Comparing Proportions: A modern solution to a classical problem

Presented By: Savita Upadhyay

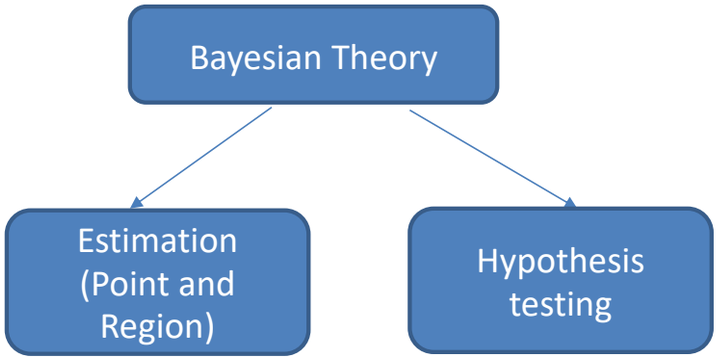
Thailand HIV Vaccine Trials

Hypothesis Testing

- Vaccine group: r1 = 51 (infected), n1 = 8197
- Placebo group: r2 = 74 (infected), n2 = 8198
- VE (vaccine efficacy) = $1 - \frac{\mathrm{r1/n1}}{r2/n2}$ = 0.31. It was reported that HIV vaccine conferred about 30% protection against HIV acquisition.
- Using Conventional frequentist testing, two corresponding binomial params θ_1, θ_2 were said to be significantly different with p-value = 0.04.
- Is this correct ?

Introduction

- Probability model: $M_z = \{p(z|\omega), z \in Z, \omega \in \Omega\}$.
- All that can be said about any function $\theta(z|\omega) \in \Theta$ of vector ω is contained in its posterior distribution $p(\theta\,|z)$.



- One would expect the same prior $p(\omega)$ could be used to derive both type of summaries(estimation and hypothesis testing).
- However, since Theory Of Probability (Jeffrey's), Bayesian methods have made use of radically difference types of prior

Key Ideas of the paper

It is argued that this is certainly not necessary use two radically different types of prior for estimation and some for hypothesis testing. A coherent solution to both problems using the same prior is possible within the standard framework of Bayesian decision theory.

The paper specifies:

- Point estimation, Region estimation and precise hypothesis testing results are dependent of on the choice of both the loss function and prior distribution
- Proposes the required properties of a loss function (intrinsic log loss) as a self – calibrated information bases continuous loss function for Hypothesis Testing
- Shows that by using a common prior and intrinsic log loss on Normal and binomial distributions we can do both estimation and Hypothesis testing
- Application to HIV trial data from Thailand

Integrated Bayesian analysis:

1. A decision theoretic formulation for estimation(point and region) and hypothesis testing highly depend on the choices of both the loss function and the prior distribution.

1. It is argued that their construction is better made with a coherent decision theoretical framework, making use of the same prior distribution in all cases.

2. Let $M_z = \{p(z|\omega), z \in Z, \omega \in \Omega\}$ be data generating model where Z = X^n and $\theta(\omega)$ is vector of interest. M_z can be expressed as:
   $M_z = \{p(z|\theta, \lambda), z \in Z, \theta \in \Theta, \lambda \in \Lambda\}$, where λ is some appropriately chosen nuisance parameter.

4 The loss function for (unknown) parameter values (θ, λ) if working with model M_z and using θ_o as proxy for θ is given as : $l\{\boldsymbol{\theta_o}, (\boldsymbol{\theta}, \boldsymbol{\lambda})\}$

5. Posterior expected loss is given as:

$$l(\theta_o|z) = \int_{\Theta}\int_{\Lambda} l\{\theta_o, (\theta, \lambda)\}p(\theta, \lambda|z)d(\theta)d(\lambda)$$

Where $p(\theta, \lambda|z)$ is the reference posterior distribution of θ.

Point Estimation, Region Estimation and Hypothesis Testing

1.Point Estimation:

Point estimate of θ is given by: $\theta^*(z) = \arg\{inf\}_{\theta_o \in \Theta} l(\theta_o|z)$

2. Region Estimation:

A Bayes q-credible region $\Theta^*{}_q(z) \subset \Theta$ is given as:

$$\forall \theta_i \in \Theta^*{}_q(z), \forall \theta_j \notin \Theta^*{}_q(z), l(\theta_i|z) < l(\theta_j|z)$$

3.Hypothesis testing:

- $H_o: \theta = \theta_o$
- Action space = { a_0: accept H_o, a_1:reject H_o} loss function for accepting or rejecting $H_o = l_h\{a_i, (\theta, \lambda)\}$
- We reject H_o if : $\int_\Theta \int_\Lambda (l_h\{a_o, (\theta, \lambda)\} - l_h\{a_1, (\theta, \lambda)\})\, p(\theta, \lambda|z) d(\theta) d(\lambda) > 0$
- Marginal loss for θ_o is given as:
  $\Delta l_h\{\theta_o, (\theta, \lambda)\} = l_h\{a_o, (\theta, \lambda)\} - l_h\{a_1, (\theta, \lambda)\}$
- Let $l_h\{a_1, (\theta, \lambda)\} = l_o$. Since $l_h\{\theta_o, (\theta_o, \lambda)\}$= 0, therefore $\Delta l_h\{\theta_o, (\theta, \lambda)\} = -l_o$

Therefore we reject H_o iff: $\boldsymbol{l(\theta_o|z) > l_o}$

Loss Function

The threshold l_o is part of specification of the decision problem and should be context dependent.

However, a judicial choice of loss function leads to self-calibrated losses.

Since It may naively appear that what is needed is just some measure of the discrepancy between θ_o and θ. However, since all parameterizations are arbitrary, what is really required is some measure of the discrepancy between the models labelled by θ_o and θ. By construction, such a discrepancy measure will be independent of the particular parameterization used. C. P. Robert [17] coined the word intrinsic to refer to these model-based loss functions; by construction, they are always invariant under one-to-one reparameterizations.

The intrinsic logarithmic loss

1. The intrinsic logarithmic loss works for both discrete and continuous function. It is invariant and leads to self-calibrated losses.
2. Let $M_z = p(z|\omega), z \in Z, \omega \in \omega$ be data generating model where Z = X^n and $\theta(\omega)$ is vector of interest. M_z can be expressed as: $M_z = \{p(z|\theta,\lambda), z \in Z,\}$ be the model.
3. $\kappa\{p_z(.|\omega_o)p_z(.|\theta,\lambda)\} = \int_Z p\,(z|\theta,\lambda) log\,\frac{p(z|\theta,\lambda)}{p(z|\omega_o)}\,dz$; KL directed divergence Average log-likelihood ratio against the alternative model. This is known to be nonnegative and zero iff $p(z|\omega_o)$ = $p(z|\theta,\lambda)$ almost everywhere.
4. The intrinsic log-loss function assuming H_o is the minimum average under sampling of the log-likelihood ratio against an element of M_0:

   $\delta\{H_o|\theta,\lambda,M_z\} = inf_{\lambda_o \in \Lambda_o} \int_Z p\,(z|\theta,\lambda) log\,\frac{p(z|\theta,\lambda)}{p(z|\omega_o)} dz$ …(using point 2)

5. With this structure we reject $H_o: \theta = \theta_o$ if and only if the reference expected loss is greater than l_o

$d(H_o|z) = \int_\Theta \int_\Lambda \delta\{H_o|\theta,\lambda,M_z\} p(\theta,\lambda|z) d\theta d\lambda > l_o$

Example: Normal variance

We want to test the variance σ^2.

1. $H_o: \sigma^2 = {\sigma_o}^2$

2. The Kullback-Leibler discrepancy of p($z|\mu_o, \sigma_o$)from p($z|\mu, \sigma$) is given by:
$\kappa N_z(.|\mu_o, \sigma_o) N_z(.|\mu, \sigma) = n \int_R N(x|\mu, \sigma) log \frac{N(x|\mu,\sigma)}{N(x|\mu_o,\sigma_o)} dx = \frac{n}{2}[log \frac{{\sigma_o}^2}{\sigma^2} + \frac{\sigma^2}{{\sigma_o}^2} - 1 + \frac{(\mu-\mu_o)^2}{{\sigma_o}^2}]$
(proof in the appendix)

3. The intrinsic logarithmic loss function of p($z|\mu_o, \sigma_o$)from p($z|\mu, \sigma$) is given by:

$$\delta\{H_o|\sigma, \nu, M_z\} = \frac{n}{2}\left[log \frac{{\sigma_o}^2}{\sigma^2} + \frac{\sigma^2}{{\sigma_o}^2} - 1\right]$$

$\kappa\{N_z(.|\mu_o, \sigma_o) N_z(.|\mu, \sigma)\}$is minimized at $\mu = \mu_o$

4 .The reference prior is $\pi(\mu, \sigma) = \sigma^{-1}$. The corresponding posterior density is given as:
$\pi(\sigma|z) = \pi(\sigma|n, s^2) = G^{\frac{-1}{2}}\left(\frac{n-1}{2}, \frac{ns^2}{2}\right)$
where $s^2 = n^{-1} \sum_{i=1}^{n} (x_j - x)^2$

5 .The reference posterior expected loss from using σ_o as a proxy of σ, given a random sample of size n is:
$d(\sigma_o|z) = \int_0^\infty n\, \delta\{\sigma_o\, |\sigma, \mu\}\, \pi(\sigma|z)\, d\sigma$

6.Hypothesis H_o will be rejected if :
$d(H_o|z) = d(\sigma_o|z) > l_o$

where conventional choices for $l_o$are {log20, log100,log1000}

plot

HIV Trials Revisited: Objective Bayesian Analysis

1. Reference prior $\pi(\theta_1, \theta_2)$ = Be(θ_1| 0.5 , 0.5) Be(θ_2|0.5 , 0.5)
2. Joint reference posterior:
   $\pi(\theta_1, \theta_2 | z) = \pi(\theta_1, \theta_2 | r_1, r_2, n_1, n_2)$ = Be(θ_1| $r_1 + 0.5, n_1 - r_1 + 0.5$) Be($\theta_2$|$r_2, n_2 - r_2 + 0.5$)
3. $$\Pr[\theta_1 < \theta_2 \,|\, \boldsymbol{z}] = \int_0^1 \int_0^{\theta_2} \pi(\theta_1, \theta_2 \,|\, \boldsymbol{z})\, d\theta_1\, d\theta_2 = 0.981,$$

4. Intrinsic logarithmic loss function is: $$\delta\{H_0 \,|\, \boldsymbol{\theta}, n_1, n_2\} = \inf_{\boldsymbol{\theta}_0 \in H_0} E_{\boldsymbol{z}\,|\,\boldsymbol{\theta}} \left[\log \frac{p(\boldsymbol{z} \,|\, \boldsymbol{\theta})}{p(\boldsymbol{z} \,|\, \boldsymbol{\theta}_0)} \right]$$

5 . Posterior expected logarithmic intrinsic loss is: $$d(H_0 \,|\, \boldsymbol{z}) = \int_0^1 \int_0^1 \delta\{H_0 \,|\, \boldsymbol{\theta}, n_1, n_2\}\, \pi(\theta_1, \theta_2 \,|\, \boldsymbol{z})\, d\theta_1\, d\theta_2 = 2.624 = \log[13.8].$$

6. Reference posterior for actual efficacy $\emptyset$ of vaccine can be: $\pi(\emptyset | z)$, where $\emptyset(\theta_1, \theta_2)$ = 1- (θ_1/θ_2)

7. $$d(\phi_0 \,|\, \boldsymbol{z}) = \int_{-\infty}^{\infty} \int_0^1 \delta\{\phi_0 \,|\, \phi, \theta_2, n_1, n_2\}\, \pi(\phi, \theta_2 \,|\, \boldsymbol{z})\, d\theta_2\, d\phi.$$

8. $\emptyset$ is minimized at 0.297 and the values within (-0.071, 0.544) is less than ζ_0 = log[20]
9. The intrinsic reference 0.95 credible region is (-0.009, 0.514), which also contains 0.

All these elaborate on the basic conclusion provided by $d(H_o|z)$ =log[13.8]. Weak evidence against the Hypothesis H0 that there

There is no difference between the two hazard rates. There is some suggestion of a vaccine efficacy about 30%. But the efficacy could anywhere between 1% to 50% against the firm conclusion of an existing difference between the parameters implied by the p = 0.04 frequentist significance

Bayesian Implies: More information is really necessary for final answers

HIV Trials : Reference posterior efficacy of vaccince

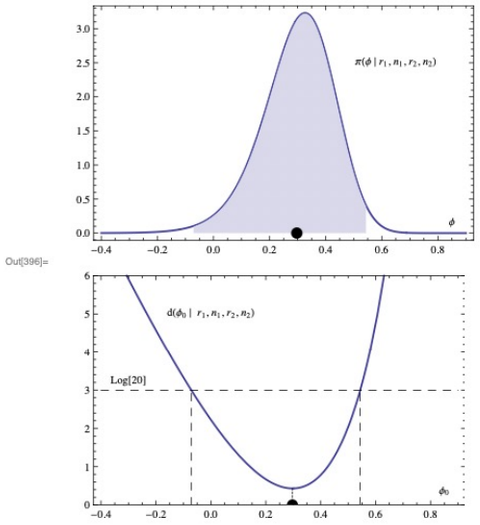


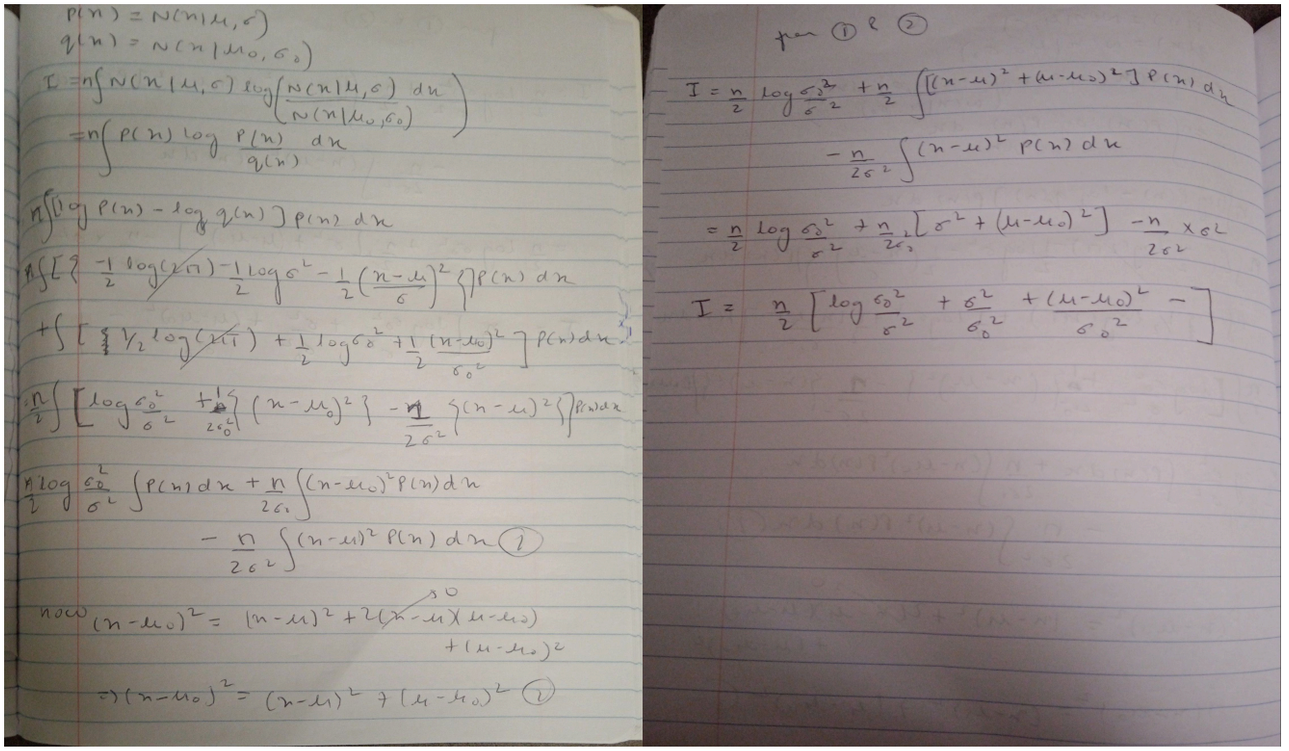
FIGURE 1.4
Reference posterior analysis of the vaccine efficacy $1 - (\theta_1/\theta_2)$ *for the RV144 vaccine efficacy trial data.*

Appendix

Very short summary of Jeffreys proposal

- There are a few reasons for having different priors for testing and estimation.
- One of them is that some priors commonly used for estimation have a terrible behavior when used for testing (such as converging to the null model with probability 1 as n—> infinity regardless of what the truth is, one example is using a flat prior on the real line for the regression coefficients, this one is commonly used for estimation but has this terrible behavior for testing). This is for example the case of the Jeffreys' prior for the multivariate normal model, which can be used for estimation but not for testing.
- Also, there is the issue that the author mentions in the paper, if you are testing a sharp null hypothesis you want that event to have probability greater than 0, and that is not the case if you have a continuous parameter
.

Proof



$$p(x) = N(x|\mu,\sigma)$$
$$q(x) = N(x|\mu_0,\sigma_0)$$
$$I = n\int N(x|\mu,\sigma)\log\left(\frac{N(x|\mu,\sigma)}{N(x|\mu_0,\sigma_0)}\,dx\right)$$
$$= n\int p(x)\log\frac{p(x)}{q(x)}\,dx$$
$$n\int[\log p(x) - \log q(x)]\,p(x)\,dx$$
$$n\int\left[\left\{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}p(x)\,dx\right.$$
$$\left.+\int\left[\left\{\tfrac{1}{2}\log(2\pi) + \frac{1}{2}\log\sigma_0^2 + \frac{1}{2}\frac{(x-\mu_0)^2}{\sigma_0^2}\right]p(x)\,dx\right.$$
$$= \frac{n}{2}\int\left[\log\frac{\sigma_0^2}{\sigma^2} + \frac{1}{2\sigma_0^2}\left\{(x-\mu_0)^2\right\} - \frac{n}{2\sigma^2}\left\{(x-\mu)^2\right\}\right]p(x)\,dx$$
$$\frac{n}{2}\log\frac{\sigma_0^2}{\sigma^2}\int p(x)\,dx + \frac{n}{2\sigma_0}\int(x-\mu_0)^2 p(x)\,dx$$
$$- \frac{n}{2\sigma^2}\int(x-\mu)^2 p(x)\,dx \quad (1)$$
$$\text{now } (x-\mu_0)^2 = (x-\mu)^2 + 2(x-\mu)(\mu-\mu_0) + (\mu-\mu_0)^2$$
$$\Rightarrow (x-\mu_0)^2 = (x-\mu)^2 + (\mu-\mu_0)^2 \quad (2)$$

per (1) & (2)

$$I = \frac{n}{2}\log\frac{\sigma_0^2}{\sigma^2} + \frac{n}{2}\int\left[(x-\mu)^2 + (\mu-\mu_0)^2\right]p(x)\,dx$$
$$- \frac{n}{2\sigma^2}\int(x-\mu)^2 p(x)\,dx$$
$$= \frac{n}{2}\log\frac{\sigma_0^2}{\sigma^2} + \frac{n}{2\sigma_0^2}\left[\sigma^2 + (\mu-\mu_0)^2\right] - \frac{n}{2\sigma^2}\times\sigma^2$$
$$I = \frac{n}{2}\left[\log\frac{\sigma_0^2}{\sigma^2} + \frac{\sigma^2}{\sigma_0^2} + \frac{(\mu-\mu_0)^2}{\sigma_0^2} - \right]$$

Normal Intrinsic loss

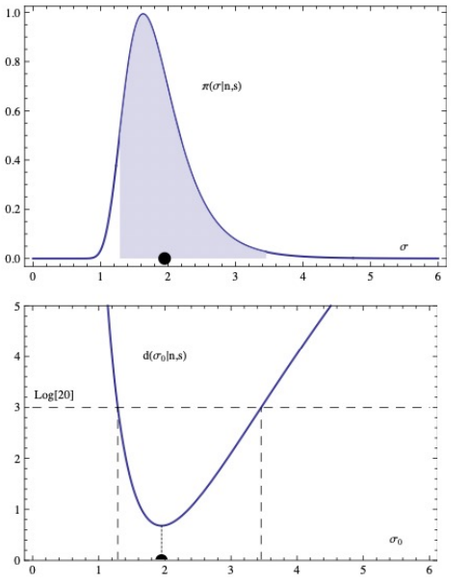


FIGURE 1.1
Intrinsic reference analysis for the standard deviation of a normal distribution, given a random sample of size $n = 5$*, with* $s = 1.631$.