Bayesian Inference and Computation

Billy Junqi Tang

School of Mathematics

Welcome

- Hi everyone — I'm **Billy Junqi Tang**, the lecturer for **Bayesian Inference and Computation (4BIC)**.
- This is a **Master's-level statistics module**.
- It is worth **20 credits** and lectures run up to **Easter**.

Today (Lecture 1)

- Quick admin: where materials are, timetable, assessment
- Big picture: what Bayesian inference is and why we care

Module materials: Canvas

Everything you need is on **Canvas**:

- Lecture notes link
- Lab sheets + guidance
- Assessment schedule + feedback information
- Recommended reading list
- Reference page for common probability distributions

If you can't find something

Check Canvas first — then email me if it's genuinely missing.

Lecture notes

All the lecture notes are written in **Bookdown**.

Why Bookdown is useful

- More interactive than a static PDF
- You can change the font size
- Built-in search function
- Download options:
  - PDF

Copy-and-paste coding workflow

A nice feature of Bookdown notes is that you can **copy and paste code directly**.

Why this matters

- We will do **a lot of coding in R** throughout the module.
- You can copy code chunks straight into R / RStudio.
- This helps you focus on:
  - understanding what the code does
  - modifying it
  - running experiments

Computer lab: key information

You must bring a device

Please bring a laptop (or something similar) that can run **R**.

Time

Friday at 1pm.

Practical reality

You cannot really complete the module well without doing the coding.

Weekly structure

Each week has the same structure:

- **3 lectures**
- **1 computer lab**
- **1 guided study session**

Why this structure?

- Lectures: theory + intuition + modelling ideas
- Labs: turn theory into practice using R
- Guided study: consolidate, discuss, ask questions

Assessment: overview

This module is assessed **50/50**:

50% Exam + 50% Coursework

Exam

- Summer exam period
- Tests understanding + ability to reason with Bayesian ideas

Coursework

- Coursework is **coding-focused**
- You'll analyse data by implementing Bayesian ideas in R

Coursework schedule (Canvas)

On Canvas:

`Assessment & Feedback` → Coursework timetable

Weights

- Coursework 1: **10%**
- Coursework 2: **20%**
- Coursework 3: **20%**

Release weeks

Courseworks are released on the **Monday** of:

Week 3, Week 7, Week 10.

Coursework 1: coding quality matters

Important message

Coursework 1 is designed to encourage **good coding practice**.

- The statistics may be relatively straightforward.
- But there are **marks for writing clear code**:
  - coherent structure
  - readable formatting
  - sensible variable names
  - clear logic and outputs

Why?

In real work (academia or industry), messy code costs time, money, and errors.

Coursework guidance + feedback

Coursework briefings

On the Monday of Weeks 3, 7, and 10:

- I'll explain what the coursework is asking for
- What's expected
- How the mark scheme works

Feedback

- There are deadlines for when feedback must be returned.

Extensions and wellbeing

Important

If you need an extension due to wellbeing or other issues:

- Please apply via the **wellbeing team**.
- Do **not** apply to me directly (I cannot grant extensions).

General comment

This is the same policy across all modules.

Office hours and communication

- Tuesday 2pm-4pm.

Good habit

Ask questions early rather than waiting until deadline week.

Recommended books (optional)

You don't need to buy books — everything required is in the notes.

If you want extra support

1. **Bayesian Statistics** (short, dry, definition-focused)
2. **Bayesian Data Analysis** (long, comprehensive, opinionated)
3. **Statistical Rethinking** (intuitive, example-driven, very readable)

Videos: Statistical Rethinking

Extra resource

The author of *Statistical Rethinking* has a full lecture series on YouTube.

- Videos are very good for intuition.
- The author is an anthropologist, not a "traditional statistician".
- Examples can be... **distracting** (in a fun way).

Warning

You may end up in a Wikipedia rabbit hole about anthropology.

Reference distributions on Canvas

On Canvas you will also find a reference sheet for distributions you should know:

- Normal
- Beta
- Gamma
- Exponential
- Binomial

Why?

We use these repeatedly, so having them in one place is helpful.

Module roadmap

Why this module exists

Goal

Bayesian Inference and Computation gives you a **new set of tools** for statistical inference, beyond the **classical / frequentist** methods you have learned so far.

- We will learn how to **reason under uncertainty** using probability.
- We will focus on **probabilistic modelling** and **computation**.
- We will use **R and RStudio** throughout.

What is statistical inference?

Informal definition

Statistical inference means: using data to draw conclusions about quantities that are **unobserved**.

Type 1: Prediction

Unobserved events or events that haven't happened yet.

- clinical trials: future patient outcomes
- insurance: future payouts and premiums
- forecasting: demand, risk, trends

Type 2: Unobservable quantities

Quantities that cannot be directly observed.

- regression parameters
- latent variables
- hidden states in a model

Where you are starting from: classical inference

So far, most of your statistics has been **classical / frequentist**.

Typical frequentist workflow

1. Assume a data-generating model
2. Use long-run frequency arguments
3. Estimate parameters via **maximum likelihood** (MLE) and variants
4. Quantify uncertainty using **confidence intervals** and tests

In this module

We are **not** going to focus on classical inference workflows.

The Bayesian shift in one sentence

Key idea

Bayesian inference uses a definition of probability that supports **subjective uncertainty**, and uses **Bayes' theorem** to infer:

model parameters given observed data

Bayes' theorem as a modelling tool

We start with a model and a likelihood, just like classical inference.

Likelihood (data model)

$$\pi(y \mid \theta) \quad \text{(probability of data given parameters)}$$

Bayesian inference uses **Bayes' theorem**:

$$\pi(\theta \mid y) = \frac{\pi(y \mid \theta)\, \pi(\theta)}{\pi(y)}.$$

- $\pi(\theta)$: **prior** (belief / information before data)
- $\pi(\theta \mid y)$: **posterior** (belief after seeing data)
- $\pi(y)$: **evidence / marginal likelihood** (normalising constant)

Classical vs Bayesian: what changes?

Classical inference

- Main object:

$$\pi(y \mid \theta)$$

- Estimate θ via:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} \pi(y \mid \theta)$$

- Uncertainty via:
  - confidence intervals
  - tests / p-values

Bayesian inference

- Main object:

$$\pi(\theta \mid y)$$

- Output is a **distribution** over θ
- Uncertainty is direct:
  - posterior means/medians
  - credible intervals
  - posterior predictive checks

Why do we want a distribution?

A point estimate like $\hat{\theta}$ is useful, but it hides uncertainty.

Bayesian outputs are uncertainty-aware

From $\pi(\theta \mid y)$ we can compute:

- Most plausible values of θ (mean/median/mode)
- Uncertainty intervals (credible intervals)
- Probabilities of hypotheses:

$$\Pr(\theta > 0 \mid y), \quad \Pr(\theta \in A \mid y)$$

- Predictions for new data:

$$\pi(y_{\text{new}} \mid y) = \int \pi(y_{\text{new}} \mid \theta)\, \pi(\theta \mid y)\, d\theta$$

The module has two connected parts

Part 1: Theory (lectures)

- Bayesian probability
- Bayes' theorem for inference
- Modelling assumptions + interpretation
- Sampling theory and computation

Part 2: Practice (labs)

- Implement methods in **R / RStudio**
- Work with real data
- Compute posterior summaries + uncertainty
- Build good coding habits

Key message

Bayesian inference is powerful — but often **computational**.

Why computation matters in Bayesian inference

Bayes' theorem is simple to write, but hard to compute in practice.

The issue

The posterior can be:

- high-dimensional, non-Gaussian, multi-modal
- analytically intractable

So you usually cannot "solve" it by hand.

Solution

We compute with **sampling-based methods**:

$$\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(N)} \sim \pi(\theta \mid y)$$

and use samples to approximate posterior quantities.

R and RStudio

Practical skills

All labs and coursework will be in **R**, using **RStudio**.

- We start with **R workshops** early on.
- If you are new to R: we will build up slowly and support you.
- If you already know R: you will learn how to implement inference cleanly.

Module roadmap

What we will cover

1. Fundamentals of Bayesian inference
2. Bayesian modelling in practice
3. Subjectivity in modelling (priors + assumptions)
4. Sampling methods (core computation tools)
5. Markov Chain Monte Carlo (MCMC)
6. Advanced computation topics

Main questions

- What is **Bayesian probability**?
- How do we interpret probability as **uncertainty**?
- How does Bayes' theorem build **posterior beliefs**?
- What does it mean to "learn" from data?

Takeaway

Bayesian inference formalises learning by updating beliefs:

$$\text{prior} \longrightarrow \text{posterior}.$$

2) Bayesian inference in practice

Once we can write down $\pi(\theta \mid y)$, we can solve real data problems.

We will ask

- Given a particular dataset, what is a sensible model?
- What likelihood should we use?
- What does a posterior tell us about parameters?
- How do we quantify uncertainty and make predictions?

3) The subjective nature of Bayesian inference

A key feature of Bayesian inference is that **models involve choices**.

Subjective ingredients

- Prior distributions: $\pi(\theta)$
- Likelihood assumptions: $\pi(y \mid \theta)$
- Independence assumptions, noise models, constraints

Core questions

- Which assumptions are reasonable in a given context?
- Which assumptions are dangerous or misleading?
- How sensitive are results to these choices?

4) Sampling methods: getting started

We need to generate random samples from distributions.

Foundational sampling techniques

- Inverse transform sampling
- Rejection sampling

Why these matter

They teach core principles that later scale up to Bayesian computation:

- how to sample from complicated distributions
- how to validate sampling algorithms
- how to turn probability into computation

5) Markov Chain Monte Carlo (MCMC)

MCMC is the workhorse method for Bayesian computation.

Why MCMC?

- In realistic models, $\pi(\theta \mid y)$ is complex and high-dimensional.
- Direct sampling is impossible.
- MCMC constructs a **Markov chain** whose stationary distribution is:

$$\pi(\theta \mid y).$$

Checking MCMC works (diagnostics)

You should not blindly trust an algorithm just because it runs.

Practical questions

- Has the chain **converged** to stationarity?
- Is it **mixing** well (exploring the posterior efficiently)?
- Are there signs of **autocorrelation** or poor exploration?

Why it matters

We do not want algorithms that take **days** to generate useful samples.

Making sampling algorithms faster

Sampling performance depends on algorithm design + tuning.

What we can adjust

- proposal distributions (step sizes, scales)
- acceptance rates
- reparameterisations
- computational tricks (vectorisation in R)

Outcome

Efficient computation ⇒ practical Bayesian inference on real datasets.

6) Advanced topic: data augmentation (missing data)

Missing data is common in real-world problems.

Classical difficulty

Classical approaches often require detailed assumptions like:

- missing completely at random (MCAR)
- missing at random (MAR)
- missing not at random (MNAR)

Bayesian viewpoint

Treat missing values as additional unknowns and infer them:

$$\pi(\theta, y_{\text{miss}} \mid y_{\text{obs}}).$$

Why data augmentation is powerful

Examples

- Clinical trials: patients drop out $\Rightarrow$ outcomes unobserved
- Crime/forensics: events may go unrecorded or partially observed

Core idea

Instead of classifying missingness only, we can **learn plausible missing values** from the structure in the observed data.

7) Advanced topic: Bayesian nonparametric regression

Sometimes you do not want to assume a fixed parametric form.

Classical mindset

Choose a model like:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

(linear predictor with finite-dimensional parameters).

Nonparametric mindset

Assume:

$$y = f(x) + \varepsilon,$$

where f is **unknown**. Bayesian inference allows us to place a distribution over functions:

$$f \sim \Pi(\cdot).$$

What it means to "sample a function"

In Bayesian nonparametrics, uncertainty lives in a random function.

Key shift

Instead of sampling parameters $\theta \in \mathbb{R}^d$, we may sample:

$$f(\cdot) \text{ (a random function).}$$

- This yields flexible regression models.
- Uncertainty bands come naturally from posterior samples.
- Often easier to reason about than complex classical alternatives.

What you will be able to do by the end

Skills

- Build Bayesian models for real datasets
- Use Bayes' theorem to form posteriors
- Compute posterior summaries + uncertainty quantification
- Implement sampling algorithms in R
- Diagnose and improve MCMC performance
- Extend Bayesian inference to missing data and flexible regression

Module roadmap

Big picture: what is Bayesian inference?

Before we start the technical content, we need a big-picture overview:

- What is Bayesian inference?
- How is it different from what you've done before?
- Why do Bayesians think they're right? (yes, we are opinionated)

Quick check

How many of you have done Bayesian inference before?

Your previous experience: frequentist inference

So far in your maths/stats background, you've mainly seen:

Frequentist / classical inference

Inference built around the **likelihood function**:

1. Choose a model
2. Write down the likelihood
3. Collect data
4. Use likelihood to estimate parameters and do inference

Typical tools

- maximum likelihood estimation (MLE)
- hypothesis tests
- confidence intervals

Axioms: algebra vs probability

In first-year algebra/analysis you learn axioms like:

$$a + b = b + a.$$

You do not prove them — you **accept** them.

Probability is the same

Probability has axioms and definitions too, but people rarely stop and ask:

"What does probability actually mean?"

This week

We begin by discussing different definitions of probability.

Frequentist probability: long-run frequency

Definition (informal)

In frequentist probability:

$$P(A) = \text{long-run relative frequency of event } A.$$

Dice example

Roll a fair die many times:

$$\widehat{P}(\text{roll a 6}) = \frac{\#\{\text{6's}\}}{\#\{\text{rolls}\}} \longrightarrow \frac{1}{6}.$$

Relative frequency plot intuition

When you repeat the experiment many times:

- After the first roll:
  - if you get a 6, the frequency is 1 (100%)
- After two rolls:
  - if only one 6, the frequency is 1/2 (50%)
- As the number of rolls grows:

$$\text{relative frequency} \to \frac{1}{6}.$$

Why it feels natural

It is simple enough to understand in primary school.

The problem: repetition is baked in

The long-run definition of probability:

- makes the foundations intuitive,
- but forces "repeat the experiment" into everything.

Consequence

Later concepts become:

- technical
- unintuitive
- hard to explain in plain language

Confidence intervals: the unintuitive definition

Frequentist definition (idea)

A 95% confidence interval means:

> *If you repeated the experiment many times and built an interval each time, 95% of those intervals would contain the true parameter.*

What people want it to mean

$$\Pr(\theta \text{ is in this interval}) = 0.95.$$

But that is **not** the frequentist interpretation.

Why this matters (outside statistics)

Many users of statistics are not statisticians:

- medicine
- biology
- psychology
- social sciences

Common issue

People naturally interpret confidence intervals and p-values in the "probability of parameter" way. This mismatch leads to misunderstandings and bad decisions.

Likelihood: important, but not everything

Frequentist focus

Likelihood is treated as the centre of inference:

$$\pi(y \mid \theta).$$

- Likelihood contains lots of information.
- But it is not the *only* information available.

Notation note

You might have seen:

$$L(\theta; y), \quad f(y \mid \theta), \quad \pi(y \mid \theta).$$

In this module, I write **all distributions** as $\pi(\cdot)$.

The "wrong way round" feeling

The likelihood asks:

$$\pi(y \mid \theta) \quad \text{(data given parameters).}$$

But what do we actually observe?

We observe the data y. We do **not** observe θ.

What we want

$$\pi(\theta \mid y) \quad \text{(parameters given data).}$$

This is the object Bayesian inference is built around.

Bayes' theorem connects them

Bayes' theorem gives:

$$\pi(\theta \mid y) = \frac{\pi(y \mid \theta)\,\pi(\theta)}{\pi(y)}.$$

Interpretation

- $\pi(y \mid \theta)$: likelihood (model + data)
- $\pi(\theta)$: **prior** (before observing data)
- $\pi(\theta \mid y)$: **posterior** (after observing data)
- $\pi(y)$: marginal likelihood / evidence (normalisation)

Why priors upset frequentists

Frequentist viewpoint:

- θ is a fixed unknown constant (not random).
- randomness comes from the data-generating process.

So they say

"You can't put a probability distribution on a fixed parameter."

Bayesian viewpoint

A prior represents uncertainty about θ before seeing data. It is a modelling choice, like the likelihood.

Bayesian probability: the scary definition

Definition (informal)

Bayesian probability is the **subjective belief** that an event occurs.

Initial reaction

- "That sounds unscientific."
- "Can you just make probabilities up?"

Promise

Over the next 10 weeks, we will make this **sensible, scientific, and useful**.

Why Bayesian probability is useful

Some events do not have a meaningful "repeat forever" interpretation.

Example 1: individual event

Probability that **I personally** vote Labour in the next election.

- One person
- One election
- One vote

Frequentist workaround

Model a group of "similar people" and treat randomness as population variability. This addresses groups, but not a single individual directly.

Example 2: astronomy / space photography

You take a photo (e.g. distant part of the galaxy) and ask:

- probability a star has a planet
- probability of life, etc.

Why long-run frequency fails

- repeating the photo gives the same result (idealised)
- no meaningful "repeat experiment" variation

Bayesian benefit

You can express uncertainty and update beliefs using available information.

Example 3: forensics / crime evidence

Footprint evidence:

- often only one trace sample
- no repetition possible

Frequentist workaround

Again, you model a population of "similar events" and infer at that level.

But Bayesian probability lets you reason directly about uncertainty even when repetition is not realistic.

The plan: build everything from scratch

From this point onward, we start properly.

How the lectures will work

- We follow the lecture notes closely
- I will add extra explanation and intuition
- You can read the notes alone, but:
  - you'll miss the coding experience
  - you'll miss interactive explanations

Reset your brain

Forget maximum likelihood for a moment. We rebuild probability + inference carefully.

Definition: inference

Core definition

Inference is drawing conclusions from numerical data about quantities that are not observed.

Two types

1. Future / unobserved outcomes (prediction)
2. Parameters or quantities that cannot be directly observed

Inference type 1: predicting the future

Examples of **prediction** problems:

- **Clinical trials:** predict patient outcomes under treatments
- **Finance/insurance:** forecast future payouts and risks
- **Actuarial work:** set premiums from historical data

Key idea

You use data about the past to make probabilistic statements about the future.

Inference type 2: unobservable parameters

Examples of inference for quantities you cannot observe:

- Linear regression:

$$Y = mX + c + \varepsilon$$

- You can observe X and Y, but not m and c directly.
- They exist as **model parameters**, not “physical objects”.

Inference goal

Use data to learn plausible values of parameters and their uncertainty.

Frequentist probability (again)

Definition 1.1 (informal)

The probability of an event is the **long-run relative frequency** with which the event occurs.

What this forces

- repetition
- fixed sample size
- sampling distributions drive uncertainty statements

Sampling distribution of a statistic

Key idea

If you repeat an experiment many times using a fixed sample size, your statistic changes from run to run due to randomness and noise.

- This distribution explains variability in estimates.
- Sample size n heavily controls uncertainty.

Clinical trial intuition

Different groups of patients $\Rightarrow$ slightly different outcomes, even if the underlying effect is the same.

Likelihood and maximum likelihood estimation

Likelihood

$$\pi(y \mid \theta)$$

Frequentist approach:

- choose θ that makes observed data most likely

$$\hat{\theta}_{\mathrm{MLE}} = \arg\max_{\theta} \pi(y \mid \theta)$$

- then quantify uncertainty via sampling distributions

Many variants exist

method of moments, penalised likelihood, quasi-likelihood, ...

Confidence intervals vs credible intervals

Confidence interval (frequentist)

Coverage property under repetition:

$$\Pr\left(U(X) \leq \theta \leq V(X)\right) = 1 - \alpha.$$

Credible interval (Bayesian)

Direct probability statement about θ:

$$\Pr(\theta \in [a, b] \mid y) = 1 - \alpha.$$

Takeaway

Bayesian intervals usually match what people *want* uncertainty statements to mean.

What we really care about

Bayesian target object

$$\boxed{\pi(\theta \mid y)}$$

The distribution of model parameters given observed data.

- This object is central to everything in the module.
- It allows:
  - uncertainty quantification
  - prediction
  - decision-making under uncertainty

Bayesian inference: belief updating

Bayes' theorem

$$\pi(\theta \mid y) = \frac{\pi(y \mid \theta)\,\pi(\theta)}{\pi(y)}.$$

- Start with prior belief $\pi(\theta)$
- Observe data y
- Update to posterior belief $\pi(\theta \mid y)$

Slogan

Posterior $\propto$ **Likelihood** $\times$ **Prior**.

Bayesian probability: definition 1.5 (informal)

Definition

Probability = subjective belief about an event occurring.

- This is the major philosophical jump.
- It is also where Bayesian inference gets its flexibility.

Important

We will learn how to choose priors **sensibly** and how to justify modelling decisions scientifically.

End of lecture: focus shifts to labs

The last part of Lecture 1 is about the first lab.

Quick poll

Who has programmed in R before?

- Some will be experienced.
- Some will be brand new.
- The module is designed to support both.

Why R?

R is the main language for statistics

If you haven't used it before:

- it's very widely used in statistics
- it has great libraries and plotting tools
- it's ideal for implementing Bayesian computation

RStudio

I recommend using **RStudio** as it makes working in R easier.

Lab notes: Chapter 2

Where to find lab material

Go to the lecture notes and open **Chapter 2**.

- Chapter 2 covers the first two labs.
- It is about becoming comfortable in R.
- It is strongly focused on **sampling and simulation**.

Lab 1: sampling in R

Main theme

Generate samples from distributions in R.

- Download/install R
- (Recommended) install RStudio
- Copy/paste starter code from the notes
- Learn how to run simulations

Why sampling?

Sampling is the computational foundation of Bayesian inference.

Exercises: three difficulty levels

Each exercise comes with multiple support levels.

Options

- **Hard:** just the question
- **Medium:** some structure / guidance
- **Easy:** fill in the blanks (code mostly provided)

How to choose

- New to R $\Rightarrow$ start with Easy/Medium
- Comfortable with R $\Rightarrow$ start with Hard

A key skill: for loops

Why we practice loops

Many statistical ideas require repetition:

- simulation
- Monte Carlo approximations
- repeated sampling experiments

Core concept

Repeat something many times to understand randomness.

This is useful both in classical and Bayesian settings.

Advice for learning R

1) Google is very helpful

- Lots of good answers exist online (e.g. Stack Overflow).
- If you're stuck, search for the error message.

2) Use ChatGPT wisely

- Great for debugging code
- Great for explaining error messages
- Great for translating Python $\rightarrow$ R
- Not great if it writes everything for you (you won't learn!)

Lab preparation checklist

Before the first lab:

1. Install **R**
2. Install **RStudio** (recommended)
3. Open the notes at **Chapter 2**
4. Try running a simple script in R
5. Bring your laptop to **Friday 1pm**

If you hit problems

Bring them to the lab — we can fix them together.

Lecture 1 summary

Admin

- Canvas contains notes, timetable, assessments
- Notes are in Bookdown (copy/paste code!)
- Weekly schedule: 3 lectures + lab + guided study
- Assessment: 50% exam + 50% coding coursework

Big picture

- Frequentist inference: likelihood + long-run frequency interpretation
- Bayesian inference: focus on $\pi(\theta \mid y)$ via Bayes' theorem
- Bayesian probability: subjective belief (made rigorous over the module)

Next lecture

- We start building probability and inference from scratch.
- We will formalise Bayes' theorem and core definitions.
- Keep thinking about:

"What should probability mean?"

Questions?

Admin / assessment / labs / Bayesian vs frequentist?