Assignment 5: Probabilistic ML

Marcus Hansen

December 2024

Original Assignment:

Exercise 1 (25%)

Consider the Johnson's Su distribution with PDF:

$$p(x) = \frac{\sqrt{2}}{\sqrt{\pi\left(1 + (x-1)^2\right)}} \exp\left(-\frac{1}{2}\left(3 + 2\sinh^{-1}(x-1)\right)^2\right)$$

The corresponding Python code for this PDF can be implemented as follows using `Numpy`:

```
np.sqrt(2) / np.sqrt(np.pi * (1 + (x - 1) ** 2)) *
np.exp(-0.5 * (3 + 2 * np.arcsinh(x - 1)) ** 2)
```

Q1: Implement an importance sampler with a standard normal distribution (mean = 0, variance = 1) as proposal to generate (weighted) samples from p(x). Draw L = 1000 samples.

See Jupyter Notebook "Assignment5_Exercise_1_v1.ipynb"!

Q2: Plot a histogram of the samples together with the target PDF and the proposal PDF and inspect your obtained Monte Carlo approximation. What can you say about the quality of the approximation you have obtained? How does it relate to the shape of the proposal?

The approximation we obtained captures some of the features of the target distribution. With 1,000 samples, we can conclude that the approximation is acceptable. It is clear that our proposal distribution has significant overlap with the target distribution, which contributes to a better approximation. For instance, we are better at approximating the right tail, as our proposal distribution overlaps more with that part of the target distribution compared to the left tail, where the approximation is more challenging. Additionally, we can see that the peak of the distribution is reasonably well approximated. However, this could be further improved by increasing the number of samples or by using a proposal distribution with greater overlap.

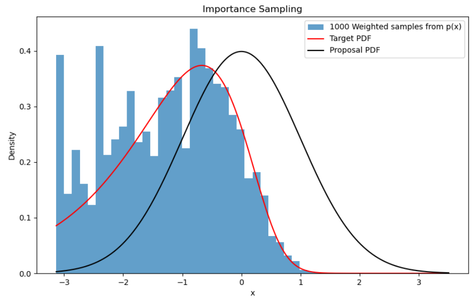


Figure 1: 1000 Weighted samples together with target PDF and proposal PDF.

Q3: Try using more samples L. Does the Monte Carlo approximation improve?

As we can see in Figure 2, our Monte Carlo approximation improved as we increased the sample size. It is now much more similar to the target distribution.

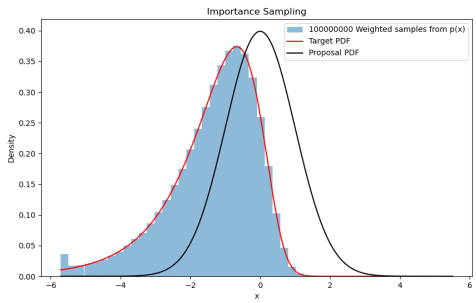


Figure 2: 1000000 Weighted samples together with target PDF and proposal PDF.

Q4: Limit yourself to L = 1000 samples again, and try to instead adjust the proposal to improve the approximation.

To improve the approximation, we aimed to increase the overlap between the target distribution and the proposal distribution. This was achieved by visually analyzing the results from Q2 and adjusting the mean to -0.8 while increasing the variance to 4, allowing us to better capture both the left and right tails. In Figure 3, we can observe that this adjustment led to a more accurate approximation of the target distribution. We successfully captured more of the target distribution's shape, improving the approximation of the tails while still capturing the density of the peak. However, the right tail of the target distribution has a steeper decay, suggesting that a skewed Gaussian distribution might further enhance the approximation. Overall, the result is improved by modifying the proposal distribution, even with only 1,000 samples.

Q5: Use the samples to estimate the mean and variance of the target. (One can analytically show that the mean is -1.41 and the variance 1.98). Explore how the quality of these estimates changes with different L.

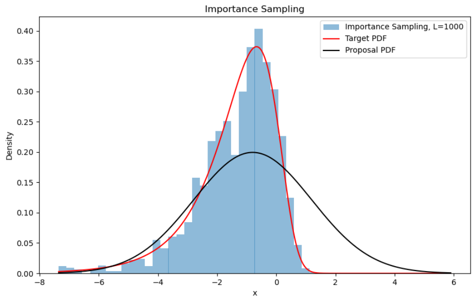


Figure 3: 1000 Weighted samples together with target PDF and updated proposal PDF.

In this case, we used the samples to estimate the mean and variance for different sample sizes. As shown in Table 1, we can clearly observe that the error between the true mean and variance decreases as the number of samples increases. To achieve the best approximations, we should use between 10^6 and 10^8 samples. It's also interesting to note that the approximation of the distribution remains quite good even with a smaller number of samples. This is likely because the Gaussian distribution is a good proposal distribution that closely resembles the target distribution.

Samples	Estimated Mean	Mean error	Estimated Variance	Variance Error
100	-1.0684	0.3416	0.7319	1.2481
1000	-1.0604	0.3496	0.8599	1.1201
10000	-1.2228	0.1872	1.1466	0.8334
100000	-1.2647	0.1453	1.2685	0.7115
1000000	-1.3963	0.0137	1.7263	0.2537
10000000	-1.3507	0.0593	1.5816	0.3984
100000000	-1.3914	0.0186	1.7626	0.2174

Table 1: Estimated Means and Variances with Errors for Different Sample Sizes

Q6: Change the proposal to a uniform distribution on the interval [-4, 1]. Is it still a valid importance sampler?

There are two key considerations when choosing a proposal distribution to achieve a valid importance sampler. The first is that the proposal distribution, $q(z)$, must be nonnegative for all values of z. This is satisfied by the uniform distribution between $[-4, 1]$. The second consideration is that the proposal distribution should be able to generate samples over the support of the target distribution. Our target distribution has support over the entire real line, while our proposal distribution assigns zero to values outside the interval. However, in our case, this is acceptable because the target distribution has very small density in the tails. As shown in Figure 4, we are able to approximate the target distribution quite well with a small number of samples. Therefore, the uniform distribution between $[-4, 1]$ is a valid importance sampler.

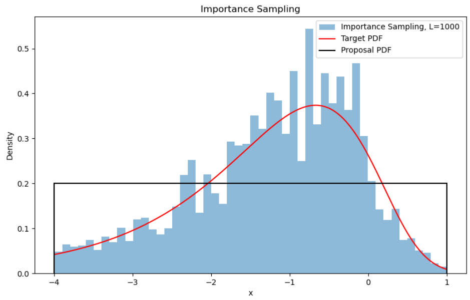


Figure 4: 1000 Weighted samples together with target PDF and updated proposal PDF (uniform).

Exercise 2 (25%) Markov Chain

The given equation is:

$$x[k+1] = 0.9x[k] + v[k], \quad v[k] \sim \mathcal{N}(0, 0.19).$$

Q1: Verify that the given equation is a Markov chain

To verify that this equation defines a Markov chain, we need to show that the future state $x[k+1]$ depends only on the current state $x[k]$ and not on earlier states.

The term $v[k]$ is an independent noise sampled from $\mathcal{N}(0, 0.19)$, and $x[k+1]$ is computed solely from $x[k]$ and $v[k]$. Hence, the transition probability satisfies the defeniton of fisrt order markov chain [1]:

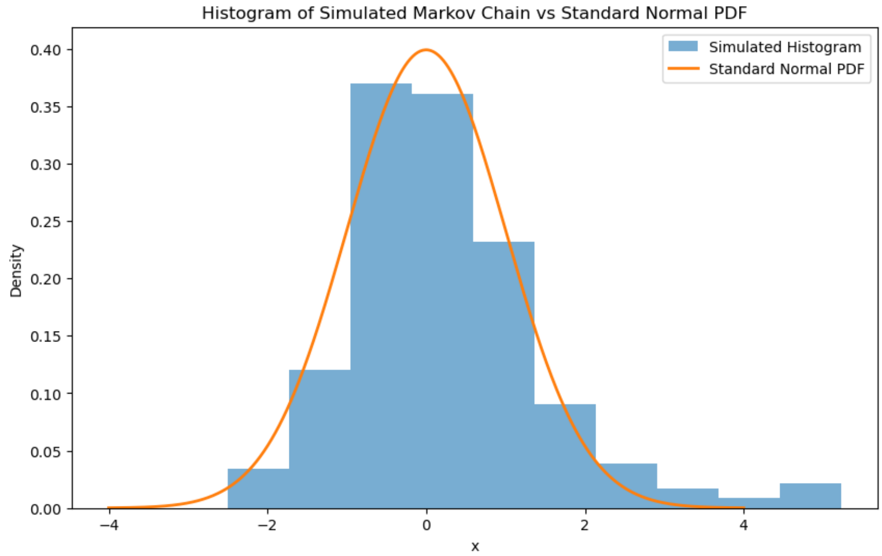
$$P(\mathbf{z}^{(m+1)} \mid \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}) = P(\mathbf{z}^{(m+1)} \mid \mathbf{z}^{(m)}).$$

$$P(x[k+1]|x[1], \dots, x[k]) = P(x[k+1]|x[k]).$$

This confirms that the given equation satisfies the Markov property.

Q2: Simulating the Markov Chain

When plotting the histogram of the Markov chain, we observe that the distribution of x converges to a Gaussian distribution with a mean of 0 and a variance of 1, as shown in Figure 5. However, we also notice that the distribution is slightly right-skewed. This is expected because the initial value $x[1] = 5$ is far from the stationary mean of 0, and the Markov chain takes some time to fully converge.



[!h]

Figure 5: Histogram correpsonding to Exercise 2 Q2. A histogram over x when using the given markov chain and intialize with x[0]=5.

Q3: Burn-in Period markov chain

The burn in period of this markov chain seeams to be around 30 steps. This is shown in the plot in Figure 6. The value of x seems to converge to a value around 0 after 20 steps.

Q4 Stationary Distribution

To prove that the stationary distribution of the Markov chain is $\mathcal{N}(0,1)$, we check whether $x[k+1]$ and $x[k]$ have the same distribution under stationarity.

Substituting $x[k] \sim \mathcal{N}(0,1)$ into the recurrence relation:

$$x[k+1] = 0.9x[k] + v[k],$$

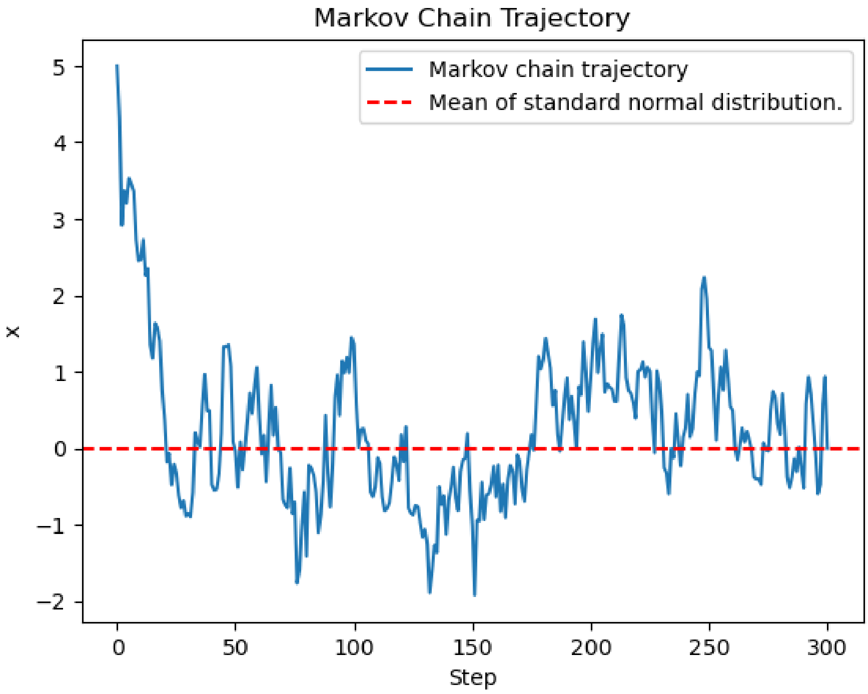
where $x[k] \sim \mathcal{N}(0,1)$ which gives that $0.9x[k] \sim \mathcal{N}(0, 0.9^2)$ and $v[k] \sim \mathcal{N}(0, 0.19)$. Since the sum of two independent Gaussians is also Gaussian, we have:

$$x[k+1] \sim \mathcal{N}(0, 0.9^2 + 0.19).$$

The variance of $x[k+1]$ becomes:

$$\mathrm{Var}(x[k+1]) = 0.9^2 + 0.19 = 0.81 + 0.19 = 1.$$

This matches the variance of $\mathcal{N}(0,1)$, confirming that the stationary distribution is indeed $\mathcal{N}(0,1)$.



[!h]

Figure 6: Plot correpsonding to Exercise 2 Q3. The value of x over increasing time (step) when using the given markov chain and intialize with x[0]=5.

Exercise 3 (25%)Kullback-Leibler divergence

Kullback-Leibler divergence is defined as

$$KL(p(x)\|q(x)) = \int_{-\infty}^{\infty} p(x) \ln \frac{p(x)}{q(x)}\, dx$$

Q1: KL Divergence Between Two Scalar Gaussians

where

$$\begin{aligned} p(x) =& \mathcal{N}(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ q(x) =& \mathcal{N}(x;m,s^2) = \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(x-m)^2}{2s^2}\right), \end{aligned}$$

First evaluate the logarithmic term of the integral in the Kullback-Leibler divergence

$$ln\frac{p(x)}{q(x)} = ln\left(\frac{\sqrt{s^2}}{\sqrt{\sigma^2}}\right) + ln\left(\frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(x-m)^2}{2s^2}\right)}\right)$$

Using the fact that s and σ are positive, and the logritmc rules we get:

$$ln\frac{p(x)}{q(x)} = ln\left(\frac{s}{\sigma}\right) + \left(\frac{(x-m)^2}{2s^2}\right) - \left(\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Then take the integral:

$$KL(p(x)\|q(x)) = \int_{-\infty}^{\infty} p(x)\left[ln\left(\frac{s}{\sigma}\right) + \left(\frac{(x-m)^2}{2s^2}\right) - \left(\frac{(x-\mu)^2}{2\sigma^2}\right)\right] dx$$

As the integral is linear we can split it up into three integrals:

$$KL(p(x)\|q(x)) = \int_{-\infty}^{\infty} p(x) ln\left(\frac{s}{\sigma}\right) dx + \int_{-\infty}^{\infty} p(x)\left(\frac{(x-m)^2}{2s^2}\right) dx - \int_{-\infty}^{\infty} p(x)\left(\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

For the first term as s and σ is independet of x and that $p(x)$ is a distributtion we get:

$$\int_{-\infty}^{\infty} p(x) ln\left(\frac{s}{\sigma}\right) dx = ln\left(\frac{s}{\sigma}\right)\int_{-\infty}^{\infty} p(x)\, dx = ln\left(\frac{s}{\sigma}\right)$$

For the second term use the expectation under a probability $p(x)$:

$$\begin{aligned} \int_{-\infty}^{\infty} p(x)\left(\frac{(x-m)^2}{2s^2}\right) dx =& \frac{1}{2s^2}\mathbb{E}_{p(x)}[(x-m)^2] \\ =& \frac{1}{2s^2}\left(Var_{p(x)}[x] + (\mathbb{E}_{p(x)}[x]-m)^2\right) \\ =& \frac{1}{2s^2}\left(\sigma^2 + (\mu-m)^2\right) \end{aligned}$$

And using the same for the third term we get:

$$\begin{aligned}\int_{-\infty}^{\infty} p(x)\left(\frac{(x-\mu)^2}{2\sigma^2}\right)dx =& \frac{1}{2\sigma^2}\mathbb{E}_{p(x)}[(x-\mu)^2]\\ =&\frac{1}{2\sigma^2}\left(Var_{p(x)}[x]+(\mathbb{E}_{p(x)}[x]-\mu)^2\right)\\ =&\frac{1}{2\sigma^2}\left(\sigma^2+(\mu-\mu)^2\right)\\ =&\frac{\sigma^2}{2\sigma^2}\\ =&\frac{1}{2}\end{aligned}$$

Combineing this result we get:

$$KL(p(x)\|q(x)) = ln\left(\frac{s}{\sigma}\right)+\frac{1}{2s^2}\left(\sigma^2+(\mu-m)^2\right)-\frac{1}{2}$$

Q2: KL Divergence Between Two Multivariate Gaussians

The solution to Q2 follows many of the step from Q1 just with matrices instead of scalars. Therefore some of obvious derivations that has been derived in Q1 is not derived in Q2.

Given two multivariate Gaussian distributions with corepsonding PDFs

$$\begin{aligned}p(x) = \mathcal{N}(x;\mu,\Sigma) &= \frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right),\\ q(x) = \mathcal{N}(x;m,S) &= \frac{1}{(2\pi)^{k/2}|S|^{1/2}}\exp\left(-\frac{1}{2}(x-m)^T S^{-1}(x-m)\right).\end{aligned}$$

Evaluate the logarithmic term of the integral in the Kullback-Leibler divergence

$$\begin{aligned}ln\frac{p(x)}{q(x)} =& ln\left(\frac{|S|^{1/2}}{|\Sigma|^{1/2}}\right)+ln\left(\frac{\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)}{\exp\left(-\frac{1}{2}(x-m)^T S^{-1}(x-m)\right)}\right)\\ =& ln\left(\frac{|S|^{1/2}}{|\Sigma|^{1/2}}\right)+\frac{1}{2}\left((x-m)^T S^{-1}(x-m)-(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\end{aligned}$$

Then take the integral and use the fact that the integral operation is linear

$$\begin{aligned}KL(p(x)\|q(x)) =& \int p(x)\left[\ln\frac{|S|^{1/2}}{|\Sigma|^{1/2}}+\frac{1}{2}\left((x-m)^T S^{-1}(x-m)-(x-\mu)^T\Sigma^{-1}(x-\mu)\right)\right]dx.\\ =&\frac{1}{2}\ln\frac{\det(S)}{\det(\Sigma)}+\frac{1}{2}\int p(x)(x-m)^T S^{-1}(x-m)\,dx-\frac{1}{2}\int p(x)(x-\mu)^T\Sigma^{-1}(x-\mu)\,dx.\end{aligned}$$

Solving the integral term by term The first term is:

$$\frac{1}{2}\ln\frac{\det(S)}{\det(\Sigma)}.$$

The second term.

$$\begin{aligned}\int p(x)(x-m)^T S^{-1}(x-m)\,dx =&\mathbb{E}_{p(x)}\left[(x-m)^T S^{-1}(x-m)\right]\\ &\text{Using the hint}\\ =&(\mu-m)^T S^{-1}(\mu-m)+\mathrm{tr}(S^{-1}\Sigma).\end{aligned}$$

The third term.

$$\begin{aligned}\int p(x)(x-\mu)^T \Sigma^{-1}(x-\mu)\,dx =&\mathbb{E}_{p(x)}\left[(x-\mu)^T \Sigma^{-1}(x-\mu)\right]\\ =&tr(\Sigma^{-1}\Sigma)=tr(I_k)\\ =&k.\end{aligned}$$

Where k is the dimensionality of the multivariate Gaussian.
Combining tis reresults we get the final expression for the KL divergence between two multivariate Gaussians

$$KL(p(x)\|q(x))=\frac{1}{2}\left(\ln\frac{\det(S)}{\det(\Sigma)}+\mathrm{tr}(S^{-1}\Sigma)+(\mu-m)^T S^{-1}(\mu-m)-k\right).$$

Assignment V2:

Exercise 1: Gibbs Sampling

1. Derive the Conditional Distributions

1. For the Gaussian model:

$$y_i \sim N(\mu, \tau^{-1}),$$

where $\mu \sim N(0, \omega^{-1})$ and $\tau \sim$ Gamma(α, β). Prove the following:

$$\mu \mid y, \tau \sim N\left(\frac{\tau}{n\tau + \omega} \sum y_i, \frac{1}{n\tau + \omega}\right) \tag{1}$$

$$\tau \mid y, \mu \sim \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (y_i - \mu)^2\right) \tag{2}$$

We have the likelihood function of observing y_i as:

$$p(y_i \mid \mu, \tau) = \mathcal{N}(y_i; \mu, \tau^{-1})$$

and assuming IID samples the likelihood function for all y_i, y is given by:

$$\begin{aligned} p(y \mid \mu, \tau) &= \prod_{i=1}^{n} \mathcal{N}(y_i; \mu, \tau^{-1}) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\tau^{-1}}} \exp\left[\frac{-1}{2\tau^{-1}} (y_i - \mu)^2\right] \\ &= \frac{1}{(2\pi\tau^{-1})^{\frac{n}{2}}} \exp\left[\frac{-1}{2\tau^{-1}} \sum_{i=1}^{n} (y_i - \mu)^2\right] \end{aligned} \tag{3}$$

and from the assignment we have the prior pdf of μ as:

$$\begin{aligned} p(\mu) &= p(\mu; 0, \omega^{-1}) \\ &= \frac{1}{\sqrt{2\pi\omega^{-1}}} \exp\left[\frac{-1}{2\omega^{-1}} \mu^2\right] \end{aligned} \tag{4}$$

and the prior pdf of τ as:

$$\begin{aligned} p(\tau) &= p(\tau; \alpha, \beta) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha - 1} \exp\left[-\beta\tau\right], \quad \alpha, \beta, \tau > 0 \end{aligned} \tag{5}$$

Starting with equation 1, from Bayes theorem we have that the posterior distribution of μ is:

$$
\begin{aligned}
p(\mu \mid y, \tau) &\propto p(y \mid \mu, \tau) p(\mu) = \\
&= \frac{1}{(2\pi\tau^{-1})^{\frac{n}{2}}} \exp\left[\frac{-1}{2\tau^{-1}} \sum_{i=1}^{n} (y_i - \mu)^2\right] \frac{1}{\sqrt{2\pi\omega^{-1}}} \exp\left[\frac{-1}{2\omega^{-1}} \mu^2\right] \\
&= \frac{1}{(2\pi\tau^{-1})^{\frac{n}{2}}} \frac{1}{\sqrt{2\pi\omega^{-1}}} \exp\left[\frac{-1}{2}\left(\tau \sum_{i=1}^{n} (y_i - \mu)^2 + \omega\mu^2\right)\right] \\
&\propto \exp\left(\frac{-1}{2}\left[\tau\left(\sum_{i=1}^{n} y_i - 2\mu \sum_{i=1}^{n} y_i + n\mu^2\right) + \omega\mu^2\right]\right) \\
&= \exp\left(\frac{-1}{2(n\tau + \omega)}\left[\mu^2 + -2\mu \frac{\tau}{n\tau + \omega} \sum_{i=1}^{n} y_i + \frac{\tau}{n\tau + \omega} \sum_{i=1}^{n} y_i\right]\right) \\
&= \exp\left(\frac{-1}{2(n\tau + \omega)}\left[\left(\mu - \frac{\tau}{n\tau + \omega} \sum_{i=1}^{n} y_i\right)^2\right]\right) \exp\left(\frac{1}{2(n\tau + \omega)}\left[\left(\frac{\tau}{n\tau - \omega} \sum_{i=1}^{n} y_i\right)^2 + \frac{\tau}{n\tau + \omega} \sum_{i=1}^{n} y_i\right]\right) \\
&\propto \exp\left(\frac{-1}{2(n\tau + \omega)}\left[\left(\mu - \frac{\tau}{n\tau + \omega} \sum_{i=1}^{n} y_i\right)^2\right]\right)
\end{aligned}
$$

which up to a constant is a Gaussian distribution of μ with mean $\frac{\tau}{n\tau+\omega} \sum_{i=1}^{n} y_i$ and variance $\frac{1}{(n\tau+\omega)}$. Hence we have shown equation 1, i.e. that:

$$
\mu \mid y, \tau \sim \mathcal{N}\left(\frac{\tau}{n\tau + \omega} \sum y_i, \frac{1}{n\tau + \omega}\right)
$$

Now for the posterior of distribution of τ, equation 2, we also use Bayes theorem:

$$
\begin{aligned}
p(\tau \mid y, \mu) &\propto p(y \mid \mu, \tau) p(\tau) = \\
&= \frac{1}{(2\pi\tau^{-1})^{\frac{n}{2}}} \exp\left[\frac{-1}{2\tau^{-1}} \sum_{i=1}^{n} (y_i - \mu)^2\right] \frac{\beta^\alpha}{\Gamma(\alpha)} \tau^{\alpha - 1} \exp\left[-\beta\tau\right] \\
&= \frac{\tau^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} \frac{\beta^\alpha}{\Gamma(\alpha)} \exp\left[-\left(\beta + \frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)^2\right)\tau\right] \tau^{\alpha - 1} \\
&\propto \tau^{\overbrace{\frac{n}{2} + \alpha}^{\alpha'} - 1} \exp\left[-\underbrace{\left(\beta + \frac{1}{2} \sum_{i=1}^{n} (y_i - \mu)^2\right)}_{\beta'} \tau\right]
\end{aligned}
$$

By comparing the expression above to the gamma distribution in equation 5 we identify that this, up to a constant, is a Gamma distribution with parameters α' and β', and hence we have proven expression 2, i.e. that:

$$
\tau \mid y, \mu \sim \text{Gamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (y_i - \mu)^2\right)
$$

2. For the Poisson model:

$$y_i \sim \text{Poisson}(\mu),$$

where $\mu \sim$ Gamma$(2, \beta)$ and $\beta \sim$ Exponential(1). Prove the following:

$$\mu \mid y, \beta \sim \text{Gamma}\left(2 + \sum_i y_i, n + \beta\right) \tag{6}$$

$$\beta \mid y, \mu \sim \text{Gamma}\left(3, 1 + \mu\right) \tag{7}$$

We this time have the likelihood of observing y as:

$$\begin{aligned} p(y \mid \mu, \beta) &= \prod_{i=1}^{n} \frac{\mu^{y_i} e^{-\mu}}{y_i!} \\ &= \frac{\mu^{\sum_{i=1}^{n} y_i} e^{-n\mu}}{\prod_{i=1}^{n} y_i!} \end{aligned}$$

the prior pdf for μ:

$$p(\mu) = p(\mu; 2, \beta) = \frac{\beta^2}{\Gamma(2)} \mu e^{-\beta\mu}$$

and the prior pdf for β:

$$p(\beta) = p(\beta; 1) = e^{-\beta}$$

Starting with equation 6, again from Bayes theorem we have that the posterior of μ given y and β is proportional to:

$$\begin{aligned} p(\mu \mid y, \beta) &\propto p(y \mid \mu, \beta) p(\mu) \\ &\propto \mu^{\sum_{i=1}^{n} y_i} e^{-n\mu} \mu e^{-\beta\mu} \\ &= \mu^{1+\sum_{i=1}^{n} y_i} e^{-(n+\beta)\mu} \\ &= \mu^{2+\sum_{i=1}^{n} y_i - 1} e^{-(n+\beta)\mu} \end{aligned}$$

which we recognize as a Gamma distribution with parameters $\alpha = 2 + \sum_{i=1}^{n} y_i$ and $\beta = (n + \beta)$, and hence we have shown expression 6.

For expression 7, we use Bayes theorem which yields that the posterior of β given μ and y is proportional to:

$$\begin{aligned} p(\beta \mid y, \mu) &\propto p(\mu \mid y, \beta) p(\mu \mid y), \quad y \text{ not depending directly on } \beta \\ &= p(\mu \mid \beta) p(\mu) \\ &\propto \beta^2 e^{-\beta\mu} e^{-\beta} \\ &= \beta^{3-1} e^{-(1+\mu)\beta} \end{aligned} \tag{8}$$

which we recognize up to a constant to be a Gamma distribution with parameters $\alpha' = 3$ and $\beta' = (1 + \mu)$, hence we have shown expression 7.

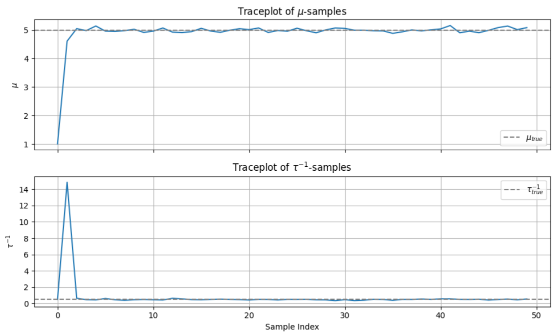
2. Implement Gibbs Sampler

Q1 Gaussian Model:

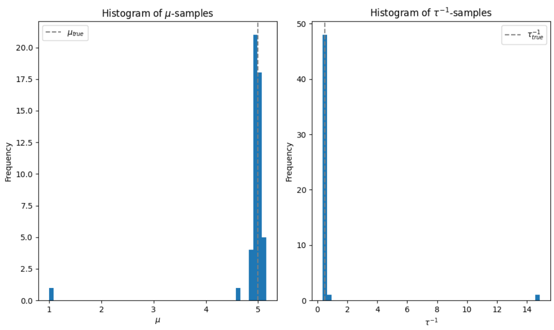
- **Use the Python code to implement Gibbs sampling for the Gaussian model.**
- **Simulate $n = 100$ observations from $y_i \sim \mathcal{N}(5, 2^{-1})$.**

See "Assignment5_Exercise_1_v2.ipynb" for code and simulation of 100 observations from a Gaussian with mean 5 and variance 0.5.

- **Plot traceplots and histograms for μ and τ^{-1} (variance). Compare with true values.**



(a)



(b)

Figure 7: Traceplots (a) and histograms (b) of posterior samples of mean μ, and variance τ^{-1}, together with true mean and variance (used when creating samples y_i).

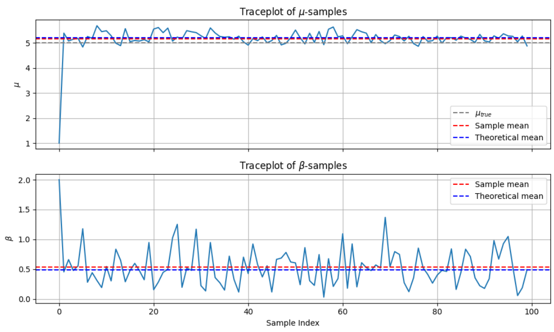
In figure 7 we see that the drawn samples align really well with the true mean and variance of the distribution since we oscillate around the true mean and variance in the traceplots, figure 7a. From this plot we also see that the algorithm converges very fast, after just a couple of samples. That the samples are good is also evident in the histogram, figure 7b, since most samples are around the true values of mean and variance.

Q2 Poisson Model:

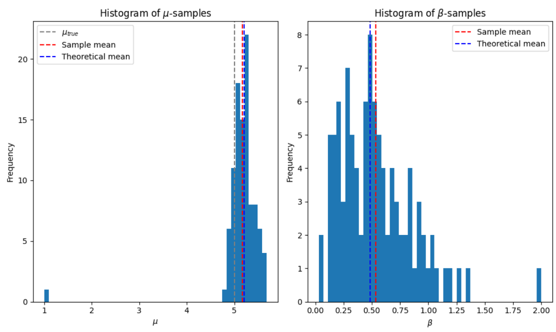
- **Extend the Gibbs sampler to the Poisson model with n = 100 observations from $y_i \sim$ Poisson(5).**

See "Q1.ipynb" for code and simulation of 100 observations from a Poisson distribution with rate parameter $\mu = 5$.

- **Plot traceplots and histograms for μ and β. Compare with theoretical expectations.**



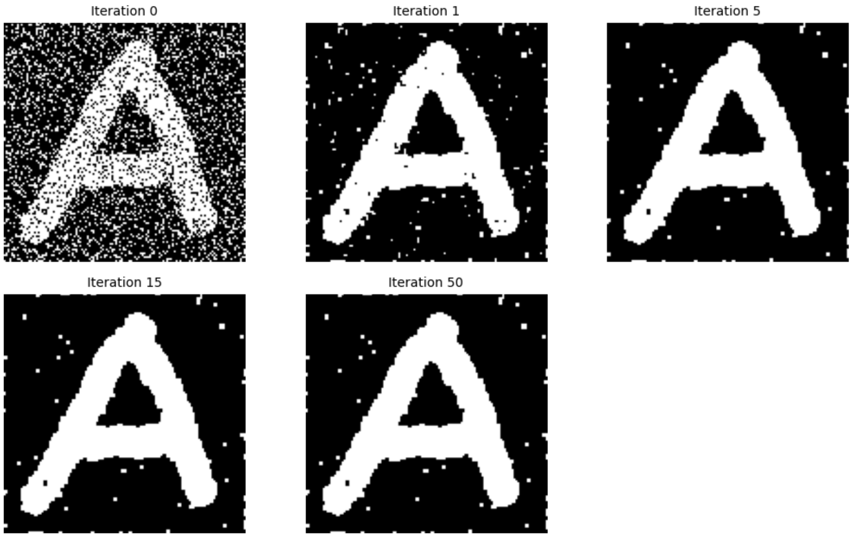
(a)



(b)

Figure 8: Traceplots (a) and histogram (b) of the posterior samples of the rate μ of the Poisson distribution, and rate β of the Gamma distribution (describing parameter μ).

Focusing on the mean, μ of the Poisson distribution we in figure 8 see that we are a bit off from the true mean (the one used when creating the data y), but the sample mean and theoretical mean are quite close to each other. This is visible both in the traceplot 8a and in the histogram 8b. Same goes for the rate parameter β, although we do not have the true parameter beta to compare with in this case.



[!h]

Figure 9: Denoised image plots from iteration: 0, 1, 5, 15, 50.

Exercise 2

One application of Markov Random Fields (MRFs) is to denoise noisy images. In this task, we aim to minimize the following energy function using Gibbs Sampling and Mean Field Approximation:

...

Q1 Implement Gibbs Sampling for Image Denoising:

- Simulate noisy images with varying noise levels (e.g., Gaussian noise or salt-and-pepper noise).
- Implement Gibbs Sampling to minimize the energy function.
- Apply the algorithm to the noisy image and produce a denoised image.
- Plot the denoised image after iterations from the following list: [0, 1, 5, 15, 50].

See "Assignment5_Exercise_2_v2.ipynb" for code. See figure 9 for plot.

References

[1] C. M. Bishop, *Pattern Recognition and Machine Learning.* Berlin, Heidelberg: Springer-Verlag, 2006, Available for free download. [Online]. Available: `https://www.microsoft.`

com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf.