MAXIMUM A POSTERIORI ESTIMATION

Overview—

In Bayesian statistics, a maximum a posteriori probability (MAP) estimate is an estimate of an unknown quality, that equals the mode of the posterior distribution. The MAP can be used to obtain a point estimate of an unobserved quantity on the basis of empirical data. It is closely related to the method of maximum likelihood (ML) estimation, but employs an augmented optimization objective which incorporates a prior distribution (that quantifies the additional information available through prior knowledge of a related event) over the quantity one wants to estimate. MAP estimation can therefore be seen as a regularization of ML estimation.

Description—

Assume that we want to estimate an unobserved population parameter θ on the basis of observations x. Let f be the sampling distribution of x, so that $f(x|\theta)$ is the probability of x when the underlying population parameter is θ. Then the function $\theta \mapsto f(x|\theta)$ is known as the likelihood function and the estimate $\hat{\theta}_{MLE}(x) = arg\ max_\theta f(x|\theta)$ is the maximum likelihood estimate of θ.

Now assume that a prior distribution g over θ exists. This allows us to treat θ as a random variable as in Bayesian statistics. We can calculate the posterior distribution of θ using Bayes' theorem:

$\theta \mapsto f(\theta|x) = \frac{f(x|\theta)g(\theta)}{\int_\Theta f(x|\vartheta)g(\vartheta)d\vartheta}$ where g is density function of θ. Θ is the domain of g.

The method of maximum a posteriori estimation then estimates θ as the mode of the posterior distribution of this random variable

$$\hat{\theta}_{MAP}(x) = arg\ max_\theta f(\theta|x) = arg\ max_\theta \frac{f(x|\theta)g(\theta)}{\int_\theta f(x|\vartheta)g(\vartheta)d\vartheta} = arg\ max_\theta f(x|\theta)g(\theta)$$

The denominator of the posterior distribution (so-called marginal likelihood) is always positive and does not depend on θ and therefore plays no role in the optimization. Observe that the MAP estimate of θ coincides with the ML estimate when the prior g is uniform (that is, a constant

function).

When the loss function is of the form

$$f(n) = \begin{cases} 0, & \text{if } |a - \theta| < c \\ 1, & \text{otherwise,} \end{cases}$$

as c goes to 0, the Bayes estimator approaches the MAP estimator, provided that the distribution of θ is quasi-concave. But generally a MAP estimator is not a Bayes estimator unless θ is discrete.

Computation—

MAP estimates can be computed in several ways:

- Analytically, when the models of the posterior distribution can be given in closed form. This is the case when conjugate priors are used.
- Via numerical optimization such as the conjugate gradient method or Newton;s method. This usually requires first or second derivatives, which have to be evaluated analytically or numerically.
- Via a modification of an EM algorithm. This does not require derivatives of the posterior density.
- Via a Monte Carlo method using simulated annealing.