Introduction to Bayesian statistics

Practical

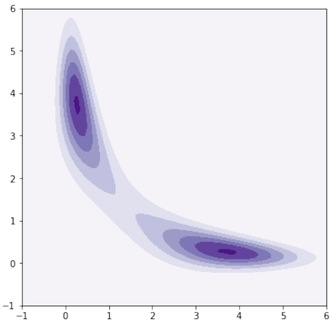
30th January 2023

2D visualisation of MCMC algorithms on a toy example

Usually, in Bayesian inference, MCMC techniques are applied to sample from a posterior distribution (which arises in a particular model after conditioning on data).

In this exercise, we have been given the following two-dimensional target distribution whose density $p(x, y)$ is known only up to a normalising constant C,

$$p(x, y) = \frac{1}{C} \exp\left(-\frac{(x^2\, y^2 + x^2 + y^2 - 8x - 8y)}{2}\right)$$



Our goal is to visualise and compare the behaviour of two sampling algorithms (**Metropolis-Hastings** and **Gibbs**) on this example. See notebook practical .ipynb for more details.

Posterior inference for a Gaussian distribution

Suppose we have observations $D = \{x_1, \ldots, x_N\}$ and we have decided to model them using a Gaussian distribution, i.e. we assume

$$x_1, \ldots, x_N \sim \mathcal{N}(\mu, \sigma^2)\,.$$

Our goal is to perform posterior inference over both of these parameters. Often it is easier to work with the precision (i.e. inverse variance) of the Gaussian instead, so we denote $\lambda := 1/\sigma^2$ and work with λ instead.

We place the following prior over the parameters (μ, λ)

$$p(\mu, \lambda) = \mathcal{N}\left(\mu \,|\, m_0, \frac{1}{\kappa_0 \lambda}\right) \mathrm{Gamma}(\lambda | \alpha_0, \beta_0)$$

with fixed constants $m_0 = 0, \kappa_0 = 1$ and $\alpha_0 = 1, \beta_0 = 1$.

Task 1 To get a sense what the prior looks like, generate some pairs of values $(\mu^*, \lambda^*) \sim p(\mu, \lambda)$ and visualise the corresponding distributions

Task 2 Next, our goal is to perform posterior inference via Gibbs sampling. For this,

1. Write down the likelihood $p(D|\mu, \lambda)$.
2. For Gibbs sampling, we need to be able to generate samples from the full conditional distributions. For this, derive the expressions for $p(\mu|D, \lambda)$ and $p(\lambda|D, \mu)$.

   Tip: The product of two Gaussian PDFs is a Gaussian PDF

3. Implement a Gibbs sampler (in your chosen programming language) to generate samples from the posterior
4. Run the Gibbs sampler, assuming we have 5 observations $[2.7, 0.4, -0.1, 3.2, -0.9]$
5. How well did your Gibbs sampler work? Comment on the convergence and mixing of the chain.

Task 3 Analogously to task 3, visualise samples from the posterior, by drawing pairs of values $(\mu^*, \lambda^*) \sim p(\mu, \lambda|D)$

Task 4 (optional) Instead of writing your own sampler, alternatively one could have used a probabilistic programming framework such as Stan. Following an example such as https://mc-stan.org/docs/stan-users-guide/linear-regression.html, can you code the above model in Stan?

[Optional] Posterior inference for Student's t-distribution

[Optional] Part 1: Student's t-distribution as a scale mixture of Gaussians

Student's t-distribution is known to be more heavy-tailed alternative to a Gaussian distribution that is more robust towards outliers. But where does it come from? In fact, t-distribution can be expressed as an (infinite) mixture of Gaussians[1]

Here is one possible construction for a t-distribution with ν degrees of freedom:

$$t_\nu(y|\mu, \sigma^2) = \int_0^\infty \mathcal{N}(y \,|\, \mu, \sigma^2/\kappa)\, \mathrm{Gamma}(\kappa \,|\, \nu/2, \nu/2) d\kappa$$

We can use this construction for sampling. In order to sample

$$y \sim t_\nu(\mu, \sigma^2)$$

we can equivalently sample an *auxiliary variable* κ, and then conditional on κ we can sample from a Gaussian:

$$\kappa \sim \mathrm{Gamma}(s \,|\, \frac{\nu}{2}, \frac{\nu}{2})$$
$$y|\kappa \sim \mathcal{N}(\mu, \sigma^2/\kappa)$$

Visualise this mixture, by sampling κ^* values, and see if the resulting distribution of y^* values matches the t-distribution density.

[Optional] Part 2: Posterior inference

Suppose we decided that a Gaussian distribution did not provide a good fit to our data, and it would be better to use a more heavy-tailed distribution $t_\nu(\mu, \sigma^2)$ with $\nu = 3$,

$$x_1, \ldots, x_N \sim t_\nu(\mu, \sigma^2)\,.$$

Can you write a Gibbs sampler via introducing the auxiliary variables from the previous scale-mixture question?

[1] https://www.sumsar.net/blog/2013/12/t-as-a-mixture-of-normals/.