Markov Chain Monte Carlo (MCMC) and Bayesian Inference with an Intractable Posterior Distribution

In Bayesian inference, our goal is to find the **posterior distribution**, which represents our updated beliefs about the parameters of a model after observing data. Mathematically, it's given by **Bayes' Theorem**:

$P(\theta|D)=P(D)P(D|\theta)P(\theta)$

Where:

- $P(\theta|D)$ is the **posterior distribution** of the parameters θ given the data D.
- $P(D|\theta)$ is the **likelihood function**, representing the probability of observing the data D given the parameters θ.
- $P(\theta)$ is the **prior distribution**, representing our beliefs about the parameters θ before observing the data.
- $P(D)=\int P(D|\theta)P(\theta)d\theta$ is the **evidence** or marginal likelihood, which is the probability of observing the data D marginalized over all possible values of θ. This acts as a normalization constant, ensuring the posterior distribution integrates to 1.

The Challenge: Intractable Posterior Distributions

The **evidence term, P(D)**, is often the culprit in making the posterior distribution **intractable**. Intractability means that:

1. **Analytical Integration is Impossible:** The integral $\int P(D|\theta)P(\theta)d\theta$ cannot be solved in closed form using standard mathematical functions. This often arises when the likelihood and/or prior distributions are complex and their product doesn't lead to a recognizable, analytically integrable form.
2. **Normalization Constant is Unknown:** Consequently, we cannot directly calculate the normalization constant P(D), and therefore we cannot obtain the exact functional form of the posterior distribution $P(\theta|D)$.

Why MCMC is Needed:

When the posterior distribution is intractable, we cannot directly sample from it. This is where Markov Chain Monte Carlo (MCMC) methods come into play. MCMC algorithms allow us to obtain samples from the posterior distribution *without needing to explicitly calculate the intractable normalization constant P(D)*.

MCMC works by constructing a Markov chain whose stationary distribution is the desired posterior distribution. By running the chain for a sufficiently long time, the generated samples will approximately follow the posterior distribution. The acceptance/rejection steps in MCMC algorithms (like Metropolis-Hastings or Gibbs sampling) rely on the *ratio* of the unnormalized posterior density, $P(D|\theta t)P(\theta t)P(D|\theta *)P(\theta *)$, where the P(D) terms cancel out. This means we only need to be able to evaluate the product of the likelihood and the prior, which is usually feasible even when their integral is not.

Example of an Intractable Posterior Distribution: Logistic Regression with Complex Priors

Consider a binary logistic regression model where we want to predict a binary outcome $y\in\{0,1\}$ based on a set of predictors x. The likelihood function for N independent observations is:

$P(D|\beta)=i=1\prod Np(xi)yi(1-p(xi))1-yi$

where $p(xi)=1+\exp(-\beta Txi)1$ is the probability of $yi=1$, and β is the vector of regression coefficients we want to estimate.

Now, let's introduce a complex prior distribution for the coefficients β. Instead of using simple independent Gaussian priors, suppose we believe that the coefficients might be related in a complex way, and we define a prior distribution that reflects this dependency, for example, a hierarchical prior with multiple levels or a prior based on a complex, non-standard distribution.

$P(\beta)=\text{Complex, Non-Standard Distribution}(\beta\mid\text{hyperparameters})$

The posterior distribution is then:

P(β∣D)=∫(∏i=1N(1+exp(−βTxi)1)yi(1−1+exp(−βTxi)1)1−yi)×Complex, Non-Standard Distribution(β∣hyperparameters)dβ(∏i=1N(1+exp(−βTxi)1)yi(1−1+exp(−βTxi)1)1−yi)×Complex, Non-Standard Distribution(β∣hyperparameters)

Intractability:

The integral in the denominator, $P(D)$, is likely to be **analytically intractable** due to the non-linear form of the logistic likelihood and the complexity of the prior distribution. There is no closed-form solution for this integral in most cases involving complex priors.

Solution with MCMC:

Even though we cannot calculate $P(D)$, we can still evaluate the numerator, which is proportional to the posterior:

$P(\beta\mid D)\propto P(D\mid\beta)P(\beta)$

MCMC algorithms like Metropolis-Hastings can be used to sample from this posterior distribution. We would:

1. Start with an initial guess for β.
2. Propose a new value β_*.
3. Calculate the acceptance ratio based on the ratio of the unnormalized posterior density at β_* and the current value of β.
4. Accept or reject the proposed β_* based on this ratio.
5. Repeat this process to generate a sequence of samples that approximate the posterior distribution of β.

In summary, an intractable posterior distribution arises when the normalization constant (evidence) cannot be calculated analytically. Complex likelihood functions combined with complex prior distributions often lead to such intractability in Bayesian models like logistic regression with sophisticated prior structures. MCMC methods provide a powerful way to sidestep the need to calculate this constant by directly sampling from the unnormalized posterior density.