Bayesian inference

CMED6040 – Session 9

Tim Tsang(matklab@hku.hk)

School of Public Health
The University of Hong Kong

18 July 2023

Session 9 learning objectives

After this session, students should be able to

- Use Bayes' theorem to synthesize prior information with observed data to form a posterior distribution
- Describe alternative specifications for prior distributions
- Compare and contrast maximum likelihood versus Bayesian estimation

Bayes' theorem

Three interpretations of probability

Classical	Equiprobable events
Frequentist	Repeat many times and observe long-run frequency
Subjective	Probability represents degrees of personal belief and uncertainty

Three interpretations of probability

When I say "The probability of the coin landing with heads up is $\frac{1}{2}$", what does that mean?

Classical	Heads and tails are equiprobable events, hence the probability of each event is $\frac{1}{2}$
Frequentist	If I toss the coin many, many times under equivalent circumstances, half the time it will come up heads
Subjective	I believe that the probability of heads is $\frac{1}{2}$ and I will behave accordingly

The likelihood function

- The likelihood function describes the probability of the data under different hypothesized parameter values

$$l(\theta|x) = p(x|\theta)$$

- How can we use the likelihood function to make inferences about plausible values of parameters?

- We have used the maximum of the likelihood function, but this does not take into account information we had before we observed x.

Prior and posterior distributions

- We may have some prior information or conviction about θ

- This can be represented using a probability distribution $p(\theta)$, called the **prior distribution** of θ

- When we observe new data x, they tell us something about θ via the likelihood function $l(\theta|x) = p(x|\theta)$

- We can combine them to create $p(\theta|x)$, the **posterior distribution** of θ

Bayes' theorem

- Bayes theorem:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{l(\theta|x)p(\theta)}{p(x)}$$

- We can define:
  - the ‘prior’ information about θ is $p(\theta)$
  - the likelihood is $l(\theta|x)$ or equivalently $p(x|\theta)$
  - the ‘posterior’ information is $p(\theta|x)$

Bayes' theorem

- Bayes' theorem:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{l(\theta|x)p(\theta)}{p(x)}$$

- As much of the time we only care about relative probabilities of different values of θ, it is sufficient to use the following simpler version of the theorem:

$$p(\theta|x) \propto l(\theta|x)p(\theta)$$

i.e. the posterior is proportional to the likelihood multiplied by the prior.

Subjective priors

- How can we set our prior distribution?
- **Elicitation** of the prior distribution involves converting expert knowledge about θ into probabilities
- This can be done with betting games, for example

- If θ is discrete, prior probabilities can be ascertained for each individual possible value of θ
- If θ is continuous, we could try to assign probability weights for various intervals

Subjective prior – example

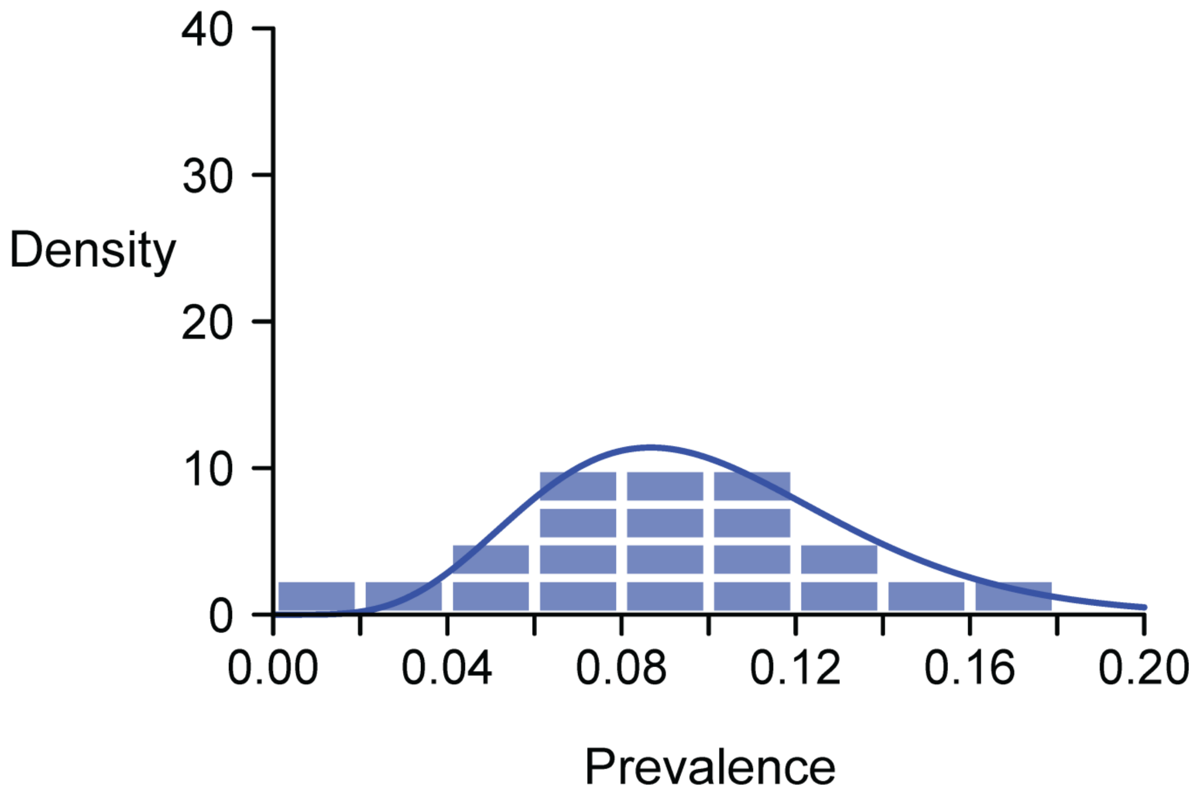

Figure: Subjective prior (based on expert knowledge) for prevalence of asthma in children in Hong Kong, and fitted beta(6.6, 60) distribution.

Using specific functional forms

- Prior knowledge about θ can be expressed using parametric distributions (e.g. Normal, beta, gamma etc).

- For example we could specify a Normal prior if we believe
  - θ is symmetrically distributed with respect to the mode;
  - Its density decays fast when far away from the mode;
  - Intervals far from the mode have negligible probabilities.

Recap – the binomial likelihood

- Let θ represent the probability of a dichotomous outcome variable, and X is the number of "positive" outcomes in *n* trials, i.e. $X \sim Bin(n, \theta)$.

- The likelihood function for $X = x$ is

$$l(\theta|x) \propto \theta^x (1-\theta)^{n-x}$$

The beta-binomial model

- Let the prior information about θ be represented by a beta(α, β) distribution, $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$, $0 \le \theta \le 1$.
- Then the posterior distribution for θ is:

$$
\begin{aligned}
p(\theta|x) &\propto l(\theta|x)p(\theta) \\
&\propto \theta^{x}(1-\theta)^{n-x}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&\propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}
\end{aligned}
$$

- Therefore $(\theta|x) \sim beta(\alpha+x,\ \beta+n-x)$

Prior and posterior – example

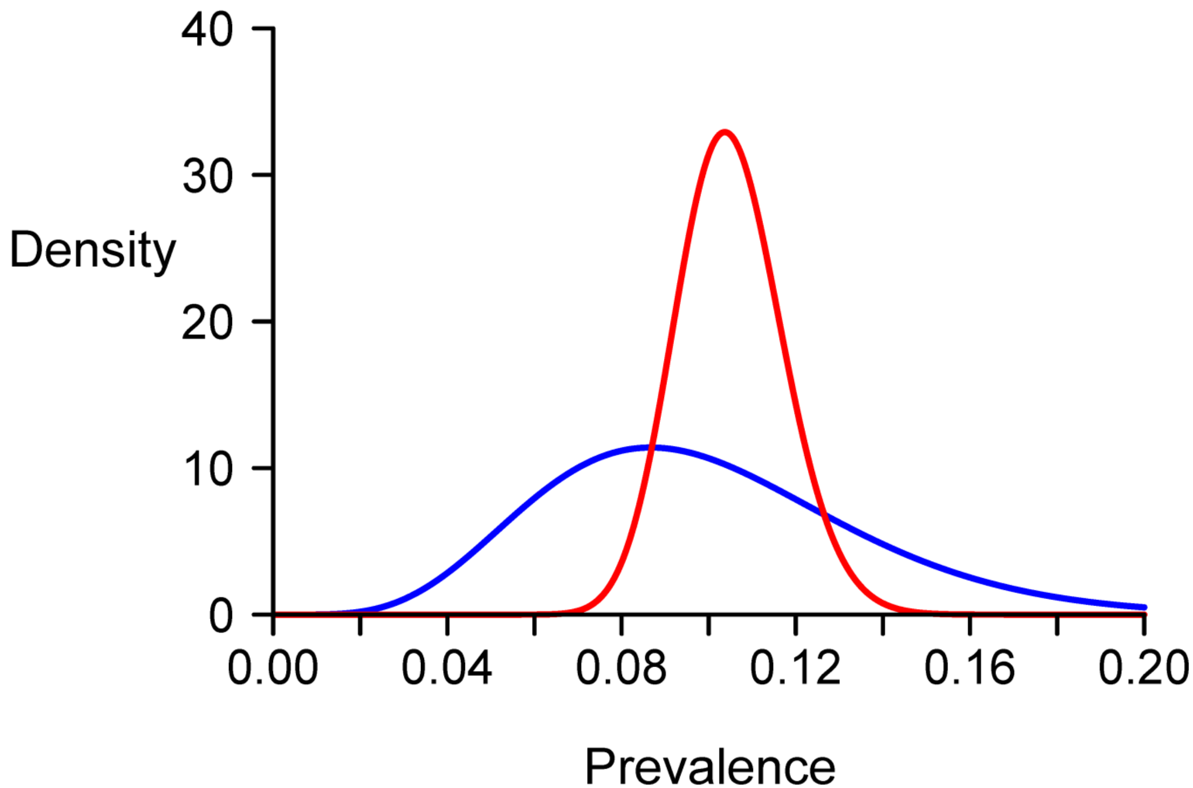

Figure: Subjective beta(6.6, 60) prior for prevalence of asthma, and posterior beta(66.6, 568) distribution based on observed 60/568 cases.

Prior and posterior – example

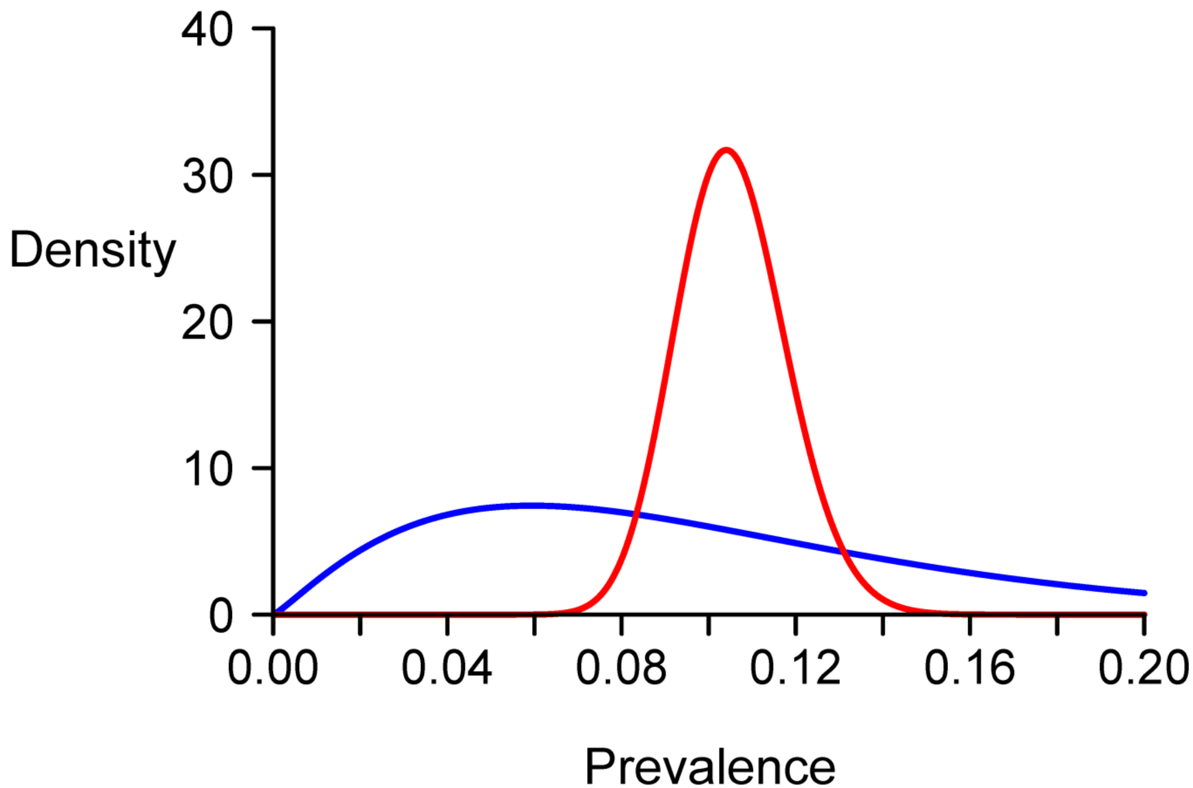

Figure: Subjective beta(3.3, 30) prior for prevalence of asthma, and posterior based on observed 60/568 cases.

Conjugate priors

- For a binomial likelihood, the beta prior is known as a conjugate prior.
- A conjugate prior for a likelihood will ensure the posterior has the same form.
- Other examples of conjugate priors:
  - Gamma prior for Poisson likelihood.
  - Normal prior for Normal likelihood.
  - Gamma prior for Exponential likelihood.
  - More examples: https://en.wikipedia.org/wiki/Conjugate_prior

Non-informative priors

- One criticism of the Bayesian paradigm is that the choice of prior distribution is arbitrary
- "Science should be objective, not subjective"
- One solution is to specify **non-informative priors** that minimise the prior information included in the analysis
- These could be used as reference priors to compare how much changing the priors changes the result

Uniform priors

- One obvious non-informative prior is the uniform distribution

$$p(\theta) = k, \text{ for all } \theta$$

  where k is chosen so that the total of the probabilities equals 1

- If there are n possible values of θ then $k = \frac{1}{n}$
- If θ is continuous between a and b (where $a < b$) then

$$p(\theta) = \frac{1}{b-a} \text{ for } a \leq \theta \leq b$$

- But if θ is unbounded this is not technically possible
- If $p(\theta) \propto 1$ for unbounded θ, then the prior is **improper**
  - Posterior can still be proper, even if the prior is not.

Uniform prior – example

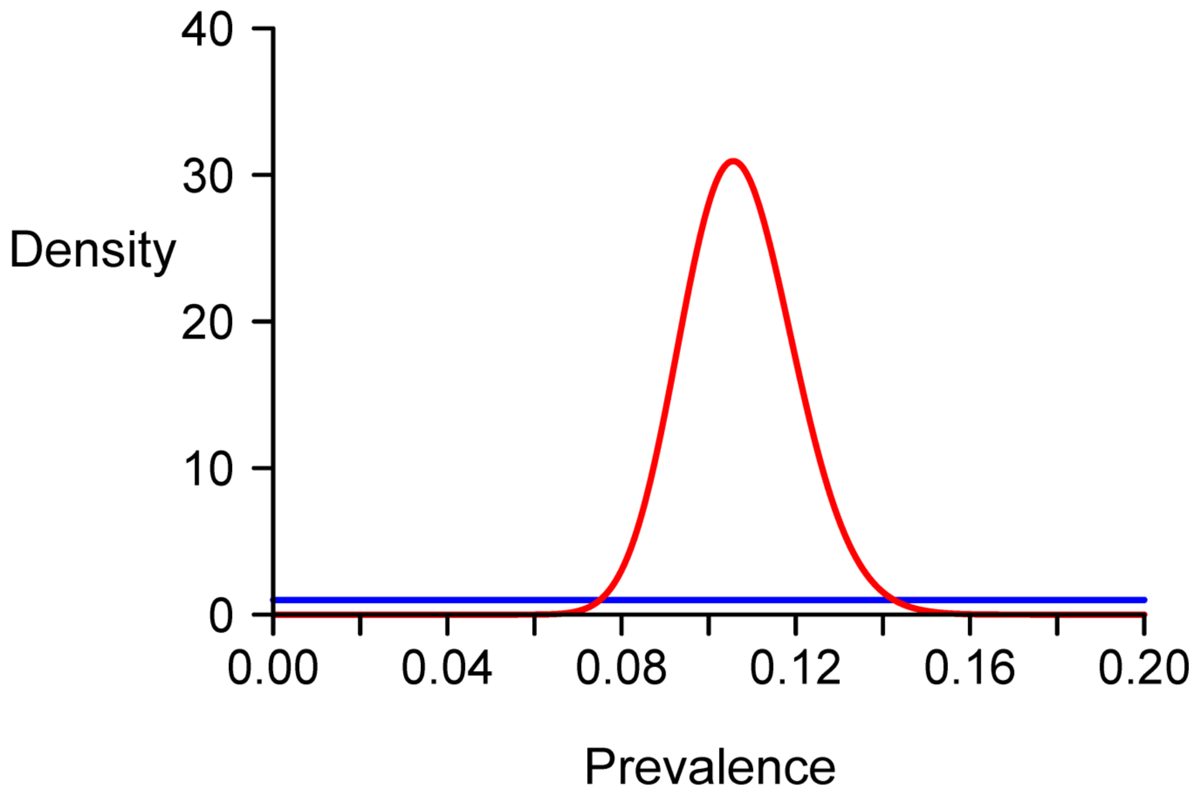

Figure: Flat uniform prior for prevalence of asthma between 0 and 1 (equivalent to beta(1,1) distribution), and beta(61, 509) posterior based on observed 60/568 cases.

Review of priors

- Priors describe our knowledge about parameter(s) θ before observing data.
- Alternative ways to code prior information:
  - Entirely subjective specification
  - Functional forms (particularly conjugate priors)
  - Non-informative (reference) priors

Bayesian inference

- There are two main differences between the Bayesian approach to statistical inference and the classical approach.
  - Subjective interpretation of probability.
  - Use of prior distributions.

- The subjective definition of probability has a close connection with the use of prior distributions in an analysis.

- In Bayesian approach, unknown parameters follow a distribution. In classical approach, unknown parameters are fixed values.

Bayesian analysis

Example – asthma prevalence

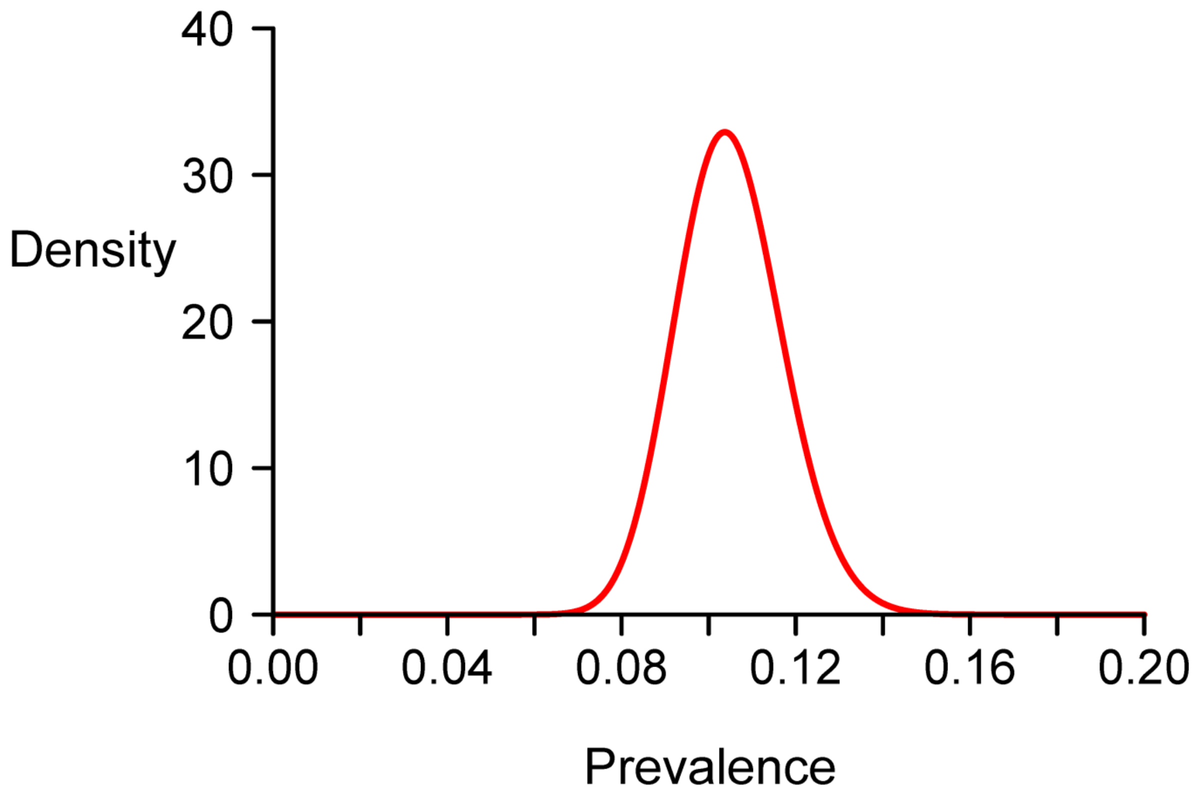

Figure: For the posterior distribution above, how could we summarise the information about the prevalence θ into a single number?

Summarizing a distribution

- Is it even wise to summarise θ with a single number? . . .
- . . . sometimes it may be useful to provide a ‘best guess’ of θ
- We could choose the mean, median or mode of the posterior distribution as our point estimate of the parameter.
  - Usually the mean.

Example – point estimation

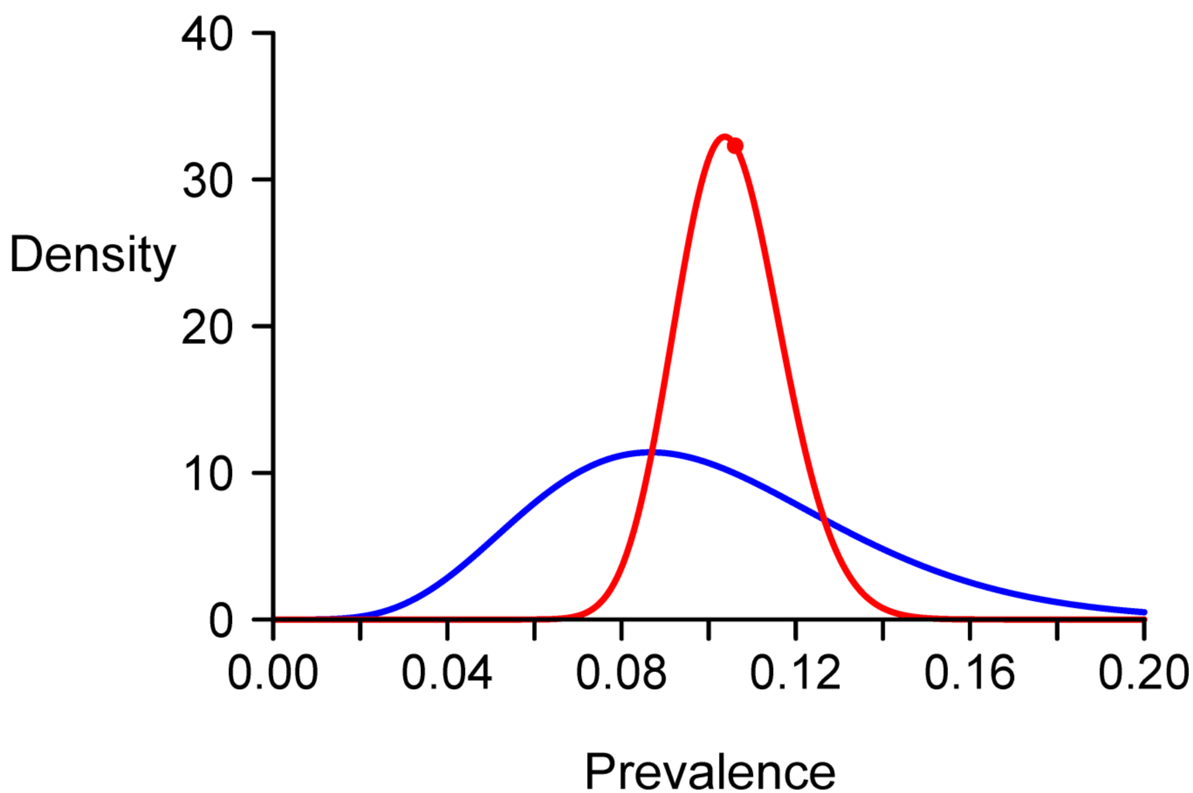

Figure: Mean of the beta(66.6, 568) posterior distribution is 66.6 / 634.6 = 0.105.

Two alternative posteriors

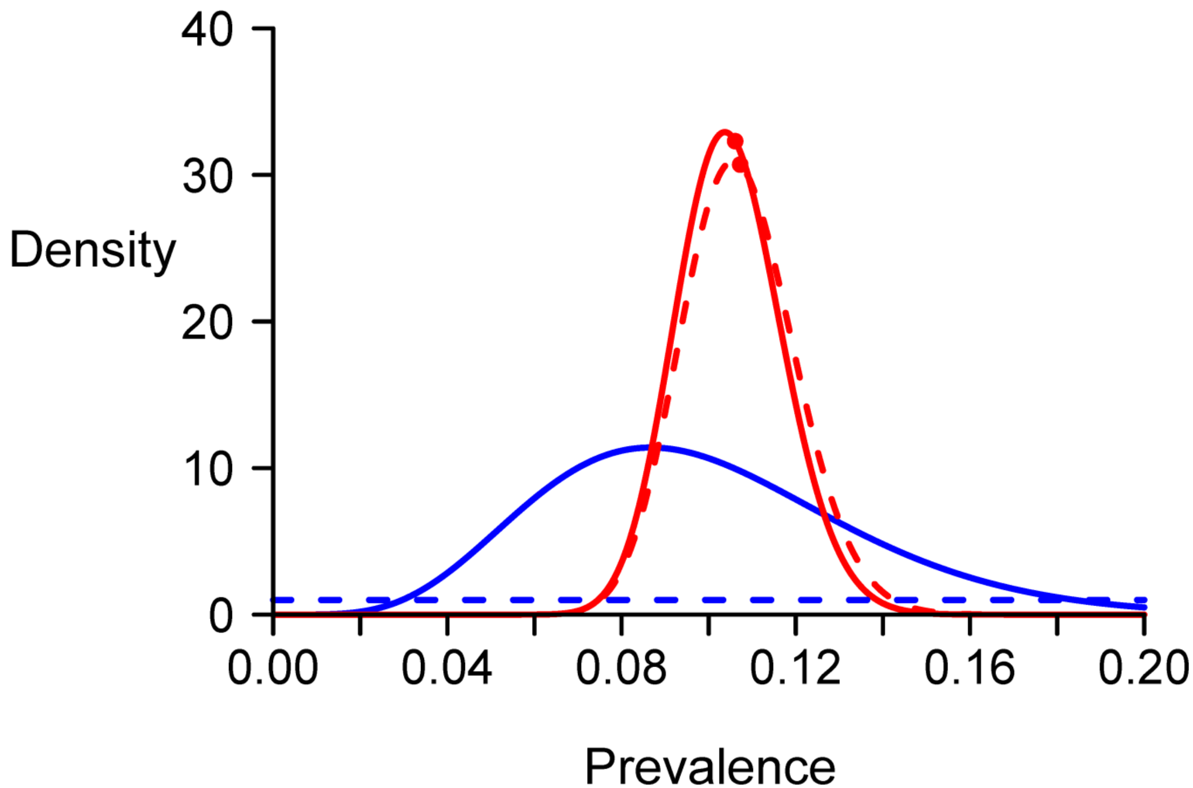

Figure: Posterior point estimates are 0.105 with subjective prior (solid line) and 0.107 with flat prior (dashed line).

Bayesian interval estimates

- Posterior distribution describes $p(\theta|x)$.
- This is the probability of θ for various possible values, after taking into account the observed data (and any prior information).
- We can derive a range of values of θ which have higher probability – a range of 'most probable' values.

A posterior distribution

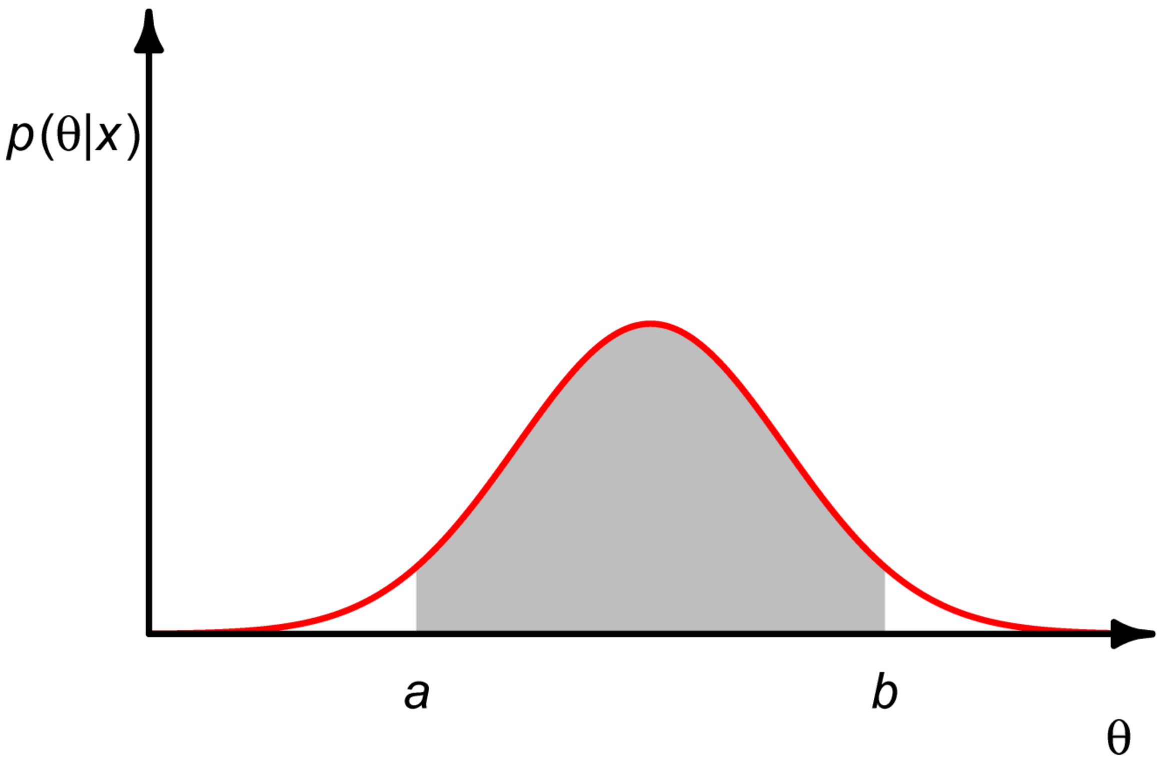

Figure: 'Most' of the most probable values of θ lie between *a* and *b*.

Another Bayesian interval estimate

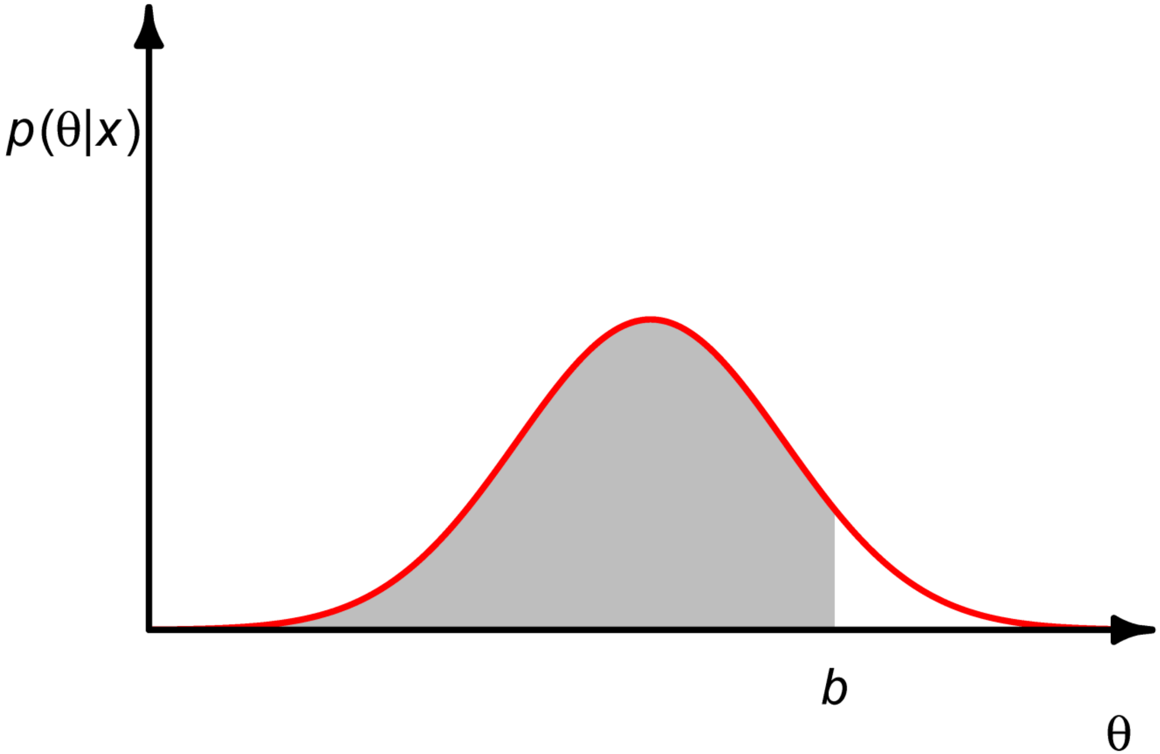

Figure: 'Most' of the most probable values of θ also lie below b.

Bayesian credible intervals

- An interval $[a, b]$ is a $100(1-\alpha)\%$ Bayesian confidence interval or **credible interval** for θ if $Pr(a \leq \theta \leq b|x) \geq 1-\alpha$.
- In this case, $1-\alpha$ is called the credible or confidence level.
- Usually we refer to these as credible intervals to distinguish them from frequentist confidence intervals.
- Unlike frequentist confidence intervals, these can be interpreted as containing the "true" value of θ with a particular probability, namely $1-\alpha$

Bayesian credible intervals

- The Bayesian equal tail CrI method gives threshold values of the posterior distribution that represent an interval with the probability of interest (e.g., 95%) of the distribution mass around the center of the distribution.
- The lower limit of the 95% equal tail CrI is the quantile representing a probability of 0.025 (or the 2.5% percentile) of the posterior distribution.
- The upper limit of the equal tail CrI is the quantile representing a probability 0.975 (or the 97.5% percentile) of the posterior distribution.

Bayesian credible intervals

- Definition can be extended to multivariate parameters (credible regions rather than intervals).
- Could be a set of intervals e.g. $[c_1, c_2] \cup [c_3, c_4]$.
- Credible intervals are invariant under 1-1 transformations e.g. log$(\cdot)$, exp$(\cdot)$
  - For example if a 95% credible interval for θ is $[a, b]$ then 95% credible interval for $\exp(\theta)$ is $[\exp(a)$, $\exp(b)]$.

Shortest intervals

- Usually we would like an interval that is as small as possible for a fixed credible level
  - The first rather than the second example on the previous slides.
- Can find alternative $100(1-\alpha)\%$ credible intervals and choose the shortest one.
- But note that the shortest $100(1-\alpha)\%$ credible interval for θ will ***not*** for example be the shortest $100(1-\alpha)\%$ credible interval for $\exp(\theta)$.

Bayesian interval – interpretation

- If a 95% credible interval for θ is $[a, b]$, we can interpret this as follows:
- "There is a 95% probability that θ is between a and b, given the observed data (and given the model and the prior information.)"
- Or "we are 95% sure that θ is between a and b, given the observed data (and given the model and the prior information.)"

Example – Asthma prevalence

In a recent survey, 60 of 568 sampled children in Hong Kong were found to have asthma. What is the population prevalence?

- With a non-informative beta(1, 1) prior, the posterior distribution is beta(61, 509).
- Point estimate (posterior mean) of the prevalence is $\frac{61}{61+509} = 10.7\%$.
- A 95% credible interval for the prevalence is (8.3%, 13.4%).
  - Lower limit a found using `qbeta(0.025, 61, 509)`.
  - Upper limit b found using `qbeta(0.975, 61, 509)`.

Example – Bayesian credible interval

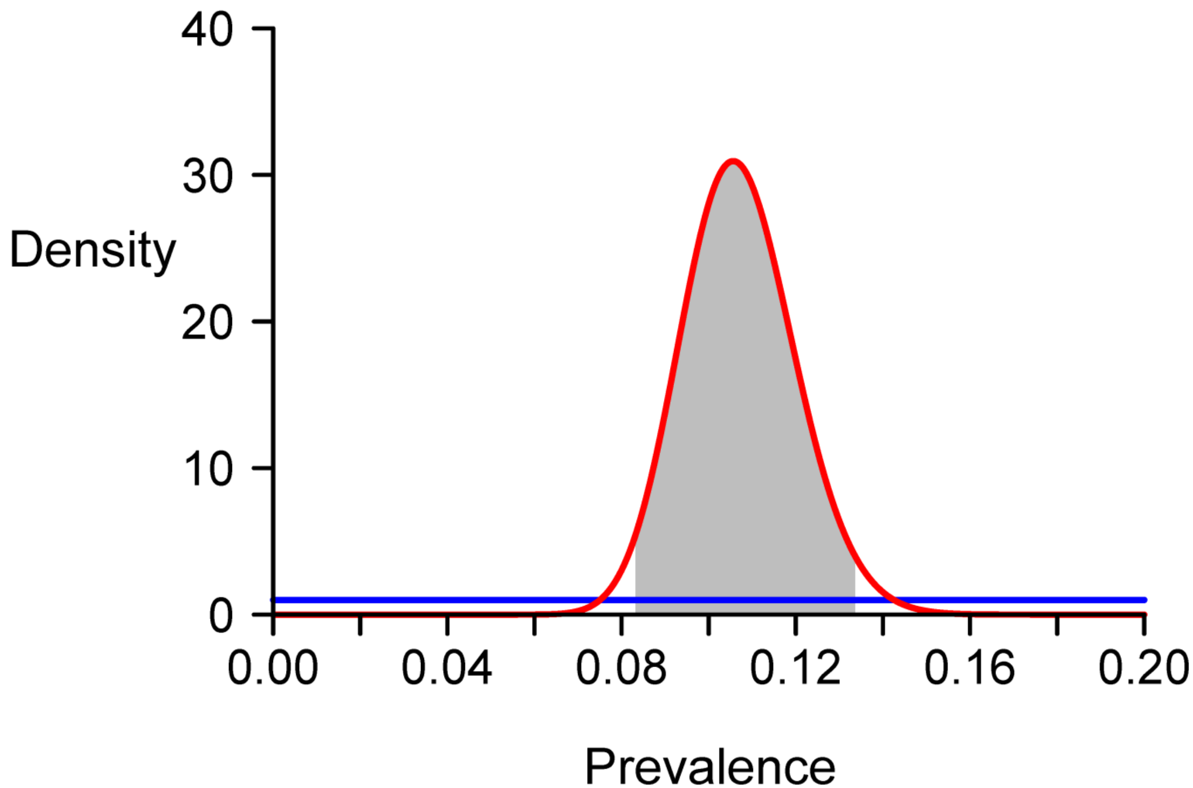

Figure: Flat prior for prevalence of asthma, and posterior distribution with 95% credible interval (8.3%, 13.4%) shaded.

Example – Informative prior

In the above asthma prevalence example,

- With an informative beta(3.3, 30) prior, the posterior distribution is beta(63.3, 538).
- Point estimate (posterior mean) of the prevalence is 10.5%.
- A 95% credible interval for the prevalence is (8.2%, 13.1%).
  - Lower limit a found via $Pr(\theta < a) = 0.025$ using `qbeta(0.025, 63.3, 538)`.
  - Upper limit b found via $Pr(\theta > b) = 0.025$ using `qbeta(0.975, 63.3, 538)`.

Assignment 3 and Exam format

- Assignment 3 will be released after this lecture and the deadline is after final exam
- Exam date: August 1, 2023 (Tuesday)
- Time: 6:30 – 8:30pm
- Venue: 3SR-LT2
- Open book
- Use your own laptop
  - allow access to R help
  - suggest to pre-install the packages
- Choose 2 out of 3 questions:
  - structured questions on analyzing dataset using R
  - correct choice of method and interpretation of the results, clear presentation

Final exam 2023

1. You will be asked to analyze datasets and present your results. Please make sure your computer is ready with R and relevant packages installed (listed under the section "Final Examination" in Moodle). Also, it is important that you are familiar with how to read data from a downloaded dataset.

2. You will be asked to choose 2 out of 3 structured questions, similar to the assignments and tutorials. High marks will be given for correct choice of methods and interpretation of the results, and also clear presentation. R script will not be marked, but you could submit the script for our reference (optional).

Final exam 2023

3. Some answers will depend on earlier parts of the structured questions. Please be reassured that earlier errors will not be brought over to later parts in the marking. In case of any technical problem, please capture the screen and email us for our reference.

4. Please arrive 15 minutes before the exam starts at 6:30pm.

5. Submit your answer files through the Moodle system. **Only** the last submitted version will be graded.

6. You are advised to submit your files five minutes before the end of the examination.

Final exam 2023

7. You may refer to your printed or electronic lecture notes, practical, assignments, solutions and online R help pages during the examination. However, **access to other materials through internet and/or discussions is NOT allowed.**

Anything other than R, R studio, pdf and word file on your screen may be suspected to be cheating and reported to HKU examination office

8. No cell phone. Using cell phone as calculators **is NOT allowed.**

Final exam 2023

Before the examination:

- You should check if your laptop and R work properly.
- Quit all communication / messaging programs on your laptop.

During the examination:

- You are STRONGLY advised to save your work frequently in your laptop / Moodle to avoid lost in case of computer problem.

After submission:

- Please check again you have submitted your files to the Moodle and both files are readable.