**Bayesian method**

33 PI:

Bayesian inference

* Infer p(Y | X, D) from:
  + a given parametric model or likelihood function p(Y | X, Θ)
  + a prior distribution p(Θ) over the model parameters
  + a data set D ≡ (X,y)
* calculate posterior parameter distribution p(Θ|D)
  + “inference”
  + Often ass
  + 2 Methods: Variational inference, Monte Carlo integration
  + BTYDplus uses Monte Carlo integration: pnbd.mcmc.DrawParameters
* based on this parameter distribution, calculate the outcome distribution p(y|x,D)
  + “prediction”
  + mcmc.DrawFutureTransactions
* point estimate: E(y|x,D)
* intervals: quantile(y|x,D)

48 PI:

* “The resultant prediction intervals also differed according to the adopted prior distributions. […] Thus, if frequentist accuracy is required, Bayesian prediction intervals should be used cautiously in practice.”

49 PI:

* They use non-informative Bayesian prediction intervals and bootstrap
* Bayesian PIs perform better than bootstrap Pis
* Bootstrap is frequentist approach
* Bayesian is not, it uses an uninformative prior distribution
* Bayesian good for small sample sizes, at least in this case (conclusion)

51PI:

* “This contrasts with the Bayesian approach, discussed more fully below, where is a random variable.”
* “a parameter such as p is assumed to have a probability distribution so that it now makes sense to say that the probability of p falling within a given interval is 95%”
* “quantifies knowledge of, and more specifically uncertainty about, p before the data are observed”
* “The prior distribution is multiplied by the likelihood function (and normalized by the marginal distribution of the data) to give a posterior distribution.”
* Advantage: “Bayes intervals are appealing because they have an easier interpretation than the usual frequentist confidence intervals”
* Disadvantage: “However, their main drawback is that it is necessary to provide a prior distribution for p. If good, quantifiable, prior knowledge is available, then Bayes intervals are a sound choice; otherwise, different researchers may come up with different prior distributions. The fact that these different prior distributions may lead to different intervals, as we see below, can be a serious drawback. Prior ignorance is often represented by an uninformative prior, but even here different choices are possible for what is meant by “uninformative,””

42PI:

* Difference between Bayesian and frequentist approach

Used by

* 48 PI: Biometrical data, implement different methods for Bayesian interval estimation
* 49 PI: Constituent in river water

**Ensemble methods**

33PI, 45PI

* Naïve approach:
* Train x models, take the distribution from the results
* Calculate desired metrics from this distribution
* Useful when larger number of models
* Not applicable/useful, here

**Quantile regression**

33PI

* Direct interval estimation
* “This loss function tries to balance the number of data points below and above the (estimated) quantile.”
* Not really a loss function but a replacement for it
* “this definition is equivalent to that of the minimizer of the average of all absolute residuals”
* “High quality principle”? not implemented

47PI

* “Just as we can define the sample mean as the solution to the problem of minimizing a sum of squared residuals, we can define the median as the solution to the problem of minimizing a sum of absolute residuals.”
* “simply giving differing weights to positive and negative residuals—would yield the quantiles”? not implemented
* Origin of quantile regression: Koenker and Bassett 1978

54PI

* Origin of quantile regression

**Method 1**

Sventunkov

* “basic conventional statistical ways of capturing uncertainty about estimates of parameters is via the calculation of the covariance matrix of parameters”
* (can be derived from Hessian matrix or via bootstrap)
* Rectified normal distribution because of cutting parameters on the boundary?
* “simulation-based approach […] that relies on the selected model form” (they are doing it for additive and multiplicative models but for multiplicative models, it is barely possible to estimate the uncertainty (-propagation), as is the case for the pnbd-model

**Conformal prediction**

Introduction

Classification

Regression

Time series

Implementation