Objective Bayesian Analysis and Frequentist Statistics

- Subjective vs. Objective Bayes
- Objective Bayes vs. Frequentist
- Scientific / Statistical Modeling

History

Paradigm	Age (Years)
Objective Bayes	200
Frequentist statistics	80
Subjective Bayes	60

Subjective vs. Objective Bayes
– My View –

- When possible, subjective Bayes is great. (But also do objective Bayes as part of a sensitivity study.)
- If one cannot – or does not want to – do subjective Bayesian analysis, do objective Bayes.
- Most subjective Bayes analyses partly utilize objective Bayes methodology.

Example: Measuring Weak Counting Signa[...]

$X \sim$ Poisson$(\theta + \xi)$

mean signal rate   mean backgro[...]

$Y \sim$ Poisson(ξ)

Objective Bayes: Laplace would have chosen prior $\pi(\theta, \xi) = 1$.

<u>(Formal) posterior density of</u> (θ, ξ):

$$\pi(\theta, \xi \mid x, y) \propto p(x \mid \theta+\xi)\, p(y \mid \xi) \cdot 1$$

observed data   Poisson density

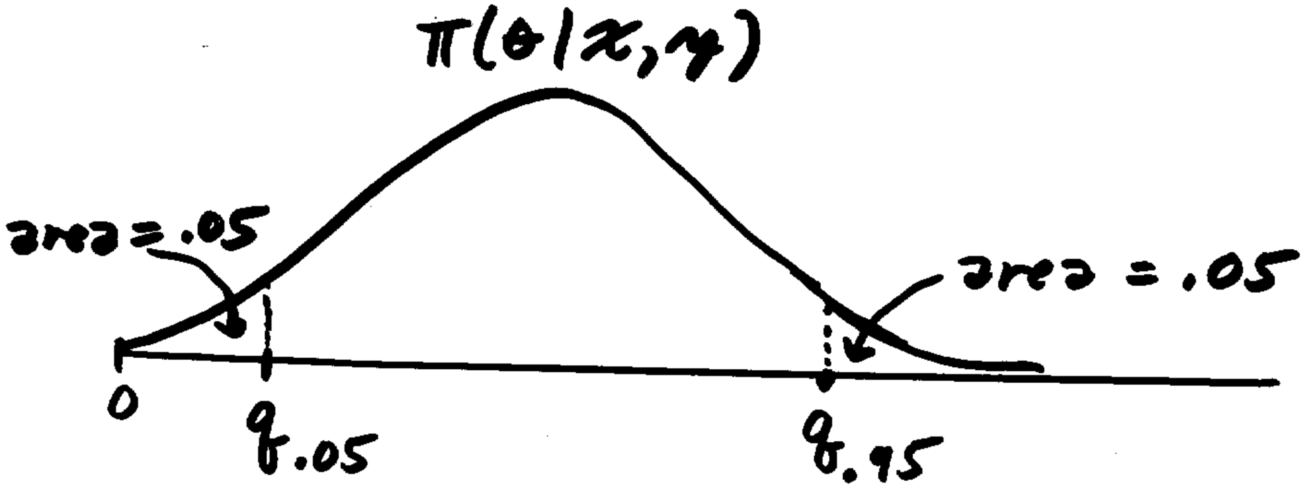
<u>Posterior density of</u> θ:

$$\pi(\theta \mid x, y) \propto \int_0^\infty p(x \mid \theta+\xi)\, p(y \mid \xi)\, d\xi$$

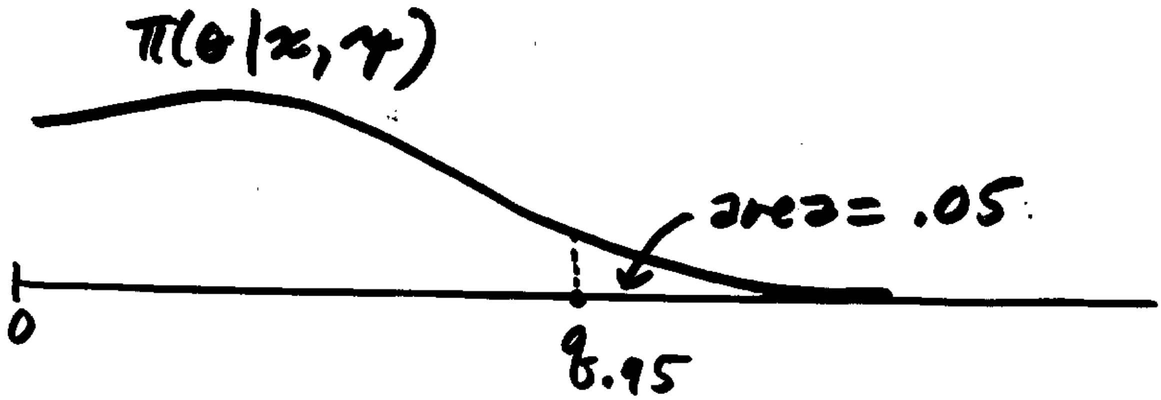
<u>Confidence limits:</u> The confidence limit, q_α, is defined by

$$\alpha = \Pr(\theta \le q_\alpha \mid x, y) = \int_0^{q_\alpha} \pi(\theta \mid x, y)\, d\theta.$$

Case 1. $x >> y$



Case 2. $x \approx y$



Note: If a confidence interval at level $1-\alpha$ is desired, it is typical to give the equal-tailed interval $(q_{\alpha/2}, q_{1-\frac{\alpha}{2}})$; in case 2, one might well prefer $(0, q_{1-\alpha})$, however.

(Partially) Subjective Bayes:

Past measurements of the variable mean background rate ξ suggested that ξ has mean 12 and std. deviation 6.

We might then choose the prior for ξ

$$\pi(\xi) = \frac{1}{486}\, \xi^3 e^{-\xi/3}$$

(the Gamma density with the given mean and standard deviation).

But little may be known about θ, so the objective prior $\pi(\theta)=1$ is used. Thus

$$\pi(\theta \mid x, y) \propto \int_0^\infty p(x \mid \theta+\xi)\, p(y \mid \xi)\, \xi^3 e^{-\xi/3}\, d\xi .$$

Objective Bayes and Frequentist Statisti

My operational (not foundational) goals for non-subjective statistical procedures:

Goal 1. Easy interpretability is desirable.

Goal 2. The procedure should have good conditional performance (i.e., be sensible for the data that is actually obtained).

Goal 3. The procedure should have good frequentist performance.

(I won't give the Bayesian motivations for this.)

Goal 1. Easy interpretability is desirable

Example: $C = (3.2, 9.6)$ is the obtained 95% confidence set.

<u>Bayesian:</u> The probability that θ is between 3.2 and 9.6 is 0.95.

<u>Frequentist:</u> Consider the confidence sets, $C(X)$, that would arise from any possible data X. Then, if one repeatedly used $C(X)$ for data generated from the distribution of X given θ, $C(X)$ would contain θ with probability 0.95.

Goal 2. Satisfactory Conditional Perform

One wants to ensure that the answer is sensible for the data at hand.

Example: $X_i = \begin{cases} \theta+1 & \text{with prob. } \frac{1}{2} \\ \theta-1 & \text{with prob. } \frac{1}{2} \end{cases}, \quad i=1,2.$

An optimal "point" confidence set is

$$C(x_1, x_2) = \begin{cases} (x_1+x_2)/2 & \text{if } x_1 \neq x_2 \\ x_1 - 1 & \text{if } x_1 = x_2 \end{cases}$$

<u>(Textbook) Frequentist Coverage</u>:

$$Pr\left(C(X_1, X_2) \text{ equals } \theta \mid \theta \right) = 0.75.$$

<u>Objective Bayes</u>: Choose $\pi(\theta)=1$,

$$Pr\left(\theta \text{ equals } C(x_1,x_2) \mid x_1, x_2\right) = \begin{cases} 1 & \text{if } x_1 \neq x_2 \\ \frac{1}{2} & \text{if } x_1 = x_2 \end{cases}$$

- The frequentist coverage measures the quality of the experiment.
- The objective Bayesian answer is clearly the one to be reported.
- One <u>must</u> tailor the answer to "lucky" and to "unlucky" data.

Note: Objective Bayesian analysis automatically sorts this out.

An Alternative: Conditional Frequentist

Example: Compute the coverage prob. conditional on the statistic $S = |X_1 - X_2|$.

$$Pr\left(C(X_1, X_2) \text{ equals } \theta \mid S, \theta\right) = \begin{cases} 1 & \text{if } S = 2 \\ \frac{1}{2} & \text{if } S = 0. \end{cases}$$

Note: When cond. freq. analysis is done right, answers typically close to obj. Bayes (Roe and Woodroofe, ...) but, in general

- finding a good S is very hard (much harder than finding a good objective prior)
- this must be handled in concert with dealing with the problem that coverage can be highly variable in θ.

Example: Suppose, also, that $0 \le \theta \le 10$.

- obj. Bayes again trivially gives right answer
- I'm not sure it is possible to find a good conditional frequentist analysis.

Goal 3. Satisfactory Frequentist Performa

In repeated use, a (say) nominal 95% confidence set should contain the unknown θ about 95% of the time (perhaps conditional on S).

Claim: The easiest way to get good frequentist confidence limits is to find "good" objective Bayes confidence limit

- use "optimal" objective priors;
- if a two-sided interval is desired, typically use the equal-tailed interval $(q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}})$.

Note: Good conditional performance is a "free" side benefit.

Example: $X \sim \text{Binomial}(n, \theta)$

X: # successes; n: # trials; θ: success probability

Textbook confidence interval for θ:

$$\hat{\theta} \pm z_{(1-\frac{\alpha}{2})} \sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}, \quad \hat{\theta} = \frac{X}{n}$$

$z_{(1-\frac{\alpha}{2})}$: standard normal quantile

Objective Bayes conf. interval for θ:

Choose $\pi(\theta) \propto \dfrac{1}{\sqrt{\theta(1-\theta)}}$.

Use the equal-tailed Bayesian interval

$$C(x) = \left(q_{\frac{\alpha}{2}}(x), \; q_{1-\frac{\alpha}{2}}(x) \right).$$

Brown, Cai, and DasGupta (1999) studied the frequentist performance of these and six other procedures, in terms of coverage probability and size.

Question: What are "optimal" objective priors?

My View: those that give the best frequentist answers.

<u>The Contenders:</u> let $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_p)$

Laplace: $\pi(\underset{\sim}{\theta}) \propto 1$

My evaluation: rarely gives bad answers.

Jeffreys-rule prior: $\pi(\underset{\sim}{\theta}) \propto \sqrt{\det(I(\underset{\sim}{\theta}))}$,

$I(\underset{\sim}{\theta})$ is the Fisher information matrix having (i, j) entry

$$I_{ij}(\underset{\sim}{\theta}) = E\left[\left(\frac{\partial}{\partial \theta_i} \log f(\underset{\sim}{x} \mid \theta)\right)\left(\frac{\partial}{\partial \theta_j} \log f(\underset{\sim}{x} \mid \theta)\right) \Bigg| \theta\right].$$

Binomial example: $I(\theta) = \dfrac{n}{\theta(1-\theta)}$.

$$\pi(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}}.$$

Poisson example:

$X \sim \text{Poisson}(\theta + \xi)$, $Y \sim \text{Poisson}(\xi)$,

$$I(\theta, \xi) = \begin{pmatrix} \frac{1}{(\theta+\xi)} & -\frac{1}{(\theta+\xi)} \\ -\frac{1}{(\theta+\xi)} & \left[\frac{1}{(\theta+\xi)} + \frac{1}{\xi}\right] \end{pmatrix}.$$

$$\pi(\theta, \xi) \propto \sqrt{\det(I(\theta,\xi))} = \frac{1}{\sqrt{\xi(\theta+\xi)}}.$$

Note: Jeffreys himself recommended $\pi^{0}(\theta, \xi) = \frac{1}{(\theta+\xi)\,\xi}$, but I think that's wors

My evaluation of the Jeffreys-rule prior

It is optimal for real-valued θ, but suspect for vector $\underset{\sim}{\theta}$.

Frequentist Matching Priors:

<u>Case 1.</u> Real-valued θ.

<u>Theorem</u> (Welch and Peers, 1965)

(Continuous) i.i.d. data $\underset{\sim}{x} = (x_1, \ldots, x_n)$.

$C(\underset{\sim}{x}) = (-\infty, q_{1-\alpha}(\underset{\sim}{x}))$ is the one-sided Bayesian confidence set at level $1-\alpha$.

(i) As $n \to \infty$ and for any prior $\pi(\theta) > 0$,

$$Pr\left(C(\underset{\sim}{X}) \text{ contains } \theta \,\middle|\, \theta \right) = 1 - \alpha + \frac{C(\theta)}{\sqrt{n}} + O\left(\frac{1}{n}\right).$$

(ii) $C(\theta) = 0$ if and only if $\pi(\theta) = \sqrt{I(\theta)}$.

My View: $\pi(\theta) = \sqrt{I(\theta)}$ is optimal here; objective Bayesians <u>and</u> frequentists cannot do any better.

<u>Case 2.</u> Vector-valued $\underset{\sim}{\theta}$.

- lots of theory and solutions
- matching priors are not unique

My View: to date, provides little practical guidance

Reference Priors (Bernardo (1979), B+B (90

Rough idea:

- Write $\pi(\underset{\sim}{\theta}) = \pi(\theta_1)\,\pi(\theta_2 \mid \theta_1) \cdots \pi(\theta_p \mid \theta_1, \dots \theta_{p-1})$
- Determine a suitable one-dimensional Jeffreys-rule prior for each $\pi(\theta_j \mid \theta_1, \dots \theta_{j-1})$
- After appropriate renormalization, multiply these together.

My evaluation: Reference priors yield procedures with stunning performance, from Bayesian, conditional, and freq. views. But: they are not easy to derive.

Other Objective Priors:

- Maximum entropy priors are usually very nice.
- In problems invariant w.r.t. some group action, right-Haar priors are the holy grail of objective priors.
- Dozens of others, ranging from okay to ridiculous.

Ad-hoc Objective Priors (other than $\pi(\underset{\sim}{\theta})=1$)

Often one can use known 'optimal' objective priors for components of a larger problem (but not automatically).

Example: Medical Diagnosis (Mossman, 1999)

$$\Pr(\text{Disease} \mid \text{test result}) = \frac{\theta_0 \theta_1}{\theta_0\theta_1 + (1-\theta_0)\theta_2},$$

$$X_0 \sim \text{Binomial}(n_0, \theta_0)$$
$$X_1 \sim \text{Binomial}(n_1, \theta_1)$$
$$X_2 \sim \text{Binomial}(n_2, \theta_2).$$

Goal: Find a confidence interval.

My suggestion: This is a case where the 'optimal' priors for each θ_i should work well together; so use

$$\pi(\theta_0, \theta_1, \theta_2) \propto \frac{1}{\sqrt{\theta_0(1-\theta_0)} \cdot \sqrt{\theta_1(1-\theta_1)} \cdot \sqrt{\theta_2(1-\theta_2)}}$$

and compute $\left(q_{\frac{\alpha}{2}},\ q_{1-\frac{\alpha}{2}}\right)$.

Numerical study: Comparison of $C_{Bayes} = (q_{\frac{\alpha}{2}}, q_{1-\frac{\alpha}{2}})$ with four sophisticat non-Bayesian confidence intervals.

Results for $n_0 = n_1 = n_2 = 20$; nominal level is 95%; presented are the proportions of misses in each tail (should be 0.025).

	$(\theta_0, \theta_1, \theta_2)$		
	(0.25, 0.75, 0.75)	(0.1, 0.9, 0.1)	(0.5, 0.9, 0.1)
C_{Bayes}	.034, .023	.024, .034	.024, .022
$C_{1/2}$	.014, .019	.000, .000	.000, .0[illegible]
C_S	.015, .014	.000, .000	.000, .02[illegible]
C_{GN}	.029, .017	.027, .009	.018, .02[illegible]
C_δ	.025, .028	.008, .008	.011, .014

Note: The C_{Bayes} intervals were, on the whole, considerably narrower.

III. Scientific / Statistical Modeling

- With limited data, one should not separate the processes of scientific and statistical modeling.
- To accurately estimate uncertainty in an answer, one should not
  - replace unknown parameters by estimates
  - ignore measurement errors

    ⋮
- In complex situations, use of some prior information is often necessary (the data cannot do everything).

My view: The only available methodology for any of these points is the Bayesian approach.

Example: Distance to Cepheid Stars
(with William Jefferys and Thomas Barnes)

- Radial velocities (unknown physics)
  - actual velocities modeled as Fourier or wavelet series
    - of <u>unknown</u> order K
    - with <u>unknown</u> phase ϑ
  - measured velocities
    - have variances known up to an unknown "inflation" factor
    - distributions can be Gaussian or t-distributions (for robustness)
- Photometry (mostly known physics)
  - nonlinear physical model relating photometry to radial velocity, depending on
    - unknown star angular diameter, ϕ_0
    - unknown distance to star, s
  - measurment errors modeled as above

- Prior distributions for unknown parameters
  - prior distribution for $s \propto s^\alpha$ $(\alpha \cong 1)$ (Lutz - Kelker bias)
  - other parameters (as many as 30) handled via objective priors.
- Computation
  - <u>no</u> approximations to models or simplifications needed.
  - done via reversible jump MCMC.
  - produces estimates of all desired quantities, together with measures of accuracy.

M. Kendall, giving the 'old' frequentist viewpoint of Bayesian analysis;

"If they [Bayesians] would only do as he [Bayes] did and publish posthumously, we should all be saved a lot of trouble."

What should be the view today;

Objective Bayesian analysis is the best frequentist tool around.