

# Bayesian Inference for Species Distribution Modelling with Gaussian Processes

Isaac William Caruso

Amherst College  
Department of Computer Science

2021-03-30



# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Probability, Bayes' theorem, and Bayesian inference</b>   | <b>5</b>  |
| 1.1      | Random Variables . . . . .   | 6         |
| 1.2      | Cumulative Distribution Functions . . . . .  | 7         |
| 1.3      | Probability Mass and Density Functions . . . . .   | 8         |
| 1.4      | Expectation . . . . .  | 11        |
| 1.5      | Conditional Probability . . . . .  | 13        |
| 1.6      | Bayes' Theorem . . . . .   | 14        |
| 1.7      | Bayesian Inference . . . . .   | 14        |
| <b>2</b> | <b>Approximation algorithms for inference on complex systems</b>                                       | <b>15</b> |
| <b>3</b> | <b>Modern Bayesian inference with STAN</b>   | <b>17</b> |
| <b>4</b> | <b>Bayesian applications in Biology: Hybrid species distribution modelling with gaussian processes</b> | <b>19</b> |



# Chapter 1

## Probability, Bayes' theorem, and Bayesian inference

The objective of this chapter is to provide an introduction and brief summary of the probability terms and concepts necessary for conceptualizing and approaching Bayesian inference. Simply put, Bayesian inference is a statistical technique for iteratively updating a probability quantity of interest upon the observation of data. Specifically, the statistical technique Bayesian inference relies on is known as Bayes' theorem and the probability quantity of interest is most often a probability density function, though in introductory examples is oftentimes a single value. Before embarking on this exposition of Bayesian statistics, one must first gain a basic appreciation for a few key elements of probability theory which will provide us with the foundational terms to discuss Bayesian inference and its applications in the broad field of machine learning—random variables and probability functions. A brief discussion of random variables will naturally lead to the introduction of cumulative distribution functions and probability functions/ distributions. An understanding of probability functions will provide the fundamental terminology necessary for introducing a discussion of the concepts expected value and conditional probability. This chapter will then conclude by

introducing Bayes' theorem and Bayesian inference in demonstrating the steps for performing a simple Bayesian update.

## 1.1 Random Variables

Imagine for a moment you are tossing a coin. There are many experiments you could perform by tossing a coin, but for the sake of simplicity let us consider our quantity of interest to be the number of times our coin lands on a tails. It is clear that the number of tails, the outcome of our experiment, is dependent on the eventual realization of some random (stochastic) process, or at the very least not clearly deterministic. In countless scenarios like the coin toss experiment, random variables provide a convenient and uniform terminology for discussing sample spaces and their connection to some sequence of observations that is the data. In essence, a random variable is a variable whose value is dependent on the outcome(s) of a stochastic phenomenon. A random variable is commonly expressed as  $X$ , and a realized value of  $X$  is  $x$ .

In the canonical example of tossing a coin, where the data is some sequence of coin tosses  $[H, T, T, \dots]$ , we could define a random variable  $X$  to be the number of tails. In this toy example, as noted above, the realized value of  $X$ , is dependent on the outcome(s) of a random event—a coin toss. In the case of tossing a coin twice consecutively,  $X$ —the number of tails observations—has three possible realized states,  $x$ , depending on the outcome of this stochastic experiment:  $x = 0$ ,  $x = 1$ , or  $x = 2$ . Additionally, assuming the coin is a fair coin the probability that our random variable  $X$  is realized as  $x$ , some real number of tails, can be assigned to each possible overall outcome of this coin toss experiment as represented in Table 1.1

In reality, this is an example of a specific type of random variable known as a discrete random variable. Discrete random variables can, as their name implies, only assume discrete values.

Table 1.1:  $P(X = x)$  for two tosses

| $x$ | $P(X = x)$ |
|-----|------------|
| 0   | 0.25       |
| 1   | 0.50       |
| 2   | 0.25       |

To the contrary, continuous random variables and mixed random variables are useful for describing sample spaces with continuous and mixed outcomes, respectively. For example, a continuous random variable may be used to describe an experiment measuring blossoming heights of flowers, where the data is a sequence of observations of heights at which different flowers blossomed. In this case, the outcome of our blossoming experiment is an infinite number of real values that is properly represented in a continuous random variable.

## 1.2 Cumulative Distribution Functions

The concept of a distribution follows intuitively from our previous discussion of random variables. In the previous example we represented the probability of various outcomes of a coin toss experiment in a tabular format. Another way to represent this distribution of probabilities is as a cumulative distribution function:

**Definition 1.1** (Cumulative distribution function 'CDF'). *The cumulative distribution function is defined as a function where  $F_X \in [0, 1]$ :*

$$F_X(x) = P(X \leq x)$$

As the name implies, the cumulative distribution function simply represents the probability that a random variable  $X$  is realized to be less than or equal to  $x$  for each possible input value of  $x$ . Figure 1.1 depicts a graphical representation of the CDF for our coin tossing

experiment.

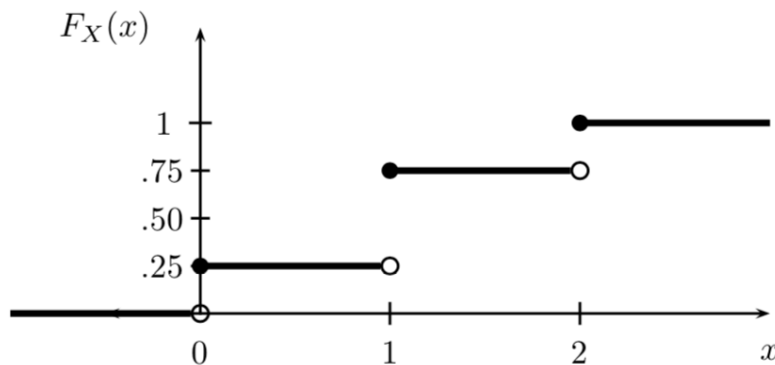


Figure 1.1: CDF for tossing a coin twice (from All of Stats, should I make my own?)

Here, for every value of  $x$ —the number of tails in two coin tosses—the probability that  $X$  is equal to or less than this value is represented. In this discrete example, we see that our CDF is represented by several non-decreasing discrete lines defined for all  $x$ . In the example of a continuous random variable, this function is a continuous, non-decreasing distribution also defined for all  $x$ .

### 1.3 Probability Mass and Density Functions

Now that we have a basic understanding of random variables and their association with the CDF, we can begin discussing probability functions for both continuous and discrete random variables. Broadly speaking, the probability mass function and probability density function are two methods for calculating a probability over a sample space, which find their use with discrete and continuous random variables respectively.

In the discrete setting, a probability mass function yields the probability of an outcome for every possible outcome in any given discrete schema. To reiterate the previous discussion of discrete versus continuous random variables, a random variable  $X$  is considered discrete “if it takes countably many values  $\{x_1, x_2, \dots\}$ ” (Wasserman, 2004). In this case, a definition



for the probability mass function follows:

**Definition 1.2** (Probability mass function 'PMF'). *The probability function for a discrete random variable  $X$ —the probability mass function for  $X$ —is defined as a function*

$$F_X(x) = P(X = x)$$

Here, the PMF has a few key attributes. Namely  $P(X = x) > 0$  for every  $x$  in the sample space of  $X(x \in S_X)$ , and  $\sum_{x \in S_X} f(x) = 1$ . With these features in mind, the probability mass function of  $X$  follows logically from the cumulative distribution function of  $X$  insofar as the CDF is the sum of the PMF for all  $x_i \leq x$ . This relationship in the discrete scenario between the PMF and CDF can be expressed formally as:

$$F_X(x) = P(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$$

In the continuous setting where these attributes no longer hold, a probability density function (PDF) is instead used to represent probability of  $x$  for all  $x \in S_X$ . In the case where the random variable  $X$  is continuous, a PDF is defined as follows:

**Definition 1.3** (Probability density function 'PDF'). *The probability function for a continuous random variable  $X$ —the probability density function for  $X$ —is defined as a function  $f(x)$  where  $a$  and  $b$  are two real numbers and every  $a \leq b$ , so*

$$P(a < X < b) = \int_a^b f_X(x) dx$$

In other words, the probability that the realized value  $x$  of our continuous random variable  $X$  is between two numbers  $a$  and  $b$  is equal to the integral of the probability density function of  $x$  from  $x = a$  to  $x = b$ . This formalization of the PDF  $f_X(x)$  allows a natural comparison

to be drawn to the CDF  $F_X(x)$  of a continuous random variable  $X$ , that is:

$$F_X(x) = \int_{-\infty}^x f_X(x)dx$$

Specifically, this implies that  $F'_X(x) = f_X(x)$  for all differentiable points in  $F_X$ . In plain English this signifies that the derivative of the CDF is the PDF, in all cases involving a continuous random variable.

Probability density and mass functions come in a wide variety of forms, each with their own properties and methodologies for defining their shape and characteristics. Some are useful for defining binary data, others are useful for defining arbitrarily large sample spaces with any number of localized concentrations of density. Still others boast properties, such as being closed under conditioning and marginalization, that make them useful for specific statistical applications. Figure 1.2 displays some of the most common continuous and discrete probability distributions.

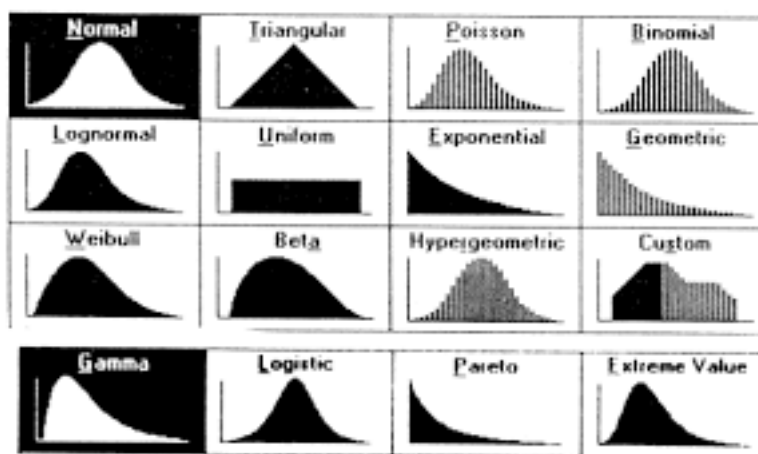


Figure 1.2: Common probability distributions

## 1.4 Expectation

One of the final core statistical concepts necessary to approach Bayesian statistics on sure footing is the idea of expectation or expected value. Simply put, the expectation of a random variable  $X$  is the average value of  $X$ .

**Definition 1.4** (Expectation 'E'). *The expectation, expected value, or mean of a random variable  $X$  is*

$$E(X) = EX = \begin{cases} \sum_x xf(x) & \text{if } X \text{ is discrete} \\ \int xf(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

While on first glance the expectation may appear unfamiliar and dissimilar to the common understanding of a mean, key characteristics of probability distributions will foster an intuitive understanding of their similarities. To return to a familiar example, consider a random variable  $X$  to represent the number of tails in 6 coin tosses. Figure 1.3 depicts the PMF for  $X$  using the binomial distribution, which represents the probability of observing a specific number of successes in a success-failure experiment.

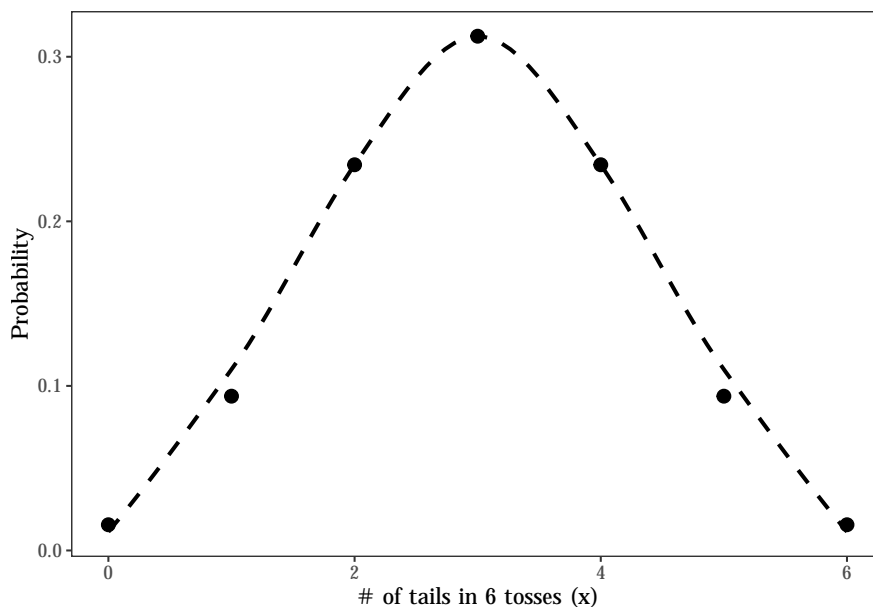


Figure 1.3: The binomial probability mass function for the 6 trial coin toss experiment

In this case, the x-axis represents each particular outcome ( $x$ ) of  $X$  and the y-axis is the relative probability of that outcome. Table 1.2 presents the output of the binomial PMF for every  $x \in X$ .

Table 1.2:  $P(X = x)$  for six tosses

| x | y     |
|---|-------|
| 0 | 0.016 |
| 1 | 0.094 |
| 2 | 0.234 |
| 3 | 0.312 |
| 4 | 0.234 |
| 5 | 0.094 |
| 6 | 0.016 |

Computing  $EX$  given the values in this table is demonstrated using the discrete case of Definition 1.4 in Example 1.1.

**Example 1.1.** The expected value of a discrete random variable  $X \sim B(6, 0.5)$ , which represents the number of tails in 6 coin tosses and has a binomial distribution, is calculated as follows:

$$\begin{aligned}
 E(X) &= \sum_x x f(x) \\
 &= (0 * 0.16) + (1 * 0.094) + (2 * 0.234) + (3 * 0.312) + (4 * 0.234) + (5 * 0.094) + (6 * 0.016) \\
 &= 3
 \end{aligned}$$

Naturally, the expected value of  $X$  is 3 in Example 1.1. Importantly, while in this case  $EX$  corresponds well to the peak in the PMF for this experiment this should not be assumed to be the case unilaterally, as the same expectation would result from any distribution symmetrical about  $x = 3$ . While surfically the expected value of a random variable may appear to be a trivial reframing of the calculating of a mean, it is crucial to note that expectation is calculated using the relative probability of  $x$  in lieu of the number of observations in

the sample space. In the continuous case, this is the property that makes computing the expectation possible. To summarize, in the field of probability the expected value provides a crucial, unifying term for the characterization of probability distributions. This term will prove useful in many situations, including normalization in the context of Bayes' theorem.

## 1.5 Conditional Probability

Conditional probability provides a uniform, comprehensible format for considering the probability of events which are dependent on the outcome of a different event. Conditional probability, as presented in Definition 1.5, requires only that the event on which some other event depends has some nonzero probability of occurring.

**Definition 1.5** (Conditional Probability). *Assuming  $P(B) > 0$ ,*

$$P(A|B) = \frac{P(AB)}{P(B)}$$

Conditional probability asserts that the probability of event A occurring given that event B occurs is equivalent to the probability of both A and B occurring divided by the probability that B occurs. Example 1.2 explains how to use conditional probability to calculate the probability of drawing an ace from a deck given I drew a king. While it may seem logical to conclude that  $P(A|B) = P(B|A)$ , this is not the case generally speaking. For example, the probability that I can swim given that I am in the water is clearly not the same as the probability that I am in the water given that I can swim.

**Example 1.2.** Conditional probability can be used to determine the probability of drawing an ace from a standard deck of cards without replacement, given I first drew a king from

that deck. This scenario can be represented as:

$$A = \text{Draws ace}, P(A) = 4/51$$

$$B = \text{Draws king}, P(B) = 4/52$$

$$\begin{aligned} P(ace|king) = P(A|B) &= \frac{P(AB)}{P(B)} \\ &= \frac{\frac{4}{51} \frac{4}{52}}{\frac{4}{52}} \\ &= \frac{4}{663} = 0.00603 \end{aligned}$$

## 1.6 Bayes' Theorem

Under construction... (Kotta et al., 2019)

## 1.7 Bayesian Inference

## Chapter 2

# Approximation algorithms for inference on complex systems





## Chapter 3

# Modern Bayesian inference with STAN



## Chapter 4

Bayesian applications in Biology:

Hybrid species distribution modelling  
with gaussian processes



# Bibliography

- Kotta, J., Vanhatalo, J., Jänes, H., Orav-Kotta, H., Rugiu, L., Jormalainen, V., Bobsien, I., Viitasalo, M., Virtanen, E., Sandman, A. N., Isaeus, M., Leidenberger, S., Jonsson, P. R., and Johannesson, K. (2019). Integrating experimental and distribution data to predict future species patterns. *Scientific Reports*, 9:1–14.
- Wasserman, L. (2004). *All of Statistics*. Springer.