Bayesian Statistics: course overview

- Half theory / lectures — half hands-on practical work
- How to learn from different kinds of data using Bayesian modelling
- How to do it in practice using R and JAGS software

1. Expressing uncertainty with probability
2. Bayesian inference
3. Regression models
4. Critiquing and comparing Bayesian models
5. Hierarchical models
6. Modelling missing and censored data
7. Integrating multiple sources of data

Session 1: Expressing uncertainty with probability

What makes a method "Bayesian"

Expressing uncertainty about knowledge using probability

1. Basic definitions: (unknowns, data, predictions, decisions...)
   Expressing judgements as probability distributions
   Practical session (1): probability judgements [1h 10min]
2. Making predictions and decisions given probabilistic judgements
   Practical session (2): predictions and decisions [1h 10min]
3. Introduction to probabilistic programming using JAGS
   Practical session (3): introduction to JAGS [40 min]

General approach to implementing Bayesian methods

Computation using random sampling

Session 1: Expressing uncertainty with probability

What makes a method "Bayesian"

Expressing uncertainty about knowledge using probability

1. Basic definitions: (unknowns, data, predictions, decisions...)
   Expressing judgements as probability distributions
   Practical session (1): probability judgements [1h 10min]
2. Making predictions and decisions given probabilistic judgements
   Practical session (2): predictions and decisions [1h 10min]
3. Introduction to probabilistic programming using JAGS
   Practical session (3): introduction to JAGS [40 min]

General approach to implementing Bayesian methods

Computation using random sampling

Session 1: Expressing uncertainty with probability

What makes a method "Bayesian"

Expressing uncertainty about knowledge using probability

1. Basic definitions: (unknowns, data, predictions, decisions...)
   Expressing judgements as probability distributions
   Practical session (1): probability judgements [1h 10min]
2. Making predictions and decisions given probabilistic judgements
   Practical session (2): predictions and decisions [1h 10min]
3. Introduction to probabilistic programming using JAGS
   Practical session (3): introduction to JAGS [40 min]

General approach to implementing Bayesian methods

Computation using random sampling

Session 1: Expressing uncertainty with probability

What makes a method "Bayesian"

Expressing uncertainty about knowledge using probability

1. Basic definitions: (unknowns, data, predictions, decisions. . .)
   Expressing judgements as probability distributions
   Practical session (1): probability judgements [1h 10min]
2. Making predictions and decisions given probabilistic judgements
   Practical session (2): predictions and decisions [1h 10min]
3. Introduction to probabilistic programming using JAGS
   Practical session (3): introduction to JAGS [40 min]

General approach to implementing Bayesian methods

Computation using random sampling

Session 1: Expressing uncertainty with probability

What makes a method "Bayesian"

Expressing uncertainty about knowledge using probability

1. Basic definitions: (unknowns, data, predictions, decisions...)
   Expressing judgements as probability distributions
   Practical session (1): probability judgements [1h 10min]
2. Making predictions and decisions given probabilistic judgements
   Practical session (2): predictions and decisions [1h 10min]
3. Introduction to probabilistic programming using JAGS
   Practical session (3): introduction to JAGS [40 min]

General approach to implementing Bayesian methods

Computation using random sampling

Concepts in Bayesian statistics

What is Bayesian statistics

Unknowns

Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

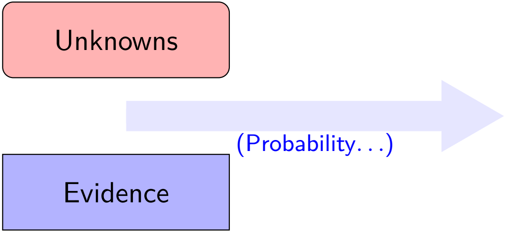
Unknowns

Evidence

Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

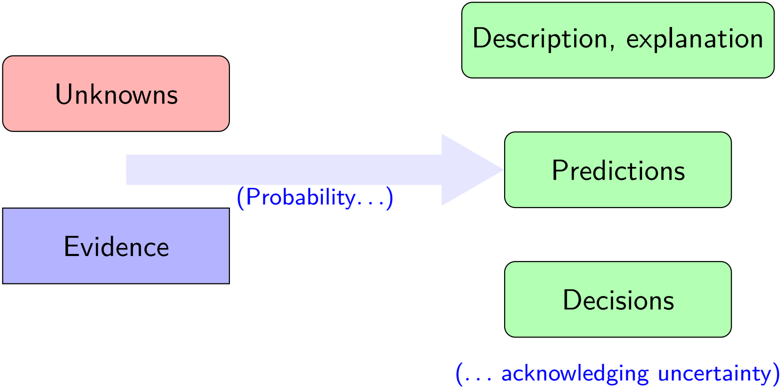
What is Bayesian statistics



Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

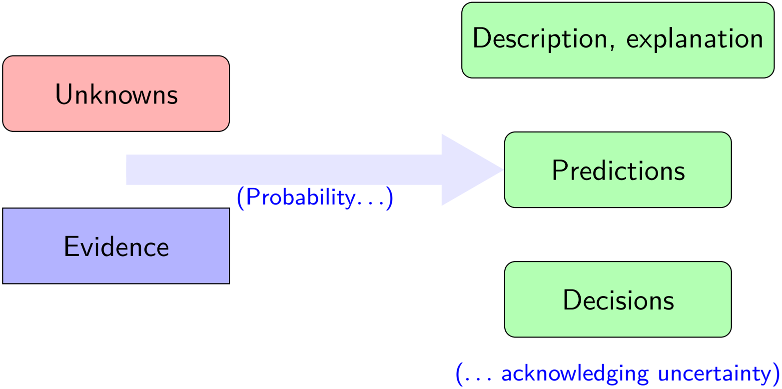
What is Bayesian statistics



Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

What is Bayesian statistics



Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

What is Bayesian statistics

Unknowns

Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

Examples of unknowns

(a) Do I have COVID-19?

Binary (yes/no)

We could know this given evidence from a perfect test

(b) What is the prevalence of COVID-19 in the region?

Proportion (between 0 and 1)

Could know this, given evidence from testing everyone with a perfect test

Quantities like (b) are usually treated as descriptions of an infinite population, and will be called parameters.

A Bayesian describes uncertainty about any unknowns by probability. We will give unknowns the symbol θ.

Examples of unknowns

(a) Do I have COVID-19?

Binary (yes/no)

We could know this given evidence from a perfect test

(b) What is the prevalence of COVID-19 in the region?

Proportion (between 0 and 1)

Could know this, given evidence from testing everyone with a perfect test

Quantities like (b) are usually treated as descriptions of an infinite population, and will be called parameters.

A Bayesian describes uncertainty about any unknowns by probability. We will give unknowns the symbol θ.

Examples of unknowns

(a) Do I have COVID-19?

Binary (yes/no)

We could know this given evidence from a perfect test

(b) What is the prevalence of COVID-19 in the region?

Proportion (between 0 and 1)

Could know this, given evidence from testing everyone with a perfect test

Quantities like (b) are usually treated as descriptions of an infinite population, and will be called parameters.

A Bayesian describes uncertainty about any unknowns by probability. We will give unknowns the symbol θ.

Expressing uncertainty about unknowns using probability

(a) $\theta = \begin{cases} 1 & \text{if I have COVID 19} \\ 0 & \text{otherwise} \end{cases}$

Belief fully described by $p = P(\theta = 1)$
probability that I have COVID-19

(b) θ: Prevalence of COVID-19

Example of a belief as a probability: "I am 90% sure that the prevalence is less than 50%"

A full probability distribution for beliefs about θ would describe $P(\theta < c)$ for any $c \in [0, 1]$

Expressing uncertainty about unknowns using probability

(a) $\theta = \begin{cases} 1 & \text{if I have COVID 19} \\ 0 & \text{otherwise} \end{cases}$

Belief fully described by $p = P(\theta = 1)$
probability that I have COVID-19

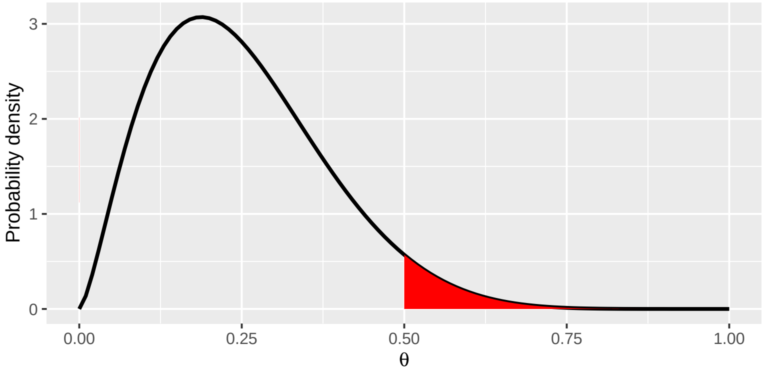
(b) θ: Prevalence of COVID-19

Example of a belief as a probability: "I am 90% sure that the prevalence is less than 50%"

A full probability distribution for beliefs about θ would describe $P(\theta < c)$ for any $c \in [0, 1]$

A full probability distribution for an unknown

Probability density $p(\theta)$ for disease prevalence
Area under curve between c_1, c_2 is $\int_{c_1}^{c_2} p(u)du = P(c_1 < \theta < c_2)$



Red area is $P(\theta > 0.5)$, our "certainty" that prevalence is $> 50\%$.

See later for how to define specific distributions based on judgements

What is Bayesian statistics

Unknowns

Evidence

Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

Learning from observed data

Combine probabilistic judgements about unknowns (prior distributions), with evidence from data.

(a) $\theta =$ do I have COVID-19?
beliefs: (e.g. symptoms, recent contacts, prevalence in area)
data: (e.g. negative lateral flow test)
$P(\theta = 0)$ may not be exactly 0 or 1 given beliefs+data

(b) $\theta =$ COVID-19 prevalence in area
beliefs: (complete ignorance? or prevalence in other areas)
data: Do a survey: test a sample of the population.

Update beliefs given new evidence, using laws of conditional probability ("Bayes' theorem"), producing a posterior distribution. Covered in Session 2

Could be multiple sources of belief and evidence (see Session 7)

Learning from observed data

Combine probabilistic judgements about unknowns (prior distributions), with evidence from data.

(a) $\theta =$ do I have COVID-19?
  beliefs: (e.g. symptoms, recent contacts, prevalence in area)
  data: (e.g. negative lateral flow test)
  $P(\theta = 0)$ may not be exactly 0 or 1 given beliefs+data

(b) $\theta =$ COVID-19 prevalence in area
  beliefs: (complete ignorance? or prevalence in other areas)
  data: Do a survey: test a sample of the population.

Update beliefs given new evidence, using laws of conditional probability ("Bayes' theorem"), producing a posterior distribution. Covered in Session 2

Could be multiple sources of belief and evidence (see Session 7)

Learning from observed data

Combine probabilistic judgements about unknowns (prior distributions), with evidence from data.

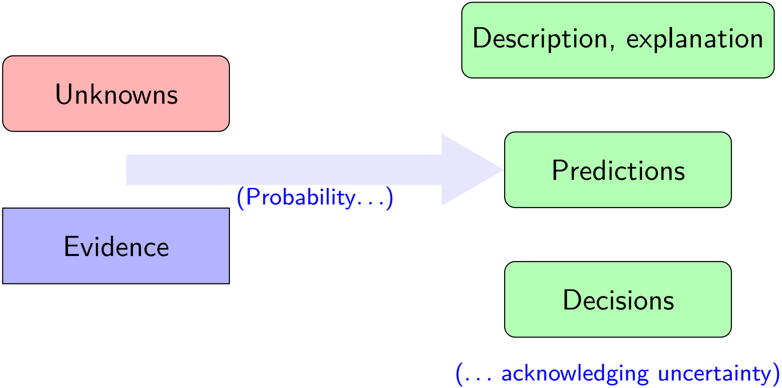
(a) $\theta =$ do I have COVID-19?
  beliefs: (e.g. symptoms, recent contacts, prevalence in area)
  data: (e.g. negative lateral flow test)
  $P(\theta = 0)$ may not be exactly 0 or 1 given beliefs+data

(b) $\theta =$ COVID-19 prevalence in area
  beliefs: (complete ignorance? or prevalence in other areas)
  data: Do a survey: test a sample of the population.

Update beliefs given new evidence, using laws of conditional probability ("Bayes' theorem"), producing a posterior distribution. Covered in Session 2

Could be multiple sources of belief and evidence (see Session 7)

What is Bayesian statistics



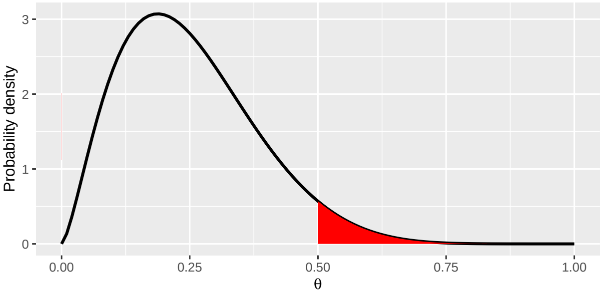
Expressing beliefs about unknowns , and combining evidence , using probability.

What do we mean by these things?

Predicting observable outcomes given current beliefs

Alternative example: Drug

$\theta =$ probability that a patient responds to a drug.
We have a distribution representing beliefs about this.



If we gave the drug to 20 people

- how many would we expect to respond?
- what is our uncertainty about this prediction?

See later for how to formalise this mathematically

Make decisions given current evidence

A new treatment is more effective, but more expensive than the current standard of care

- Should the health service pay for it? (see later)

Given test results (e.g. positive LFT + negative PCR) I think I am COVID-positive with probability p

- Should I restrict social contacts?
  (theoretical example, not developed!)

Requires additional information: our "loss" or "utility" for different consequences of actions given true θ.

See "Decision theory" later.

Make decisions given current evidence

A new treatment is more effective, but more expensive than the current standard of care

- Should the health service pay for it? (see later)

Given test results (e.g. positive LFT + negative PCR) I think I am COVID-positive with probability p

- Should I restrict social contacts?
  (theoretical example, not developed!)

Requires additional information: our "loss" or "utility" for different consequences of actions given true θ.

See "Decision theory" later.

Make decisions given current evidence

A new treatment is more effective, but more expensive than the current standard of care

- Should the health service pay for it? (see later)

Given test results (e.g. positive LFT + negative PCR) I think I am COVID-positive with probability p

- Should I restrict social contacts?
  (theoretical example, not developed!)

Requires additional information: our "loss" or "utility" for different consequences of actions given true θ.

See "Decision theory" later.

Bayesian versus frequentist view of probability

Bayesians and frequentists both use probability models to describe observable quantities, e.g.

(a) an individual's COVID status

(b) proportion of positive tests r/n we would observe from a survey of n people

But they conceive uncertainty about model parameters θ differently.

- Bayesian: place a probability distribution on θ, update given data
- Frequentist: θ is fixed, but given data, we can, e.g.
  - test hypotheses about different values of θ
  - construct a parameter estimator e.g. $\hat{\theta} = r/n$
  - represent uncertainty through the standard error of the estimator: variability with respect to repeated samples of size n.

Bayesian versus frequentist view of probability

Bayesians and frequentists both use probability models to describe observable quantities, e.g.

(a) an individual's COVID status

(b) proportion of positive tests r/n we would observe from a survey of n people

But they conceive uncertainty about model parameters θ differently.

- Bayesian: place a probability distribution on θ, update given data
- Frequentist: θ is fixed, but given data, we can, e.g.
  - test hypotheses about different values of θ
  - construct a parameter estimator e.g. $\hat{\theta} = r/n$
  - represent uncertainty through the standard error of the estimator: variability with respect to repeated samples of size n.

Frequentist versus Bayesian inference from data

Frequentist inference only based on the current study data

- $\text{var}(\hat{\theta})$ describes variance of an estimator over repeated samples
- Uncertainty arises from the finite nature of the sample n

Bayesian inference interprets data in context of external evidence about quantities of interest

- $\text{var}(\theta)$ is the variance of the probability distribution that describes our belief
- Measures uncertainty about our belief, given whatever evidence has been supplied (prior beliefs, observed data...)

Expressing judgements about unknowns as probability distributions

Expressing uncertainty using probability

In this session we will show how to make predictions/decisions based on substantive information, alone, which we will call a prior.

- In session 2 this will be combined with data (and we will define the posterior...).

How do we express judgements using probability?

First determine class of distributions to use based on range of θ

- binary: individual disease status, $\theta = 0$ or 1
  a single number $P(\theta = 1)$
- proportion: disease prevalence $0 \leq \theta \leq 1$
  $\theta \sim$ Beta, uniform, logit-normal...
- positive: mean survival time $\theta > 0$
  $\theta \sim$ exponential, gamma, Weibull...
- unrestricted: $\theta =$ mean change in cholesterol $-\infty < \theta < \infty$
  $\theta \sim$ normal, t...

Expressing uncertainty using probability

In this session we will show how to make predictions/decisions based on substantive information, alone, which we will call a prior.

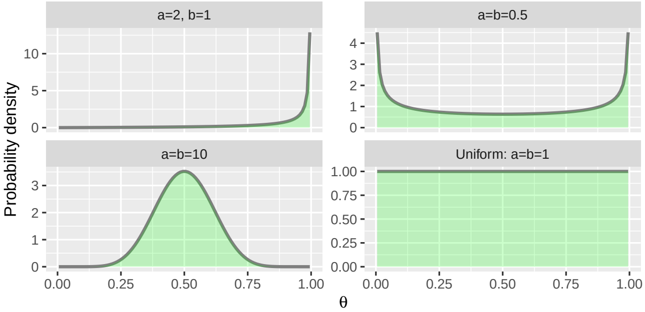
- In session 2 this will be combined with data (and we will define the posterior...).

How do we express judgements using probability?

First determine class of distributions to use based on range of θ

- binary: individual disease status, $\theta = 0$ or 1
  a single number $P(\theta = 1)$
- proportion: disease prevalence $0 \leq \theta \leq 1$
  $\theta \sim$ Beta, uniform, logit-normal...
- positive: mean survival time $\theta > 0$
  $\theta \sim$ exponential, gamma, Weibull...
- unrestricted: $\theta =$ mean change in cholesterol $-\infty < \theta < \infty$
  $\theta \sim$ normal, t...

Beta distribution for probabilities



Flexible distribution that can express a range of beliefs about a quantity between 0 and 1

Defined by two "shape" parameters a and b

Probability density is $p(\theta) = \theta^{a-1}(1-\theta)^{b-1}/B(a,b)$ (where $B()$ is the "beta function"). Won't usually use this formula directly.

How to translate beliefs to specific a and b...?

Heuristics to obtain Beta distributions from beliefs

Example. Drug: θ is the probability that a patient responds to a drug. We have knowledge about similar compounds that treat similar conditions:

- a guess at most likely value θ_0
- a credible interval $\theta^{(L)}, \theta^{(U)}$ describing plausible range of values

Beta(a, b) distribution has mean $\mu = a/(a+b)$
standard deviation $\sigma = \sqrt{ab}/((a+b)\sqrt{a+b+1})$

Set mean to $\mu = \theta_0$
Obtain σ from width of interval (based on normal distribution)

- e.g. 95% credible interval has width $\theta^{(U)} - \theta^{(L)} = 4\sigma$

Solving for a, b gives
$a = (\mu(1-\mu)/\sigma^2 - 1)\mu, \quad b = (\mu(1-\mu)/\sigma^2 - 1)(1-\mu)$

Heuristics to obtain Beta distributions from beliefs

Example. Drug: θ is the probability that a patient responds to a drug. We have knowledge about similar compounds that treat similar conditions:

- a guess at most likely value θ_0
- a credible interval $\theta^{(L)}, \theta^{(U)}$ describing plausible range of values

Beta(a, b) distribution has mean $\mu = a/(a+b)$
standard deviation $\sigma = \sqrt{ab}/((a+b)\sqrt{a+b+1})$

Set mean to $\mu = \theta_0$
Obtain σ from width of interval (based on normal distribution)

- e.g. 95% credible interval has width $\theta^{(U)} - \theta^{(L)} = 4\sigma$

Solving for a, b gives
$a = (\mu(1-\mu)/\sigma^2 - 1)\mu, \quad b = (\mu(1-\mu)/\sigma^2 - 1)(1-\mu)$

Heuristics to obtain Beta distributions from beliefs

Example. Drug: θ is the probability that a patient responds to a drug. We have knowledge about similar compounds that treat similar conditions:

- a guess at most likely value θ_0
- a credible interval $\theta^{(L)}, \theta^{(U)}$ describing plausible range of values

Beta(a, b) distribution has mean $\mu = a/(a+b)$
standard deviation $\sigma = \sqrt{ab}/((a+b)\sqrt{a+b+1})$

Set mean to $\mu = \theta_0$
Obtain σ from width of interval (based on normal distribution)

- e.g. 95% credible interval has width $\theta^{(U)} - \theta^{(L)} = 4\sigma$

Solving for a, b gives
$a = (\mu(1-\mu)/\sigma^2 - 1)\mu, \quad b = (\mu(1-\mu)/\sigma^2 - 1)(1-\mu)$

Heuristics to obtain Beta distributions from beliefs

Example. Drug: θ is the probability that a patient responds to a drug. We have knowledge about similar compounds that treat similar conditions:

- a guess at most likely value θ_0
- a credible interval $\theta^{(L)}, \theta^{(U)}$ describing plausible range of values

Beta(a, b) distribution has mean $\mu = a/(a+b)$
standard deviation $\sigma = \sqrt{ab}/((a+b)\sqrt{a+b+1})$

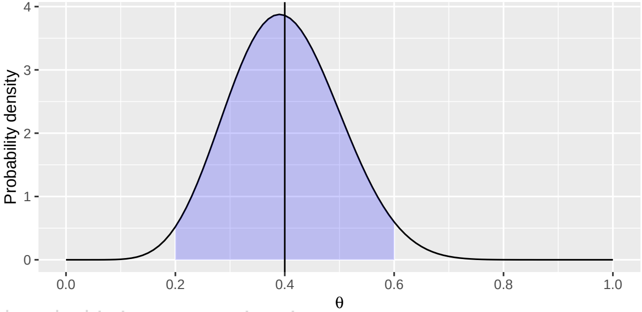
Set mean to $\mu = \theta_0$
Obtain σ from width of interval (based on normal distribution)

- e.g. 95% credible interval has width $\theta^{(U)} - \theta^{(L)} = 4\sigma$

Solving for a, b gives
$a = (\mu(1-\mu)/\sigma^2 - 1)\mu, \quad b = (\mu(1-\mu)/\sigma^2 - 1)(1-\mu)$

Simple elicitation of a Beta distribution: Drug example

Prior guess expected value 0.4
95% credible interval 0.2 to 0.6
Obtain Beta($a = 9.2, b = 13.8$) using that method

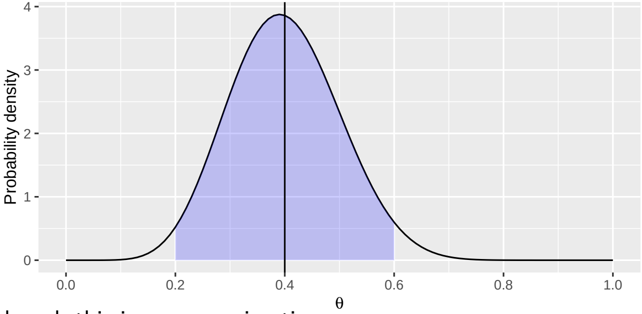


though this is an approximation:

```
> 9.2/(9.2+13.8)
[1] 0.4
> qbeta(c(0.025, 0.975), 9.2, 13.8)
[1] 0.2143175 0.6021595
```

Simple elicitation of a Beta distribution: Drug example

Prior guess expected value 0.4
95% credible interval 0.2 to 0.6
Obtain Beta($a = 9.2, b = 13.8$) using that method

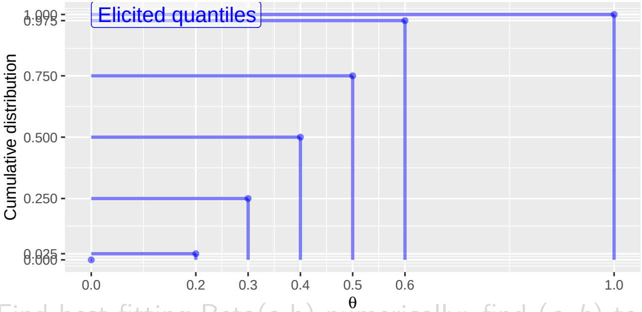


though this is an approximation:

```
> 9.2/(9.2+13.8)
[1] 0.4
> qbeta(c(0.025, 0.975), 9.2, 13.8)
[1] 0.2143175 0.6021595
```

General method to fit a distribution to elicited quantiles

Suppose we can judge quantiles c_i for θ, where we are $100p_i\%$ confident that $\theta < c_i$ for a series of $i = 1, 2, \ldots$, e.g.



Find best-fitting Beta(a,b) numerically: find (a, b) to minimise sum of squared error between Beta quantile and our guess

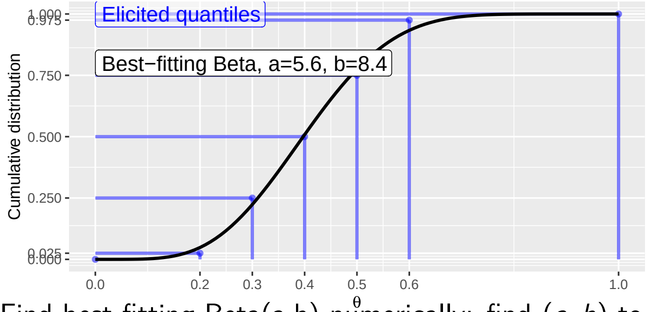
$$\sum_i (qbeta(p_i, a, b) - c_i)^2$$

where if $X \sim Beta(a, b)$, $P(X < qbeta(p_i, a, b)) = p_i$

Similar procedure in the `SHELF` R package (function `fitdist`)

General method to fit a distribution to elicited quantiles

Suppose we can judge quantiles c_i for θ, where we are $100p_i\%$ confident that $\theta < c_i$ for a series of $i = 1, 2, \ldots$, e.g.



Find best-fitting Beta(a,b) numerically: find (a, b) to minimise sum of squared error between Beta quantile and our guess

$$\sum_i (qbeta(p_i, a, b) - c_i)^2$$

where if $X \sim Beta(a, b)$, $P(X < qbeta(p_i, a, b)) = p_i$

Similar procedure in the `SHELF` R package (function `fitdist`)

Defining a Beta distribution from an "effective sample size"

A $Beta(a, b)$ distribution for a proportion θ is roughly equivalent to:

the information from a successes, with success probability θ, from an effective sample size of $a + b$ binary outcomes.

(justification from Bayes' theorem, see Session 2)

So we can elicit a Beta distribution using

- a guess at the mean $a/(a + b)$
- an effective sample size measuring the confidence in this guess.

Examples of Beta distributions with a mean of $a/(a + b) = 0.4$

a	b	95% credible interval	Effective sample size
4	6	(0.14, 0.70)	10
40	60	(0.31, 0.50)	100
400	600	(0.37, 0.43)	1000

Defining a Beta distribution from an "effective sample size"

A $Beta(a, b)$ distribution for a proportion θ is roughly equivalent to:

the information from a successes, with success probability θ, from an effective sample size of $a + b$ binary outcomes.

(justification from Bayes' theorem, see Session 2)

So we can elicit a Beta distribution using

- a guess at the mean $a/(a + b)$
- an effective sample size measuring the confidence in this guess.

Examples of Beta distributions with a mean of $a/(a + b) = 0.4$

a	b	95% credible interval	Effective sample size
4	6	(0.14, 0.70)	10
40	60	(0.31, 0.50)	100
400	600	(0.37, 0.43)	1000

Practical session (1): probability judgements

- Defining Beta distributions to represent uncertainty about a proportion, based on point estimates and credible intervals.
- Checking and visualising the defined distributions
- Writing generalizable code in R to do this

Note: "probability is perfect, but we can't elicit it perfectly"
(quote from O'Hagan and Oakley, 2004)

Making predictions and decisions in Bayesian analyses

Prediction under uncertainty

Example

- We have information about the monthly incidence of a disease
  Probability θ that a person gets the disease within a month
  Expectation 0.1, 95% CI (0.05, 0.15) $\rightarrow$ Beta(14, 128)
- In a population of size $n = 1000$, how many people would we expect to get the disease in the next month?
- What is the chance that the healthcare system will be overwhelmed?
  - e.g. predicted cases > 200

Describe uncertainty about predictions using probability

Predictive distributions. Example: Beta-binomial

Given population size n, common incidence $\theta \sim Beta(a, b)$

- sampling distribution for number of cases in a month: $y \sim \text{Binomial}(n, \theta)$ (assuming a person can't get the disease twice)
- However we are uncertain about θ
- We need the predictive distribution of y, with probability mass function $p_y(y) = \int p_{y|\theta}(y|\theta)p_\theta(\theta)d\theta$

If $\theta \sim Beta(a, b)$, and $y \sim Binomial(n, \theta)$ this distribution is called the beta-binomial with parameters (a, b, n)

$p_y(y; a, b, n)$ has a closed form involving beta functions

https://en.wikipedia.org/wiki/Beta-binomial_distribution

How to compute

- Mean or median and credible interval for cases y?
- $P(y > 200)$: probability that we will exceed capacity?

Predictive distributions. Example: Beta-binomial

Given population size n, common incidence $\theta \sim Beta(a, b)$

- sampling distribution for number of cases in a month: $y \sim \text{Binomial}(n, \theta)$ (assuming a person can't get the disease twice)
- However we are uncertain about θ
- We need the predictive distribution of y, with probability mass function $p_y(y) = \int p_{y|\theta}(y|\theta) p_\theta(\theta) d\theta$

If $\theta \sim Beta(a, b)$, and $y \sim Binomial(n, \theta)$ this distribution is called the beta-binomial with parameters (a, b, n)

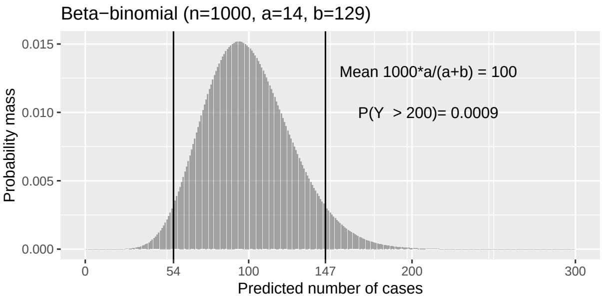
$p_y(y; a, b, n)$ has a closed form involving beta functions

https://en.wikipedia.org/wiki/Beta-binomial_distribution

How to compute

- Mean or median and credible interval for cases y?
- $P(y > 200)$: probability that we will exceed capacity?

Beta-binomial predictive distribution: computation

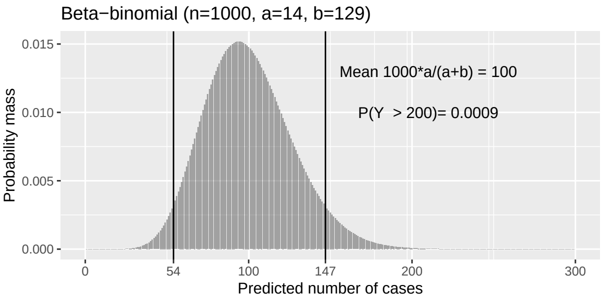


- $P(y > 200)$? From cumulative distribution function (CDF). Could program by hand
- 95% credible limits? Quantile function. No closed form
  - Could compute by using numerical methods to invert the CDF (as implemented in the R package used in the practical)

More general numerical method for summarising distributions

Draw a sample from the distribution, and summarise it.

Beta-binomial predictive distribution: computation

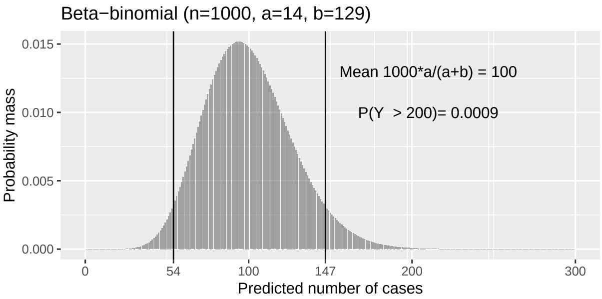


- $P(y > 200)$? From cumulative distribution function (CDF). Could program by hand
- 95% credible limits? Quantile function. No closed form
  - Could compute by using numerical methods to invert the CDF (as implemented in the R package used in the practical)

More general numerical method for summarising distributions

Draw a sample from the distribution, and summarise it.

Beta-binomial predictive distribution: computation



- $P(y > 200)$? From cumulative distribution function (CDF). Could program by hand
- 95% credible limits? Quantile function. No closed form
  - Could compute by using numerical methods to invert the CDF (as implemented in the R package used in the practical)

More general numerical method for summarising distributions

Draw a sample from the distribution, and summarise it.

Sampling from prior and predictive distributions

"Monte Carlo" integration:

- Draw a large sample of values $\theta_r : i = 1, \ldots, R$ from the Beta(a, b) prior distribution
- For each r, draw a sample y_r from the Binomial(n, θ_r) distribution
- The resulting $y_1, y_2, \ldots$ are samples from the Beta-Binomial(n, a, b) distribution

Easy in R using "vectorised" code:

```
R <- 10000
theta <- rbeta(R, 2, 3)    # vector of length R
y <- rbinom(R, size=n, prob=theta)
```

Computing summaries of a distribution from a sample

Easy to compute any desired summary of a distribution, given a sample from it, e.g.

$P(y > 200)$: proportion of values in the sample that are > 200

Median and 95% credible interval for y: 50%, 2.5% and 97.5% quantiles of the sample (finding c such that p% of sampled values are $< c$)

Probability density function: histogram or kernel density estimate of the sampled values

(Mathematically: all examples of "Monte Carlo integration". Integration achieved by sampling and summarising. Explained more formally in Session 2)

In R:

```
mean(y)         # mean of sample estimates mean of distribution
mean(y > 200)   # proportion of sample > 200
                # Exercise: explain why this works in R
quantile(y, c(0.025, 0.5, 0.975))
plot(density(y))
```

How many samples to draw in a Monte Carlo estimate?

Estimate converges to the true value as the number of samples R increases:

	Estimate of mean
R=10	108.5
R=100	102.0
R=1000	99.89
True	100

- Easy answer: "draw as many samples as possible"
- More thoughtful answer: "draw as many as you need for the required precision"
  - 2-3 significant figures usually enough for communication.

How to judge how precisely a summary of a Monte Carlo sample estimates the true summary of a distribution (e.g. the mean)?

How many samples to draw in a Monte Carlo estimate?

Estimate converges to the true value as the number of samples R increases.

(example)	Estimate of mean	Monte Carlo SE
R=10	108.5	6.6
R=100	102.0	2.6
R=1000	99.89	0.9
True	100	

Accuracy of a Monte Carlo estimate of the mean?

- With independent samples y_r, $\frac{1}{R}\sum_r y_r$ normally distributed with mean $E(Y)$ and SD $sd(Y)/\sqrt{R}$ (asymptotically)
- SD of sample divided by $\sqrt{R}$: "Monte Carlo standard error"

Monte Carlo SE estimators available for other summaries (e.g. quantiles, see `posterior` R package).

Rough approach: choose R big enough that so that repeated samples agree within desired accuracy.

How many samples to draw in a Monte Carlo estimate?

Estimate converges to the true value as the number of samples R increases.

(example)	Estimate of mean	Monte Carlo SE
R=10	108.5	6.6
R=100	102.0	2.6
R=1000	99.89	0.9
True	100	

Accuracy of a Monte Carlo estimate of the mean?

- With independent samples y_r, $\frac{1}{R}\sum_r y_r$ normally distributed with mean $E(Y)$ and SD $sd(Y)/\sqrt{R}$ (asymptotically)
- SD of sample divided by $\sqrt{R}$: "Monte Carlo standard error"

Monte Carlo SE estimators available for other summaries (e.g. quantiles, see `posterior` R package).

Rough approach: choose R big enough that so that repeated samples agree within desired accuracy.

Normal distribution with an uncertain mean

Example: Measured level of some pollutant X in an area. Unknown true level is μ. Measurement instrument has a random error, distributed as $N(0, \sigma^2)$, with σ^2 known.

Sampling model:

- $X|\mu \sim N(\mu, \sigma^2)$

Uncertainty distribution $p_\mu()$ ("prior") from background knowledge, e.g. pollution levels in other areas

- $\mu \sim N(m_\mu, s_\mu^2)$

What do we expect to see before we actually measure X?

Predictive distribution for X

- $p_X(x) = \int p(x|\mu) p_\mu(\mu) d\mu$

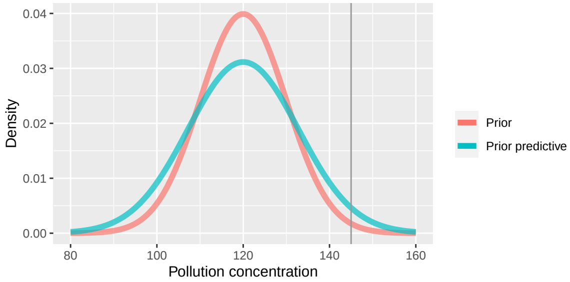
(Prior) predictive distribution for a normal outcome

Normal outcome, fixed variance, normal prior on mean
Predictive distribution is $X \sim N(m_\mu, s_\mu^2 + \sigma^2)$
(can compute more generally with simulation)

Expected outcome $E(X)$ is simply the prior mean m_μ.

Predictive distribution is wider than the prior: uncertainty includes both the prior variance s_μ^2 and the measurement error variance σ^2

Example: $m_\mu = 120$, $s_\mu = 10$, $\sigma = 8$



Used for model checking: if actual measured value of X is extreme compared to the prior predictive distribution, then our model assumptions may be wrong. See Session 4.

Reminder: different kinds of distributions

Sampling distribution: $p(x|\theta)$ describes variability in an (observable) outcome x, given fixed parameter values θ

Prior distribution: $p(\theta)$ describes uncertainty about a parameter θ

Predictive distribution: $p(x) = \int p(x|\theta)p(\theta)d\theta$ describes variability in an outcome, given uncertain parameters

- In this session, all our predictive distributions are prior predictive distributions (knowledge of θ described by the prior)

In Session 2 you will see posterior and posterior predictive distributions...

Decision making under uncertainty

Bayesian decision theory

- Choose between actions $a = 1, 2, \ldots$, given knowledge of (vector of) parameters $\boldsymbol{\theta}$ with uncertainty distribution $f_{\boldsymbol{\theta}}()$
- Define a net benefit function $NB(a, \boldsymbol{\theta})$
- Choose action a that maximises $E_{\boldsymbol{\theta}}(NB(a, \boldsymbol{\theta}))$, expectation with respect to the uncertain parameters (Bernardo and Smith *Bayesian Theory* 1994)

Can implement using Monte Carlo simulation

- simulate $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_R$
- compute by $NB(a, \boldsymbol{\theta}_1), \ldots, NB(a, \boldsymbol{\theta}_R)$ for each a
- determine $E_{\boldsymbol{\theta}}(NB(a, \boldsymbol{\theta}))$ as mean of the ath sample

Example of Bayesian decision making

Example: health economic evaluations in the UK (and similar systems) are informed by Bayesian decision-theoretic models.

- $NB(a, \boldsymbol{\theta})$ defined to trade off the health benefits and costs of a treatment a, and the decision-maker's willingness to pay.
- Parameters $\boldsymbol{\theta}$ include risks of disease-related outcomes, healthcare costs and treatment effects on these.
- All available/relevant evidence about $\boldsymbol{\theta}$ obtained, and expressed as probability distributions
- Distributions of $NB(a, \boldsymbol{\theta})$ deduced by Monte Carlo methods $\rightarrow$ decision-making.

Details in, e.g. Baio, G., (2013) *Bayesian Methods in Health Economics*

This usually requires multiple sources of uncertain evidence (see Session 7).

Practical session (2): predictions and decisions

Predicting from Binomial distributions with uncertain outcomes using the Beta-Binomial distribution

- Exact calculations using analytic properties of the distribution (using R, with a specialised package)
- Calculations by simulation in R
- Decision-making between two treatments with binary outcomes

Predicting from Normal distributions with an uncertain mean

- What if the standard deviation is uncertain as well as the mean

Introduction to probabilistic programming using JAGS

Introduction to probabilistic programming

Probabilistic programming software allows the values of variables to be uncertain and described by probability distributions.

e.g. instead of defining `p = 0.4`, we can define $\mathtt{p} \sim \mathit{Beta}(2, 1)$

Designed to make it easy to

1. Predict uncertain quantities based on probabilistic judgements
   - Examples in this session. Though can just as easily use basic R for these
2. Learn about uncertain quantities from observed data using Bayesian inference (next session)
   - This is much harder to do in general without specialised packages

Examples of probabilistic programming software

BUGS: `http://www.mrc-bsu.cam.ac.uk/bugs`

- Various software packages (WinBUGS, OpenBUGS, MultiBUGS, NIMBLE) based on a similar language
- We will use a program called JAGS, `https://mcmc-jags.sourceforge.io`
  Uses a version of the BUGS language.
  Can be controlled from R.

Stan `https://mc-stan.org` also widely used and recommended.

- Can be controlled from R, or Python, or from command line

PyMC `https://www.pymc.io` is also popular with Python users

Turing.jl `https://turing.ml` in the Julia language

Example: Drug model in JAGS language

Define a model in the JAGS language

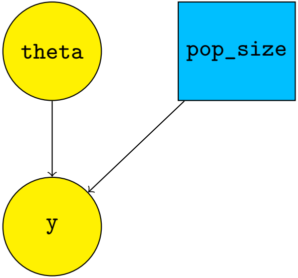
- describes relation between known and uncertain quantities
- via statistical models or probabilistic judgements

Beta / Binomial prediction example:

```
model {
  # Defining a fixed quantity
  pop_size <- 1000

  # Defining a prior distribution
  # for a disease incidence
  theta ~ dbeta(9.2, 13.8)

  # Predicting number of cases
  # given uncertain incidence
  y ~ dbinom(theta, pop_size)
}
```



A directed acyclic graph:
unknowns in yellow circles,
knowns as blue squares
Arrows: model definitions

Key features of JAGS syntax

`<-` means a deterministic definition.

Here we are defining the size of the population to predict as 1000

We can also use mathematical functions here, e.g.

`x <- exp(log(10000) + 2)`

`~` means a stochastic definition.

Defines a distribution for a random variable

This can be used in various subtly different contexts:

(a) Defining prior distributions of parameters.
    Here `theta` is given a Beta prior

(b) Predicting data from models.
    Here `y` is predicted from a Binomial model

(c) Defining a statistical model to fit to observed data
    Discussed in next session. There is no observed data in this example.

What running JAGS does

The model is "run" by calling an R function

Draws a sample from the joint distribution of all unknown parameters, given any observed data

- This session. No observed data. All information specified as probability judgements on parameters.
  - JAGS samples from prior distributions (for parameters) or prior predictive distributions (for observable quantities)
  - Uses standard Monte Carlo methods. Just as easy in basic R
- Next session. There is observed data.
  - JAGS automatically constructs and simulates from posterior distributions for unknowns given data
  - Uses Markov Chain Monte Carlo. Harder — what JAGS was designed for

What running JAGS does

The model is "run" by calling an R function

Draws a sample from the joint distribution of all unknown parameters, given any observed data

- This session. No observed data. All information specified as probability judgements on parameters.
  - JAGS samples from prior distributions (for parameters) or prior predictive distributions (for observable quantities)
  - Uses standard Monte Carlo methods. Just as easy in basic R
- Next session. There is observed data.
  - JAGS automatically constructs and simulates from posterior distributions for unknowns given data
  - Uses Markov Chain Monte Carlo. Harder — what JAGS was designed for

What running JAGS does

The model is "run" by calling an R function

Draws a sample from the joint distribution of all unknown parameters, given any observed data

- This session. No observed data. All information specified as probability judgements on parameters.
  - JAGS samples from prior distributions (for parameters) or prior predictive distributions (for observable quantities)
  - Uses standard Monte Carlo methods. Just as easy in basic R
- Next session. There is observed data.
  - JAGS automatically constructs and simulates from posterior distributions for unknowns given data
  - Uses Markov Chain Monte Carlo. Harder — what JAGS was designed for

Practical session (3): introduction to JAGS

A worked example of running JAGS via the `rjags` R package.

- Drug example: prediction from a binomial distribution with a Beta prior on the probability
- Learning basic steps
  1. writing models in JAGS code
  2. calling JAGS to generate random samples from distributions
  3. summarising the sampled output

  Explanation provided for what each function and line of code does
- Demonstration that we get the same results as with Monte Carlo simulation in R (see Practical 2)

Looking forward to later sessions

Next session will introduce observed data, and Bayes' theorem for learning from combinations of

- observed data
- probability judgements about unknowns ("priors")

Subsequent sessions will cover how to fit, check and compare models for different kinds of data

- hence make inferences, predictions or decisions, as we saw here

There will be more on how to choose prior distributions for particular kinds of model parameters

- Prior judgements are more important when the data are weaker