

Week 1: Introduction to Bayesian Analysis

An Overview of Bayesian Thinking

welcome!

I'm excited to be here.

- I'm learning the material along with you.
- Drawing on materials and resources made by a master (Richard McElreath)

Goals of the class¹:

- Develop an intuition for statistical modeling using a Bayesian framework.
- Learn how to execute Bayesian statistics using R.

My commitments:

- Be here², be excited, be patient, be flexible.

COURSE FORMAT: Flipped class style

- Before class (2x/week)
 - watch video lectures by McElreath
- During class (2x/week)
 - comprehension quiz (5% of grade)
 - review major concepts
 - practice code
- After class (1x/week)
 - weekly homework assignment (95% of grade) (due Mondays)
 - graded on completion, not accuracy
 - one-week grace period
 - solutions posted on Tuesday – up to you to score your assignment and review missed questions

Workspace setup:

```
1 library(tidyverse)
2 library(cowplot)
```

probability

For today, we'll build a foundation by reviewing probability (covered in PSY 611 – remember?) and connecting these ideas to Bayesian frameworks for calculating and thinking about probability.

I'll draw on an article, [Introduction to Bayesian Inference](#) by Alex Etz and Joachim Vandekerckove (2018, *Psychon Bull Review*).

Interpretation

- **Epistemic:** probability is the *degree of belief*.
 - A number between 0 and 1 that quantifies how strongly we think something is true based on relevant information.
 - There is no such thing as *the* probability. There is only *your* probability.
 - BUT probability is not *arbitrary*.
- **Aleatory:** probability is a statement of the expected frequency over many repetitions of a procedure.
 - Cannot speak to singular events.
 - Assumes independence among repetitions.
 - Can be a valid conceptual interpretation but is rarely ever an operational one.

In the vast majority of cases, psychologists are trying to make statements about singular events:
- *this* theory is true or not. - *this* effect is positive or negative. - *this* model or *that* model is more likely.

Notation

- $P(A)$ is the probability of event A.
- $P(A, B)$ is the probability that both A and B happen.
- $P(B|A)$ is the probability of event B given that A is true.

Product Rule: Basics

- Formula:

$$P(A, B) = P(A)P(B|A) = P(B)P(A|B)$$

- Meaning:

- The probability of **(A)** and **(B)** occurring together – **P(A, B)** – can be calculated using:
 - **P(A)**: Probability of **(A)**.
 - **P(B|A)**: Probability of **(B)** given **(A)**, or vice versa.

Product Rule: Example

- Scenario:
 - Toss a coin twice.
 - (A): First toss is heads.
 - (B): Second toss is heads.
- Given:
 - $P(A) = 0.5$ (fair coin).
 - $P(B|A) = 0.5$ (independent tosses).
- Using the product rule:

$$P(A, B) = P(A)P(B|A) = 0.5 \times 0.5 = 0.25$$

Product Rule: Intuition

- The joint probability – $P(A, B)$ – reflects:
 - The likelihood of one event.
 - Adjusted by how the second event depends on the first.
- For Independent Events:

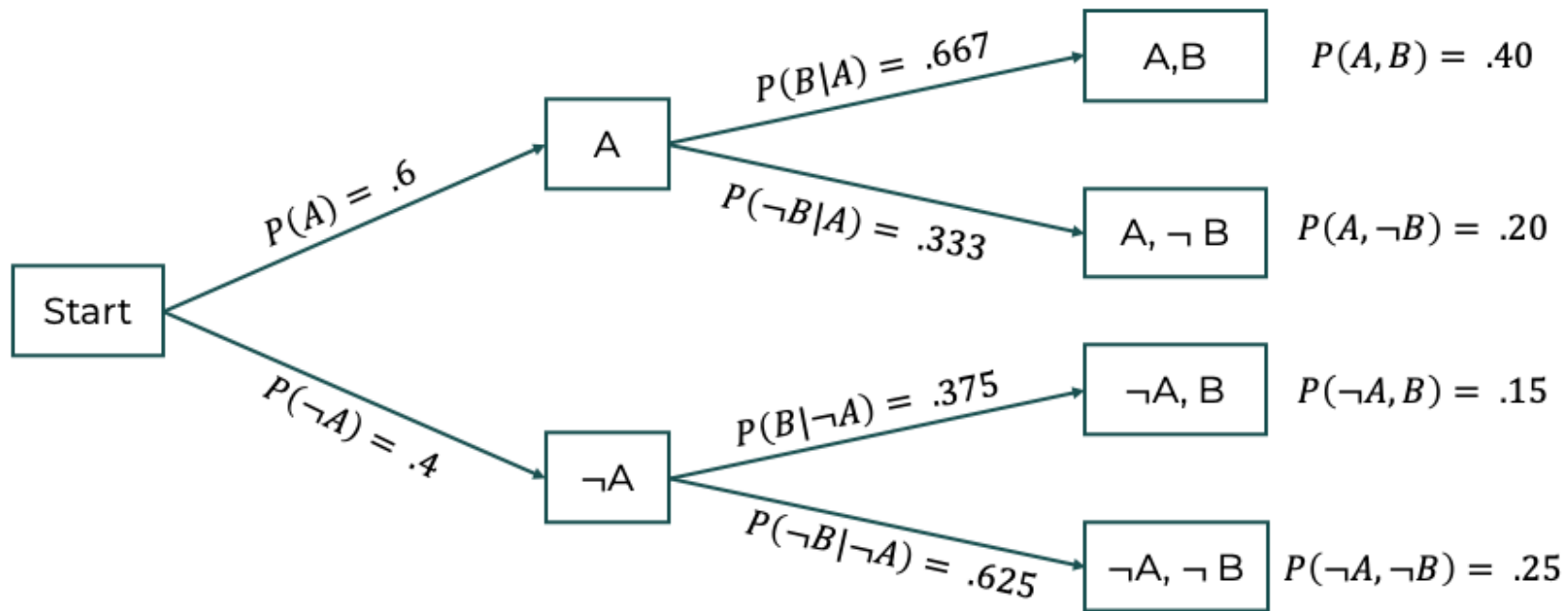
$$P(A, B) = P(A)P(B)$$

- For Dependent Events:

$$P(A, B) = P(A)P(B|A)$$

Let's say A is the event that it rains today and B is the event that it rains tomorrow. There's a 60% chance it will rain today. If it does rain today, it'll probably rain tomorrow (let's say 2/3 chance). But if it doesn't rain today, it probably won't rain tomorrow ($p = .625$).

- $P(A) = .6$
- $P(B|A) = .667$
- $P(\neg B|\neg A) = .625$



The probability of the joint events are found by multiplying the values along a path.

Sum Rule: Basics

- Formula:

$$P(A) = P(A, B) + P(A, \neg B)$$

- Meaning:

- The probability of (**A**) happening is the sum of:
 - $P(A, B)$: Probability of (**A**) and (**B**) both happening.
 - $P(A, \neg B)$: Probability of (**A**) happening without (**B**).

- Disjoint set:

- A collection of mutually exclusive events.

Sum Rule: Example

- Scenario:

- Drawing a card from a deck.
- (A): The card is red.
- (B): The card is a heart.
- ($\neg B$): The card is a diamond.

- Using the Sum Rule:

$$P(A) = P(A, B) + P(A, \neg B)$$

- Given:

- $P(A, B) = \frac{13}{52}$ (hearts).
- $P(A, \neg B) = \frac{13}{52}$ (diamonds).

- Result:

$$P(A) = \frac{13}{52} + \frac{13}{52} = 0.5$$

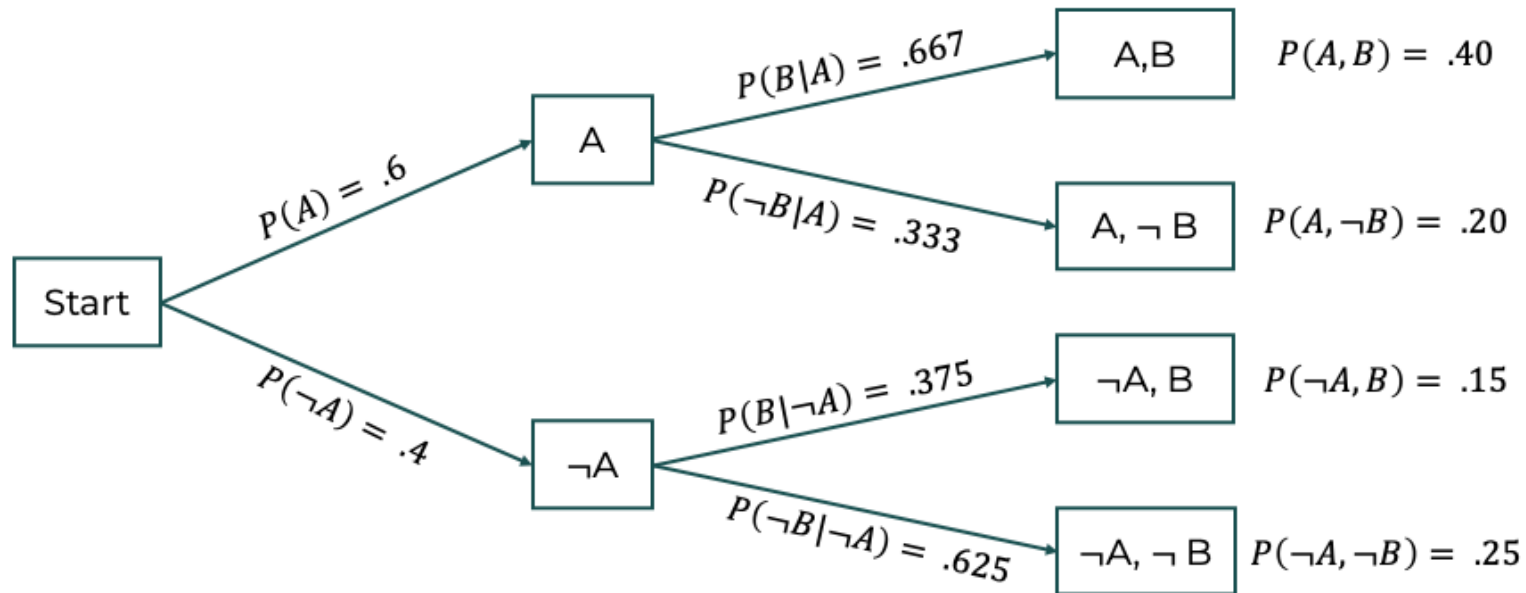
Sum Rule: Intuition

- The sum rule finds **total probability** by accounting for all disjoint ways (**A**) can occur.
- **General Formula:**

$$P(A) = \sum_i P(A, B_i)$$

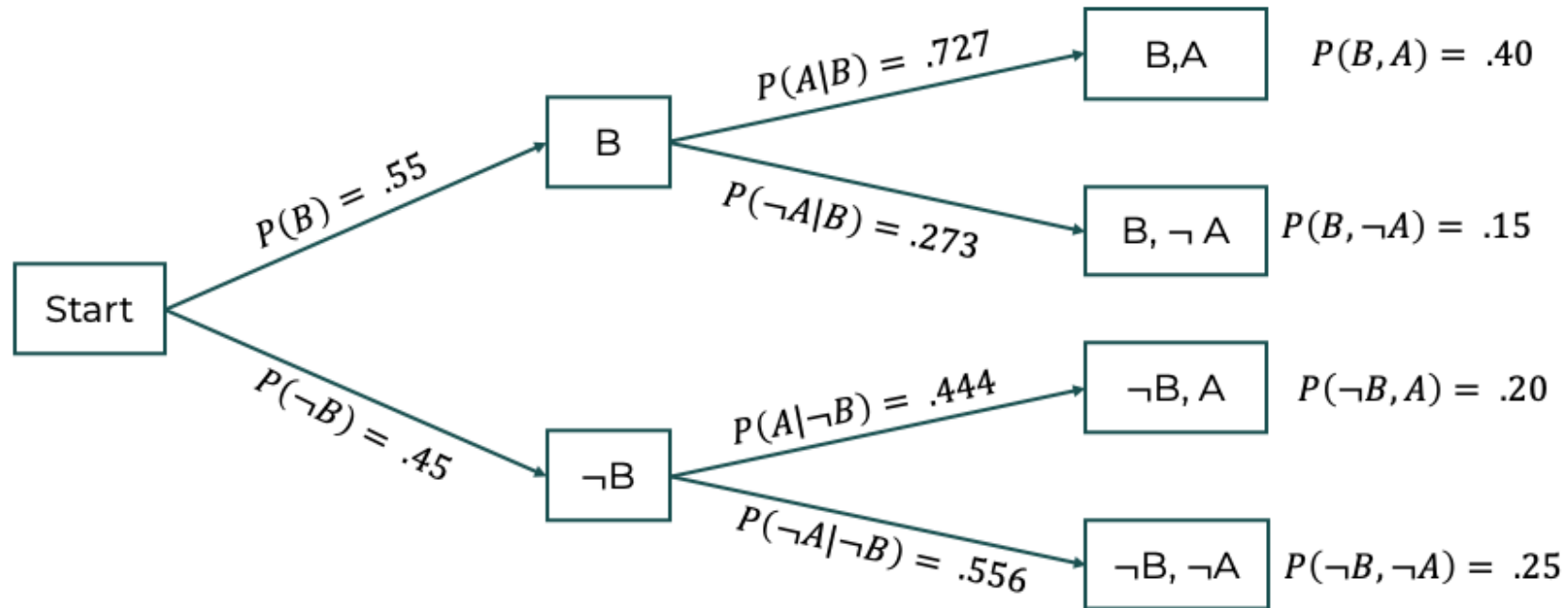
- Ensures that no possibilities are overlooked.

Exercise



Construct the equivalent path diagram starting on the left with a fork that depends on event B, instead of event A.

Exercise: Solution



Bayesian inference

Bayesian inference is the application of the product and sum rules to real problems of inference.

- Consider \mathbf{H} to be a hypothesis and $\neg\mathbf{H}$ to be the competing hypothesis.
- Before any data are collected, the researcher has some belief in these hypotheses. These are priors or **prior probabilities**, $P(\mathbf{H})$ and $P(\neg\mathbf{H})$.
- These hypotheses are well-defined if they make a specific prediction about each experimental outcome (D) through a *likelihood function* like $P(\mathbf{D}|\mathbf{H})$ and $P(\mathbf{D}|\neg\mathbf{H})$.
 - Likelihoods are how strongly data (D) are implied by a hypothesis.
 - Think NHST: $P(\mathbf{D}|\mathbf{H}_0)$.

Bayes' Rule

If

$$P(H, D) = P(D)P(H|D) = P(H)P(D|H)$$

then:

$$P(H|D) = \frac{P(H)P(D|H)}{P(D)}$$

This is Bayes' Rule!

- prior probability: $P(H)$
- likelihood function: $P(D|H)$
- posterior probabilities: $P(H|D)$

Prior predictive probability $P(D)$

How do we calculate this probability? We use the Sum Rule.

$$\begin{aligned} P(D) &= P(D, H) + P(D, \neg H) \\ &= P(H)P(D|H) + P(\neg H)P(D|\neg H) \end{aligned}$$

Now, we can rewrite Bayes' Rule using only prior probabilities and likelihoods.

$$P(H|D) = \frac{P(H)P(D|H)}{P(H)P(D|H) + P(\neg H)P(D|\neg H)}$$

And we can express our posterior in any case with K competing and mutually-exclusive hypotheses.

$$P(H|D) = \frac{P(H)P(D|H)}{\sum_{k=1}^K P(H_k)P(D|H_k)}$$

Bayes' Rule Terminology

- **prior probabilities:** $P(H)$ and $P(\neg H)$
 - Probability prior to seeing data.
- **likelihood functions:** $P(D|H)$ and $P(D|\neg H)$
 - function showing how likely an outcome/data are given a specific hypothesis.
- **posterior probabilities:** $P(H|D)$ and $P(\neg H|D)$
 - probability of a hypothesis given data. A combination of prior probabilities and likelihood functions.

Quantifying evidence using Bayes' rule

We form a ratio of relative belief in one hypothesis vis-a-vis another by comparing their posterior odds:

$$\frac{P(H|D)}{P(\neg H|D)}$$

We can insert the equations for posterior odds and find that this reduces to:

$$\frac{P(H)}{P(\neg H)} \times \frac{P(D|H)}{P(D|\neg H)}$$

The first part is called the *prior odds* and the second is called *Bayes factor*.

Bayes Factor: the extent to which the data sway our relative belief from one hypothesis to the other.

Bayes factors are **not** the same as posterior probabilities.

- Bayes factors: a “learning factor” – tells us how much evidence the data have delivered.
- posterior probabilities: our total belief after taking into account the data and our prior beliefs.

Example: Professor Sprout

At Hogwarts, professor Sprout leads the Herbology Department. In the Department's greenhouses, she cultivates a magical plant that when consumed causes a witch or wizard to feel euphoric and relaxed. Professor Trelawney, the professor of Divination, is an avid user of this plant and frequently visits Professor Sprout's laboratory to sample the latest harvest.

However, it has turned out that one in a thousand codacle plants is afflicted with a mutation that changes its effects: Consuming mutated plants causes unpleasant side effects such as paranoia, anxiety, and spontaneous levitation.

In order to evaluate the quality of her crops, Professor Sprout has developed a mutation-detecting spell. The new spell has a 99% chance to accurately detect an existing mutation, but also has a 2% chance to falsely indicate that a healthy plant is a mutant. When Professor Sprout presents her results at a School colloquium, Trelawney asks two questions: What is the probability that a plant is a mutant, when your spell says that it is? And what is the probability the plant is a mutant, when your spell says that it is healthy?

- The Professor Sprout example illustrates **Bayesian reasoning**.
- We calculate probabilities of plant mutations based on:
 - Prior probability of mutation.
 - Diagnostic spell's sensitivity and specificity.
- Key questions:
 1. What is the probability the plant is a mutant, given a “mutant” diagnosis, $P(M|D)$?
 2. What is the probability the plant is a mutant, given a “not mutant” diagnosis, $P(M|\neg D)$?

Question 1: $P(M|D)$

- Bayes' Rule:

$$P(M|D) = \frac{P(M)P(D|M)}{P(M)P(D|M) + P(\neg M)P(D|\neg M)}$$

- Values:

- $P(M) = 0.001$, $P(D|M) = 0.99$
- $P(\neg M) = 0.999$, $P(D|\neg M) = 0.02$

- Plugging in the values:

$$P(M|D) = \frac{0.001 \times 0.99}{(0.001 \times 0.99) + (0.999 \times 0.02)}$$

- Despite the spell's high sensitivity and specificity, $P(M|D) \approx 4.7\%$.
- Why? **Mutations are extremely rare.**
 - Even with an accurate test, most positive results are false positives.

An intuitive understanding

Instead of reporting probabilities, let's summarize the problem another way:

1. In a greenhouse of 100,000 plants, 100 are mutants.
2. Of the 100 mutants, 99 will be detected.
3. Of the 99,900 healthy plants, 1,998 will be falsely detected by the spell.

$$\frac{99}{99 + 1998} \approx .047$$

Question 2:

$$P(M|\neg D)$$

- Bayes' Rule:

$$P(M|\neg D) = \frac{P(M)P(\neg D|M)}{P(M)P(\neg D|M) + P(\neg M)P(\neg D|\neg M)}$$

- Values:

- $P(\neg D|M) = 1 - P(D|M) = 0.01$
- $P(\neg D|\neg M) = 0.98$

- Plugging in the values:

$$P(M|\neg D) = \frac{0.001 \times 0.01}{(0.001 \times 0.01) + (0.999 \times 0.98)}$$

- If the spell indicates “not mutant,” the plant is almost certainly not a mutant:

$$P(M|\neg D) \approx 0.001\%$$

- Why? The specificity ($P(\neg D|\neg M) = 98\%$) ensures most negatives are true negatives.

Exercise

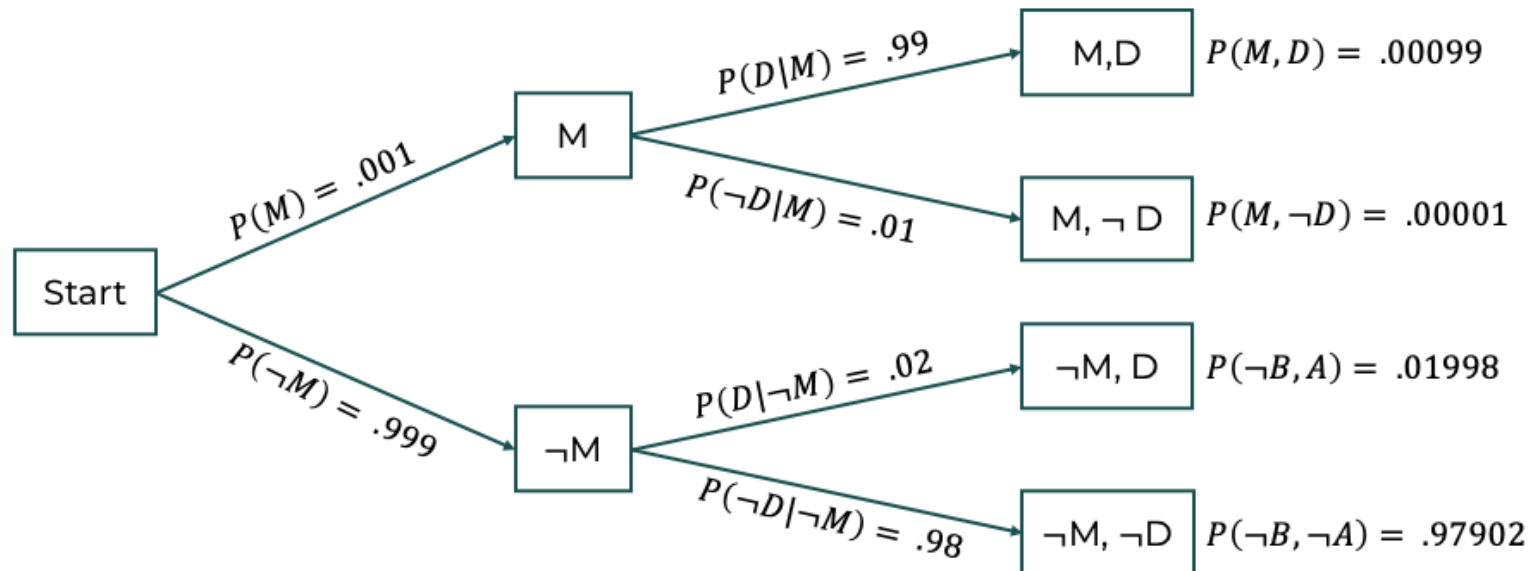
Diagram this example (similar to the rain/no rain diagram).

- $P(M) = 0.001, P(D|M) = 0.99$
- $P(\neg M) = 0.999, P(D|\neg M) = 0.02$

Exercise: Solution

Diagram this example (similar to the rain/no rain diagram).

- $P(M) = 0.001$, $P(D|M) = 0.99$
- $P(\neg M) = 0.999$, $P(D|\neg M) = 0.02$



Exercise

Suppose, however, that Trelawney knows that Professor Sprout's diagnosis (H_S) is statistically independent from the diagnosis of her talented research associate Neville Longbottom (D_L) — meaning that for any given state of nature M or $\neg M$, Longbottom's diagnosis does not depend on Sprout's. Further suppose that both Sprout and Longbottom return the mutant diagnosis (and for simplicity we also assume Longbottom's spells are equally as accurate as Sprout's). To find the posterior probability the plant is a mutant after two independent mutant diagnoses, $P(M|H_S, D_L)$, Trelawney can apply a fundamental principle in Bayesian inference: Yesterday's posterior is today's prior.

What is the probability the plant is mutant after two independent mutant diagnoses?

Exercise: Solution

Because diagnosis H_S and diagnosis D_L are independent, we know that:

$$P(D_L|M, H_S) = P(D_L|M)$$

and

$$P(D_L|\neg M, H_S) = P(D_L|\neg M)$$

therefore

$$\begin{aligned} P(M|H_S, D_L) &= \frac{P(M|H_S)P(D_L|M)}{P(M|H_S)P(D_L|M) + P(\neg M|H_S)P(D_L|\neg M)} \\ &= \frac{.047 \times .99}{.047 \times .99 + .953 \times .02} \approx .71 \end{aligned}$$

Key Takeaways

1. Posterior probabilities depend heavily on prior probabilities.

- Rare conditions remain unlikely even with positive test results.
- Accurate tests reduce uncertainty but don't guarantee certainty.

2. Bayes' Rule updates prior beliefs with evidence:

$$P(H|D) \propto P(H) \times P(D|H)$$

3. There is value in multiple independent sources of evidence:

- a plant only once diagnosed has a small chance of being a mutant.
- a plant that has twice been independently diagnosed as a mutant is quite likely to be one.

4. Practical applications:

- Diagnostic testing in medicine.
- Risk assessment in rare-event scenarios.

Example: Sorting Hat

- Hogwarts' Sorting Hat was damaged by a curse during the battle against You-Know-Who.
- It now assigns students to **Slytherin** 40% of the time, regardless of their true house.
 - 40% of students are assigned to Slytherin.
 - Only 20% each to Gryffindor, Ravenclaw, and Hufflepuff.
- Professor Binns develops a **Diagnostic Test**:
 - PARSEL test (Placement Accuracy Remedy for Students Erroneously Labeled):
 - Scores predict true house:
 - Excellent (E): Likely Slytherin.
 - Other scores (Outstanding, Acceptable, Poor) indicate other houses.

Data from the PARSEL Test

- **Benchmark results** from students sorted before the Battle of Hogwarts:
 - Probability of test scores by true house:

Score	Slytherin	Gryffindor	Ravenclaw	Hufflepuff
Excellent	0.80	0.05	0.05	0.00
Outstanding	0.10	0.20	0.80	0.10
Acceptable	0.05	0.70	0.15	0.25
Poor	0.05	0.05	0.00	0.65

Question: Is the Student a Gryffindor?

Professor McGonigall wants to know how likely it would be that a true Gryffindor will end up in Slytherin after taking this test.

- Goal:
 - Calculate $P(\text{Gryffindor} | H_S, S_E)$.

Step 1: Bayes' Rule for Gryffindor

$$P(\text{Gryffindor} | H_S, S_E) = \frac{P(\text{Gryffindor})P(H_S | \text{Gryffindor})P(S_E | \text{Gryffindor})}{P(H_S, S_E)}$$

- Known probabilities:
 - $P(\text{Gryffindor}) = 0.25$ (prior probability of Gryffindor).
 - $P(H_S | \text{Gryffindor}) = 0.20$ (damaged Hat assigns 20% of Gryffindors to Slytherin).
 - $P(S_E | \text{Gryffindor}) = 0.05$.

Step 2: Calculate Marginal Probability

What is the probability a student is sorted into Slytherin and also scores Excellent?

$$P(H_S, S_E) = \sum_i P(\text{House}_i)P(H_S | \text{House}_i)P(S_E | \text{House}_i)$$

- Contribution from all houses:

House	Prior P(House)	P(H _S given House)	P(S _E given House)	Joint P(H _S , S _E)
Slytherin	0.25	1.00	0.80	0.2000
Gryffindor	0.25	0.20	0.05	0.0025
Ravenclaw	0.25	0.20	0.05	0.0025
Hufflepuff	0.25	0.20	0.00	0.0000

Step 3: Calculate Posterior for Gryffindor

- Marginal probability:

$$P(\text{Slytherin, Excellent}) = 0.2000 + 0.0025 + 0.0025 + 0.0000 = 0.2050$$

- Posterior probability:

$$P(\text{Gryffindor} | (H_S, S_E)) = \frac{P(\text{Gryffindor})P(H_S | \text{Gryffindor})P(S_E | \text{Gryffindor})}{P(H_S, S_E)}$$

- Calculation:

$$P(\text{Gryffindor} | (H_S, S_E)) = \frac{0.25 \times 0.20 \times 0.05}{0.2050} = \frac{0.0025}{0.2050} \approx 0.0122$$

Interpretation

- The student has only about a **1.2% probability** of being a true Gryffindor.
- Most students sorted into Slytherin with an Excellent PARSEL score are true Slytherins:
 - $P(\text{Slytherin} | (H_S, S_E)) \approx 0.975$.

Bayes in the continuous case

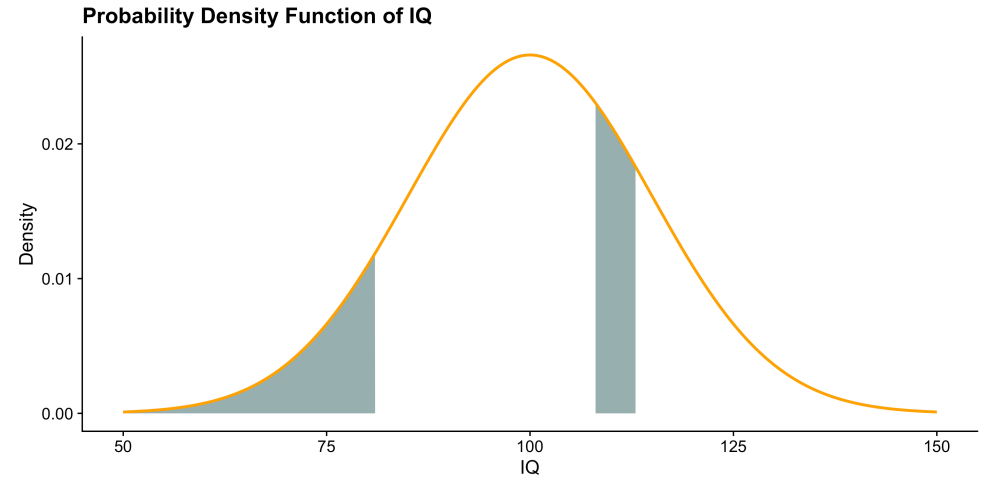
Most of our research questions are about parameters in the continuous case. For this, we make use of probability *density* functions. Densities:

- express how much a probability exists “near” a particular value of a , while the probability of any particular value is zero.
- probability is the integral of the density function over a certain interval:

$$P(a_1 < A < a_2) = \int_{a_1}^{a_2} p(a) da$$

- the total area under the density curve is 1.

$$P(- < A < a_2) = \int_{a_1}^{a_2} p(a) da$$



The product and sum rules have continuous analogues:

- Product rule:

$$p(\mathbf{a}, \mathbf{b}) = p(\mathbf{a})p(\mathbf{b}|\mathbf{a})$$

Where $p(\mathbf{a})$ is the density of the continuous parameter a , and $p(\mathbf{b}|\mathbf{a})$ is the conditional density of b (assuming a value of a).

- Sum rule:

$$p(\mathbf{a}) = \int_{\mathbf{B}} p(\mathbf{a}, \mathbf{b}) d\mathbf{b}$$

$d\mathbf{b}$ – the *differential* – represents an infinitesimally small interval of the variable \mathbf{b} . It is part of the integral that sums (or integrates) the joint probability $p(\mathbf{a}, \mathbf{b})$ over all possible values of \mathbf{b} within the range \mathbf{B} . - In other words, the $d\mathbf{b}$ indicates that we're integrating over values of \mathbf{b} (not values of \mathbf{a}).

Bayes' Rule in Continuous Case

- Derived from the product rule:

$$p(a|b) = \frac{p(a, b)}{p(b)} = \frac{p(a)p(b|a)}{p(b)}$$

- Numerator:
 - Combines prior $p(a)$ and likelihood $p(b|a)$.
- Denominator:
 - Marginal likelihood $p(b)$, ensuring the posterior is a proper probability density:

$$p(b) = \int_A p(a)p(b|a)da$$

Bayesian Parameter Estimation

- Posterior density:

$$p(\theta|\mathbf{x}) = \frac{p(\theta)p(\mathbf{x}|\theta)}{\int_{\Theta} p(\theta)p(\mathbf{x}|\theta) d\theta}$$

- θ : Parameter of interest.
- $p(\theta)$: Prior belief.
- $p(\mathbf{x}|\theta)$: Likelihood based on data \mathbf{x} .
- Numerator:
 - $p(\theta)p(\mathbf{x}|\theta)$: Unnormalized posterior.
- Denominator:
 - Normalizing constant (ensures posterior sums to 1).

Example: Puking pastilles

- Context:

- George Weasley is developing a new gag product: Puking Pastilles.
- The pastilles cause multiple “expulsion events” (vomiting) over time.
- George wants to estimate the **expulsion rate** (λ) more precisely.

- Objective:

- Use Bayesian parameter estimation to determine the most plausible values of λ .
- Incorporate prior beliefs and observed data.

Model: Poisson Distribution

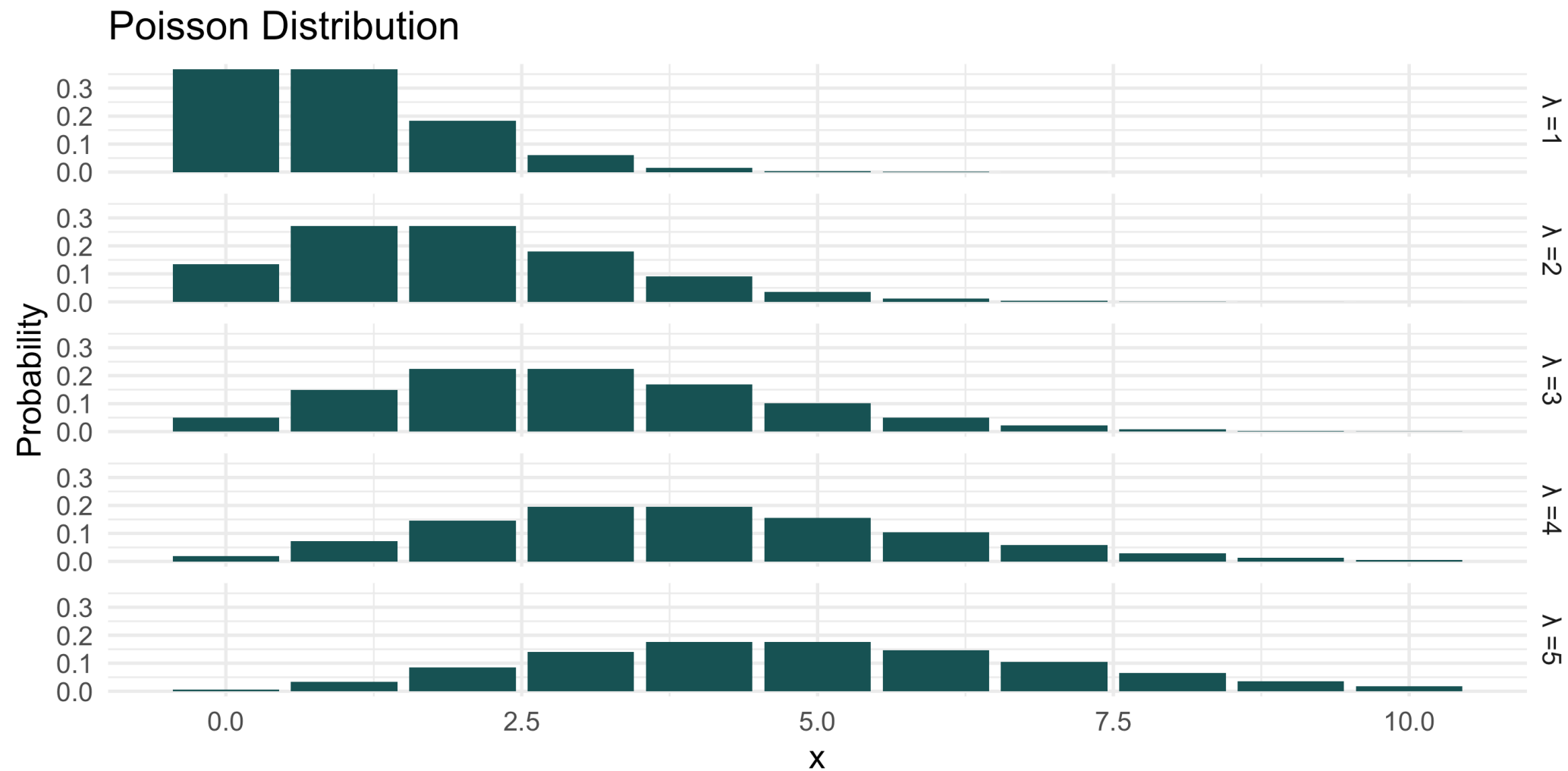
- Expulsion events are modeled with a **Poisson distribution**¹:

$$p(x|\lambda) = \frac{1}{x!} \lambda^x \exp(-\lambda)$$

- λ : The expected number of events per time interval. - x : Observed number of events in an interval.

- **Key property:**
 - The Poisson distribution is parameterized by λ , representing the **rate**.

Example Poisson distributions



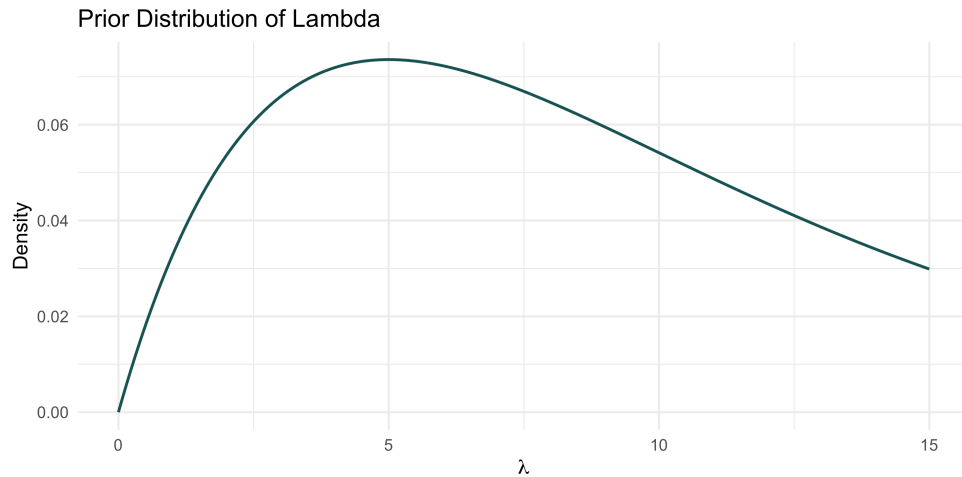
Prior Belief

- George's prior belief about λ :
 - Based on customer feedback, expulsion rates likely range between **3 and 5** events/hour.
 - Chosen prior distribution: **Gamma distribution**:

$$p(\lambda|a, b) = \frac{b^a}{\Gamma(a)} \exp(-b\lambda) \lambda^{a-1}$$

Prior distribution

- Parameters: $a = 2$, $b = 0.2$. - How were these chosen? Some knowledge about the gamma distribution and some trial and error.
 - The Gamma distribution ensures $\lambda > 0$ and is conjugate to the Poisson likelihood.



- priors are not point estimates.
 - “I think lambda is 5.” – this is *not* a prior.
- priors are distributions: the relative likelihood of every single possible value of the parameter.
 - lambda is most likely between 3 and 7, unlikely to be 1 or 2, possibly above 10...

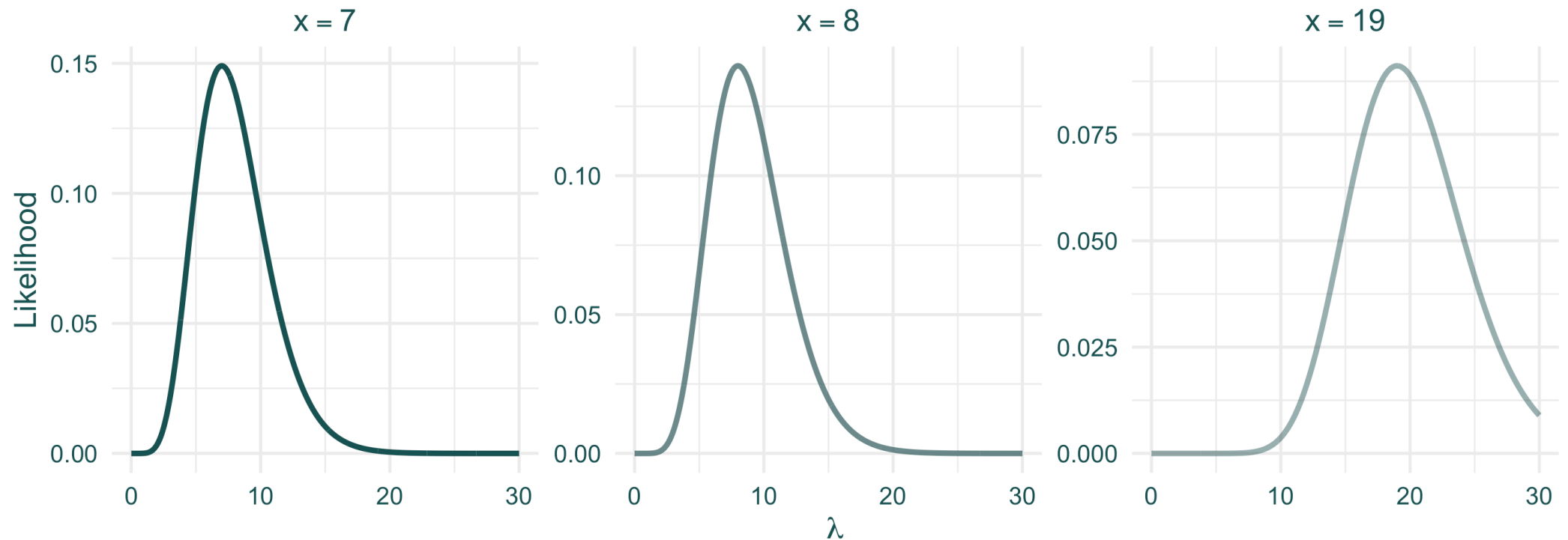
George collects data from three experiments:

- $x_1 = 7, x_2 = 8, x_3 = 19$.

Likelihood for the data:

$$p(\mathbf{X}_n | \lambda) = \prod_{i=1}^n \frac{1}{x_i!} \lambda^{x_i} \exp(-\lambda)$$

Poisson Likelihoods for Different Observed Counts



Posterior Distribution

Bayes' Rule:

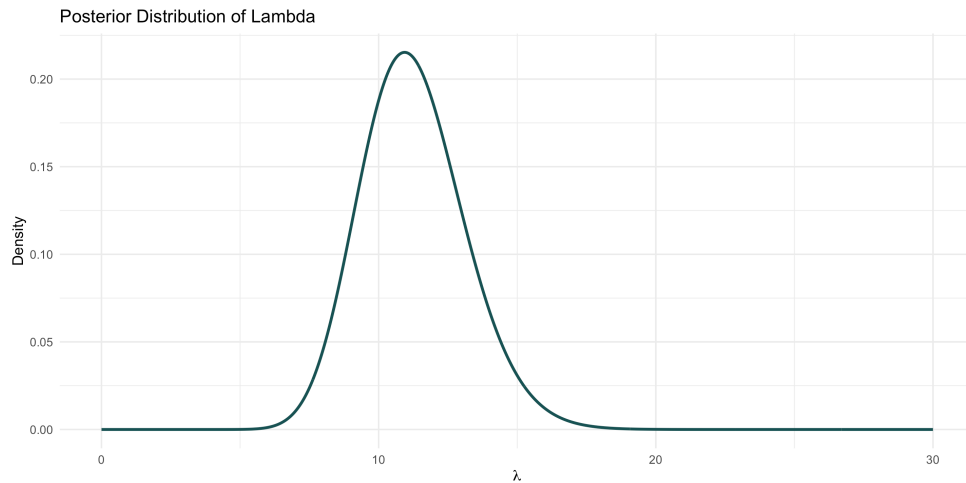
$$p(\lambda|X_n) \propto p(\lambda)p(X_n|\lambda)$$

Conjugacy simplifies the posterior to another Gamma distribution:

$$p(\lambda|X_n) = \text{Gamma} \left(a + \sum x_i, b + n \right)$$

- Parameters of the posterior:
- $a_{\text{post}} = a + \sum x_i = 2 + (7 + 8 + 19) = 36$
- $b_{\text{post}} = b + n = 0.2 + 3 = 3.2$

Results



Posterior Mean:

$$\mathbb{E}[\lambda | \mathbf{X}_n] = \frac{a_{\text{post}}}{b_{\text{post}}} = \frac{36}{3.2} \approx 11.25$$

Posterior Mode:

$$\lambda_{\text{mode}} = \frac{a_{\text{post}} - 1}{b_{\text{post}}} = \frac{36 - 1}{3.2} \approx 10.94$$

- 90% credible interval:
- Calculated from the Gamma distribution:
- Lower bound: **8.3**
- Upper bound: **14.5**

Formulas and equations

Bayesian analysis is the act of applying product and sum rules to probability.

Is that not your cup of tea?

Don't worry! Most of this term, we'll be abandoning the formulas altogether – there are easier and more intuitive ways to fit and use these models.

So why did we just go through all of this? - So you can appreciate the work done by others to make this more approachable. - Because you will need to get familiar with different probability distributions (Poisson, gamma, Cauchy, etc). These form the basis of your prior distributions, so knowing what they look like and how to set good priors using them is essential.