When comparing Maximum Likelihood Estimation (MLE) and Bayesian methods for predicting probabilities, you can consider the following:

Bias
MLE can be biased when there is minimal data. For example, if you flip a fair coin three times and each time it lands on heads, MLE would estimate that the probability of heads is 1 and the probability of tails is 0.

Prior knowledge
Bayesian methods can incorporate prior knowledge to help produce more accurate models. For example, if you know a coin is fair, you can use that prior knowledge in a Bayesian method to produce a more accurate model.

Output
MLE and MAP are point estimators that produce a single fixed value. Bayesian inference produces a probability density function or probability mass function, rather than a single value.

Here are some other differences between MLE and Bayesian methods:

Performance
Bayesian methods can be more accurate and have better coverage than MLE, especially for small sample sizes.

Convergence
Both methods can struggle to converge, but incorporating prior knowledge can help Bayesian methods converge toward realistic solutions.

- [convergence is when a model reaches a stable state where its predictions stop improving and its error rate stays constant. This means that the model's parameters, such as its weights and biases, have settled on values that can accurately predict the training data.]
-

Implementation

Both methods can be implemented using the Markov chain Monte Carlo algorithm.

Q. Compare and contrast Maximum Likelihood Estimation with Bayesian methods for predicting probabilities.

Key Differences between MLE and Bayesian Estimation

While both, Maximum Likelihood Estimation and Bayesian Estimation , are parameter estimation techniques based on probability distribution, There are some key differences between the two.

Factors	MLE	Bayesian Estimation
Predictions	In MLE, We make use of latent variables for making estimations.	We use Posterior distribution, prior probability, and evidence to make estimations.
Ideal scenario to use with	Can become biased with minimal data	Since it uses previous knowledge previous makes it a bit more reliable.
Complexity is	Relatively less complex. Maximum likelihood estimation is computationally less expensive since it requires differential calculus techniques or gradient descent for theta.	More complex since we need likelihood, prior, and evidence to make the computations. The computation may require complex multi-dimensional integration.
Interpretability	MLE is easier to understand since it chooses the single best model among all the provided models.	Bayesian can be slightly less intuitive since it gives the weighted average of all the models.

Feature	**(MLE)**	**Bayesian Methods**
Nature of Parameters	Parameters are fixed values	Parameters are random variables
Prior Knowledge	No prior knowledge included	Incorporates prior knowledge explicitly
Result	Point estimate	Posterior distribution
Uncertainty	Uncertainty estimated indirectly	Uncertainty directly captured in posterior
Computational Complexity	Typically simpler, closed-form solutions	Computationally more complex, often requiring approximation
Interpretation	Best estimate of parameters based on data	Full probabilistic interpretation, including prior beliefs and data
Flexibility	Can struggle with complex models or small datasets	Handles small datasets or prior information well

Contrast:

- MLE is typically more straightforward and computationally efficient, providing a single point estimate for the parameters. However, it does not account for prior information or provide a full picture of uncertainty.
- Bayesian methods offer a richer framework that includes prior knowledge and a more comprehensive treatment of uncertainty through the posterior distribution but are often more computationally intensive.

MLE is often preferred for simplicity and efficiency, especially when prior information is not available or not relevant. Bayesian methods are more flexible, especially when prior knowledge or uncertainty plays a critical role, but at the cost of more complex computations.

Markov Chain Monte Carlo (MCMC) is used in statistics & various scientific fields to sample from complex probability distributions.

What Is Markov Chain Monte Carlo?

Markov Chain Monte Carlo (MCMC) is a powerful technique used in statistics and various scientific fields to sample from complex probability distributions. It is particularly useful when directly sampling from the distribution is difficult or impossible. Here is a breakdown of the name:

- Monte Carlo: This refers to a general approach using randomness to solve problems, drawing inspiration from the element of chance involved in casino games.
- Markov Chain: This is a sequence of random events where the probability of the next event depends only on the current event, not the history leading up to it.

MCMC combines constructing a Markov chain and recording samples from the chain. The chain is designed to spend more time in regions with higher probability according to the target distribution. Then, by recording states from the chain after it has 'warmed up' and reached a stable state, you effectively get samples from the target distribution.

How Is Markov Chain Monte Carlo Used In Machine Learning?

MCMC plays a crucial role in various aspects of machine learning, particularly when dealing with complex probabilistic models or situations where direct sampling is difficult. Here are some key ways it's utilised:

- **Bayesian Inference:** Machine learning often involves estimating unknown parameters in models based on observed data. In the Bayesian framework, these parameters are treated as random variables with prior probability distributions. MCMC helps sample from the posterior distribution, which combines the prior information with the likelihood of the data, allowing for a better understanding of the parameter uncertainties and making predictions with appropriate confidence intervals.
- **Model Selection:** When choosing between different models, MCMC can be used to compare their posterior probabilities by integrating over the parameter space. This helps identify the model that best fits the data and accounts for the model's complexity.
- **Latent Variable Models:** These models involve hidden variables that are not directly observed but influence the observed data. MCMC is used to infer the

posterior distribution of these latent variables, providing insights into the underlying structure of the data. This is crucial in techniques like dimensionality reduction and topic modelling.

- **Variational Inference (VI):** While not directly using Markov Chain Monte Carlo, some machine learning algorithms like Variational Inference (VI) borrow ideas from MCMC. VI approximates the posterior distribution through an optimisation process inspired by MCMC, making it applicable when exact MCMC sampling might be computationally expensive.
- **Deep Learning**: Markov Chain Monte Carlo can be integrated with deep learning techniques, particularly in Bayesian deep learning, where MCMC helps sample from the posterior distribution of the network weights, enabling learning and uncertainty quantification.

Markov property

A random sequence is a Markov chain if and only if, given the current value , the future observations are conditionally independent of the past values , for any positive integers and .

This property, known as the Markov property, says that the process is memoryless: the probability distribution of the future values of the chain depends only on its current value , regardless of how the value was reached (i.e., regardless of the path followed by the chain in the past).

Markov property

$$P(X_{t+n} = x | X_t, X_{t-1}, \dots, X_{t-k}) = P(X_{t+n} = x | X_t)$$

Conditional and unconditional distributions

Thanks to the Markov property, we basically need only two things to analyse a chain:

1. the conditional probabilities (or densities) that $X_{t+n} = x_{t+n}$, given that $X_t = x_t$, denoted by

   $f(x_{t+n}|x_t)$

2. the unconditional probabilities that $X_t = x_t$, denoted by $f(x_t)$

What Are The Advantages & Disadvantages Of Markov Chain Monte Carlo?

As with each machine learning technique, MCMC also has its benefits and drawbacks —

Advantages Of MCMC:

- **Handles Complex Distributions:** MCMC excels at sampling from intricate probability distributions, even when direct sampling is impossible or inefficient. This makes it invaluable for various applications in statistics, machine learning, and scientific simulations.
- **No Analytical Solutions Required:** Unlike some methods that require deriving analytical solutions, MCMC can operate even when such solutions are unavailable. This provides a flexible and robust approach when dealing with challenging problems.
- **Provides Uncertainty Quantification:** MCMC enables the generation of samples from the posterior distribution, allowing for the estimation of uncertainty associated with parameters or predictions. This is crucial for building reliable and interpretable models in various AI applications.
- **Widely Applicable:** MCMC finds use in diverse fields like Bayesian inference, machine learning, physics, economics, and finance. Its versatility makes it a powerful tool for tackling problems across various domains.
- **Relatively Easy Implementation:** Compared to some other advanced statistical techniques, MCMC algorithms can be relatively straightforward to implement, especially with readily available software libraries.

Disadvantages Of MCMC:

- **Computational Cost:** MCMC simulations can be computationally expensive, especially when dealing with high-dimensional distributions or requiring high accuracy. This can limit its applicability in situations with limited computational resources.
- **Convergence Issues:** Ensuring proper convergence of the Markov chain to the target distribution is crucial. This can be challenging and requires careful monitoring and diagnostics to avoid obtaining biased results.
- **Sensitivity To Starting Point:** The initial state of the Markov chain can impact the convergence process. Choosing an inappropriate starting point can

lead to slow convergence or even getting stuck in irrelevant regions of the distribution.
- **Difficulties In Assessing Convergence:** Evaluating the convergence of the Markov chain can be complex and subjective. Different tests and diagnostics are available but they might not always provide a definitive answer and require careful interpretation.
- **Not Always The Best Option:** Depending on the specific problem and available resources, other methods like gradient-based optimisation might be more efficient or suitable alternatives to MCMC.