Number of Book
II nd used

B.S. 2071

TRIBHUVAN UNIVERSITY

INSTITUTE OF SCIENCE & TECHNOLOGY

No. 34787

EXAMINATION BOARD

KIRTIPUR

Book I

Students are required to write their answers on both sides & a margin of about $1^1/_4$ inches should be left on each page.

Master's Level

Year/Part II

Centre SMS TU

Roll No. 06

Roll No. In Words Six

Arjun Saplota

Reg. No. ....

Subject Monte Carlo Methods

Paper ....

Script English.

Date 2081/02/12

INSTRUCTIONS TO CANDIDATES

[1] Each Candidate will write legibly on the title page his or her ROLL NO. REGISTERED NO. AND THE SCRIPT in which answers are written but not his or her name and the name of the campus from which he or she appears. This should be done in each answer - book used before beginning to write inside.

[2] No loose papers will be provided for scribbling and no other paper should be brought in for this purpose. Any candidate found with loose paper in his or her possession WILL BE EXPELLED. All work must be done in the book provided and pages MUST NOT BE TORN OUT. The book provided CANNOT BE REPLACED BY ANOTHER but, if necessary, an additional book will be given. All work intended for the examiner must be written ON BOTH SIDES of the paper. Anything cancelled will not be looked into. Should a torn leaf be discovered inside an answer book, it should not be removed but crossed out and folded after bringing it to the notice of the invigilator.

[3] Candidate is forbidden to write answers or anything else on the question - paper.

[4] No Candidate will be allowed to leave the room until one hour has passed from the time when the papers are distributed.

[5] Candidate who uses two or more answer-books should see, before handing over to the invigilator, that they are properly stitched together.

INSTRUCTIONS TO THE EXAMINER

1. Aggregate marks of each question should be placed in the given appropriate box.
2. Marks for parts of a qusetion should be totalled under the question inside the answer- book.

MARKS OBTAINED

1st Q	3	11th Q	
2nd Q	3	12th Q	
3rd Q	3	13th Q	
4th Q	3	14 th Q	
5th Q	2	15th Q	
6th Q	6	16th Q	
7th Q	6	17th Q	
8th Q	6	18th Q	
9th Q	5	19th Q	
10thQ	6	20th Q	

43 TOTAL

Signature of Scrutiny Board

Signature of Examiner

Group 'A':

1. Bayesian Inference.

Bayesian Inference is one of the type of statistical approach in statistics which incorporates the prior information of an event. It completely depends on Bayes Theorem which is given as,

$$P(A|B) = \frac{P(B/A).P(A)}{P(B)}$$

where,

$P(A|B)$ = posterior probability
$P(A)$ = prior probability
$P(B/A)$: likelihood.
$P(B)$ = marginalization factor

3

In Bayesian Inference, there is always a prior information of an event in order to get the future happening probability.

Example:

Let us consider a fair coin, if we want to get the probability of getting head then we must have a prior assumption on the event in order to get the posterior probability. Let 50% be the assumed probability of getting head, then with this prior information we can easily get the probability of getting head with Bayes Theorem.

Thus, Bayes Inference is very important and useful approach in modern Data Analysis techniques.

CamScanner

2. The motivating vignetts of Statistical inferences are:..

① personal probability
② Missing Data
③ Bio cell Identification, etc.

① Personal probability

With the statistical inferences, we can obtain the personal probability values. let us consider a scholar whose journal should be published, in initial attempt he/she assume that it will get published with 30% chance and later it get published later in second research journal he/she assumes 100%. it will get published in this way the probability of publishing changes each time. this is a simple illustration of personal probability.

② Missing Data.

Missing data can be identified with the help of statistical inference. Example:

2 6 7 13 * 19

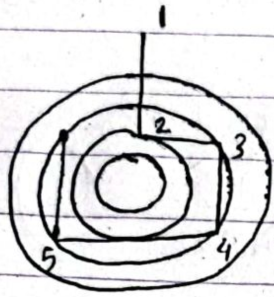
Here, we can find the missing '*' with the help of statistics. If we know the total sum or we have a pattern of certain numbers then we can easily get the missing data.

③ Bio cell identification

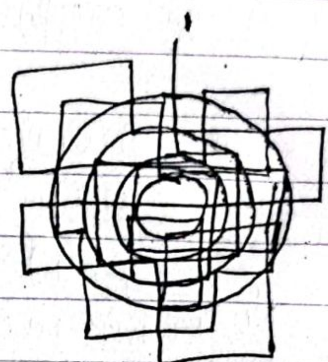
Identifying the bio cells, Medical treatments can be done with the data of Bio incorporated with the statistical inference we can do many more statistical analysis when we have data.

3. Reparametrization.

Reparametrization in Gibbs sampling is the technique which is used to solve the bottleneck problem. Optimizing the gibbs sampling there comes an important role of reparametrization let us consider.



After 5 iteration



After 50 iteration.

In 50th iteration of Gibbs sampling the convergence becomes slow and when the iteration increases the convergence will be slow and the computational complexity will also gets increases. To solve this bottleneck issue, we reparametrize the variables.

Here the circle variables X and Y in their specified range of uniform distribution gets reparametrized into another variables U and V. The range of X will be reparametrized to the range of U and similarly Y will reparametrized to V. The reparametrized variables are now performed with the gibbs sampling. And the result will be interesting the problem of slow convergence gets solved here with reparametrization.

Therefore, reparametrization helps to optimize the gibbs sampling by increasing the convergence speed in the sampling process.

4. Prior and posterior distribution of Bayesian Statistics

We have from Bayes theorem.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

i.e. $$\pi(\theta|x) = \frac{f(x|\theta) \cdot \pi(\theta)}{m(x)}$$

Here, $\pi(\theta|x)$ is the posterior distribution
$\pi(\theta)$ is the prior distribution.
$f(x|\theta)$ is the likelihood function

3

Now Differenciability,

Prior Distribution	Posterior distribution.
(i) It is the distribution of the prior belief	It is the distribution after the event has been happened.
(ii) It is based on assumption of prior Information.	It is the computed values with the help of prior information
(iii) From Bayes theorem above, $\pi(\theta)$ gives the probability distribution of prior belief.	from above Bayes Theorem $\pi(\theta|x)$ is the posterior probability distribution.
(iv) Assumed value or prior belief is simply a prior distribution.	It requires prior assumed value, likelihood & marginal total probability to get the posterior distribution.

5. Gibbs Sampling

Gibbs sampling is the sampling technique which is used when there is difficult to perform the normal sampling. Gibbs sampling is originated from the Image processing where the multivariable data needs to be sampled.

In the case of $P(x, y)$ it is difficult to sample the multivariable distribution. But in conditional distribution like

$P(x|y)$ } it will be easier to use the gibbs
$P(y|x)$ } sampling technique here.

Sampling the conditional distribution of x given y and sampling the conditional distribution of y given x, these can be easily performed in Gibb's sampling.

2

In Gibbs sampling, we sample the interest of distribution $\pi(\theta)$. where $\theta = \theta_1, \theta_2 \ldots \theta_d$. with the conditional distribution. Here we iterate the conditional distribution samples until we get the convergence.

Gibbs sampling can be applied in various statistical models, Machine learning Models to predict the data. Linear regression, Logistic regression, hierarchical model are some application models of Gibbs sampling.

Group 'B' : LIT

10. Given,

claim = 5% defective

inspector feels 10% defective

We have to decide wheather to accept or reject the shipment based on θ (proportion of defective parts)

prior assumption = 50-50 chance to both suggested θ

i.e. prior $(\pi(\theta)) = 50\% = 0.5$

i.e. $\pi(0.05) = \pi(0.10) = 0.5$.

sample (n) = 20 , defective = 3

Now

posterior distribution of θ = ?

First we get the likelihood values as,

$f(x|\theta=0.05) = F(3|\theta=0.05) - F(2|\theta=0.05)$

$= F_1 - F_2$ — 0.1886

$= a$

$=$

assuming variable since no probability given in question

Similarly,

$f(x|\theta=0.10) = F(3|\theta=0.10) - F(2|\theta=0.10)$

$= F_3 - F_4$

$= b$

same as abov

Again we calculate the total marginal probability as,

$$m(x=3) = f(x|\theta=0.05)\,\pi(0.05) + f(x|\theta=0.10)\,\pi(0.10)$$
$$= a \times 0.5 + b \times 0.5.$$
$$= 0.05(a+b).$$

finally, the posterior distribution of θ is,

$$\pi(\theta|x) = \frac{f(x|\theta=0.05).\,\pi(0.05)}{m(3)}$$

i.e. $$\pi(\theta=0.05|x=3) = \frac{f(x|\theta=0.05).\,\pi(0.05)}{m(3)}$$
$$= \frac{a.0.5}{0.05(a+b)}.$$

6

$$= \frac{10a}{a+b}.$$

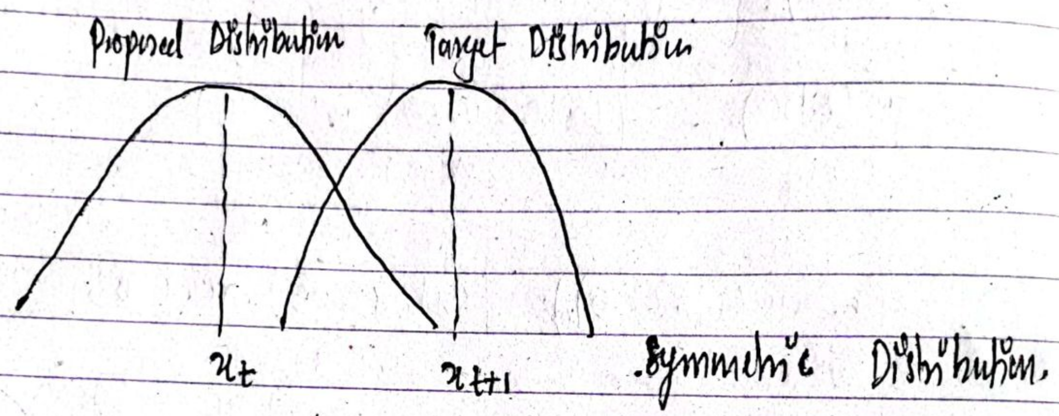
And,

$$\pi(\theta=0.10|x=3) = \frac{f(x|\theta=0.10)\,\pi(0.10)}{m(3)}$$
$$= \frac{b.\,0.5}{0.05.(a+b)}.$$
$$= \frac{10\,b}{a+b}.$$

Thus
$\pi(\theta=0.05|x=3)$ and $\pi(\theta=0.10|x=3)$ are the required posterior distribution.

6. Changes carried out by Hasting in Metropolis Algorithm

Metropolis algorithm was first introduced by the metropolis where we can sample multivariate distribution problem easily with the help of Target distribution.
i.e.
Let us consider a distribution plot.

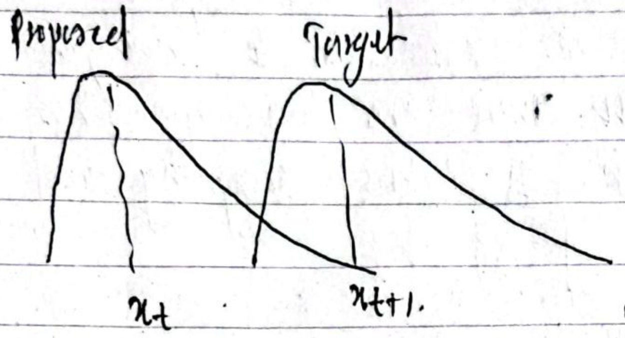


In Metropolis we can predict the sample with the parameters of Target distribution x_{t+1} for getting x_t. Here the distribution should be symmetrical and it is more restrictive sampling due to symmetricity of the distribution.

The acceptance criteria is $\min(1, \alpha)$ where the acceptance ratio, $\alpha = \frac{p(x^{(i)})}{p(x^{(t)})}$

Here, $p(x^{(i)})$ is the candidate value and $p(x^{(t)})$ is the target value.
Based on this the new samples gets generated.

However, Hasting has made some changes in the metropolis Algorithm. Here the distribution must not required to be in the symmetric, it can be done in asymmetric distribution as well. this has speed up the sampling technique.
i.e.



Asymmetric Distribution

In Metropolis Hasting, we can sample the states based on the target asymmetric distribution parameter x_{t+1} to get x_t. Hasting has made the algorithm more flexible and the samples will get generated faster than the prior Metropolis algorithm.

The acceptance criteria also gets updated by Hasting as $\min(1,\alpha)$ where

$$\text{acceptance ratio } (\alpha) = \frac{P(x^{(t)})\, q(x^{(c)})}{P(x^{(c)})\, q(x^{(t)})}$$

6

Here, $x^{(t)}$ are the target values and $x^{(c)}$ are the candidate values.

These are the changes done by Hasting which has increased the performance of the algorithms.

7. Convergence criteria in Gibbs sampling.

Gibb's sampling can be used to model the large number of iterations in multivariable distribution. But this will increase the computational cost so the samples also needs time to converge at the certain region. There are two approaches in the convergence criteria in Gibbs sampling and they are as follows:

(i) Theoritical ~~[illegible]~~ approach
(ii) Statistical assumption.
(iii) Graphical method.

(i) Theoritical approach
In this approach to test the convergence of the Gibb's sampling we measure the distance between the samples. Once the distance is measured then we can compare the distances in each ergodic process. When new samples generated by the gibbs sampling then the computed distance are compared and the convergence is finalized with the comparative analysis of the distance variation between the successive samples generated. This is a bit complex approach then the other ones.

(ii) Statistical assumption.
In order to test the convergence with this

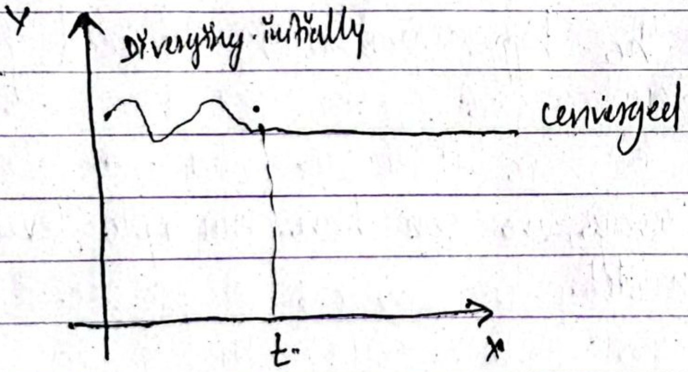
method. the statistical hypothesis are made on the observed data to test the convergence. This method will not guarantees the convergence. There it is only assumed on the basis of observed data so, it is simple but not as stronger than the theoretical approach of converge diagnostics.

(iii) Graphical Method,

We also have another visual method in which we can identify the convergence of the samples. Here we plot the graph of the samples to test for the convergence,

Let us consider a graph

6



From above graph we can easily visualize the fluctuations of the samples. if there are minimal fluctuations as the samples iteration are increasing then we can conclude it as a convergence of the samples. from above graph after time t the samples are getting converged.

Thus, we can measure the convergence in the above explained method in Gibb's Sampling.

8.

Algorithm to apply Gibb's Sampling to study Hierarchical model.

In Hierarchical model, we may have various level of models in there distribution linked with another distribution as parameter or hyper-parameters.

let us consider a model.

$$x \sim f(x|\theta)$$

here, θ is the parameter having the following distribution as,

$$\theta \sim g(\theta|\lambda)$$

where,

λ is the hyper-parameter obtained from the prior distribution.

In the same way, we can have multiple levels in this hierarchical model.

For the application of Gibbs' sampling in this hierarchical model, we first consider the level upto two in the model.

Algorithm can be designed as follows for the Gibb's sampling in hierarchical model!

Steps.

1. First of all we can initialize the initial values of the variables of the hierarchical model.

2. Then we can iterate the sample generation with the help of conditional distribution of the variable of the hierarchical model.

Each variables from another conditional variable distribution are sampled.

Here we update the sample data to collect it.

3. Step 2 is repeated for the large number of samples for multiple iteration

4. After convergence, the burn-in period samples are discarded and the remaining samples are taken for further analysis.

6

In this way we can apply Gibbs sampling to study hierarchical model.

Here $x \sim f(x|\theta)$ and $\theta \sim g(\theta|\lambda)$ with this level of hierarchy we developed the samples with the help of Gibb's sampling. This helps to converges the samples faster in short time.

9. Convergence of a Metropolis - Hasting Sampler

Convergence checking condition for metropolis sampler is same as the condition that we use in Gibbs sampling. All the formal and informal methods of convergence diagnosis are applied here in the metropolis-Hasting Algorithm.

Metropolis Hasting algorithm is the sampling algorithm which is used when we know about the target distribution. Based on the acceptance criteria we sample the data with the help of this metropolis hasting Algorithm. Metropolis Hasting Algorithm works on the asymmetrical distribution of the sample data so it is more flexible and it converges faster as well.

In Metropolis Hasting Algorithm we do not need the normalizing constant as we required in Gibbs sampling. The samples can be drawn easily with the Markov process in a chain without the involvement of the normalizing constant. The Markov chain Monte Carlo property can be applied to draw the samples and we can get them upto the convergence time.
In order to check the convergence in this metropolis hasting algorithm we use the following given techniques.

No. 157001

Code No. :

Code No. :

Code No. :

TRIBHUVAN UNIVERSITY

Faculty of Humanities and Social Sciences

OFFICE OF THE DEAN

Code No. ....

EXAMINATION ANSWER BOOK II

Students are required to write their answers on BOTH SIDES of each page and clearly mention the answer number.

Particulars below must be filled up immediately after the receipt of this book. This book must be stitched to the 1st answer sheet before submitting the answer copy.

Name - Arpan Saphuta
Roll No - 06 (stx)

Department ....

Semester 3rd.

Symbol No. 06.

EXAMINATION ....

Course Title Monte Carlo Methods

Course Code No. MDS 607

....

Signature of Invigilator

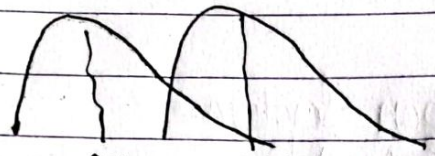
P.T.O.

Metropolis-Hasting Algorithms convergence can be checked as.

(i) Theoritical Study:

The samples generated with the Metropolis-hasting sampler can be analyzed by computing the distance between the samples.

Let us consider a deep inside scenario. Example:



Here distance between the two samples x_t & x_{t+1} i.e. $x_t - x_{t+1}$ can be computed to the ergodic process for all the samples. The obtained distance can be compared with other distances and can observe the variance in the distance. If the Variance are increasing then the samples has not been conversed else the samples has been conversed.

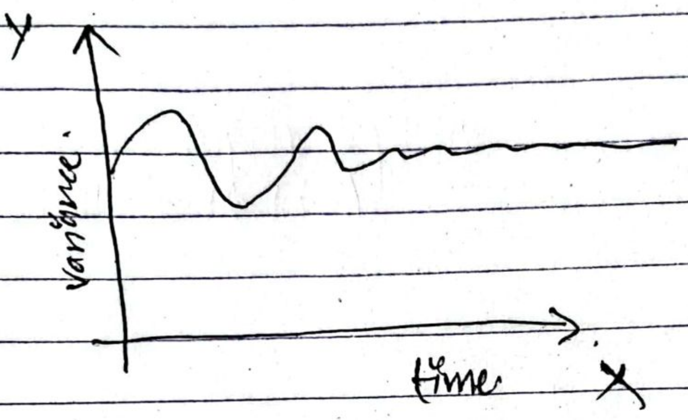
(ii) Statistical Analysis.

We can analyze the samples statistically. Statistical inference can be done by setting up the hypothesis and then the hypothesis can be tested. This method of checking convergence will not be 100%. sure that the samples are conversed because it is based on the present observed data only. But we can somehow get an idea of convergence with the statistical analysis.

(iii) Graphical Visualization.

With the visualization plot of the samples obtained from Metropolis hasting can be tested for the convergence.

Example:



Here the variance are decreasing as the time increases, so we can confirm it as the sample are getting conversed at time t.

In this way, we can test the convergence of the samples generated from Metropolis hasting algorithm. This is also similar to the case for Gibbs Algorithm.