Reporting Conclusions from Bayesian Inference (Ch. 3.2)

Posterior summaries, MAP, credible intervals, and conjugate priors

Bayesian Inference & Computation (4BIC)

January 29, 2026

Warm-up: recognise common distributions quickly

In Bayesian inference we constantly meet standard probability models. You should be able to recognise their **functional form** and **support**.

Typical task

Given a density / mass function (possibly *up to a constant*), identify:

- the distribution name;
- its parameters;
- its support (allowed values of x);
- whether it is a **density** (continuous) or **pmf** (discrete).

Mini reference table (you should know these)

Name	Notation	Form (up to constant)	Support

Mini reference table (you should know these)

Name	Notation	Form (up to constant)	Support
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$

Mini reference table (you should know these)

Name	**Notation**	**Form (up to constant)**	**Support**
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$
Beta	$X \sim \mathrm{Beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$x \in (0, 1)$

Mini reference table (you should know these)

Name	Notation	Form (up to constant)	Support
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$
Beta	$X \sim \mathrm{Beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$x \in (0, 1)$
Gamma	$X \sim \mathrm{Gamma}(\alpha, \beta)$	$x^{\alpha-1}e^{-\beta x}$	$x \in (0, \infty)$

Mini reference table (you should know these)

Name	**Notation**	**Form (up to constant)**	**Support**
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$
Beta	$X \sim \mathrm{Beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$x \in (0,1)$
Gamma	$X \sim \mathrm{Gamma}(\alpha, \beta)$	$x^{\alpha-1}e^{-\beta x}$	$x \in (0,\infty)$
Poisson	$X \sim \mathrm{Pois}(\lambda)$	$\lambda^x e^{-\lambda}/x!$	$x \in \{0,1,2,\dots\}$

Mini reference table (you should know these)

Name	Notation	Form (up to constant)	Support
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$
Beta	$X \sim \mathrm{Beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$x \in (0, 1)$
Gamma	$X \sim \mathrm{Gamma}(\alpha, \beta)$	$x^{\alpha-1}e^{-\beta x}$	$x \in (0, \infty)$
Poisson	$X \sim \mathrm{Pois}(\lambda)$	$\lambda^x e^{-\lambda}/x!$	$x \in \{0, 1, 2, \dots\}$
Binomial	$X \sim \mathrm{Bin}(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, \dots, n\}$

Mini reference table (you should know these)

Name	Notation	Form (up to constant)	Support
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$
Beta	$X \sim \mathrm{Beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$x \in (0, 1)$
Gamma	$X \sim \mathrm{Gamma}(\alpha, \beta)$	$x^{\alpha-1}e^{-\beta x}$	$x \in (0, \infty)$
Poisson	$X \sim \mathrm{Pois}(\lambda)$	$\lambda^x e^{-\lambda}/x!$	$x \in \{0, 1, 2, \dots\}$
Binomial	$X \sim \mathrm{Bin}(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, \dots, n\}$

Exam habit

Always state **support** clearly (e.g. $x \in (0, 1)$ for Beta). It prevents many mistakes.

Mini reference table (you should know these)

Name	Notation	Form (up to constant)	Support
Normal	$X \sim \mathcal{N}(\mu, \sigma^2)$	$\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$
Beta	$X \sim \mathrm{Beta}(\alpha, \beta)$	$x^{\alpha-1}(1-x)^{\beta-1}$	$x \in (0,1)$
Gamma	$X \sim \mathrm{Gamma}(\alpha, \beta)$	$x^{\alpha-1}e^{-\beta x}$	$x \in (0, \infty)$
Poisson	$X \sim \mathrm{Pois}(\lambda)$	$\lambda^x e^{-\lambda}/x!$	$x \in \{0, 1, 2, \dots\}$
Binomial	$X \sim \mathrm{Bin}(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, \dots, n\}$

Exam habit

Always state **support** clearly (e.g. $x \in (0,1)$ for Beta). It prevents many mistakes.

What do we get at the end of Bayesian inference?

The output of Bayesian inference is the **posterior distribution**:

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

- It represents our **uncertainty about θ given data** y.
- It contains **all information we can extract from the data** under the model.

Key Bayesian idea

Once we have $\pi(\theta \mid y)$, any posterior probability is just an integral:

$$\mathbb{P}(\theta \in A \mid y) = \int_A \pi(\theta \mid y)\, d\theta.$$

Posterior probabilities answer concrete questions

Examples of posterior questions:

- "What is $\mathbb{P}(0 < \theta < 1 \mid y)$?"
- "What is $\mathbb{P}(|\theta| \leq 2 \mid y)$?"
- "What is $\mathbb{P}(\theta \geq 0 \mid y)$?"

In principle

If you know $\pi(\theta \mid y)$ **exactly**, then all such questions are straightforward.

$$\mathbb{P}(\theta \in A \mid y) = \int_A \pi(\theta \mid y)\, d\theta.$$

But in practice: we often only know posteriors “up to a constant”

In many models,

$$\pi(\theta \mid y) \;\propto\; \pi(y \mid \theta)\,\pi(\theta).$$

- The symbol $\propto$ means “proportional to”.
- There is a normalising constant Z such that

$$\pi(\theta \mid y) = \frac{1}{Z}\,\pi(y \mid \theta)\pi(\theta), \quad Z = \int \pi(y \mid \theta)\pi(\theta)\,d\theta.$$

The difficulty

Computing Z exactly is often hard, because it requires a potentially complicated integral.

Does "up to a constant" make the posterior useless? No.

Even with unnormalised posterior we can still **do Bayesian inference**.

High-level idea (sampling)

There are methods (especially later in the course) that can produce samples

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

without knowing the normalising constant Z.

- Once we can sample from $\pi(\theta \mid y)$, we can approximate integrals numerically.

Approximating posterior probabilities by Monte Carlo

Suppose we want

$$\mathbb{P}(\theta \in A \mid y) = \int_A \pi(\theta \mid y)\, d\theta.$$

If we have samples $\theta^{(1)}, \ldots, \theta^{(M)} \sim \pi(\theta \mid y)$, then:

$$\mathbb{P}(\theta \in A \mid y) \;\approx\; \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\left(\theta^{(m)} \in A\right).$$

Interpretation

This is simply "**count how many samples fall into** A".

Key message

Posterior distributions are powerful *even when known only up to a constant*.

Why summarise a posterior?

A posterior distribution is a rich object, but it can be difficult to report directly.

Common goals:

- Provide a **single best guess** (point estimate).
- Provide a measure of **uncertainty** (interval estimate).
- Make results more **interpretable** to others.

Point estimate 1: the posterior mean

A natural Bayesian point estimate is the **posterior mean**:

$$\hat{\theta}_{\text{mean}} = \mathbb{E}(\theta \mid y) = \int \theta \, \pi(\theta \mid y) \, d\theta.$$

- This is a conditional expectation.
- It is a weighted average of θ, where the weights come from the posterior.

Practical note

To compute $\mathbb{E}(\theta \mid y)$ exactly, you need the **normalised** posterior. In complex models, we approximate it using posterior samples.

Point estimate 2: the posterior mode (MAP)

Another popular point estimate is the **posterior mode**:

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

This is also called the **maximum a posteriori (MAP)** estimate.

Key advantage

MAP does **not** require the normalising constant.

MAP intuition: scaling does not change the maximiser

If a function is multiplied by a positive constant, its maximiser does not change.

Example:

$$f(x) = -(x-1)^2 \quad \Rightarrow \quad \arg\max_x f(x) = 1.$$

Now scale it by 1000:

$$g(x) = 1000\, f(x) \quad \Rightarrow \quad \arg\max_x g(x) = 1.$$

Conclusion

MAP can be computed from an **unnormalised** posterior.

MAP and MLE: connection via a “flat” prior

Recall:

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

If we choose a **uniform / non-informative prior**:

$$\pi(\theta) \propto 1,$$

then

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

Takeaway

MLE is a special case of MAP under a flat prior.

Example: Beta posterior (binomial model)

Suppose after observing data, the posterior becomes

$$\theta \mid y \sim \mathrm{Beta}(\alpha, \beta).$$

Then:

$$\mathbb{E}(\theta \mid y) = \frac{\alpha}{\alpha + \beta}, \qquad \mathrm{mode}(\theta \mid y) = \frac{\alpha - 1}{\alpha + \beta - 2} \quad (\alpha, \beta > 1).$$

Interpretation

- Posterior mean gives an "average" estimate.
- Posterior mode gives the "most probable" value (MAP).

Concrete numbers from the earlier "bot probability" example

From the previous example in the notes:

$$\theta \mid y \sim \mathrm{Beta}(4, 198).$$

- Posterior mean:

$$\mathbb{E}(\theta \mid y) = \frac{4}{4 + 198} = \frac{4}{202}.$$

- Posterior mode (since $4, 198 > 1$):

$$\hat{\theta}_{\text{MAP}} = \frac{4 - 1}{4 + 198 - 2} = \frac{3}{200}.$$

Message

Point estimates summarise a posterior with a **single number**, but we still need uncertainty!

Beyond point estimates: quantify uncertainty

In frequentist statistics, uncertainty is often reported using **confidence intervals**.

In Bayesian inference, the analogous concept is the **credible interval**.

Goal

Find an interval $[L, U]$ such that

$$\mathbb{P}(L \leq \theta \leq U \mid y) = 1 - \alpha.$$

Definition: credible interval at level $1 - \alpha$

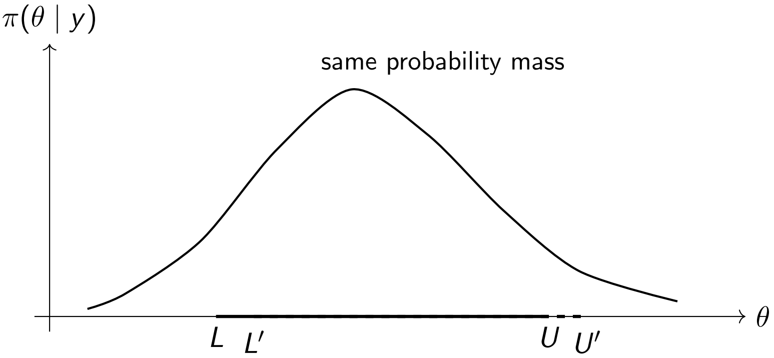
A credible interval is any interval $[L, U]$ satisfying:

$$\int_L^U \pi(\theta \mid y)\, d\theta = 1 - \alpha.$$

Important

This definition does **not** uniquely specify L and U. There can be many different intervals containing probability $1 - \alpha$.

Visual intuition: many intervals can contain 95% probability



Key point

There are infinitely many ways to pick $[L, U]$ so that the area under the posterior is $1 - \alpha$.

Most common choice: equal-tailed credible interval

A standard default is the **equal-tailed interval**, defined by:

$$\mathbb{P}(\theta \leq L \mid y) = \frac{\alpha}{2}, \qquad \mathbb{P}(\theta \geq U \mid y) = \frac{\alpha}{2}.$$

Equivalently:

$$L = F_{\theta|y}^{-1}\left(\frac{\alpha}{2}\right), \qquad U = F_{\theta|y}^{-1}\left(1 - \frac{\alpha}{2}\right),$$

where $F_{\theta|y}$ is the posterior CDF.

Interpretation

You "split" the remaining probability α equally into the two tails.

Why Bayesian credible intervals feel more intuitive

In Bayesian inference:

- θ is treated as a **random variable**.
- the interval $[L, U]$ is **fixed** once you observe the data y.

Interpretation

A $(1 - \alpha)$ credible interval means:

$$\mathbb{P}(\theta \in [L, U] \mid y) = 1 - \alpha.$$

In words:

"Given the data, there is a $100(1 - \alpha)\%$ *probability that* θ *lies in* $[L, U]$*."*

Contrast with confidence intervals (frequentist)

A frequentist $(1 - \alpha)$ confidence interval has a different meaning:

- θ is treated as a **fixed constant**.
- the interval endpoints (L, U) are **random** (they depend on random data).

Frequentist interpretation (requires repetition)

Over repeated sampling (repeating the experiment many times), the random interval contains the true fixed θ in $(1 - \alpha)$ fraction of repetitions.

Bayesian interpretation (no repetition needed)

Credible intervals directly describe probability of θ given *your observed data*.

Example: 95% credible interval for Beta posterior

Continuing the earlier example:

$$\theta \mid y \sim \mathrm{Beta}(4, 198).$$

A 95% equal-tailed credible interval is:

$$\left[F^{-1}(0.025)\,,\ F^{-1}(0.975)\right].$$

In R (built-in quantiles for Beta)

```
qbeta(c(0.025, 0.975), shape1=4, shape2=198)
```

This returns approximately:

$$[\,0.005,\ 0.040\,].$$

Interpretation

We believe there is a 95% chance that the "bot probability" lies between 0.005 and 0.040 (given the model and the observed data).

When can we find the posterior distribution exactly?

Sometimes, the posterior has a **recognisable named form**.

We have already seen examples where:

- Normal prior + Normal likelihood ⇒ Normal posterior.
- Beta prior + Binomial likelihood ⇒ Beta posterior.

Pattern

The **prior** and **posterior** belong to the same family of distributions.

Definition: conjugate prior

For a given likelihood $\pi(y \mid \theta)$, if the prior $\pi(\theta)$ and posterior $\pi(\theta \mid y)$ belong to the **same distribution family**, then:

$$\pi(\theta) \text{ is a conjugate prior for } \pi(y \mid \theta).$$

Why conjugacy is useful

- Posterior can often be written in **closed form**.
- Posterior summaries (mean, variance, credible intervals) can be computed easily.
- Great for quick analysis, intuition, and exam questions.

Example: exponential likelihood with gamma conjugate prior

Assume we observe n data points:

$$Y_1, \ldots, Y_n \mid \lambda \overset{\text{cond. i.i.d.}}{\sim} \mathrm{Exp}(\lambda), \quad \lambda > 0.$$

This means the conditional density factorises:

$$\pi(\boldsymbol{y} \mid \lambda) = \prod_{i=1}^{n} \pi(y_i \mid \lambda).$$

For the exponential:

$$\pi(y_i \mid \lambda) = \lambda e^{-\lambda y_i} \, \mathbf{1}_{(0,\infty)}(y_i).$$

Conjugate prior choice

$$\lambda \sim \mathrm{Gamma}(\alpha, \beta), \qquad \pi(\lambda) \propto \lambda^{\alpha-1} e^{-\beta\lambda} \, \mathbf{1}_{(0,\infty)}(\lambda).$$

Deriving the posterior: likelihood $\times$ prior

Posterior (up to a constant):

$$\pi(\lambda \mid \boldsymbol{y}) \propto \pi(\boldsymbol{y} \mid \lambda)\, \pi(\lambda).$$

First write the likelihood:

$$\pi(\boldsymbol{y} \mid \lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda y_i} = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} y_i\right).$$

Multiply by the gamma prior:

$$\pi(\lambda \mid \boldsymbol{y}) \propto \underbrace{\lambda^n e^{-\lambda \sum y_i}}_{\text{likelihood}} \cdot \underbrace{\lambda^{\alpha-1} e^{-\beta\lambda}}_{\text{prior}} = \lambda^{(\alpha+n)-1} \exp\Big(-(\beta + \sum y_i)\lambda\Big).$$

Recognise the form

This is exactly the kernel of a Gamma distribution.

Posterior result: Gamma again (conjugacy)

We conclude:

$$\lambda \mid \boldsymbol{y} \sim \mathrm{Gamma}\left(\alpha + n,\ \beta + \sum_{i=1}^{n} y_i\right).$$

Therefore

- Gamma prior + Exponential likelihood $\Rightarrow$ Gamma posterior.
- The Gamma prior is **conjugate** for the Exponential likelihood.

Practical meaning

Posterior uncertainty about the rate λ is updated by:

- adding n to the shape parameter (more data $\Rightarrow$ more concentration),
- adding $\sum y_i$ to the rate parameter (data-driven shift).

Summary: what you should remember

1) Posterior is the main Bayesian output

$$\pi(\theta \mid y) \quad \Rightarrow \quad \mathbb{P}(\theta \in A \mid y) = \int_A \pi(\theta \mid y)\, d\theta.$$

2) Even "up to a constant" is useful

We can sample from posteriors without normalising constants, then approximate probabilities by counting samples.

3) Common posterior summaries

- Posterior mean: $\mathbb{E}(\theta \mid y)$
- Posterior mode / MAP: $\arg\max_\theta \pi(\theta \mid y)$
- Credible interval: $\mathbb{P}(L \leq \theta \leq U \mid y) = 1 - \alpha$

Summary: conjugate priors (big exam topic)

Conjugacy

A prior is **conjugate** if the posterior stays in the same distribution family.

Examples:

- Normal–Normal conjugacy
- Beta–Binomial conjugacy
- Gamma–Exponential conjugacy

Next

We will continue this conjugate prior example tomorrow and practise similar derivations.

Questions?