Bayesian Estimation

MFM Practitioner Module:
Risk & Asset Allocation

John Dodson

February 3, 2010
Outline

Bayesian Estimation
  Prior
  Pseudo-Data

Bayesian Estimator

Determining the Prior
In Bayesian estimation we do not endow the sample with a characterization; rather, we endow the parameters with a characterization, described by hyper-parameters. This is the prior characterization. We then update the characterization by conditioning on the observed data using Bayes’ Rule, which leads to the posterior characterization from which we can build estimates.

\[ f_{\theta|Y^{(N)}}(\theta) \propto f_{\theta}(\theta)f_{Y^{(N)}|\theta}(y^{(N)}) \]

- The Bayesian approach is inherently biased. The prior is ideally based on beliefs about the results before any data have been observed.
- The Bayesian approach is appropriate when the statistician is also a subject matter expert (such as you).
Prior

In principle, the characterization of the parameter prior can be completely arbitrary.

Conjugate Prior

But a judicious choice exists, which is to choose a prior from a family that this closed under updates. Such a prior is termed the conjugate prior.

With a conjugate prior, updating can be expressed as an algebraic transformation of the hyper-parameters, involving the prior values, the sufficient statistics, and the sample size.

Improper Prior

It can be useful to consider a prior that has no information, for example a prior whose density is uniform over the sample space of the parameters. This is termed an improper prior.
Important Example

There is a conjugate prior for the univariate normal $X \sim \mathcal{N}(\mu, \sigma^2)$. It has a somewhat complicated form, but it is nonetheless very useful. It is termed the normal-inverse Gamma distribution with hyper-parameters $\mu_0, \sigma_0^{-2}, \lambda_0, \nu_0$ and it is defined by the mixture

$$
\mu | \sigma^2 \sim \mathcal{N} \left( \mu_0, \frac{\sigma^2}{\lambda_0} \right) \quad \frac{1}{\sigma^2} \sim \mathcal{G} \left( \nu_0, \frac{1}{\sigma_0^2} \right)
$$

The posterior is in the same family with updated parameters based on the sufficient statistics $t_1 = \sum_i x_i$ and $t_2 = \sum_i x_i^2$

$$
\lambda_N = \lambda_0 + N \\
\nu_N = \nu_0 + N \\
\mu_N = \frac{\lambda_0 \mu_0 + t_1}{\lambda_0 + N} \\
\sigma^2_N = \sigma_0^2 + \frac{\lambda_0 \mu_0^2 + t_2 - (\lambda_0 \mu_0 + t_1)^2}{\lambda_0 + N}
$$
Important Example

The preceding can be generalized to the the multivariate normal, \( X \sim \mathcal{N}(\mu, \Sigma) \). We call the conjugate prior for the parameters \((\mu, \Sigma)\) the normal-inverse Wishart,

\[
\mu | \Sigma \sim \mathcal{N} \left( \mu_0, \frac{1}{\lambda_0} \Sigma \right) \quad \Sigma^{-1} \sim \mathcal{W} (\nu_0, T_0)
\]

with hyper-parameters \( \mu_0 \in \mathbb{R}^M, \ T_0 \in \mathbb{R}^{M \times M} \) positive definite, \( \lambda_0 > 0 \) and \( \nu_0 > M - 1 \) with \( M = \text{dim} \ X \). The hyper-parameters conditional on a sample \( x \in \mathbb{R}^{N \times M} \) are

\[
\lambda_N = \lambda_0 + N \\
\nu_N = \nu_0 + N \\
\mu_N = \frac{\lambda_0 \mu_0 + x' \cdot 1}{\lambda_N} \\
T_N = \left( T_0^{-1} + \lambda_0 \mu_0 \cdot \mu_0' + x' \cdot x - \lambda_N \mu_N \cdot \mu_N' \right)^{-1}
\]
Sampling from the Posterior

We will see in the next chapter that a point estimate for the market vector parameters is not adequate. We will need to work with a range of possible values for the market vector parameter $\theta$ that is consistent with our objective data and our subjective view.

Gibbs Sampling

Conjugate distributions for the parameters of a characterization may be easy to update, but they can be difficult to sample from. It turns out that if the conditional characterizations of the parameters which can be easily sampled, a nested iteration scheme converges to the joint characterization.

$$\lim_{i \to \infty} \theta^{(i)} \sim \theta$$

where

$$\theta_j^{(i)} \sim \theta_j \mid \theta_1^{(i-1)}, \ldots, \theta_{j-1}^{(i-1)}, \theta_{j+1}^{(i-1)}, \ldots, Y^{(N)}$$
Pseudo-data

A useful application of the conjugate prior is to imagine that prior itself is a posterior with respect to some imaginary (random) dataset \( \tilde{Y}(\tilde{N}) = (\tilde{X}_1, \ldots, \tilde{X}_{\tilde{N}}) \) where each \( \tilde{X}_i \) are drawn independently from a known distribution representing our beliefs.

\[
f_{\theta|Y(N)}(\theta) \propto f_{Y(N)}(y) f_{\theta}(\theta)
= f_{Y(N)}(y) \left( f_{Y(\tilde{N})}(\tilde{Y}) f_{\theta}^{\text{im}}(\theta) \right)
= f_{Y(N+\tilde{N})}(y \oplus \tilde{Y}) f_{\theta}^{\text{im}}(\theta)
\]

Application

If we want to simulate from the posterior, we simply append the actual sample with a pseudo-sample of variates drawn from the characterization of \( \tilde{X} \), then proceed with classical estimation (e.g. MLE).
Bayesian Estimator

Once we know the posterior distribution $f_{\theta | Y(N)}$, we still need to provide a result. Some naïve approaches include

- mode and modal dispersion, $\hat{\theta} = \text{arg max } f_{\theta | Y(N)}$
- mean and covariance, $\hat{\theta} = E(\theta | Y(N))$

A more sophisticated approach is to find the minimum-measure region for a given level of confidence $1 - \alpha$. In the univariate setting, this would be $(\theta_0, \theta_1)$ for

$$\arg \min_{\theta_0} \theta_1(\theta_0) - \theta_0$$

and $\theta_1(\theta_0) = Q_{\theta | Y(N)} \left( F_{\theta | Y(N)}(\theta_0) + \alpha \right)$.

Bayesian Loss Function

The ideal approach is to find a parameter value that minimizes the expected value of some loss function customized to the subsequent application.
Determining the Prior

The first step towards determining our subjective model for the markets is to codify our experience. Two indirect ways of approaching is to analyze

- model portfolio allocations
- constraints on acceptable allocations

Allocation-implied characterization

The clearest expression that we can observe of the opinion of a professional asset manager is the allocation he chooses. If we can interpret this allocation as the solution to a satisfaction optimization problem with respect to some measure of performance, we can attempt to invert that solution to determine the implied market characterization.

\[ \alpha^* = \arg \max_{\alpha} S(\alpha) \]
Allocation-implied characterization

Unfortunately, there are usually many more parameters required to characterize the market vector than the single weight per holding you would observe in a model portfolio $\alpha_0$.

**Example**

Consider exponential utility with $\zeta > 0$ and $M \sim N(\mu, \Sigma)$. Here we have

$$S(\alpha) = \alpha' \cdot \mu - \frac{1}{2\zeta} \alpha' \cdot \Sigma \cdot \alpha$$

If you could fix $\Sigma = \bar{\Sigma}$ and you knew that $\alpha_0$ was optimal, then you could infer the mean of the market vector,

$$\mu = \frac{1}{\zeta} \bar{\Sigma} \cdot \alpha_0$$