

Initial vs. Final Precision

A measure of *initial precision* for a decision rule is a quantity that is computed by averaging over all possible data sets that might be observed. Such measures can be computed before observing any data.

Examples:

- Risk $R(\theta, \delta)$
- Bayes risk $r(\pi, \delta)$
- Traditional frequentist measures such as variance, mean squared error and probabilities of type I and II errors

A measure of *final precision* is one that can be computed only after observing the data. Final precision is the accuracy with which it is felt the conclusion holds *conditional* on the data.

Recall the words on p. 1 of notes:

To a Bayesian, data sets which might have been, but were not, observed are irrelevant to making inferences about the unknown parameters. *The only data set of any relevance is the one that was actually observed.*

In other words, a Bayesian feels that *final* precision is the most appropriate way to measure precision of a decision rule.

How do we reconcile the Bayes principle with the Bayesian philosophy of inference, inasmuch as the Bayes principle is based on a measure of *initial* precision?

Need to consider the *normal* and *extensive* forms of finding a Bayes rule.

Normal form of analysis

Determine $r(\pi, \delta)$ for each δ and choose the δ for which $r(\pi, \delta)$ is a minimum. In general, this method will be very difficult to apply.

Extensive form of analysis

Suppose θ is continuous.

$$\begin{aligned} r(\pi, \delta) &= \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta \\ &= \int_{\Theta} \left[\int_{\mathcal{Y}} L(\theta, \delta(\mathbf{y})) f(\mathbf{y}|\theta) d\mathbf{y} \right] \pi(\theta) d\theta \\ &= \int_{\mathcal{Y}} \left[\int_{\Theta} L(\theta, \delta(\mathbf{y})) f(\mathbf{y}|\theta) \pi(\theta) d\theta \right] d\mathbf{y}, \end{aligned}$$

with the last step following from Fubini's theorem.

If we could find a rule that minimizes the integrand of the last expression *for each* \mathbf{y} , then that rule would also minimize the integral, i.e., $r(\pi, \delta)$.

Equivalently, we could seek, for each $\mathbf{y} \in \mathcal{Y}$, an action $a = \delta(\mathbf{y})$ that minimizes

$$\int_{\Theta} L(\boldsymbol{\theta}, a) \left[\frac{f(\mathbf{y}|\boldsymbol{\theta})}{m(\mathbf{y})} \right] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} =$$

$$\int_{\Theta} L(\boldsymbol{\theta}, a) \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}.$$

The very last expression is called the *posterior risk* of the action a .

The extensive form of analysis is usually much easier to utilize than the normal form. Instead of finding a function that minimizes an integral, we only need find an a that minimizes posterior risk.

Note that posterior risk is a measure of *final precision*. Letting δ^π be the Bayes rule, Bayes principle and being a Bayesian statistician are reconciled as follows:

For any given set of data \mathbf{y} , if the Bayesian statistician chooses an action that minimizes posterior risk (a measure of final precision), then he/she is taking the action $\delta^\pi(\mathbf{y})$, and hence is acting in accordance with Bayes principle.

Importantly, the Bayesian statistician need not know the definition of the Bayes rule for every point in \mathcal{Y} . He/she needs only determine the Bayes *action* for the given data \mathbf{y} .

Single-Parameter Models

We now proceed to the subject of Chapter 2 in Gelman, Carlin, Stern and Rubin (GCSR). This chapter will familiarize us, in a simple setting, with issues that arise in any Bayesian analysis.

Will observe a value of \mathbf{Y} that has distribution $f(\mathbf{y}|\theta)$, where θ is a scalar parameter.

We'll consider four common examples of this setting:

- Binomial experiment,
- Random sample from a normal distribution,
- Random sample from an exponential distribution, and
- Poisson data.

Binomial experiment

We observe $Y \sim \text{bin}(n, \theta)$, and so

$$f(y|\theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad y = 0, 1, \dots, n.$$

It is useful to note that the factor $\binom{n}{y}$ may be ignored for purposes of computing the posterior distribution. Note that

$$\pi(\theta|y) \propto \theta^y (1 - \theta)^{n-y} \pi(\theta),$$

and

$$\pi(\theta|y) = \frac{\theta^y (1 - \theta)^{n-y} \pi(\theta)}{\int_0^1 t^y (1 - t)^{n-y} \pi(t) dt}.$$

Depending on what sort of inference is to be done, it may not even be necessary to compute $\int_0^1 t^y (1 - t)^{n-y} \pi(t) dt$. More on this shortly.

What shall we use as a prior distribution? A commonly used family of distributions for the binomial model is the *beta* family $\{\pi(\theta|a, b) : a > 0, b > 0\}$.

$$\pi(\theta|a, b) = \frac{\theta^{a-1}(1 - \theta)^{b-1}}{B(a, b)} I_{(0,1)}(\theta)$$

This family provides fair scope for expressing uncertainty the experimenter has about θ . By varying a and b , one may obtain a variety of distributional shapes.

Various beta densities

