|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | prior | | | posterior | |
|  | | | | |  | | |  | |
|  | | | | |  | | |  | |
|  | | | | |  | | |  | |
|  | | 不知, 知 | | | *,* | | |  | |
|  | | predictive | | |  | | | 是后验均值和方差 | |
|  | | 不知, 知 | | |  | | |  | |
|  | | 知, 不知 | | |  | | |  | |
|  | | 知, 不知 | | |  | | |  | |
|  | |  | | |  | | |  | |
|  | | , 不知 | | | prior |  | | | |
|  | | | | | posterior |  | | | |
| exclusive | | |  | | | | exhaustive | |  |
| Precision | | |  | | | |  | | |
| CV or  coefficient of variation | | |  | | | |  | | |
| variance | | |  | | | |  | | |
| Standard deviation | | |  | | | |
| Non-informative prior | | |  | | | |  | | |
| Hierarchical model | | | (层次模型) 参数的参数也有分布 | | | | | | |
| Specify , , , desired quantiles(1/2) for prior to get hyperparameters | | | | | | | | | |
| 优缺点: | numerical quadrature methods： are great in low dimensional settings,  not so good in high dimensional. | | | | | | | | |
|  | Gaussian 近似: behave bad if we have a small sample and posterior doesn’t look very normal | | | | | | | | |
|  | Rejection Sampling | | | a very general method that can be used for any distribution | | | | | |
|  |  | | | Relatively easy to implement | | | | | |
|  |  | | | for Bayesian inference for —don’t need normalizing constant | | | | | |
|  |  | | | If can find an envelope distribution, g(θ), where M is small, can be very efficient. | | | | | |
|  |  | | | If M is large, the rejection rate can be quite high, thus inefficient. | | | | | |
|  |  | | | difficult to find a good envelope for high dimension | | | | | |
|  | importance | | | don’t need normalizing constant | | | | | |
|  |  | | | don't discard anything. | | | | | |
|  |  | | | hard to find important sampler g(θ) in higher dimensions | | | | | |
|  | Importance: calculate integral ; Re: generate samples | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Conditional Probability | | | | | | discrete | | | | |  | | | | | | | | | | | | |
|  | | | | | | continuous | | | | |  | | | | | | | | | | | | |
| likelihood | | | | | | or | | | | | | | | | | | | | | | | | |
| posterior | | | | | |  | | | | | | | | | | | | | | | | | |
| Predictive distribution | | | | | | prior | | | | |  | | | | | | | | | | | | |
|  | | | | | | posterior | | | | |  | | | | | | | | | | | | |
| Improper prior | | | | | | are acceptable so long the resulting posterior is not improper | | | | | | | | | | | | | | | | | |
| Exponential family dist | | | | | | pdf/pmf | | | | | | |  | | | | | | | | | | |
|  | | | | | | Conjugate prior | | | | | | |  | | | | | | | | | | |
| Mixture prior | | | | | | , where and | | | | | | | | | | | | | | | | | |
|  | | | | | | , where | | | | | | | | | | | | | | | | | |
| Jeffrey’s prior  (uniformative) | | | | | | , where | | | | | | | | | | | | | | | | | |
| 换元 | |  | | | |  | | | | | | | | | | | | | | | | | |
|  | | | | | |  | | | | | | | | | |  | | | | | | | |
|  | | | | | |  | | | | | | | | | |  | | | | | | | |
|  | | | | | |  | | | | | | | | | |  | | | | | | | |
|  | | | | | |  | | | | | | | | | | , where | | | | | | | |
| Jeffrey’s prior for  multivariate parameter | | | | | | Score function:; | | | | | | | | | | | | |  | | | | |
| KL divergence | | | | | | or | | | | | | | | | | | | | | | | | |
| KL divergence between  prior and posterior | | | | | |  | | | | | | | | | | | | | | | | | |
| reference prior | | | | | | KL divergence between prior and posterior should be as large as possible | | | | | | | | | | | | | | | | | |
|  | | | | | | maximizes | | | | | | | | | | | | | | | | | |
| Point estimates | | | | | | mean平均; median中位数; mode众数pdf最高的(求导=0) | | | | | | | | | | | | | | | | | |
| Loss function | | | | | | Loss of taking action under state | | | | | | | | | | | | | | | | | |
| Risk | | | | | | Expected loss or | | | | | | | | | | | | | | | | | |
| Risk based on data | | | | | |  | | | | | | | | | | | | | | | | | |
| Posterior risk | | | | | |  | | | | | | | | | | | | | | | | | |
| Bayes estimator of | | | | | | The value of that minimizes Posterior risk: | | | | | | | | | | | | | | | | | |
| square error loss | | | | | |  | | | | | | | | | | | | 期望 | | | | | |
| absolute error loss | | | | | |  | | | | | | | | | | | | 中位数 | | | | | |
|  | | | | | | 中位数 | | | | | | | | | | | |  | | | | | |
| 0-1 loss | | | | | |  | | | | | | | | | | | | 众数 | | | | | |
| Interval estimates | | | | | |  | | | | | | | | | | | |  | | | | | |
| Bayesian Credible Intervals | | | | | | | | | | | | | | |  | | | | | | | | |
| Symmetric Credible Intervals | | | | | | | | | | | | | | |  | | | | | | | | |
| Highest posterior density intervals (HPDIs) | | | | | | | | | | | | | | |  | | | | | | | | |
| classical hypothesis testing | | | | | | | | | | | | | | | Assuming is true, if 小概率 can extreme as , reject | | | | | | | | |
|  | | | | | | | | | | | | | | | Normal: | | | | | | | | |
| Bayesian Hypothesis testing | | | | | | | | | | | | | | | | | | | | | | | |
| Simple: | | | | | |  | | | | | | | | | | | | | | | | | |
| Composite: | | | | | |  | | | | | | | | | | | | | | | | | 成立在整个 |
| S-C: | | | | | |  | | | | | | | | | | | | | | | | | |
| Bayes factor | | | | | |  | | | | | | | | | | | Posterior odds: | | | | | |  |
|  | | | | | | |  | | | | | | | | | |  | | | | | | |
|  | | | | | | |  | | | | | | | | | |  | | | | | | |
|  | | | | | | | ; ; 两个加起来是分母 | | | | | | | | | | | | | | | | |
| Bayesian Central Limit Theorem | | | | | | | |  | | | | | | | | | | | | | | | |
|  | | | | | | | | : posterior mode (MAP) | | | | | | | | | | | |  | | | |
| Monte Carlo integration method | | | | | | | |  | | | | | | | | | | | | | | | |
|  | | | | | | | |  | | | | | | | | | | | | | | | |
|  | | | | | | | |  | | | | | | | | | | | | | | | |
| Direct Sampling | | |  | | | | | | | | | | | | | | | | | | | | |
|  | | |  | | | | | | | : | | | | | | | | | | |  | | |
| Monte Carlo error | | | |  | | | | | | | | | | | | | | | | | | | |
|  | Using Central Limit Theorem: ; is MC var; is MC error | | | | | | | | | | | | | | | | | | | | | | |
| Inverse PIT | | | (continuous) | | | | | | ; where and is CDF for | | | | | | | | | | | | | | |
|  | | | (discrete) | | | | | |  | | | | | | | | | | | | | | |
| Rejection sampling | | | | | Target distribution:; envelope: , if | | | | | | | | | | | | | | | | | | |
|  | | | | | , for all | | | | | | | | | 1. 生成 from | | | | | | | | | |
|  | | | | |  | | | | | | | | | 1. 生成 from | | | | | | | | or sample from | |
|  | | | | |  | | | | | | | | | 1. 只取 | | | | | | | | or 取 | |
|  | | | | | acceptance rate: | | | | | | | | |  | | | | | | | |  | |
| Importance Sampling | | | | | |  | | | | | | | | | | | | | | | | Importance ratio: | |
|  | | | | | | ; sample from | | | | | | | | | | | | | | | | | |
| Sampling Importance Re-Sampling (SIR) | | | | | | | | | | | | 1. Choose , generate samples from | | | | | | | | | | | |
|  | | | | | | | | | | | | 1. Calculate corresponding , then | | | | | | | | | | | |
|  | | | | | | | | | | | | 1. 放回抽样 n个 , with probability | | | | | | | | | | | |
| Markov chain | | | transition probability matrix, tpm, ;one-step transition probability | | | | | | | | | | | | | | | | | | | | |
|  | | | are the same for times called *time homogenous*, otherwise *time inhomogenous* | | | | | | | | | | | | | | | | | | | | |
| Marginal probability distribution for the state at time : | | | | | | | | | | | | | | | | | | | | | | | |
| is stationary distribution for if | | | | | | | | | | | | | | | | | | | | | | | |
| If a Markov chain with tpm is irreducible and aperiodic and it has a stationary distribution , then  • is unique, i.e., there is only one stationary distribution.  • The limiting distribution is the stationary distribution: | | | | | | | | | | | | | | | | | | | | | | | |
| Metropolis-Hastings | | | | | | target distribution: ; proposal distribution: (like envelope) | | | | | | | | | | | | | | | | | |
|  | | | | | | 6. Repeat the above for N iterations and (based on diagnostics) discard the first B iterations. | | | | | | | | | | | | | | | | | |
|  | | | | | |  | | | | | | | | | | | | | | | | | |
| random walk proposal | | | | | | , where are old, are candidate, or or | | | | | | | | | | | | | | | | | |
| symmetric proposal | | | | | | ; so | | | | | | | | | | | | | | | | | |
| Multidimensional | | | | | | Entire vector update at once or one-at-a-time | | | | | | | | | | | | | | | | | |
|  | | | | | |  | | | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | | | | | | | | | |
| MCMC diagnostics | | | | | | can’t prove convergence but can check possible lack of convergence | | | | | | | | | | | | | | | | | |
|  | | | | | | • Generate multiple chains with different initial value and draw trace plots to see if the chains are mixing well and overlapping at some point.  • Calculate the BRG statistic for convergence to see if the value is close to 1.0.  assess convergence by comparing chain variation to the (average) within chain variation  need two or more chains to be generated (不收敛时大于一)  • Look at autocorrelation plots assessing potential severity of autocorrelation  • Compare the effective sample sizfactorial(14)e to the number of iterations | | | | | | | | | | | | | | | | | |
| BGR statistic | | | | | |  | | | | | | | | | | | | | | | | | |
| Effective sample size | | | | | | ; , auto-correlation for lags | | | | | | | | | | | | | | | | | |
|  | | | | | | the higher the correlation is, the smaller effective sample size is. | | | | | | | | | | | | | | | | | |
| Gibbs Sampler | | | | | | with target distribution | | | | | | | | | | | | | | | | | |
|  | | | | | | full conditional distribution for : | | | | | | | | | | | | | | | | | |
|  | | | | | | Initialise the chain with | | | | | | | | | | | | | | | | | |
|  | | | | | | At iteration given the values , generate as follows: | | | | | | | | | | | | | | | | | |
|  | | | | | | Generate from ; Generate from ; … | | | | | | | | | | | | | | | | | |
|  | | | | | | Generate from ; from ; | | | | | | | | | | | | | | | | | |
| improving performance | | | | | | | 1. Changing the proposal distribution. | | | | | | | | | | | | | | | | |
|  | | | | | | | 2. Reparameterising the model. e.g. to reduce autocorrelation | | | | | | | | | | | | | | | | |
|  | | | | | | | 3. Block updating: and are highly correlated, generate simultaneously | | | | | | | | | | | | | | | | |
| Gibbs Sampler Strengths | | | | | | | The proposal distribution is automatically defined (thus no “tuning” of a proposal) | | | | | | | | | | | | | | | | |
|  | | | | | | | Always “accept” the candidate value | | | | | | | | | | | | | | | | |
|  | | | | | | | proposal is conditional on the other parameters, so can view it as an “adaptive” algorithm | | | | | | | | | | | | | | | | |
| Gibbs Sampler weaknesse | | | | | | | The conditional distributions may not be tractable | | | | | | | | | | | | | | | | |
|  | | | | | | | sampling from the conditionals may be computationally intensive | | | | | | | | | | | | | | | | |
|  | | | | | | | 100% acceptance rate does not necessarily mean good mixing | | | | | | | | | | | | | | | | |
| M-H Strengths | | | | | | | Proposal distribution quite flexible, can be fast to sample from and  fast to evaluate MHR (particularly with symmetric proposals) | | | | | | | | | | | | | | | | |
|  | | | | | | | Don’t need to know the conditional distributions | | | | | | | | | | | | | | | | |
|  | | | | | | | Block updating can be relatively easy; e.g., multivariate t distribution as a proposal. | | | | | | | | | | | | | | | | |
| M-H Weaknesses | | | | | | | May be hard to find a good proposal, one with good mixing | | | | | | | | | | | | | | | | |
|  | | | | | | | Can take time to “tune” a proposal, e.g., a good variance value for a normal random walk proposal | | | | | | | | | | | | | | | | |
| Need of Markov chains | | | | | | | Reversible chain; detailed balance equation | | | | | | | | | | | | | | | | |
| MCMC strength | | | | | | | For bayes inference, don't need to know the normalising concept. | | | | | | | | | | | | | | | | |
|  | | | | | | | Work well in high dimensions | | | | | | | | | | | | | | | | |