King David Concepcion

CAS-05-601P

What is Bayesian Inference?
Bayesian inference is a way of making statistical inferences in which the statistician assigns subjective probabilities to the distributions that could generate the data. These subjective probabilities form the so-called prior distribution. (Taboga, 2021)

Bayesian inference is a method to figure out what the distribution of variables is. The interesting feature of Bayesian inference is that it is up to the statistician (or data scientist) to use their prior knowledge as a means to improve our guess of how the distribution looks like. (de Bastos, 2022)

Bayesian inference is all about using probability distributions to reason about unknowns! Here's the core idea:

Prior Distribution: You start with an initial belief about an unknown quantity (parameter) represented by a probability distribution. This is called the prior distribution. It reflects your existing knowledge or lack thereof.

Data Likelihood: You then observe some data relevant to the unknown quantity. This data has its own probability distribution, known as the likelihood function. It describes how probable the observed data is for different values of the unknown quantity.

Posterior Distribution: Finally, you combine the prior and the likelihood using Bayes' theorem to get a posterior distribution. This is the updated belief about the unknown quantity after considering the data. The posterior distribution reflects how your initial belief has been refined by the new information.

Choice of Distributions:

The specific probability distributions used in Bayesian inference depend on the problem at hand.

Continuous Unknown Quantities:

Normal Distribution: This approach is particularly common when the underlying process generating the data is assumed to be normally distributed, or when the Central Limit Theorem can be applied to approximate the distribution of the data as normal. A popular choice for continuous unknowns like means or proportions. For the prior, you might use a normal distribution with a wide variance reflecting weak initial belief. The likelihood could also be normal if the data is assumed to be normally distributed around the unknown means.
Here's how the normal distribution is used in Bayesian inference:

Likelihood Function: Suppose you have observed data X $\{x_1, x_2, x_3, \ldots, x_n\}$, and you want to infer the parameters of the underlying distribution from which this data was generated. If you assume that the data is generated from a normal distribution, you can write the likelihood function as:

$$P(X|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

where:

- $\theta = (\mu, \sigma^2)$ are the parameters of the normal distribution.
- μ is the mean of the normal distribution.
- σ^2 is the variance of the normal distribution.

This expression gives the probability of observing the entire dataset given the parameters θ.

Prior Distribution: Before observing the data, you specify a prior distribution for the parameters μ and σ^2. This prior distribution encodes your beliefs or assumptions about the parameters before seeing the data. Common choices for priors include the normal distribution itself (conjugate prior), or other distributions based on domain knowledge or previous data.

Posterior Distribution: Bayes' theorem is used to update the prior distribution to obtain the posterior distribution:

$$P(\theta|X) = \frac{P(X|\theta) \cdot P(\theta)}{P(X)}$$

where:

- $P(\theta|X)$ is the posterior distribution, representing our updated beliefs about the parameters after observing the data.
- $P(X|\theta)$ is the likelihood function.
- $P(\theta)$ is the prior distribution.
- $P(X)$ is the marginal likelihood, also known as the evidence, which acts as a normalization constant.

Parameter Estimation and Uncertainty Quantification: From the posterior distribution, you can derive estimates of the parameters (e.g., the mean μ and variance σ^2) and quantify uncertainties associated with these estimates. This can be done by computing summary statistics (e.g., mean, median, mode) or by constructing credible intervals.

Uniform Distribution: Used when you have minimal prior knowledge, essentially assigning equal probability to a range of values for the unknown quantity. A uniform distribution describes a scenario where all outcomes within a specific range are equally probable. It applies to both continuous and discrete data.

Continuous Uniform Distribution: Imagine a ruler. If a random point is marked on the ruler between endpoint A (say, 2) and endpoint B (say, the probability of the point landing at any specific value (3, 5.4, 7. is the same. This represents a continuous uniform distribution.

Mathematically, a continuous uniform distribution over the interval [a, b] is denoted as X ~ U(a, b). Here, a is the lower bound, b is the upper bound, and X is the random variable.

The probability density function (PDF) of a continuous uniform distribution is constant within the interval [a, b] and zero elsewhere. This means the height of the probability distribution is the same across the valid range.

Discrete Uniform Distribution: Consider a six-sided die. Each outcome (1, 2, 3, 4, 5, or 6) has an equal probability of occurring when you roll the die. This is an example of a discrete uniform distribution.

There's no well-defined formula for the PDF in a discrete case, but you can assign a probability of 1/n (n being the number of possible outcomes) to each value.

Properties of Uniform Distribution:

Symmetry: The distribution is symmetrical around the midpoint of the interval (a + b)/2 for continuous distributions.

Maximum Entropy: The uniform distribution has the maximum entropy among all distributions with the same support (the range of possible values). This implies it makes the least assumptions about the data.

Applications:

Uniform distributions are fundamental for simulating random numbers. Most random number generators use a uniform distribution as the base and then apply transformations to generate numbers following other distributions.

Uniform distributions are helpful in modeling scenarios where all outcomes within a range are equally likely, such as waiting times within a specific interval or random errors following a flat distribution.

Discrete Unknown Quantities:

Binomial Distribution: Suitable for situations where the data has binary outcomes (success/failure). The prior could be a beta distribution, which is a conjugate prior for the binomial likelihood, making calculations easier.

By using the binomial distribution in Bayesian statistics, you can effectively model binary data, estimate parameters, and make predictions while incorporating uncertainty in a principled manner.

Example: Suppose you run an e-commerce website and you're interested in estimating the conversion rate, i.e., the probability that a visitor makes a purchase. To do this, you conduct an A/B test where you randomly assign visitors to one of two versions of your website: the current version (control) or a new version (treatment). You track whether each visitor makes a purchase or not.

Data Collection: Over a period of time, you collect data on the number of visitors to each version of the website and the number of visitors who made a purchase. Let's say you conducted the test for a week and recorded the following data:

Control group (current website version): 500 visitors, 40 purchases

Treatment group (new website version): 520 visitors, 60 purchases

Modeling with Binomial Distribution: Since each visitor's action (purchase or not) can be considered a Bernoulli trial with a certain probability of success (conversion rate p), you model the likelihood of observing the data using the binomial distribution.

In this case, let pc be the conversion rate for the control group and pt be the conversion rate for the treatment group. The likelihood function for each group can be written as:

$$P(\text{Control data}|\theta) = p_c^{40}(1-p_c)^{500-40}$$

$$P(\text{Treatment data}|\theta) = p_t^{60}(1-p_t)^{520-60}$$

Prior Distribution: Before observing the data, you specify prior distributions for pc and pt. A common choice is the beta distribution, which is the conjugate prior for the binomial likelihood. Let's say you choose uninformative Beta(1, 1) priors for both conversion rates.

Posterior Distribution: Using Bayes' theorem, you update the prior distributions with the observed data to obtain the posterior distributions for pc and pt.

Parameter Estimation and Uncertainty Quantification: From the posterior distributions, you can derive estimates of the conversion rates for both website versions and quantify the uncertainties associated with these estimates, such as computing credible intervals.

Comparison and Inference: Finally, you compare the posterior distributions of the conversion rates for the control and treatment groups to determine if the new website version led to a significant improvement in the conversion rate.

Poisson Distribution: Useful for counting data (number of events). The prior could be a gamma distribution, another conjugate prior for the Poisson likelihood.

Visualizing these probability distributions throughout the Bayesian process is very helpful. You can see how the prior belief gets sharpened or shifted based on the data, reflected in the posterior distribution.

Data Collection: Over a period of time, you collect data on the number of customer arrivals per hour. Let's say you collected data for a week and recorded the following number of arrivals per hour:

Monday: 10 arrivals

Tuesday: 12 arrivals

Wednesday: 9 arrivals

Thursday: 11 arrivals

Friday: 13 arrivals

Saturday: 8 arrivals

Sunday: 10 arrivals

Modeling with Poisson Distribution: The Poisson distribution can model the probability of observing a certain number of events (customer arrivals) in a fixed interval of time given the average rate λ. In this case, let

X represent the number of customer arrivals per hour, and

λ represent the average rate of arrivals per hour. The likelihood function for the data can be written as:

$$P(X|\lambda) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

where x is the observed number of arrivals in a particular hour.

Prior Distribution: Before observing the data, you specify a prior distribution for the parameter λ. A common choice for the prior distribution of λ is the gamma distribution, which is the conjugate prior for the Poisson likelihood function. You might use historical data or domain knowledge to inform your choice of prior distribution parameters.

Posterior Distribution: Using Bayes' theorem, you update the prior distribution with the observed data to obtain the posterior distribution for λ.

Parameter Estimation and Uncertainty Quantification: From the posterior distribution, you can derive estimates of the average rate of customer arrivals per hour (λ) and quantify uncertainties associated with this estimate, such as computing credible intervals.

Resource Allocation and Planning: Finally, you use the posterior distribution of λ to make decisions about resource allocation and service planning. For example, you might use the estimated arrival rate to determine staffing levels during different hours of the day or days of the week.

References

Bastos, P. M. de. (2022, June 22). *How to use Bayesian Inference for predictions in Python*. Medium. https://towardsdatascience.com/how-to-use-bayesian-inference-for-predictions-in-python-4de5d0bc84f3

Brian. (2012, October 23). *Basics of Bayesian Statistics*. https://www.stat.cmu.edu/~brian/463-663/week09/Chapter%2003.pdf

Shen, S. (2024, March 30). *Bayesian Statistics*. Towards Data Science. https://towardsdatascience.com/tagged/bayesian-statistics

Taboga, M. (2021a). *Bayesian inference | Introduction with explained examples*. Www.statlect.com. https://www.statlect.com/fundamentals-of-statistics/Bayesian-inference#:~:text=Bayesian%20inference%20is%20a%20way

Taboga, M. (2021b). *Normal distribution - Bayesian estimation*. Www.statlect.com. https://www.statlect.com/fundamentals-of-statistics/normal-distribution-Bayesian-estimation