Palmetto Lecture in Statistics
University of South Carolina

A (VERY) SHORT COURSE ON COMPARATIVE STATISTICAL INFERENCE

Francisco J. Samaniego

University of California, Davis

March 26, 2013

I: INTRODUCTION

- Most graduate programs in Statistics offer separate courses on classical (or "frequentist") statistical methods and on Bayesian statistical methods. Courses on comparative statistical inference are rather rare. Questions like "When should one be (or not be) a Bayesian?" are rarely asked.

- My main purpose today is to raise this question and to propose a particular way of addressing it.

- In the two types of courses alluded to above, it appears that the answers would be either "always" or "never". Is there a more nuanced answer?

II: A QUICK REVIEW OF BAYESIAN ESTIMATION

Suppose one's data is governed by a probability model F_θ indexed by a scaler parameter θ. For simplicity, suppose that

$$X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta$$

For the Bayesian, the estimation of the parameter θ begins with the specification of a prior probability distribution G on θ. If the densities (or pmfs) of $X|\theta$ and of θ exist, then one uses them both to obtain the posterior density (or pmf) of $\theta|X = x$, say

$$g(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)g(\theta)d\theta}$$

A Bayes estimator of θ is obtained by minimizing $\mathbb{E}_{\theta|X=x}[L(\theta, \hat{\theta}(x)]$ relative to $\hat{\theta}(x)$ for a chosen loss function L. The most widely used loss function in classical or Bayesian estimation is *squared error loss*, that is, $L(\theta, a) = (\theta - a)^2$. Relative to SEL, the Bayes estimate of θ is $\mathbb{E}(\theta|x)$.

III: STANDARD APPOACHES TO COMPARING BAYES AND FREQUENTIST ESTIMATORS

Five Criteria for Judging B vs F:

1. Logic — B wins
2. Objectivity — F wins, but not completely
3. Admissibility — B wins, but with no practical consequences
4. Asymptotics — F wins, but in most problems, it's a tie
5. Silly Answers — Definitely a tie.

IS THERE A WINNER?

- The "debate" between Bayesians and frequentists, at least as represented by the foregoing commentary, ends up in an uncomfortably inconclusive state.

- One specific question has been left untouched. Which method stands to give "better answers" in real problems of practical interest? This is the question to which we now turn.

IV: MODELING THE TRUE STATE OF NATURE

- Let us focus on an estimation problem, given data $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta$ and a fixed loss function $L(\theta, a)$. Suppose that a frequentist statistician is prepared to estimate the unknown parameter θ by the estimator $\widehat{\theta}$ and that a Bayesian statistician is prepared to estimate θ by the estimator $\widehat{\theta}_G$, the Bayes estimator relative to his chosen prior distribution G.

- How should the "truth" be modeled? Let's consider the true value of θ to be a random variable and call its distribution G_0 the "true prior".

- Now in many problems of interest, θ is not random at all; it's just an unknown constant. In such problems, it is appropriate to take G_0 to be a degenerate distribution which gives probability one to θ_0, the true value of θ.

- In other settings, it may be appropriate to consider G_0 to be nondegenerate.

V: A CRITERION FOR COMPARING ESTIMATORS

- Let us now examine the possibility of using the Bayes risk of an estimator, relative to the true prior G_0, as a criterion for judging the superiority of one estimator over another.

- For a fixed loss function L, the Bayes risk of an estimator $\widehat{\theta}$ with respect to the true prior G_0 is given by $r(G_0, \widehat{\theta}) = E_\theta E_{\boldsymbol{X}|\theta} L(\theta, \widehat{\theta}(\boldsymbol{X}))$, where the outer expectation is taken with respect to G_0. For simplicity, let $L(\theta, a) = (\theta - a)^2$.

- The Bayes risk $r(G_0, \widehat{\theta})$ is simply the mean squared error averaged relative to the objective truth in the estimation problem of interest, and is thus a highly relevant measure of the estimator's worth.

- If the Bayesian statistician was able to discern the actual true prior G_0, then he would undoubtedly use it in estimating the parameter θ.

- Since this scenario is a virtual impossibility, the Bayesian will select a prior G, henceforth referred to as his "operational prior," in order to carry out his estimation.

- The Bayes risk $r(G, \widehat{\theta})$ only measures how well the Bayesian did relative to his prior intuition.

- The Bayesian's estimation process is not driven by the true prior G_0, but there can be no question that an impartial adjudicator would be interested in $r(G_0, \widehat{\theta})$ rather than in $r(G, \widehat{\theta})$, as it is the former measure, rather than the latter, which pertains to how well the Bayesian did in estimating the true value of θ.

- On the other hand, the frequentist also has a natural interpretation of the criterion $r(G_0, \widehat{\theta})$, as it simply represents a generalized form of his estimator's mean squared error, being the squared error of his estimator averaged over all the randomness in the problem or, in many cases, the mean squared error of his estimator evaluated at the true value of the target parameter.

- Set aside, for a moment, the fact that the true prior G_0 is generally unknown and unknowable.

VI: THE THRESHOLD PROBLEM

By the "threshold problem," we will mean the problem of determining the boundary which divides the class of priors $\mathcal{G}$ into the subclass of priors for which

$$r(G_0, \widehat{\theta}_{\mathrm{G}}) < r(G_0, \widehat{\theta}) \ ,$$

where $\widehat{\theta}$ represents a given frequentist estimator, from the subclass of priors for which

$$r(G_0, \widehat{\theta}_{\mathrm{G}}) > r(G_0, \widehat{\theta}) \ .$$

As formulated above, the threshold problem may seem entirely intractable.

A TRACTABLE VERSION OF THE THRESHOLD PROBLEM

For simplicity,

- let's assume that our data consist of a random sample from a distribution indexed by θ, that is, assume that $X_1, X_2, \ldots, X_n \overset{iid}{\sim} F_\theta$.
- let's suppose that the distribution F_θ belongs to a one-parameter exponential family,
- let L be squared error loss and
- let $\mathcal{G}$ be the class of standard conjugate priors corresponding to the distribution F_θ.

Then,

(i) We will be able to restrict attention to just one estimator $\widehat{\theta}$, the one that I will refer to as the "best frequentist estimator".

(ii) The characterization of conjugate priors G for which $\widehat{\theta}_G$ beats $\widehat{\theta}$ reduces to a search over a finite-dimensional space of prior parameters.

(iii) Under squared error loss, Bayes estimators with respect to conjugate priors take particularly simple closed-form expressions, and calculating the Bayes risk is straightforward.

(iv) Even through the true prior G_0 is unknown, we will draw useful guidance about Bayes estimators which outperform the best frequantist estimator.

VII: THE WORD-LENGTH EXPERIMENT

- Ninety-nine students in an elementary statistics class at the University of California, Davis, were asked to participate in an experiment involving an observed binomial variable with an unknown probability p of "success."
- The population from which data were to be drawn was the collection of "first words" on the 758 pages of a particular edition of Somerset Maugham's 1915 novel *Of Human Bondage*.
- Ten pages were to be sampled randomly, with replacement, and the number X of long words (i.e., words with six or more letters) was to be recorded.
- Each student was asked to provide a Bayes estimate of the unknown proportion p of long words.

- The elicitation of the students' beta priors was accomplished by obtaining each student's best guess p^* at p and the weight η he or she wished to place on the sample proportion $\widehat{p} = X/10$, with weight $(1 - \eta)$ placed on the prior guess p^*.

- It's natural to ask: how many of these nouveau Bayesians would tend to be closer to the true value of p than a statistician using the sample proportion $\widehat{p}$ as an estimator of p?

- The prior specifications $\{(p^*, \eta)\}$ obtained from the students are displayed in Figure 1.

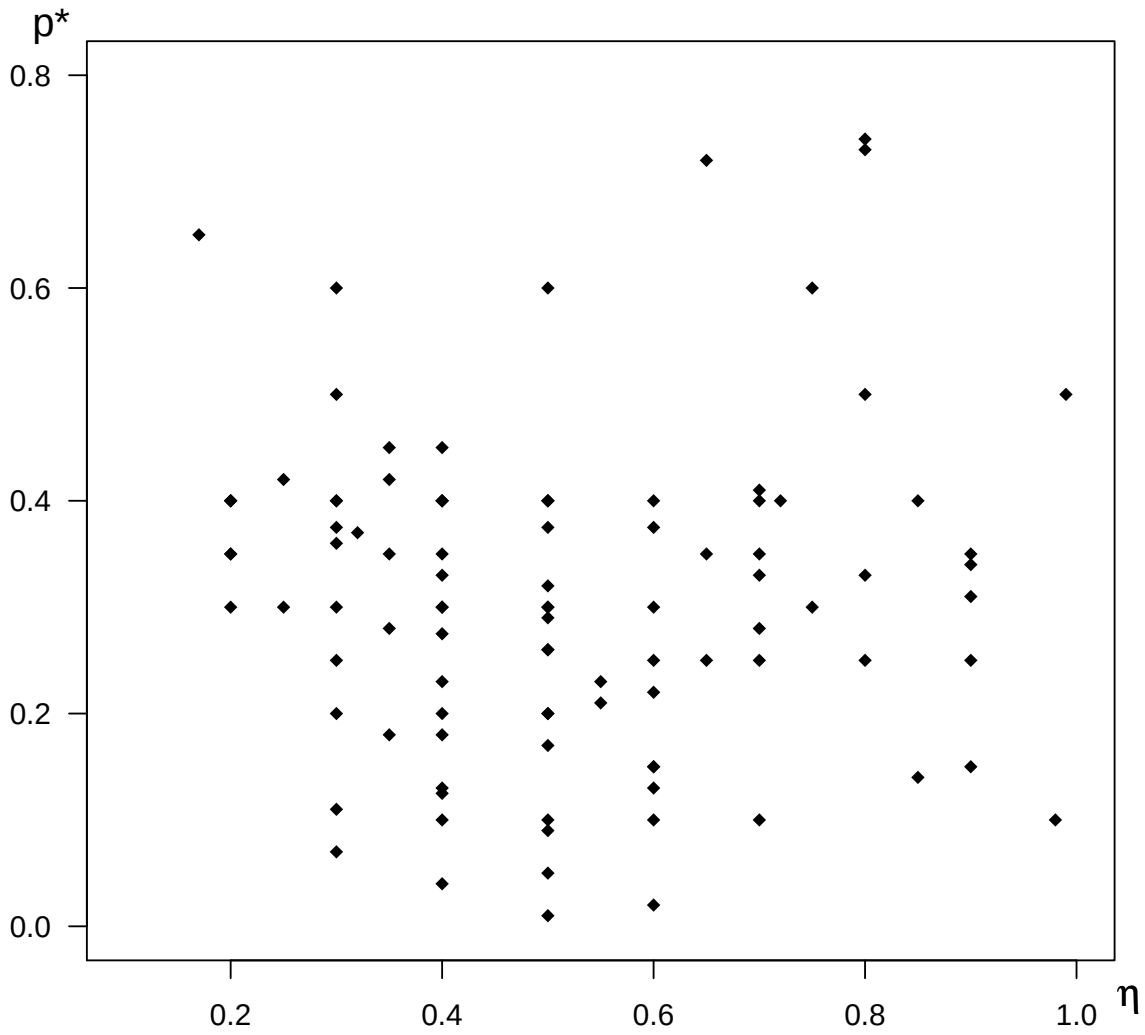


Figure 1 : Scatter plot of (η, p^*) values in the Word-Length Experiment

VIII: A THEORETICAL FRAMEWORK

Theorem 1

Assume that a random sample is drawn from a distribution F_θ. Let $\widehat{\theta}_G$ be the Bayes estimator of θ under squared error loss, relative to the operational prior G. If $\widehat{\theta}_G$ has the form

$$\widehat{\theta}_G = (1-\eta)E_G\theta + \eta\widehat{\theta}\ ,$$

where $\widehat{\theta}$ is a sufficient and unbiased estimator of θ and $\eta \in [0,1)$, then for any fixed distribution G_0 for which the expectations exist,

$$\boldsymbol{r}(G_0, \widehat{\theta}_G) \leq \boldsymbol{r}(G_0, \widehat{\theta})$$

if and only if

$$V_{G_0}(\theta) + (E_G\theta - E_{G_0}\theta)^2 \leq \frac{1+\eta}{1-\eta}\boldsymbol{r}(G_0, \widehat{\theta})\ .$$

Theorem 1 suggests that the Bayes estimator will be superior to $\widehat{\theta}$ unless the Bayesian statistician miscalculates on two fronts simultaneously.

If a Bayesian is both misguided (with a poorly centered prior) and stubborn (with a prior that is highly concentrated on the prior guess), his estimation performance will generally be quite inferior to that of the best frequentist estimator. Interestingly, neither of these negative characteristics alone will necessarily cause the Bayesian to lose his advantage.

The Bayesian's winning strategy becomes quite clear: (1) careful attention to one's prior guess is worth the effort, since when that specification is done well, you can't lose, and (2) overstating one's confidence in a prior guess can lead to inferior performance, so conservative prior modeling is advisable

Corollary 1

Under the hypotheses of Theorem 1, the Bayes estimator $\widehat{\theta}_G$ *and the frequentist estimator* $\widehat{\theta}$ *have the same Bayes risk with respect to the true prior* G_0 *for any operational prior* G *corresponding to the prior parameters* (Δ, η) *satisfying the hyperbolic equation*

$$\Delta\eta + \eta(\boldsymbol{r}(G_0, \widehat{\theta}) + V_{G_0}(\theta)) - \Delta + (\boldsymbol{r}(G_0, \widehat{\theta}) - V_{G_0}(\theta)) = 0 \ ,$$

where $\Delta = (E_G\theta - E_{G_0}\theta)^2$ *and* $\eta \in [0, 1)$ *is the weight placed on* $\hat{\theta}$.

- Earlier, I described the word-length experiment in which 99 students provided Bayes estimates, relative to their individually elicited beta priors, of the proportion of "long words" among the first words appearing on the 758 pages of a copy of Maugham's novel *Of Human Bondage*.

- We mentioned that a strong majority of the Bayes estimators tend to outperform $\widehat{p}$.

- We now examine this experiment in light of the theoretical results above. The true proportion of long words in the population was determined to be $p = 228/758 = 0.3008$.

- In the word-length experiment, Bayes estimators are superior in 88 of 99 cases.

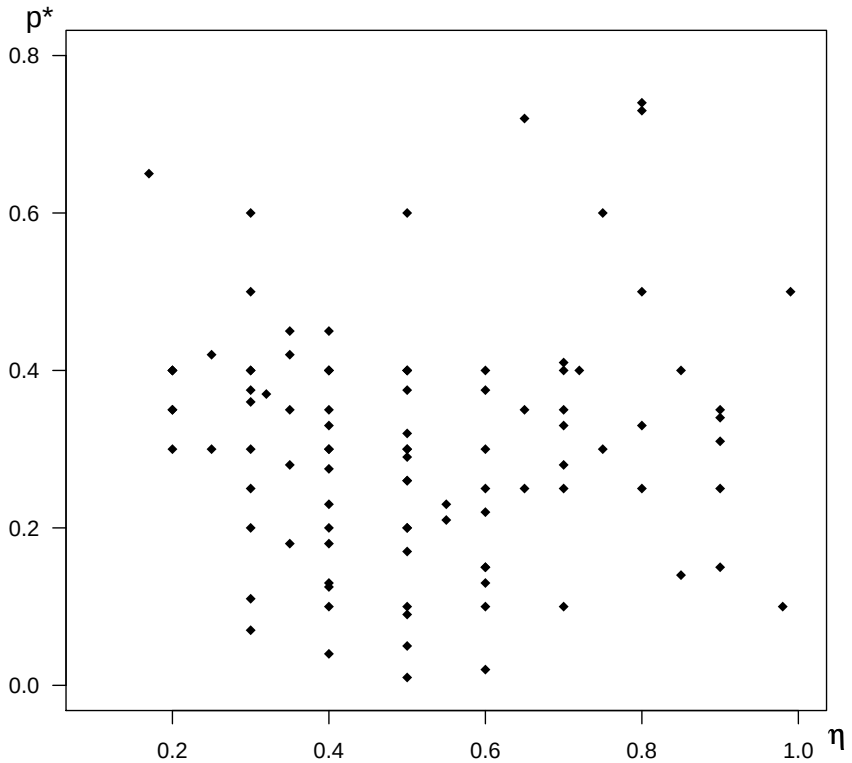


Figure 1 : Scatter plot of (η, p^*) values in the Word-Length Experiment

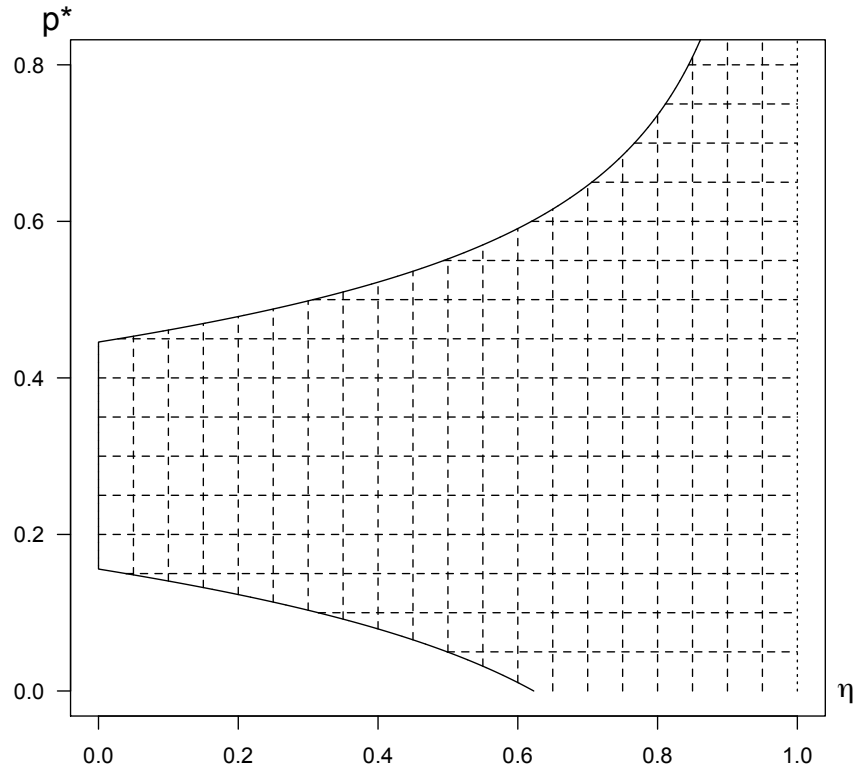


Figure 2 : Graph of the threshold, and the region of Bayesian superiority in the Word-Length Experiment

It is natural to wonder what the effect is of the chosen sample size n in the comparison of Bayes and frequentist estimators.

Table 1 : The percentage of Bayes estimators (SBEs) that are superior to the sample proportion in the word-length experiment, as a function of n

Sample size n	Number of SBEs	Percentage of SBEs
1	99	100%
5	93	94%
10	88	89%
50	83	84%
100	81	82%
1,000	80	81%
10,000	79	80%
∞	79	80%

The empirical and theoretical results we discussed support the conclusion that the Bayesian approach to estimation is surprisingly resilient, providing superior results even in cases in which the operational prior distribution used might, on the basis of some sort of impartial analysis, be considered to be quite weak. Our findings indicate that Bayes procedures work well a lot more often than we (and most other people) suspected.

IX: BAYESIAN VS. FREQUENTIST SHRINKAGE IN MULTIVARIATE NORMAL PROBLEMS

- We have focused on the comparison of Bayes and frequentist estimators of the mean $\boldsymbol{\theta}$ of a multivariate normal distribution in high dimensions. For dimension $k \geq 3$, the James–Stein estimator specified in (2.15) (and its more general form to be specified below) is usually the frequentist estimator of choice.

- The James–Stein estimator used shrinks $\overline{\boldsymbol{X}}$ toward a (possibly nonzero) distinguished point. This serves the purpose of placing the James–Stein estimator and the Bayes estimator of $\boldsymbol{\theta}$ with respect to a standard conjugate prior distribution in comparable frameworks, since the latter also shrinks $\overline{\boldsymbol{X}}$ toward a distinguished point.

- We've examined scenarios in which the threshold problem is tractable.

Conclusion

- Bayesian estimation of a high-dimensional parameter is a difficult enterprise — a fact that is not particularly surprising, given that the specification of a prior model which leads to inferences that are superior to notable frequentist alternatives is quite challenging. It becomes clear that, even in the case in which G_0 is degenerate, $\boldsymbol{\Sigma}_G = \sigma_G \boldsymbol{I}$ and $\boldsymbol{\Sigma}_X = \sigma^2 \boldsymbol{I}$, the opportunity of inferior Bayesian inference is substantial.

- Bayesian difficulties become all the more imposing once we transition to the real problem one would typically face in practice, the problem of estimating the mean of a normal distribution with a general covariance matrix $\boldsymbol{\Sigma}$.

- What remains true in all versions of this problem is the fact that there is a threshold separating good priors from bad priors.

X: COMPARING BAYESIAN AND FREQUENTIST ESTIMATORS UNDER ASYMMETRIC LOSS

Consider the Linex loss function:

$$L(\theta, \widehat{\theta}) = e^{c(\widehat{\theta}-\theta)} - c(\widehat{\theta} - \theta) - 1 \ ,$$

where c is a fixed and known constant. The Linex loss function achieves its minimum 0 when $\widehat{\theta} = \theta$ and is a convex function of the difference $\Delta = (\widehat{\theta} - \theta) \in (-\infty, \infty)$, decreasing for $\Delta \in (-\infty, 0)$ and increasing for $\Delta \in (0, \infty)$. When c is positive, Linex loss grows exponentially in positive Δ, but behaves approximately linearly for negative values of Δ. For dimension $k \geq 2$, adefine

$$L(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}) = \sum_{i=1}^{k} L(\theta_i, \widehat{\theta}_i).$$

Theorem 2

For arbitrary values of the true and operational prior means $\boldsymbol{\theta}_{G_0}$ and $\boldsymbol{\theta}_G$, the Bayes estimator $\widehat{\boldsymbol{\theta}}^G$ of a multivariate normal mean under generalized Linex loss is superior to the MLE $\overline{\boldsymbol{X}}$ if the operational prior is sufficiently diffuse.

Theorem 3

If the operational prior G is mean correct (i.e., $E_G\theta = E_{G_0}\theta$) and if the true prior G_0 is degenerate at the point $\boldsymbol{\theta}_{G_0}$, that is, $\sigma^2_{G_0} = 0$, then for all values of $\sigma^2_G > 0$,

$$r\left(G_0, \widehat{\boldsymbol{\theta}}^G\right) < r\left(G_0, \overline{\boldsymbol{X}}\right) \ .$$

XI: THE TREATMENT OF NONIDENTIFIABLE MODELS

Identifiability: If $X \sim F_{\theta_1}$, and $X \sim F_{\theta_2}$, then $\theta_1 = \theta_2$.

- In classical statistical estimation theory, the estimation of a nonidentifiable parameter is viewed, quite simply, as an ill-posed problem. While classical methods are inapplicable in treating such problems directly, there are several available options.

- Among these options are (a) placing additional restrictions on the original model, rendering the parameters of the restricted model identifiable, (b) focusing on the estimation of a function of the original parameters that is in fact identifiable and (c) expanding the model to include additional data which, together with the original data, makes the original parameters identifiable.

- In contrast with the frequentist approach to estimation in the presence of non-identifiability, the Bayesian paradigm has no difficulty in treating nonidentifiable parameters.
- A Bayesian will begin the treatment of an estimation problem by stipulating a prior distribution on the parameters of the model of interest.
- While, in a model with nonidentifiable parameters, the data available to the statistician are "defective". The data observed in such problems are still *informative* about these parameters.
- The updating of the prior distribution on the basis of the observed data is thus both feasible and meaningful, resulting in a posterior distribution on which inference can be based.
- The threshold problem here involves comparing prior and posterior inferences.

Example: Suppose that the observed data in an experiment of interest is a binomial random variable with distribution

$$X \sim \mathcal{B}(n, p_1 + p_2) \ ,$$

where $p_1 \geq 0$, $p_2 \geq 0$ and $0 \leq p_1 + p_2 \leq 1$.

The model would be appropriate in situations in which there are two mutually exclusive causes of "success" in a sequence of n Bernoulli trials, and these causes are indistinguishable without costly or infeasible follow-up.

The target of the estimation problem of interest is the pair (p_1, p_2), a parameter pair which is, of course, nonidentifiable on the basis of the available data.

Theorem 4

Let $X_n \sim \mathcal{B}(n, p_1 + p_2)$, and let (p_1^*, p_2^*) be the true but unknown value of the parameter pair (p_1, p_2). Suppose the operational prior distribution G of (p_1, p_2) is the Dirichlet distribution $\mathcal{D}(\alpha_1, \alpha_2, \alpha_3)$. As $n \to \infty$, the posterior distribution of p_1, given $X_n = x$, is a rescaled beta distribution, that is,

$$p_1 | X_n = x \xrightarrow{D} cW \ ,$$

where $W \sim Be(\alpha_1, \alpha_2)$ and $c = p_1^* + p_2^*$, and the posterior distribution of p_2, given $X_n = x$, is the complementary rescaled beta distribution, that is,

$$p_2 | X_n = x \xrightarrow{D} cV \ ,$$

where $V \sim Be(\alpha_2, \alpha_1)$ and $c = p_1^* + p_2^*$.

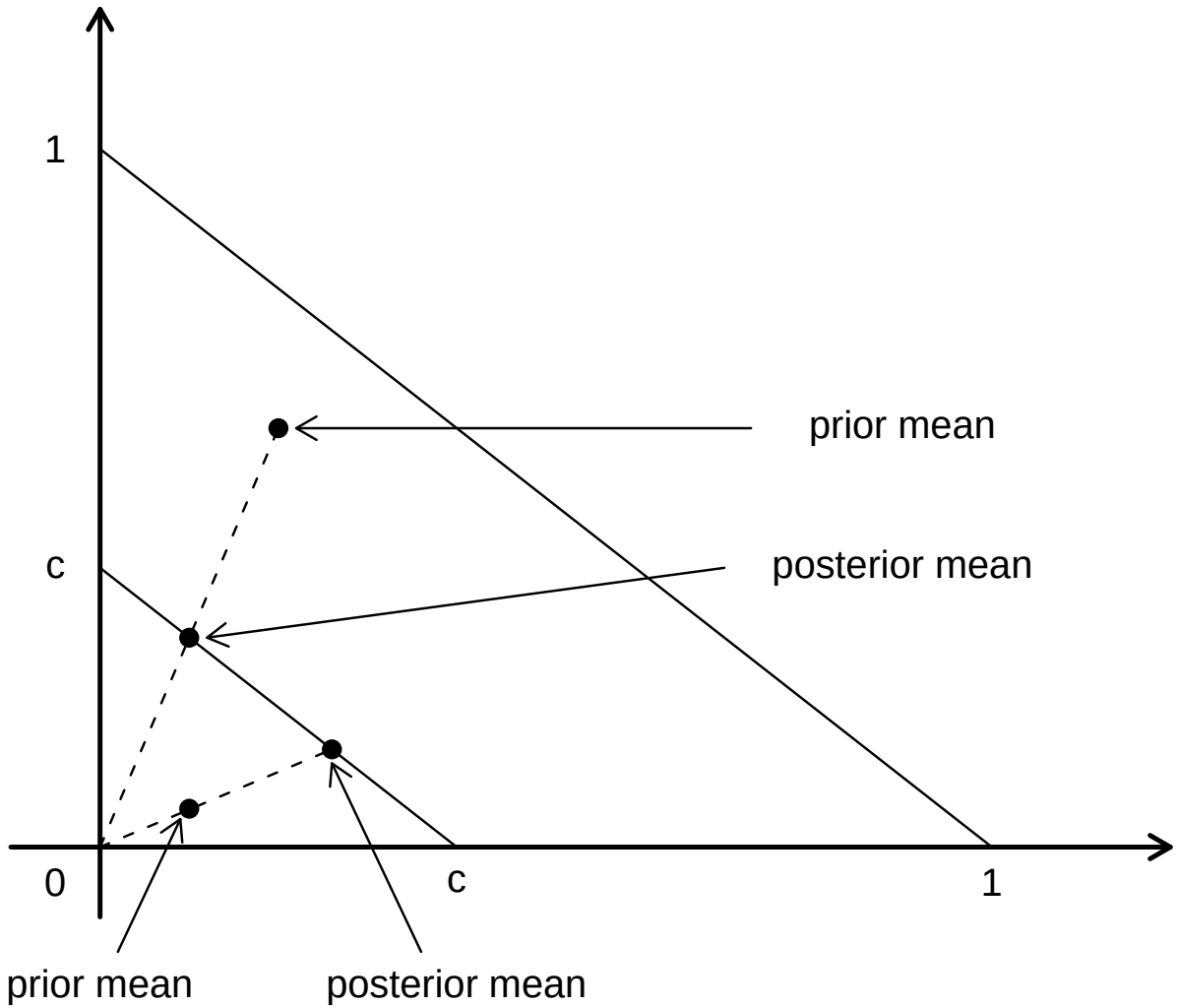


Figure 3 : Prior and limiting posterior means in the (p_1, p_2) plane, $p_1^* + p_2^* = c$.

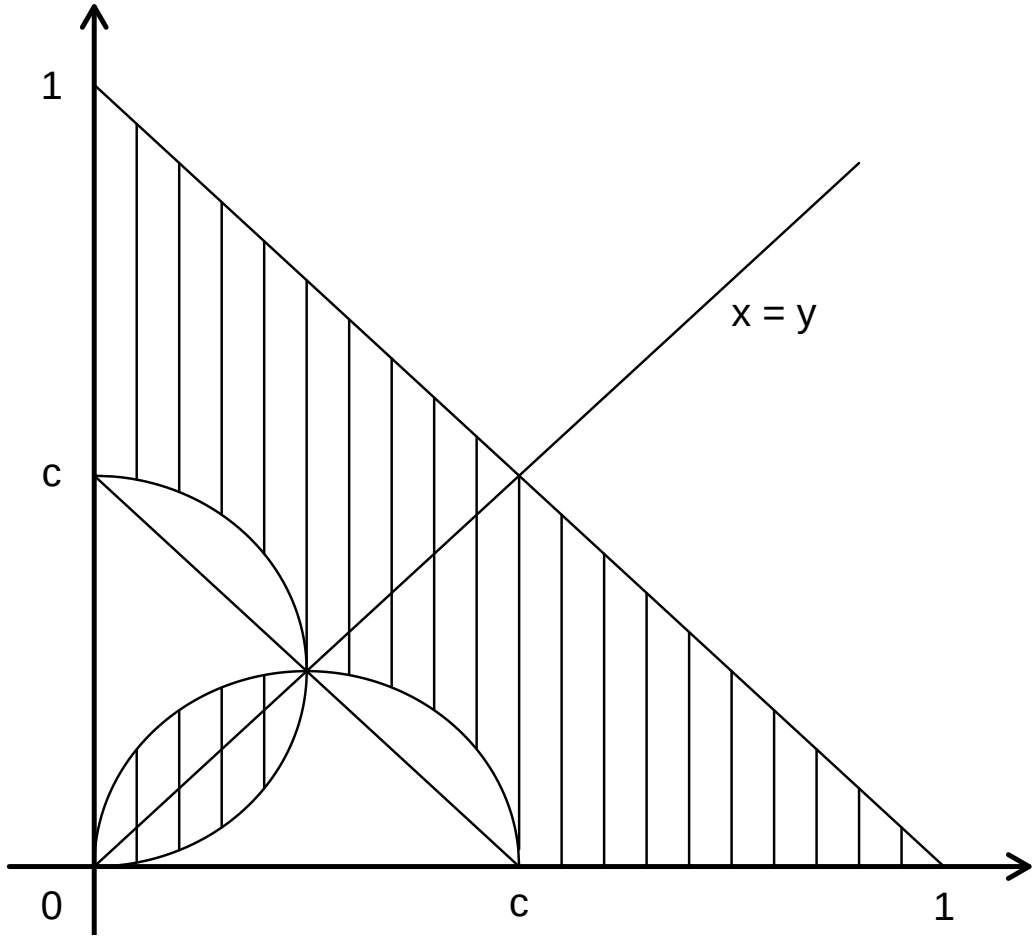


Figure 4 : Prior means (a, b) for which the limiting posterior mean $(\gamma a, \gamma b)$ is closer to all the points on the line $x + y = c$

Theorem 5

Let D_1 and D_2 be the Euclidean distances from the prior mean and the posterior mean to the true value of the paramenter pair (p_1, p_2). Then

$$\frac{\sqrt{2}}{2} \leq \frac{D_1}{D_2} \leq \infty \, ,$$

and these bounds are sharp.

XII: COMPARING BAYES AND FREQUENTIST INTERVAL ESTIMATES

- The Bayesian version of a confidence interval for θ is called a *credibility interval* for θ, and is obtained from the posterior distribution.
- For example, any interval (θ_L, θ_U) for which

$$\int_{\theta_L}^{\theta_U} g(\theta|\boldsymbol{x})d\theta = 1 - \alpha \tag{1}$$

  is a $100(1\alpha)\%$ credibility interval for θ.
- Usually, one uses the central credibility interval in which θ_L and θ_U satisfy

$$\int_{-\infty}^{\theta_L} g(\theta|\boldsymbol{x})d\theta = \frac{\alpha}{2} = \int_{\theta_U}^{\infty} g(\theta|\boldsymbol{x})d\theta.$$

- What can be said about the comparative performance of Bayesian credibility intervals and frequentist confidence intervals? Let's take a look in the binomial case. Let $X \sim B(n, p)$.

- This will give us the opportunity to take a look at the word-length experiment from the point of view of interval estimation.

- The story begins with the good old standard frequentist large-sample confidence interval for p, the interval with confidence limits

$$\hat{p} \pm \frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} Z_{\alpha/2}, \tag{2}$$

  where $\hat{p} = X/n$.

- We know that, if the binomial model is correct, the random interval from which the interval in (2) is calculated will contain p with approximate probability $1 - \alpha$.

- Interestingly, a growing number of statisticians recommend *never* using this interval! Instead, they recommend using the "plus four" confidence interval for p.

The plus-four confidence interval is a hybrid interval estimator that combines some Bayesian thinking with some frequentist thinking. To compute the plus-four interval for p, one simply replaces $\hat{p}$ in (2) by the alternative estimator $\tilde{p}$ of p given by

$$\tilde{p} = \frac{X+2}{n+4}, \tag{3}$$

yielding the "preferred" interval with confidence limits

$$\tilde{p} \pm \frac{\sqrt{\tilde{p}(1-\tilde{p})}}{\sqrt{n}} Z_{\alpha/2}. \tag{4}$$

- Note that $\tilde{p}$ in (4) is actually the Bayes estimator of p, under squared error loss, with respect to the $Be(2,2)$ prior.

- But the interval in (4) is not based on probabilities calculated from the posterior distribution of p.

- The interval in (4) is constructed by mimicking what the frequentist does, using the same form as the frequentist confidence interval, but with a substitute for $\hat{p}$.

- There's a ton of documentation that the hybrid confidence interval works better than the standard confidence interval
- The discreteness of X results in a rather unsteady probability of coverage (above and below the nominal coverage probability $1 - \alpha$) as the sample size varies.
- Why the hybrid estimator fixes this problem appears to have to do with the fact that $\tilde{p}$ shrinks $\hat{p}$ toward $1/2$.
- Because $\tilde{p}$ is closer to $1/2$ than $\hat{p}$, the confidence interval based on $\tilde{p}$ is actually slightly wider than the interval based on $\hat{p}$, so that it casts a wider net in trying to trap the unknown parameter p.

- The empirical evidence in support of $\tilde{p}$ is really quite overwhelming.

- Simulation studies have investigated the question: How large does n have to be to guarantee that the actual probability that a 95% confidence interval actually covers the true value of the parameter p with probability at least .94 for all samples of size n or larger? If $p = 0.1$, the required n for this guaranteed performance by the standard confidence interval is $n = 646$. Remarkably, the n required for this guaranteed performance by the plus-four confidence limits is $n = 11$.

- Case closed?

- Wait! There's another player here!! Let's call the interval estimates in (2) and (4) F and H.

- Let's now consider the Bayesian credibility interval B. The comparisons between F and H mentioned above are impressive but not definitive; analytical comparisons between F and H are quite challenging, and, the comparisons between B and F and B and H are even more challenging.

- I'll tell you what my colleague D. Bhattacharya and I have found out so far. It's not definitive either but I think you'll find it interesting.

- Let's look at how F, H and B did in the word-length experiment. You'll recall that 99 rag-tag Bayesians estimated the proportion p of long words in a Somerset Maughm novel. Each of them could put forward a central credibility interval for p.

- To ensure that the interval estimators F and H were not disadvantaged (by an inappropriate normal approximation) in a comparison with a given B, we increased the sample size n at which the comparison would be made to $n = 50$.

- In 1000 simulations in which $X \sim B(50, p)$ was generated, with $p = .3008$, we recorded the coverage probability P for each of the 99 intervals of type F, H and B and also recorded the width W of the resulting interval.

- Now, neither of these two measures is appropriate, by itself, to serve as a criterion for comparing interval estimates.

- A criterion that makes more sense than either of these is the ratio W/P.

- In the simulations we have done, we have estimated the ratio of the average width W divided by the frequency of coverage.

- On the basis of the criterion W/P, we found that H beat F 93 out of 99 times. Interestingly, H beat F in 100% of the cases when judged in terms of coverage probability alone.

- On the basis of the W/P criterion, B beat F 66 out of 99 times. This result is somewhat unexpected.

- Credibility intervals draw much more heavily on the posterior distribution.

- So the contest now reduces to the world-series of interval estimation, the contest between H and B.

- If you were inclined to make a wager at this point, which way do you think that comparison would come out? Although B beat F, H beat F quite a bit more emphatically. This suggests that the smart money would be backing H in this duel.

- Well, surprise of surprises, on the basis of the criterion W/P, we found that B beat H 65 out of 99 times.

- While this result might not be classified as a monumental triumph for the Bayesian, it is nonetheless provocative. After all, H is roundly accepted these days as the king of the hill!

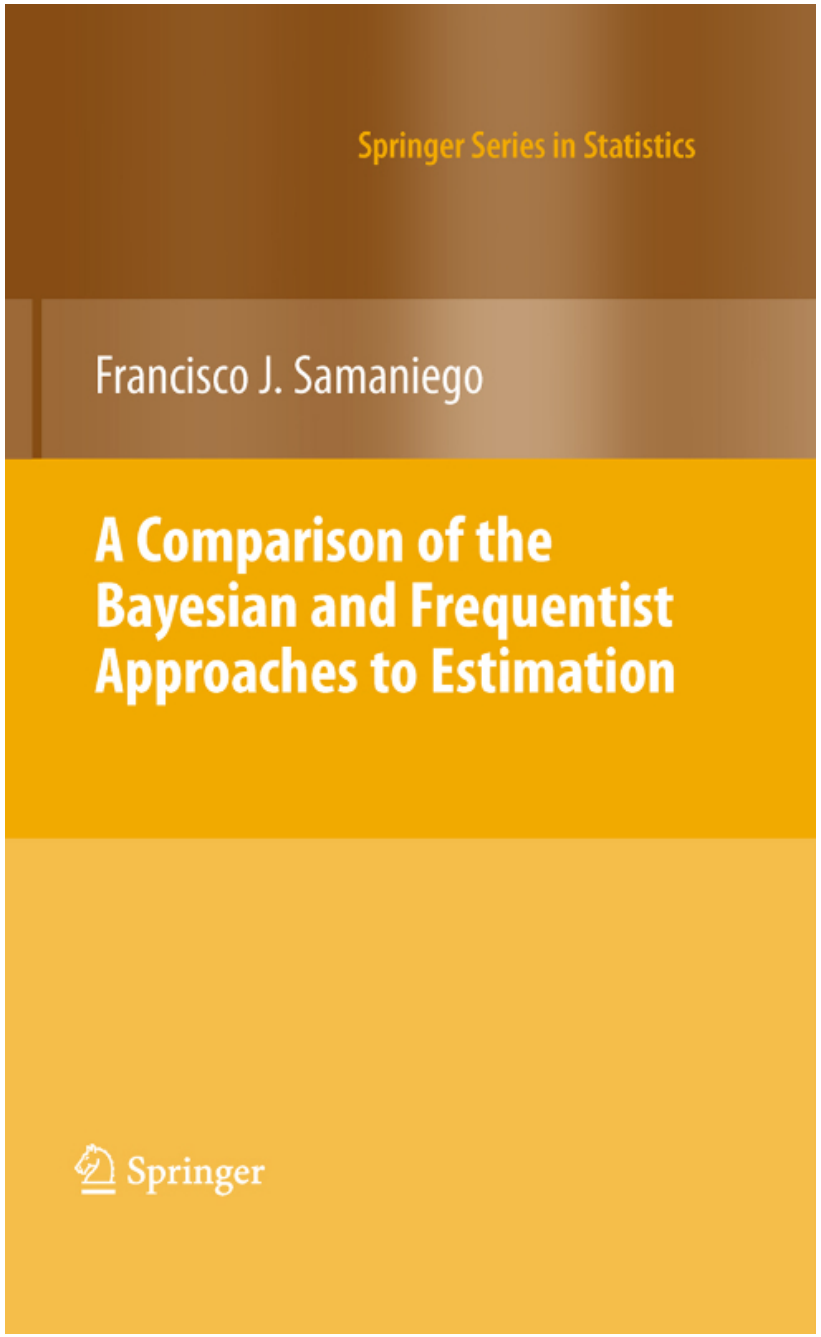
CONCLUSIION

- My own interpretation of these findings is that F is highly suspect, that H is clearly good and that B ought to be considered more seriously, as it obviously has some promise.

- Bayesian interval estimates seem to merit serious consideration and, possibly, substantially broader usage.

XIII: SUMMARY

- In general, our findings suggest that, in low dimensional problems of point estimation, Bayes procedures will often have expected performance superior to that of the best frequentist procedure.
- We have noted that, in certain contexts, the collection of Bayes estimators which outperform frequentist alternatives is substantially large, suggesting a certain natural robustness of Bayesian inference.
- When estimating high dimensional parameters, the prospects for the Bayesian are not so rosy.
- In our study of interval estimation, the Bayesian approach has surfaced as the Goliath! It's definitely worth further investigation.


Springer Series in Statistics
Francisco J. Samaniego
A Comparison of the Bayesian and Frequentist Approaches to Estimation
Springer