MAT II - MCMC

Sebastijan Trojer
MAT II 2024/25 , FRI, UL
st5804@student.uni-lj.si

I. Introduction

II. Methodology and results

A. Numerical Integration vs. Monte Carlo Integration

1) Trapezoid rule: We used the expected value $E[X] = \int_0^{100} x \cdot p(x)\,dx$ of a PERT-distributed random variable (RV). The parameters were specified as follows: $a = 0; b = 10; c = 100$. We used the trapezoid rule and iteratively refined the approximation until the error was smaller than 5×10^{-5}. The final approximation was **23.333283** and it required **271** function evaluations to converge.

2) CLT Sample Estimation for MC: We used the Central limit theorem to estimate the number of points required to obtain the estimate of $E[X]$ to 2 decimal places using Monte Carlo integration when sampling from the uniform distribution on $[a, c]$. First, we draw an initial sample of size $n_0 = 1000$ and compute $y = x \cdot p(x)$ to estimate the sample standard deviation $\hat{\sigma}$. Then, applying the formula for the confidence interval of the sample mean,

$$n \geq \left(\frac{z(c-a)\hat{\sigma}}{\text{error}} \right)^2,$$

with $z = 1.96$ for 95% confidence, we calculate the required sample size $n_{\text{req}} \approx 5.72 \times 10^7$.

3) Monte Carlo estimation verification: We verified the estimation using numerical samples. We drew 15 different samples, each using approximately 57 million draws from the uniform interval and all of them had error between 0.001 and 0.0048, confirming that our estimate was correct. The estimate was computed by sampling 57 million points from the interval, and then using them to estimate $x \cdot p(x)$. To verify the stability we did another 100 evaluations. The variance between the estimates was small, and all of the estimates were accurate up to the second decimal ($\epsilon < 0.005$)).

4) Adaptive trapezoid rule vs. MC integration: Adaptive trapezoid rule achieved higher accuracy with only 271 function evaluation, whereas the MC integration required over 57 million samples per evaluation and took longer to compute. Therefore in this case the trapezoid rule is more accurate and fast, especially since we have a smooth integrand and a simple domain. The MC integration is better suited for more complex integrations and requires more samples in order to achieve low variance.

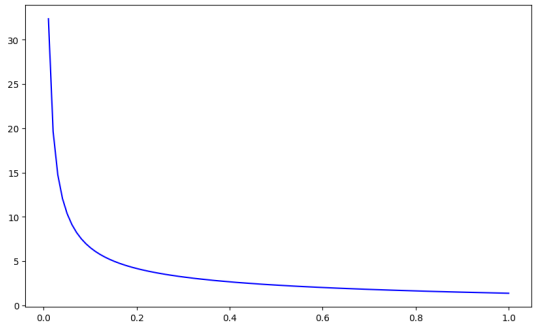

Fig. 1. Integrand $x^{-\frac{3}{4}} \cdot e^{-x}$ on interval from 0 to 1.

B. Importance sampling

We want to compute the integral $I = \int_0^1 x^{-3/4} \cdot e^{-x}\,dx$. Figure 1 shows the integrand on the interval from 0 to 1. It is a steep function that has most of the integral values around 0, with long tail.

1) Monte Carlo integration: The integral I was estimated using MC integration. We sampled from the uniform distribution with sample size of 10^7 and computed the mean and standard deviation of 10 repetitions. The mean was **3.363956 ± 0.088954 (SD)**.

2) Estimation via Importance Sampling: We estimated I using importance sampling, using proposal distribution $q(x) = cx^{-3/4}$. First we had to determine c such that $q(x)$ was a density, so we computed the integral $I = \int_0^1 x^{-3/4}\,dx$ without the constant, which evaluates to 4, meaning to make it a density, c had to be set to $\frac{1}{4}$.

In the next step, we sampled from $q(x) = \frac{1}{4}x^{-3/4}$ using sample size of 10^7. To do that we first needed to compute its CDF, which is $Q(x) = x^{1/4}$ and its inverse $Q^{-1}(p) = p^4$. Then we performed sampling as $x = p^4$ where p is uniformly distributed RV on (0, 1). The integral was estimated using the following formula: $I = \frac{1}{n}\sum_{i=1}^{n} \frac{f(x_i)}{q(x_i)}$ with f being the original integrand and q the proposal distribution. The average estimate was **3.379365 ± 0.000103**.

3) Uniform vs. inverse sampling: The variance is significantly lower in the inverse sampling case, compared to the uniform sampling method. Given that both methods had 10 estimations, we can see that the inversion sampling is much more stable, however, it requires a proposal function $q(x)$. The reason uniform sampling method underpreformed is that the function has the majority of its mass near 0,

however since we sample from the uniform interval we do not account for that, and most samples are in the areas with small values $f(x)$.

C. Metropolis-Hastings algorithm

Let X be a RV following the (α, η)-Weibull distribution with pdf $p(x|\alpha,\eta) = \alpha\eta x^{\alpha-1} exp(-x^{\alpha}\eta)$. Given a prior $\pi \propto exp(-\alpha - 2\eta)$ and observations $x = (0.3, 0.5, 0.75, 0.4)$, we would like to approximate the posterior distribution $p(\alpha,\eta|x) \propto p(x|\alpha\eta)\pi(\alpha,\eta) = f(\alpha,\eta)$. To do this we sampled from the distribution with density proportional to $f(\alpha,\eta)$. We computed the mean and the variance of α and η using Metropolis-Hastings algorithm with several multivariate normal proposals, with different covariance matrices, and a proposal distribution $q(\alpha',\eta'|\alpha,\eta) = \frac{1}{\alpha\eta} exp(-\frac{\alpha'}{\alpha} - \frac{\eta'}{\eta})$, which is a product of the densities of two exponential distributions, meaning the proposals are independent, and we can use inversion sampling to sample from each distribution independently.

For each of the above scenarios we generated 5 independent chains of 1000 samples and applied standard MCMC diagnostics, such as autocorrelation plots, ESS and trace plots. We tested three different covariance matrices for multivariate normal distribution: $\Sigma_1 = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$ and $\Sigma_3 = \begin{bmatrix} 0.1 & 0.01 \\ 0.01 & 0.1 \end{bmatrix}$, corresponding to low covariance, large covariance and correlation, respectively. Each chain was started at the point $(1, 1)$. We computed the posterior mean and variance and got the following results (using the high covariance MVN and exponential proposals):

Distribution	$E[\cdot]$	**Var**$(\cdot)$
MVN: α	1.74599	0.66866
MVN: η	1.73846	0.68305
Exp: α	1.78268	0.63512
Exp: η	2.08235	0.81352

TABLE I

POSTERIOR MEANS AND VARIANCES FOR α AND η UNDER MVN AND EXPONENTIAL PRIORS.

In the following, we take a closer look at some common MCMC diagnostics. Table II shows the ESS values for different distributions and Figure 2 shows autocorrelation plots for α, η for all the distributions, averaged across the chains. Notice that the lowest autocorrelation corresponds to the distributions with highest ESS values and the MVN with high covariance is the best prior, as it has lowest variance and autocorrelation and the largest ESS value, indicating good posterior exploration.

While both proposal distributions perform reasonable, the MCMC diagnostics and the results indicate that the normal proposal performs better with this sample size and in order to improve the performance of the exponential proposal we would have to increase the number of samples. The main downside of the normal proposal is that the covariance

Distribution	Σ	ESS α	ESS η
MVN	Σ_1	226	56
MVN	Σ_2	**570**	**391**
MVN	Σ_3	177	122
$f(x)$	N/A	436	281

TABLE II

ESS FOR α, η FOR ALL THE TESTED DISTRIBUTIONS. LARGER COVARIANCE LEADS TO HIGHEST ESS FOR BOTH PARAMETERS, MEANING A BETTER AND MORE EFFICIENT POSTERIOR EXPLORATION.

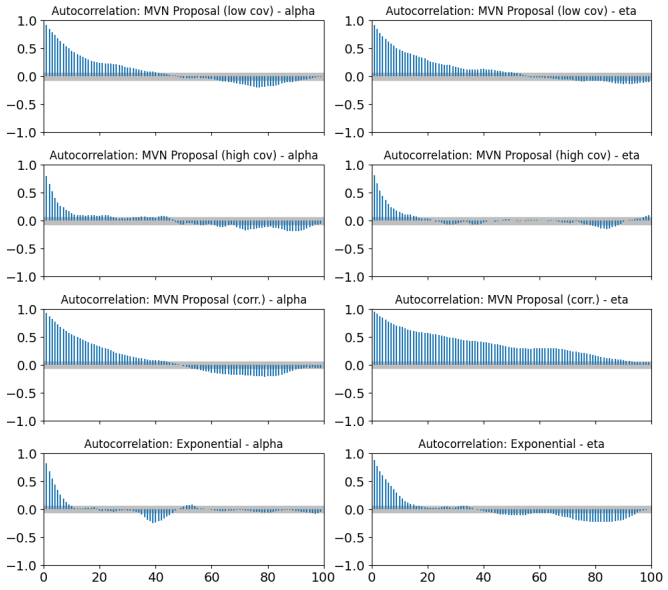

Fig. 2. Combined autocorrelation plots for α and η for each proposal, over all 5 chains (averaged). The lowest autocorrelation was achieved by the high covariance MVN and exponential proposal, corresponding to best ESS values in Table II

matrix has to be manually set to an appropriate value, or the posterior exploration quality will degrade. The autocorrelation plots became relatively stable relatively fast for both proposals, however that is less true for the exponential proposal. A downside of the exponential proposal was setting the acceptance rate, since it was more unstable and it needed some more fine tuning. An additional improvement would be to start each chain in a randomly selected point, which would better explore the posterior, however the results, as they are, indicate that even starting in a single point the estimation is feasible and

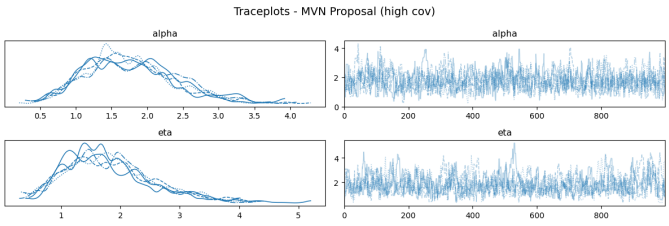

Fig. 3. Posterior and traceplot of the multivariate proposal sampling. The traceplot doesn't indicate any strong autocorrelation, which is confirmed by the autocorrelation plot on Figure 2.

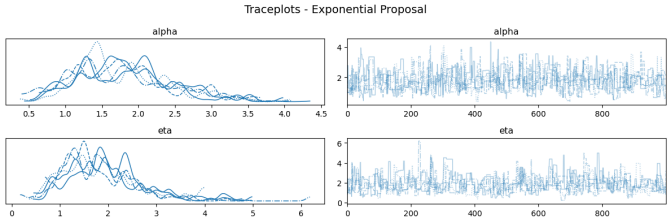

Fig. 4. Posterior and traceplot of the exponential prior. The traceplot indicates stronger autocorrelation compared to the multivariate normal and the posterior is defined more poorly (higher variance within and across chains), indicating the multivariate prior is a better choice.

starting in different points can be replaced with increasing the sample size.

1) Posterior probability on a given interval: We estimated the posterior probability $(\alpha, \eta) \in [1.3 \times \infty) \times [1.3 \times \infty)$ using two methods:

Metropolis–Hastings sampling: Based on the MCMC samples, we estimated the probability as**0.61460**.

Importance sampling: We used an exponential proposal distribution and computed a weighted average of the unnormalized posterior. This yielded an estimate of **0.56053**.

This indicates that the probability of both parameters being ≥ 1.3 is between 56 and 61%.