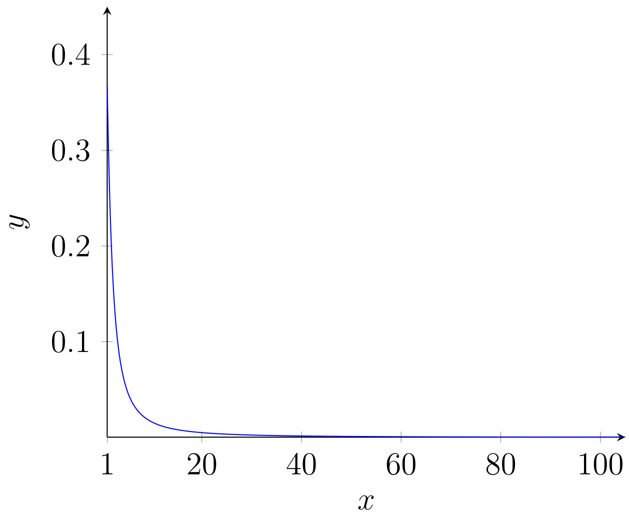
1 Importance sampling

In this problem, we would like to approximate the integral

$$I = \int_1^\infty x^{-7/4} e^{-1/x}\ dx.$$

1. First, we plot the integrand on the domain $[1, 100]$:



2. Let f be our integrand and $m = 10^7$ the number of samples used for integral estimation. First, we estimate the integral by sampling from a uniform distribution. We generate samples $x_i \sim \mathcal{U}(1, N)$ for $1 \leq i \leq m$ and compute the expected value

$$\mathbb{E}_1[f] = \mathbb{E}_{\mathcal{U}(1,N)} \left[\frac{f}{f_{\mathcal{U}(1,N)}} \right] = \frac{N-1}{m} \sum_{i=1}^{m} f(x_i).$$

This expected value approaches I almost surely. To select N for which

$$I \approx \int_1^N x^{-7/4} e^{-1/x}\ dx,$$

we compute an approximation of the error we make:

$$\int_N^{10^7} x^{-7/4} e^{-1/x}\ dx.$$

This approximation has low variance because f is very similar to a uniform distribution for higher values. From Figure 1 we can see that by choosing $N = 10^5$ we will make an error on the fourth decimal place. Higher values also do not make sense, because they increase the variance of our estimation.

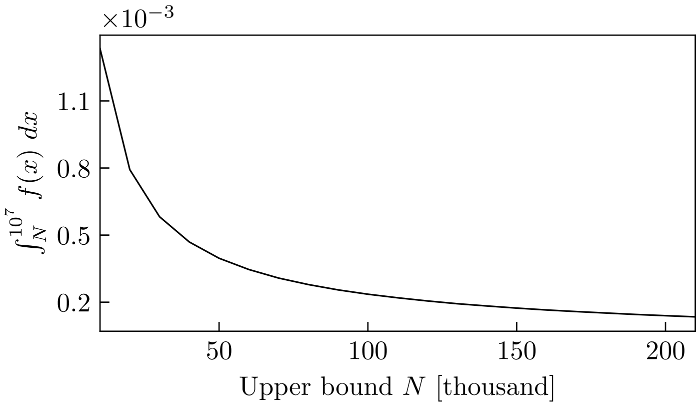


Figure 1: **Approximated error for upper bound N.**

3. Integral estimation using the uniform distribution has a high variance because its PDF is close to 0, where f has values around 0.3. Instead, we sample using

$$q(x) = Cx^{-7/4}, \qquad C = \frac{3}{4}(1 - N^{-3/4})^{-1}.$$

The normalisation constant was obtained by integrating q over its support $[1, N]$. To sample from q we need its quantile function. We compute the CDF

$$F_q(x) = -\frac{4C}{3}\left(x^{-3/4} - 1\right),$$

and its inverse (quantile in this case):

$$Q_q(x) = \left(1 - \frac{3x}{4C}\right)^{-4/3}.$$

To obtain samples $x_i \sim q$, we sample $u_i \sim \mathcal{U}(0, 1)$ and use inverse sampling:

$$x_i = Q_q(u_i).$$

Finally, the integral estimation can be computed using the expected value

$$\mathbb{E}_q\left[\frac{f}{q}\right] = \frac{1}{m}\sum_{i=1}^{m}\frac{f(x_i)}{q(x_i)}.$$

4. Repeating both approaches 10 times, we obtain results shown in Table 1. As we can see, estimation using samples from q has a much lower variance. This is expected because the shape of q is more similar to f than the uniform PDF.

Distribution	Average	SD
$\mathcal{U}(0, N)$	0.91108	0.02505
q	0.90653	0.00004

Table 1: **Integral estimation results** using samples from uniform distribution and from q. Estimates were calculated 10 times, computing their average and standard deviation, which are shown in the table.

Computing the integral analytically using a symbolical computing package `SymPy`, we get

$$I = 0.90678.$$

From this, we observe that the integral estimation calculated using samples from q is accurate up to the third decimal digit.

2 Markov chains: Countable state space

1. T_{ii} is a random variable for **recurrent** states because we may not return to w_i if it were transient. The PMF of T_{ii} is then

$$p_{T_{ii}}(n) = f_{ii}(n)$$

for $n \in \mathbb{N}$, since it sums to 1.

2. To show that the relation $\sim$ is an equivalence relation, we need to show three properties.

The relation is **reflexive** because each state is accessible from itself in zero steps, i.e., for $n = m = 0$ we have

$$P(X_0 = w_i | X_0 = w_i) > 0.$$

Let $w_i \sim w_j$ for some n and m. Then $w_j \sim w_i$ because

$$P(X_m = w_i | X_0 = w_j) > 0 \wedge P(X_n = w_j | X_0 = w_i) > 0,$$

and the relation is **symmetric**.

Finally, let $w_i \sim w_j$ for n_1 and m_1, and $w_j \sim w_k$ for n_2 and m_2. Define $n = n_1 + n_2$. Then using the Markov property we get

$$\begin{aligned} P(X_n = w_k | X_0 = w_i) &= P(X_{n_1} = w_j | X_0 = w_i) P(X_{n_1+n_2} = w_k | X_{n_1} = w_j) \\ &= P(X_{n_1} = w_j | X_0 = w_i) P(X_{n_2} = w_k | X_0 = w_j) > 0. \end{aligned}$$

Similarly we can define $m = m_1 + m_2$ and compute

$$P(X_m = w_i | X_0 = w_k) > 0,$$

meaning that $w_i \sim w_k$ and the relation is **transitive** and thus an equivalence relation.

■

To show that all states in the same class are of the same type, we assume $w_i \sim w_j$ and let w_i be a recurrent state, i.e., $\sum_{n=1}^{\infty} f_{ii}(n) = 1$. This tells us that the probability of visiting state w_i, starting in state w_i is 1, meaning an infinite number of such events happen. By the Borel-Cantelli lemma we get

$$\sum_{n=1}^{\infty} P(X_n = w_i | X_0 = w_i) = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty. \tag{1}$$

Using $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$ from $w_i \sim w_j$, we can compute

$$\sum_{k=1}^{\infty} p_{jj}^{(k)} \geq \sum_{k=n+m+1}^{\infty} p_{jj}^{(k)} = \sum_{k=1}^{\infty} p_{jj}^{(n+m+k)} \geq \sum_{k=1}^{\infty} p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)} \geq p_{ji}^{(m)} \left(\sum_{k=1}^{\infty} p_{ii}^{(k)} \right) p_{ij}^{(n)} = \infty.$$

Again, by Borel-Cantelli, $\sum_{n=1}^{\infty} f_{jj}(n) = 1$, meaning that state w_j is also recurrent. It follows that all states in the same equivalence class have to be recurrent or all transient.

■

3. Let w_i and w_j be recurrent states in the same class. Showing

$$\sum_{n=1}^{\infty} f_{ij}(n) = 1 \tag{2}$$

is the same as showing the probability of vising w_j starting in w_i is 1, i.e., infinitely many excursions occur starting in w_i and finishing in w_j by Borel-Cantelli. Because w_i is recurrent, $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$, as shown in (1), and $p_{ij}^{(n)} > 0$ for some n because $w_i \sim w_j$. Calculate

$$\sum_{k=1}^{\infty} p_{ij}^{(k)} \geq \sum_{k=1}^{\infty} p_{ii}^{(k)} p_{ij}^{(n)} = \infty.$$

By Borel-Cantelli it follows that infinitely many excursions occur, starting in w_i and finishing in w_j with a possibility of visiting both states multiple times during this walk. But in each such excursion, there exists a subexcursion from w_i to w_j without repeated visits. If there occur infinitely many excursions, there also occur infinitely many sub-excursions, meaning that the probability of visiting w_j starting in w_i is 1, proving (2).

■

4. Let us assume that the stationary distribution π exists and is equal to

$$\pi_j = \frac{N_{ij}}{\mathbb{E}[T_{ii}]}.$$

Computing the probability in the next step, we should obtain the stationary distribution once again. We start computing compute:

$$\begin{aligned}
\pi_j^{(t+1)} &= \sum_{k=1}^{\infty} \pi_k^{(t)} p_{kj} = \sum_{k=1}^{\infty} \frac{N_{ik}}{\mathbb{E}[T_{ii}]} p_{kj} \\
&= \frac{1}{\mathbb{E}[T_{ii}]} \sum_{k=1}^{\infty} \sum_{t=0}^{\infty} P(X_t = w_k, T_{ii} > t | X_0 = w_i) p_{kj} \\
&= \frac{1}{\mathbb{E}[T_{ii}]} \sum_{k=1}^{\infty} \sum_{t=0}^{T_{ii}-1} P(X_t = w_k | X_0 = w_i) p_{kj} \\
&= \frac{1}{\mathbb{E}[T_{ii}]} \sum_{t=0}^{T_{ii}-1} \sum_{k=1}^{\infty} P(X_t = w_k | X_0 = w_i) p_{kj}.
\end{aligned}$$

We used the fact that N_{ik} is the expected number of visits of the state w_k before returning to the state w_i:

$$N_{ik} = \mathbb{E}\left[\sum_{t=0}^{\infty} \mathbb{I}(X_t = w_k, T_{ii} > t | X_0 = w_i)\right] = \sum_{t=0}^{\infty} P(X_t = w_k, T_{ii} > t | X_0 = w_i),$$

moved condition $T_{ii} > t$ into the summation limit and exchanged the limits. We continue:

$$\pi_j^{(t+1)} = \frac{1}{\mathbb{E}[T_{ii}]} \sum_{t=0}^{T_{ii}-1} \sum_{k=1}^{\infty} p_{ik}^{(t)} p_{kj} = \frac{1}{\mathbb{E}[T_{ii}]} \sum_{t=0}^{T_{ii}-1} p_{ij}^{(t+1)} = \frac{1}{\mathbb{E}[T_{ii}]} \sum_{t=0}^{T_{ii}-1} p_{ij}^{(t)} = \frac{N_{ij}}{\mathbb{E}[T_{ii}]}, \quad (3)$$

replacing $t+1$ with t in the third equality, because $p_{ij}^{(0)} = p_{ij}^{(T_{ii})}$. They are either both 0 if $i \neq j$ or both 1. In the end, we obtained the same expression, so the stationary distribution exists.

■

5. Because T_{ii} is the first-passage time from w_i back to w_i,

$$\sum_{t=0}^{T_{ii}-1} p_{ii}^{(t)} = p_{ii}^{(0)} + p_{ii}^{(1)} + \ldots + p_{ii}^{(T_{ii}-1)} = 1 + 0 + \ldots + 0 = 1.$$

Using this in (3), we get

$$\pi_i = \frac{1}{\mathbb{E}[T_{ii}]}.$$

■

We show that this stationary distribution is unique in multiple steps. First, let $i = 1$ without loss of generality and $v = \{v_i : i \in \mathbb{N}\}$ be a stationary distribution with $v_1 = \frac{1}{\mathbb{E}[T_{11}]}$.
Since v is a stationary distribution, we start derivation for v_j with $j > 0$:

$$\begin{aligned} v_j &= \sum_{i_1=1}^{\infty} v_{i_1} p_{i_1 j} = v_1 p_{1j} + \sum_{i_1=2}^{\infty} v_{i_1} p_{i_1 j} = \frac{p_{1j}}{\mathbb{E}[T_{11}]} + \sum_{i_1=2}^{\infty} v_{i_1} p_{i_1 j} \\ &\geq \frac{p_{1j}}{\mathbb{E}[T_{11}]} = \frac{1}{\mathbb{E}[T_{11}]} P(X_1 = w_j, T_{11} > 0 | X_0 = w_1). \end{aligned} \quad (4)$$

We repeat a similar derivation, but this time insert the expression for v_{i_1} into the sum, instead of omitting it:

$$\begin{aligned}
v_j &= \frac{p_{1j}}{\mathbb{E}[T_{11}]} + \sum_{i_1=2}^{\infty} v_{i_1} p_{i_1 j} = \frac{p_{1j}}{\mathbb{E}[T_{11}]} + \sum_{i_1=2}^{\infty} \left(\frac{p_{1i_1}}{\mathbb{E}[T_{11}]} + \sum_{i_2=2}^{\infty} v_{i_2} p_{i_2 i_1} \right) p_{i_1 j} \\
&= \frac{p_{1j}}{\mathbb{E}[T_{11}]} + \frac{\sum_{i_1=2}^{\infty} p_{1i_1} p_{i_1 j}}{\mathbb{E}[T_{11}]} + \sum_{i_2=2}^{\infty} \sum_{i_1=2}^{\infty} v_{i_2} p_{i_2 i_1} p_{i_1 j} \\
&\geq \frac{p_{1j}}{\mathbb{E}[T_{11}]} + \frac{\sum_{i_1=2}^{\infty} p_{1i_1} p_{i_1 j}}{\mathbb{E}[T_{11}]} = \frac{1}{\mathbb{E}[T_{11}]} \left(p_{1j} + \sum_{i_1=2}^{\infty} p_{1i_1} p_{i_1 j} \right) \\
&= \frac{1}{\mathbb{E}[T_{11}]} \sum_{l=1}^{2} P(X_n = w_j, T_{11} > l | X_0 = w_i).
\end{aligned}$$

As we can see, this is similar to (4). To prove it for all k, we first assume an inductive hypothesis holds for $k-1$:

$$\begin{aligned}
v_j = &\frac{p_{1j}}{\mathbb{E}[T_{11}]} + \frac{\sum_{i_1=2}^{\infty} p_{1i_1} p_{i_1 j}}{\mathbb{E}[T_{11}]} + \ldots \\
&+ \frac{\sum_{i_{k-1}=2}^{\infty} \sum_{i_{k-2}=2}^{\infty} \cdots \sum_{i_1}^{\infty} p_{1i_{k-1}} p_{i_{k-1} i_{k-2}} \cdots p_{i_2 i_1} p_{i_1 j}}{\mathbb{E}[T_{11}]} \\
&+ \sum_{i_k=2}^{\infty} \sum_{i_{k-1}=2}^{\infty} \cdots \sum_{i_1=2}^{\infty} v_{i_k} p_{i_k i_{k-1}} \cdots p_{i_2 i_1} p_{i_1 j}.
\end{aligned}$$

Let us mark all terms except the last one with A. Inserting expression for v_{i_k} again, we do an inductive step and obtain

$$\begin{aligned}
v_j &= A + \sum_{i_k=2}^{\infty} \sum_{i_{k-1}=2}^{\infty} \cdots \sum_{i_1=2}^{\infty} \left(\frac{p_{1i_k}}{\mathbb{E}[T_{11}]} + \sum_{i_{k+1}=2}^{\infty} v_{i_{k+1}} p_{i_{k+1} i_k} \right) p_{i_k i_{k-1}} \cdots p_{i_2 i_1} p_{i_1 j} \\
&= A + \frac{\sum_{i_k=2}^{\infty} \sum_{i_{k-1}=2}^{\infty} \cdots \sum_{i_1=2}^{\infty} p_{1i_k} p_{i_k i_{k-1}} \cdots p_{i_2 i_1} p_{i_1 j}}{\mathbb{E}[T_{11}]} \\
&\quad + \sum_{i_{k+1}=2}^{\infty} \sum_{i_k=2}^{\infty} \cdots \sum_{i_1=2}^{\infty} v_{i_{k+1}} p_{i_{k+1} i_k} p_{i_k i_{k-1}} \cdots p_{i_2 i_1} p_{i_1 j}.
\end{aligned}$$

This proves that the inductive hypothesis holds for all k. Omitting the last term, we again obtain the inequality

$$\begin{aligned}
v_j &\geq \frac{1}{\mathbb{E}[T_{11}]} \left(p_{1j} + \sum_{i_1=2}^{\infty} p_{1i_1} p_{i_1 j} + \ldots + \sum_{i_k=2}^{\infty} \sum_{i_{k-1}=2}^{\infty} \cdots \sum_{i_1=2}^{\infty} p_{1i_k} p_{i_k i_{k-1}} \cdots p_{i_2 i_1} p_{i_1 j} \right) \\
&= \frac{1}{\mathbb{E}[T_{11}]} \sum_{l=1}^{k+1} P(X_l = w_j, T_{11} > l | X_0 = w_1).
\end{aligned} \tag{5}$$

Since (5) holds for all k, we get

$$v_j \geq \frac{1}{\mathbb{E}[T_{11}]} \sum_{t=0}^{\infty} P(X_t = w_j, T_{11} > t | X_0 = w_1) = \frac{N_{1j}}{\mathbb{E}[T_{11}]}.$$

We can lower the bounds to $t = 0$ because $i \neq j$. From the last inequality, it follows that $v_j \geq \pi_j$.

Now we define a distribution $u = \{u_i : i \in \mathbb{N}\}$ with $u_i = v_i - \pi_i$, which is also stationary:

$$u_j = \sum_i u_i p_{ij} = \sum_i (v_i - \pi_i) p_{ij} = \sum_i v_i p_{ij} - \sum_i \pi_i p_{ij} = v_j - \pi_j = u_j,$$

and $u_1 = 0$ because v_1 is defined to be the same as π_1. Because the Markov chain is irreducible, $\forall i, j\ \exists m \in \mathbb{N} : p_{ij}^{(m)} > 0$. Let us assume $u_j > 0$ for $j \neq 1$. Computing

$$u_1 = \sum_i u_i p_{i1}^{(m)} > 0 \neq 0 \quad \Rightarrow\!\Leftarrow$$

we get a contradiction, meaning $u_j = 0$, $v_j = \pi_j$ and proving that the stationary distribution is unique.

$\blacksquare$

3 Metropolis-Hastings algorithm

We would like to approximate the posterior distribution of parameters α and η, given five data points generated from the (α, η)-Weibull distribution. If we use the Bayes' rule,

$$p(\alpha, \eta|\mathbf{x}) = \frac{p(\mathbf{x}|\alpha, \eta)p(\alpha, \eta)}{p(\mathbf{x})} \propto p(\mathbf{x}|\alpha, \eta)p(\alpha, \eta),$$

we can see that the posterior is proportional to the product of the likelihood and the prior. Because the data points are independent, we can say

$$p(\alpha, \eta|\mathbf{x}) \propto \prod_{x_i} p(x_i|\alpha, \eta)p(\alpha, \eta).$$

We use the Metropolis-Hastings algorithm to generate samples from this distribution. For each proposal, we generate 5 chains of length 1000 and compare their statistics. Each chain has an initial state $(1, 1)$.

Multivariate normal proposal

First, we use the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\mathbf{\Sigma} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}$$

as our proposal. Diagnostics for each chain are shown in Table 2. We can see that the diagnostics for different chains do match. There is only more variety among the effective sample sizes and variances for η.

	α			η		
Chain	Average	Variance	ESS	Average	Variance	ESS
1	1.888	0.510	63	1.968	0.638	62
2	1.889	0.434	88	1.793	0.493	58
3	1.977	0.488	68	1.867	0.810	97
4	1.970	0.485	91	1.864	0.677	77
5	1.905	0.414	69	1.947	0.520	53
Average	1.926	0.466	75	1.888	0.628	69

Table 2: Diagnostic statistics for 5 chains generated using the multivariate normal proposal.

In Figure 2 we can observe the trace plots and the autocovariances. As expected, we observe similar behaviour from all 5 chains and the autocovariances tend to drop with higher k.

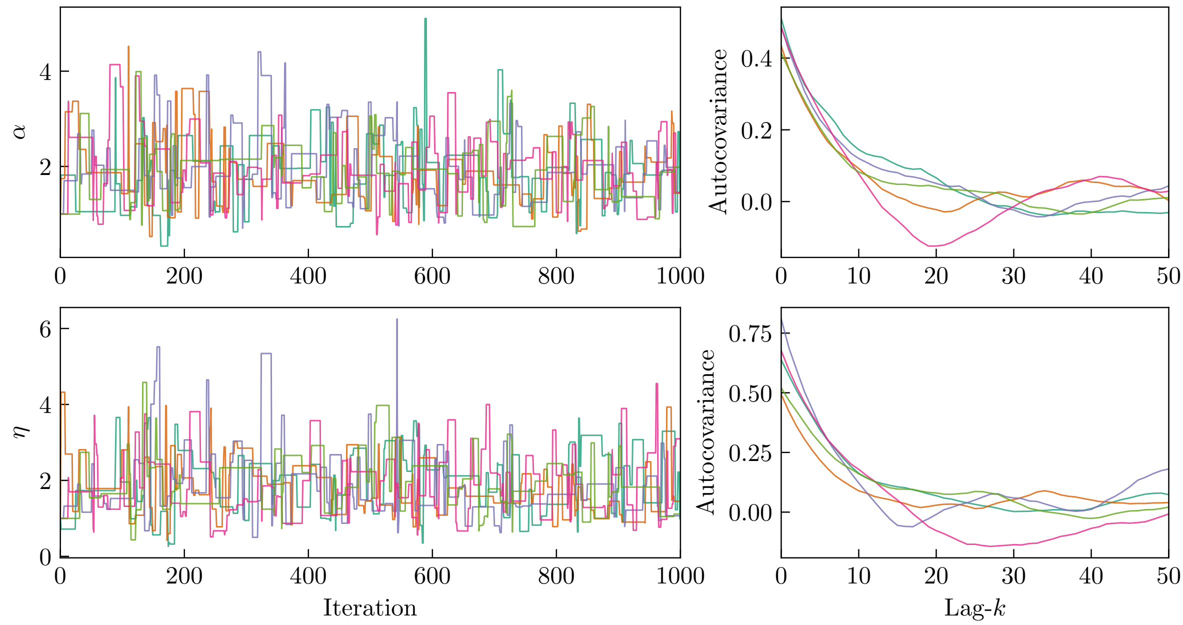


Figure 2: Trace plots for all 5 chains and parameters α and η (left) with lag-k autocovariances (right) using the multivariate normal proposal.

Exponential proposal

Next, we use the proposal distribution

$$q(\alpha', \eta' | \alpha, \eta) = \frac{1}{\alpha\eta} \exp\left(-\frac{\alpha'}{\alpha} - \frac{\eta'}{\eta}\right)$$

for our chains. This proposal can be rewritten as

$$q(\alpha', \eta' | \alpha, \eta) = \frac{1}{\alpha} \exp\left(-\frac{\alpha'}{\alpha}\right) \frac{1}{\eta} \exp\left(-\frac{\eta'}{\eta}\right),$$

meaning we can instead sample from two independent exponential distributions with $\lambda_{\alpha'} = 1/\alpha$ and $\lambda_{\eta'} = 1/\eta$. Statistics for the obtained chains are shown in Table 3. We obtain slightly higher averages and smaller variances, with ESS for α being similar, but ESS for η being smaller.

	α			η		
Chain	Average	Variance	ESS	Average	Variance	ESS
1	1.996	0.458	77	1.946	0.699	53
2	2.005	0.487	67	2.041	0.726	48
3	1.912	0.300	86	1.913	0.480	85
4	1.881	0.378	106	1.943	0.496	54
5	1.920	0.403	44	1.728	0.563	52
Average	1.943	0.405	76	1.914	0.593	58

Table 3: Diagnostic statistics for 5 chains generated using the exponential proposal.

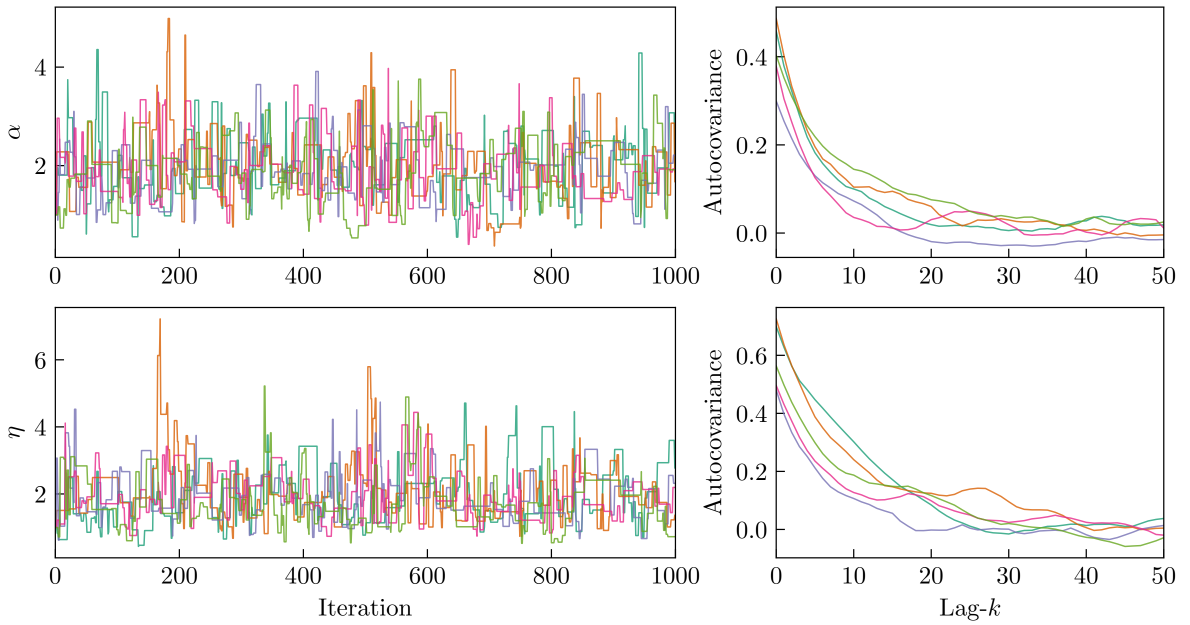


Figure 3: Trace plots for all 5 chains and parameters α and η (left) with lag-k autocovariances (right) using the exponential proposal.

Again, trace plots and autocovariances are shown in Figure 3, with results being similar to the previous proposal.

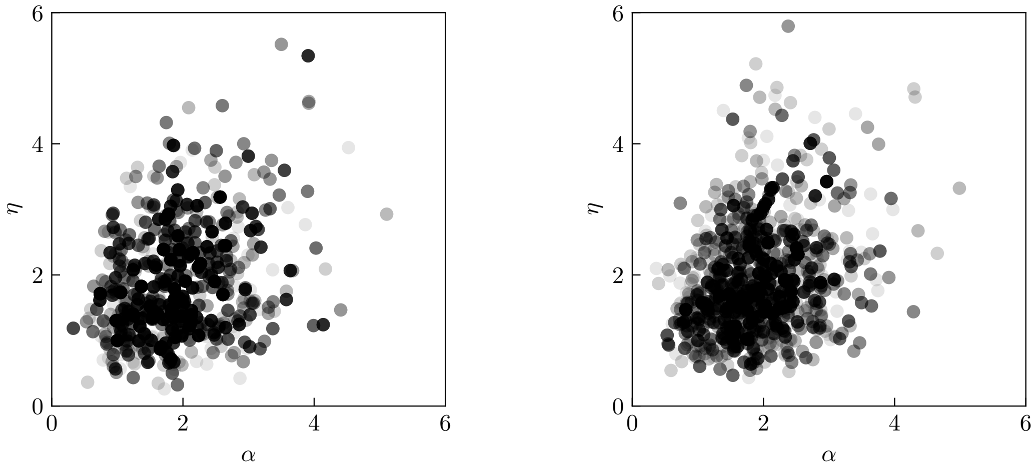


Figure 4: Samples (α, η) generated using the multivariate normal proposal (left) and the exponential proposal (right).

Discussion

Looking at the scatter plots in Figure 4 and noting the results from the tables above, we can conclude that the samples generated using the exponential distribution are slightly better. They are more densely packet at one point, but there are few more samples with high values of η.

Another difference between these two approaches is that the multivariate normal proposal has to be tuned. We have to determine the covariance matrix we will use. After trying a few such matrices, we can observe the approximate distribution of (α, η) and create a matrix with eigenvectors that form an elliptic shape similar to the distribution.

Probability estimation

With all the generated samples we now estimate the probability that the parameters $(\alpha, \eta) \in [2, \infty) \times [2, \infty)$. With the multivariate normal proposal, we get

$$P(\alpha \geq 2, \eta \geq 2) = 0.2096,$$

and with the exponential proposal,

$$P(\alpha \geq 2, \eta \geq 2) = 0.2108.$$

We can see that both proposals produce similar probability estimations. To get a better estimation, we can increase the length of the chain to 100 000 and use the first 5000 samples as burn-in. We get

$$P(\alpha \geq 2, \eta \geq 2) = 0.2311$$

for the multivariate normal proposal and

$$P(\alpha \geq 2, \eta \geq 2) = 0.2362$$

for the exponential one.