = 1/3 \quad \text{vs} \quad H_1 : y_i | \lambda \sim \text{Pn}(\lambda); \text{ with } \lambda = 2$$

1. Perform the test for $n = 2$, and $y_1 = y_2 = 0$, by using Jeffreys' scaling.
2. Perform the test for $n = 2$, and $y_1 = y_2 = 2$, by using Jeffreys' scaling.

Hint-1 Poisson distribution $x \sim \text{Pn}(\lambda)$ has PMF: $\text{Pn}(x|\lambda) = \frac{1}{x!} \lambda^x \exp(-\lambda) 1_{\mathbb{N}}(x)$, where $\mathbb{N} = \{0, 1, 2, \dots\}$ and $\lambda > 0$.

Hint-2 Negative Binomial distribution $x \sim \text{Nb}(r, \theta)$ has PMF: $\text{Nb}(x|r, \theta) = \binom{r+x-1}{r-1} \theta^r (1-\theta)^x 1_{\mathbb{N}}(x)$ with $\theta \in (0, 1)$, $r \in \mathbb{N} - \{0\}$, and $\mathbb{N} = \{0, 1, 2, \dots\}$.

Solution. This is a simple vs simple hypothesis test. I specify priors $\pi(\text{Nb}) = \pi(\text{Pn}) = 1/2$, due to the a priori ignorance about the parametric statistical model, however, we do not really need it now The Bayes factor is

$$B_{01}(y) = \frac{f_0(y)}{f_1(y)} = \frac{\prod_{i=1}^n \text{Nb}(y_i|\phi, 1)}{\prod_{i=1}^n \text{Pn}(y_i|\lambda)} = \frac{\phi^n (1-\phi)^{n\bar{y}}}{\lambda^{n\bar{y}} \exp(-n\lambda) / \prod_{i=1}^n y_i!}$$

1. Then $B_{01}(y) = \exp(4)/9 \approx 6.07$, and $\log_{10}(B_{01}(y)) \approx 0.78$. According to Jeffrey's scaling rule, H_0 is supported.
2. Then $B_{01}(y) = 4 \exp(4)/729 \approx 0.30$, and $\log_{10}(B_{01}(y)) \approx -0.54$. According to Jeffrey's scaling rule, the evidence against H_0 is substantial.

Practice

Question 21. To practice try to work on the Exercise 71 from the Exercise sheet.