LA-UR-17-30459

Approved for public release; distribution is unlimited.

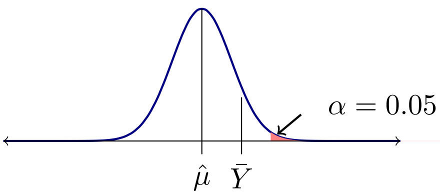
Title:	Bayesian Hypothesis Testing
Author(s):	Andrews, Stephen Arthur Sigeti, David Edward
Intended for:	UQ Project meeting
Issued:	2017-11-15

Bayesian Hypothesis Testing

Stephen A. Andrews and David E. Sigeti

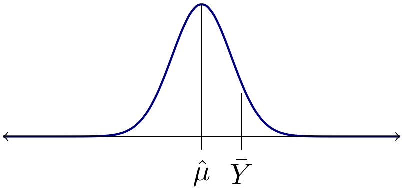
QM Project Meeting
Nov 15, 2017

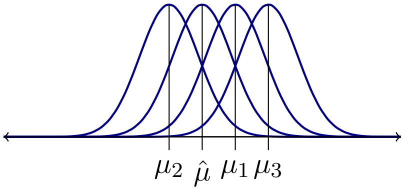
Testing a point null hypothesis



- We have data from a distribution
- We know everything about the distribution except a location parameter.
- We want to test that the location parameter of the distribution is a particular value $\hat{\mu}$
- Standard frequentist approach is a p-value based significance test.
- We are going to discuss the Bayesian approach.

Two Hypotheses





$H_0 : p(Y_i \,|\, \widehat{\mu})$ $\widehat{\mu}$ known

$H_1 : p(Y_i \,|\, \mu)$ μ unknown

- Same distribution and probability density function, p, for H_0 and H_1
- Analysis is symmetric with respect to the hypotheses.
- H_0 is the *null hypothesis*
- H_1 is the *alternative hypothesis*

Bayesian hypothesis testing

$$\underbrace{\frac{P(H_0 \mid Y)}{P(H_1 \mid Y)}}_{\text{Posterior odds ratio}} = \frac{p(Y \mid H_0)P(H_0)}{p(Y \mid H_1)P(H_1)} = \underbrace{\frac{p(Y \mid H_0)}{p(Y \mid H_1)}}_{\text{Bayes factor}} \cdot \underbrace{\frac{P(H_0)}{P(H_1)}}_{\text{Prior odds ratio} = 1}$$

- $p(Y \mid H_n)$ is the **evidence** for hypothesis n.
- P is total probability, p is probability density.
- We assume independent identically distributed data.
- We need to keep track of normalizing factors.

Evidence for the null hypothesis, H_0

Assume: $p(Y_i \mid H_0) = \mathcal{N}(Y_i \mid \hat{\mu}, \sigma^2)$

Then:

$$\underbrace{p(Y \mid H_0)}_{\text{Likelihood for all data}} = \prod_{i=1}^{n} \mathcal{N}\left(Y_i \mid \hat{\mu}, \sigma^2\right) = \mathcal{N}\left(Y \mid \hat{\mu}, \sigma^2\right)$$

$$p(Y \mid H_0) = \underbrace{\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left(\frac{-n}{2\sigma^2}\left(\overline{Y^2} - \overline{Y}^2\right)\right)}_{\text{Scaling factor } C_1} \cdot \exp\left(\frac{-n}{2\sigma^2}(\overline{Y} - \hat{\mu})^2)\right)$$

Where:

- $\overline{Y}$ Mean of the data
- $\overline{Y^2}$ Mean of the square of the data
- n The number of data
- $\hat{\mu}$ The value for the mean in the null hypothesis *i. e. null value*
- σ^2 Known variance of distribution

Evidence for the alternative hypothesis, H_1

- Using the law of total probability

$$p(Y \mid H_1) = \int d\mu \, p(Y, \mu \mid \sigma^2)$$

- Apply the definition of conditional probability.

$$p(Y \mid H_1) = \int d\mu \, p(Y \mid \mu, \sigma^2) p(\mu)$$

- As in H_0, we assume the data are normally distributed about the mean μ.

$$p(Y \mid H_1) = \int d\mu \, \mathcal{N}\left(Y \mid \mu, \sigma^2\right) p(\mu)$$

Prior for μ in H_1

- The prior must be normalizeable.
- Can't use a non-informative prior.
- We choose a normal distribution centered about $\widehat{\mu}$

$$p(\mu) = \mathcal{N}\left(\mu \,|\, \widehat{\mu}, \tau^2\right)$$

What value should we give to τ^2?

Evidence for H_1 (continued)

Using this prior, we can write the evidence for the alternative hypothesis as:

$$p(Y \mid H_1) = C_1 \frac{1}{\sqrt{2\pi\tau^2}} \int d\mu \ \exp\left(\frac{-n}{2\sigma^2} (\overline{Y} - \mu)^2 + \frac{-1}{2\tau^2} (\mu - \widehat{\mu})^2 \right)$$

- C_1 is the same normalizing factor that appeared in H_0
- The integral has a closed-form solution:

$$p(Y \mid H_1) = C_1 \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{1}{2} \frac{\left(\overline{Y} - \widehat{\mu}\right)^2}{\frac{\sigma^2}{n} + \tau^2} \right) . \sqrt{2\pi\widetilde{\tau}^2}$$

Where:

- $\frac{1}{\widetilde{\tau}^2} = \frac{1}{\tau^2} + \frac{n}{\sigma^2}$

Evaluating the Bayes ratio

$$\frac{P\left(H_0 \,|\, Y\right)}{P\left(H_1 \,|\, Y\right)} = \frac{\exp\left(\frac{-n}{2\sigma^2}(\overline{Y} - \hat{\mu})^2)\right)}{\sqrt{\frac{\widetilde{\tau}^2}{\tau^2}}\exp\left(-\frac{1}{2}\frac{\left(\overline{Y} - \widehat{\mu}\right)^2}{\frac{\sigma^2}{n} + \tau^2}\right)}$$

Where:

$\overline{Y}, n$ Given by the data

σ^2 Assumed known

$\widehat{\mu}$ Given in the problem formulation

τ^2 *Unknown from problem formulation*

Simplified expression for the Bayes ratio

Introduce:

Variance Ratio $\rho^2 \equiv \dfrac{\tau^2}{\sigma^2/n}$

- A re-scaled value for τ^2

Standard error term $\Upsilon^2 \equiv \frac{\left(\overline{Y}-\widehat{\mu}\right)^2}{\sigma^2/n}$

- The number of standard errors by which the mean of the data differ from the *null value*

Define:

$$\mathcal{G}\left(\Upsilon^2, \rho^2\right) \equiv \frac{P\left(H_0 \,|\, Y\right)}{P\left(H_1 \,|\, Y\right)}$$

To obtain:

$$\mathcal{G}\left(\Upsilon^2, \rho^2\right) = \sqrt{1+\rho^2}\,\exp\left(\frac{-\Upsilon^2}{2}\frac{\rho^2}{\rho^2+1}\right)$$

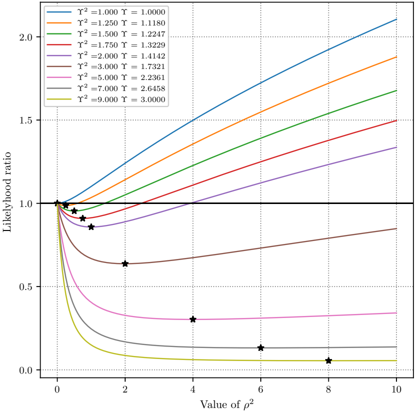
Implications of $\mathcal{G}$

$$\mathcal{G}\left(\Upsilon^2, \rho^2\right) = \sqrt{1+\rho^2}\exp\left(\frac{-\Upsilon^2}{2}\frac{\rho^2}{\rho^2+1}\right)$$

- As $\rho^2 \to 0$, Bayes ratio goes to 1
  - This is logical as H_1 is the same as H_0 when $\tau^2 = 0$
- As $\rho^2 \to \infty$, Bayes ratio goes to ∞
  - If the prior is infinitely wide, H_0 is infinitely favored
- The Bayes ratio depends critically on the width of the prior for μ.
  - Unlike in parameter estimation.
- **This is a known problem in Bayesian hypothesis testing**

Minimum of the Bayes ratio

- If $\Upsilon^2 < 1$, the minimum is at $\rho^2 = 0$
  - No value of τ^2 can make H_1 preferred.
- The function $\mathcal{G}$ has a minimum at $\rho^2 = \Upsilon^2 - 1$ when $\Upsilon^2 > 1$
- $\forall \Upsilon^2 > 1$, there is a nonzero minimum value for the Bayes ratio
  - The odds ratio in favor of H_0 cannot get lower than the value obtained with this τ^2
- *This is the state of the art*



Is there a better way to pick the width of the prior?

This approach to hypothesis testing raises some questions:

- Is the value for τ^2 that minimizes the Bayes factor a *likely* value for this variable?
- What does the *data* tell us about τ^2?

We construct a hierarchical model to:

- Examine the meaning of different choices of τ^2.
- See what value of τ^2 is suggested by the data.
- Examine the relationship between Υ^2 and τ^2

What is a reasonable value for τ^2

Based on our model for H_1, we construct a hierarchical model for the joint posterior probability of μ and τ^2 given the data.

$$
\begin{aligned}
p\big(\mu, \tau^2 \,|\, Y, \widehat{\mu}, \sigma^2\big) &\propto \underbrace{\mathcal{N}\big(Y \,|\, \mu, \sigma^2\big)}_{\text{Likelihood}} \underbrace{p(\mu,\, \tau^2)}_{\text{Joint prior}} \\
&\propto \mathcal{N}\big(Y \,|\, \mu, \sigma^2\big) \underbrace{\mathcal{N}\big(\mu \,|\, \widehat{\mu}, \tau^2\big)}_{\text{Same as } H_1} \underbrace{p(\tau^2)}_{\text{New}}
\end{aligned}
$$

This is a *hierarchical model* since the likelihood is not dependent on τ^2

Joint posterior for μ and τ^2

This expands to:

$$p(\mu, \tau^2 \,|\, \overline{Y}, \widehat{\mu}, \frac{\sigma^2}{n}) \propto \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(\frac{-1}{2}\left(\Upsilon - \xi\right)^2 + \frac{-1}{2}\frac{\frac{1}{n}\sigma^2}{\tau^2}\left(\xi\right)^2\right) p(\tau^2)$$

where:

$$\xi = \sqrt{\frac{n}{\sigma^2}}(\mu - \widehat{\mu})$$

Evaluating the posterior marginal for τ^2

- Integrate out μ from the joint posterior distribution to get the posterior marginal distribution for τ^2

$$p\big(\tau^2 \,|\, Y, \widehat{\mu}, \sigma^2\big) \propto \int_{-\infty}^{\infty} d\mu \, p(\mu, \tau^2 \,|\, \overline{Y}, \widehat{\mu}, \frac{\sigma^2}{n})$$

$$p\big(\tau^2 \,|\, Y, \widehat{\mu}, \sigma^2\big) \propto \sqrt{\frac{\frac{1}{n}\sigma^2}{\tau^2 + \frac{1}{n}\sigma^2}} \exp\left(-\frac{1}{2}\frac{(\overline{Y} - \widehat{\mu})^2}{\tau^2 + \frac{1}{n}\sigma^2}\right) p(\tau^2)$$

- We need to pick a form for $p(\tau^2)$ that will allow this distribution to be normalized

Possible forms for the hyper-prior

- The marginal posterior distribution for τ^2 looks like:

$$p\left(\tau^2 \,|\, Y, \widehat{\mu}, \sigma^2\right) \propto \frac{1}{(\tau^2 + \frac{1}{n}\sigma^2)^{\frac{1}{2}}} \exp\left(-\frac{C}{\tau^2 + \frac{1}{n}\sigma^2}\right) p(\tau^2) \quad C > 0$$

$$p(\tau^2) = 1 \qquad \text{Posterior not integrable at } \infty$$

$$p(\tau^2) = \left(\frac{1}{\tau^2}\right)^n, n \geq \frac{1}{2} \qquad \text{Posterior not integrable at } 0$$

$$p(\tau^2) = \left(\frac{1}{\tau^2 + \frac{1}{n}\sigma^2}\right)^2 \qquad \text{Posterior integrable from } 0 \text{ to } \infty$$

Including the hyper-prior

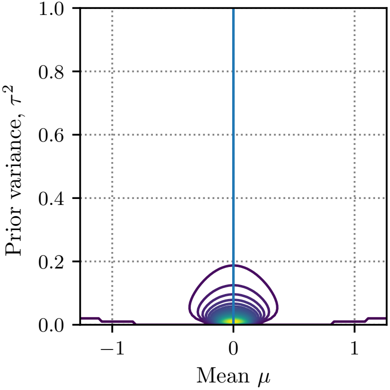
The prior probability of τ^2 is now:

$$p(\tau^2) = \left(\frac{1}{\tau^2 + \frac{1}{n}\sigma^2} \right)^2$$

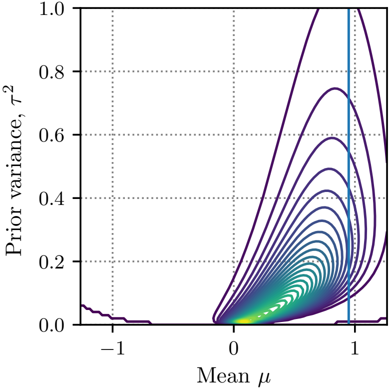
This gives us a joint posterior distribution for μ and τ^2

$$p(\mu, \tau^2 \,|\, \overline{Y}, \widehat{\mu}, \frac{\sigma^2}{n}) \propto \frac{1}{\sqrt{\tau^2}} \frac{1}{(\tau^2 + \frac{1}{n}\sigma^2)^2} \exp\left(\frac{-1}{2} \left(\Upsilon - \xi\right)^2 + \frac{-1}{2} \frac{1}{\rho^2} \left(\xi\right)^2 \right)$$

Contours of the joint probability distribution



$\Upsilon = 0$, $\overline{Y} = 0.0$



$\Upsilon = 3$, $\overline{Y} = 0.95$

Constants

$$\sigma^2 = 1 \quad n = 10 \quad \widehat{\mu} = 0$$

Change of variables in the posterior marginal for τ^2

Including the hyper-prior in the posterior marginal yields:

$$p\left(\tau^2 \,|\, Y, \widehat{\mu}, \sigma^2\right) \propto \frac{1}{\left(\tau^2 + \frac{1}{n}\sigma^2\right)^{5/2}} \exp\left(-\frac{1}{2}\frac{(\overline{Y} - \widehat{\mu})^2}{\tau^2 + \frac{1}{n}\sigma^2}\right)$$

Introduce a change of variables

$$\zeta^2 = (\tau^2 + \frac{1}{n}\sigma^2)$$

Because τ^2 must be a positive value we have:

$$p\left(\zeta^2\right) \equiv 0 \qquad \forall \zeta^2 \leq \frac{1}{n}\sigma^2$$

Marginal distribution

The marginal distribution becomes:

$$p\left(\zeta^2 \,|\, \Upsilon, \frac{1}{n}\sigma^2\right) \propto \left(\zeta^2\right)^{-\left(\frac{3}{2}+1\right)} \exp\left(-\frac{1}{2}\frac{(\overline{Y} - \widehat{\mu})^2}{\zeta^2}\right)$$

This is in the form of an un-normalized inverse gamma distribution with

$$\begin{aligned} \alpha &= \frac{3}{2} \\ \beta &= \frac{1}{2}(\overline{Y} - \widehat{\mu})^2 \\ &= \frac{1}{2}\Upsilon^2\frac{\sigma^2}{n} \end{aligned}$$

Normalizing

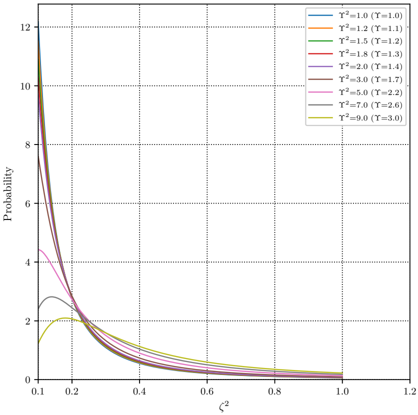
- Using the known cumulative probability function for the inverse gamma distribution

$$\int_0^x d\zeta^2 \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(\zeta^2)^{(\alpha+1)}} \exp\left(\frac{\beta}{\zeta^2}\right) = \frac{\Gamma_u(\alpha, \frac{\beta}{x})}{\Gamma(\alpha)}$$

- Where Γ_l / Γ_u is the lower / upper incomplete gamma function
- We can obtain the normalizing factor for our truncated distribution

$$\frac{\Gamma(\alpha)}{\beta^\alpha} \frac{\Gamma_l\left(\alpha, \frac{\beta}{x}\right)}{\Gamma(\alpha)} = \int_x^\infty d\zeta^2 \ \left(\zeta^2\right)^{-(\alpha+1)} \exp\left(\frac{-\beta}{\zeta^2}\right)$$

Normalized posterior marginal for ζ^2



Constants

$$\sigma^2 = 1 \quad n = 10 \quad \widehat{\mu} = 0$$

Median and Mode for posterior marginal for ζ^2

Mode

- The mode is 0 for $\Upsilon^2 < 5$
- The mode occurs at $\zeta^2 > \frac{1}{n}\sigma^2$ for $\Upsilon^2 > 5$

$$\begin{aligned}\zeta^2_{\text{mode}} &= \frac{\beta}{\alpha+1} \\ &= \frac{1}{5}\Upsilon^2\frac{1}{n}\sigma^2\end{aligned}$$

Median

$$\zeta^2_{\text{median}} = \frac{\frac{1}{2}\Upsilon^2\frac{1}{n}\sigma^2}{Q_l^{-1}\left(\alpha, \frac{1}{2}Q_l(\alpha, \frac{1}{2}\Upsilon^2)\right)}$$

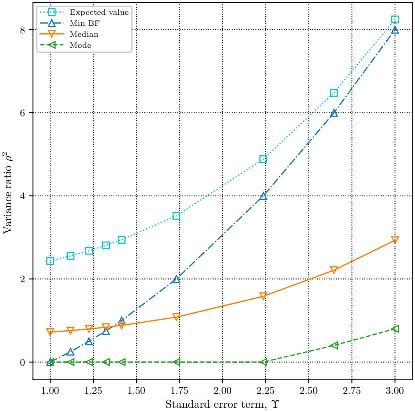
$$Q_l(\alpha, \beta) = \frac{\Gamma_l(\alpha, \beta)}{\Gamma(\alpha)}$$

Expected value of posterior marginal for ζ^2

- Since the probability density function for ζ^2 is zero from $\zeta^2 = 0$ to $\zeta^2 = \frac{1}{n}\sigma^2$, the standard expression for the mean of the inverse gamma distribution is incorrect.
- We want to determine the expected value of the posterior marginal over the domain where its probability is nonzero.

$$\begin{aligned}\mathbb{E}(\zeta^2) &= \int_{\frac{1}{n}\sigma^2}^{\infty} d\zeta^2\, \zeta^2 p(\zeta^2) \\ &= \frac{1}{2}\Upsilon^2 \frac{1}{n}\sigma^2 \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{2})} \frac{Q_l(\frac{1}{2}, \frac{1}{2}\Upsilon^2)}{Q_l(\frac{3}{2}, \frac{1}{2}\Upsilon^2)}\end{aligned}$$

Summary of the posterior marginal

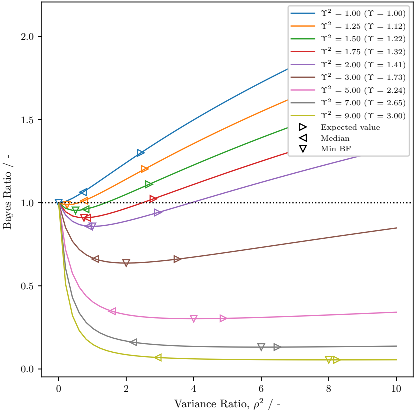


Constants

$$\sigma^2 = 1 \quad n = 10 \quad \widehat{\mu} = 0$$

Applying this to the Bayes ratio

- If Υ^2 is large, the choice of ρ^2 does not matter.
- If Υ^2 is small, the median of the posterior marginal is the closest value to that which value which minimizes the Bayes factor.



Applying this to the Bayes ratio

Υ^2	Bayes ratio				
	Minimum	Expectation	Delta	Median	Delta
1.00	1.0000	1.3003	0.3003	1.0637	0.0637
1.25	0.9867	1.2032	0.2165	1.0126	0.0259
1.50	0.9538	1.1109	0.1571	0.9612	0.0074
1.75	0.9092	1.0236	0.1144	0.9097	0.0005
2.00	0.8578	0.9413	0.0835	0.8586	0.0008
3.00	0.6372	0.6611	0.0239	0.6618	0.0246
5.00	0.3026	0.3045	0.0019	0.3472	0.0446
7.00	0.1317	0.1319	0.0001	0.1609	0.0292
9.00	0.0549	0.0550	0.0000	0.0692	0.0142

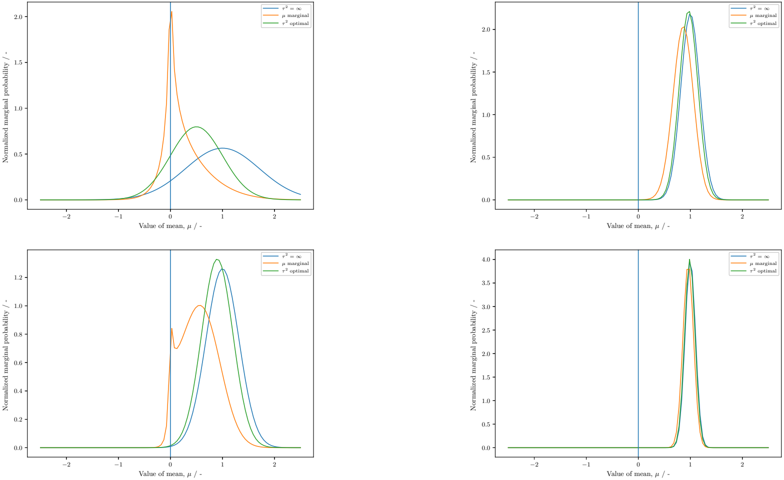
Conclusions

- The value of the Bayes factor obtained when using the median of the posterior marginal is almost the minimum value of the Bayes factor.
- The value of τ^2 which minimizes the Bayes factor is a reasonable choice for this parameter.
- This allows a likelihood ratio to be computed with is the least favorable to H_0.

Questions?

Consider the marginal distribution of μ

Marginals


$\tau^2 = \infty$
μ marginal
τ^2 optimal
Normalized marginal probability / -
Value of mean, μ / -