Day 2

Sonja Hartnack Valerie Hungerbühler

2021-07-15

Ex 3 from yesterday continued

- Example of a COVID-19 data set

1. how to prepare the data set in the correct format
   - create_data_cassaniti.R
2. how to describe the model
   - model_final.bug
3. how to run the model in JAGS with runjags
   - runjags.version.R
4. how to check convergence
5. how to analyse the data

Exercises

- Ex.3
  - Can you re-run the exercises?
  - Assess what happens if you add other covariances?
  - How many could you add and still have “meaningful results”?
  - Try different priors
  - You might also look at the runjags reference manual if ou find other things ou would like to customize?
  - Could you also extract the information for a single chain?
- Ex. 4 (Bonus)
  - Could you expand the model with a fourth test with simulated data?

Prior choice

- Posterior is proportional to likelihood and to prior
  $P(\theta|data) \propto P(data|\theta) \cdot P(\theta)$
- For the prior we need to choose a distribution and the values.

Priors for binomial data

- Likelihood $L(\pi) = \binom{n}{k}\pi^k(1-\pi)^{n-x}$ for $\pi \in (0,1)$
- log likelihood-kernel: $l(\pi) = xlog\pi + (n-x)log(1-\pi)$
- first derivative $\frac{dl(\pi)}{d\pi} = \frac{x}{\pi} - \frac{n-x}{1-\pi}$
- setting to zero gives MLE $\hat{\pi}_{ML} = \frac{x}{n}$

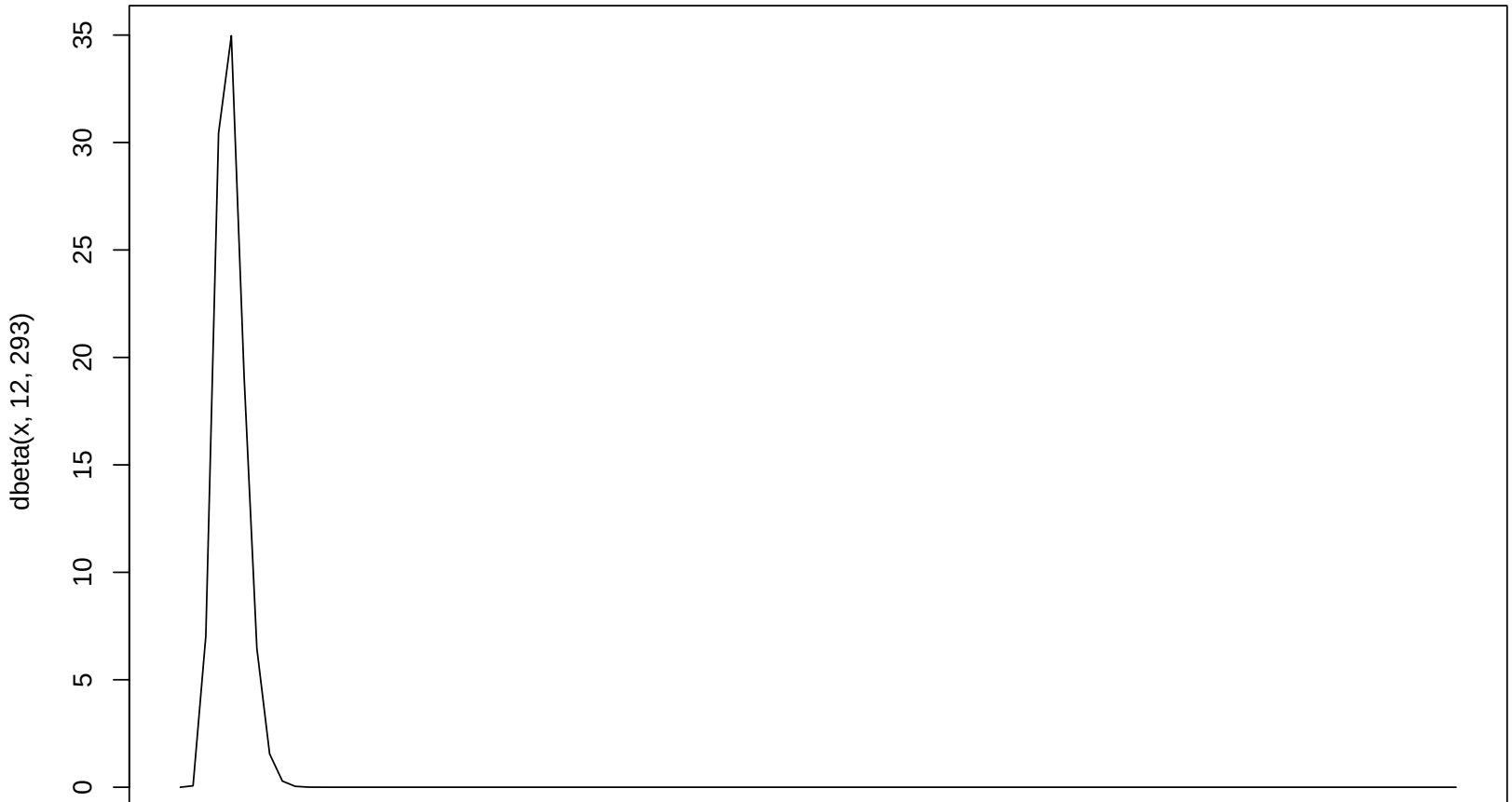
Combining binomial likelihood with a beta prior

- Likelihood: $P(data|\pi) = \binom{n}{k}\pi^k(1-\pi)^{n-x}$
  $P(data|\pi) \propto \pi^k(1-\pi)^{n-x}$
- Beta prior: $Beta(a, b)$ $P(\pi) \propto \pi^{\alpha-1}(1-\pi)^{\beta-1}$
- Posterior: $P(\pi|data) \propto \pi^k(1-\pi)^{n-x}\pi^{\alpha-1}(1-\pi)^{\beta-1}$
  $P(\pi|data) \propto \pi^{k+\alpha-1}(1-\pi)^{n-x+\beta-1}$
  $P(\pi|data) = Beta(\pi|x+\alpha, n-x+\beta)$

A pragmatic approach to choosing a prior distribution is to select a member of a specific family of distributions such that the posterior distribution belongs to the same family (conjugate prior distribution).

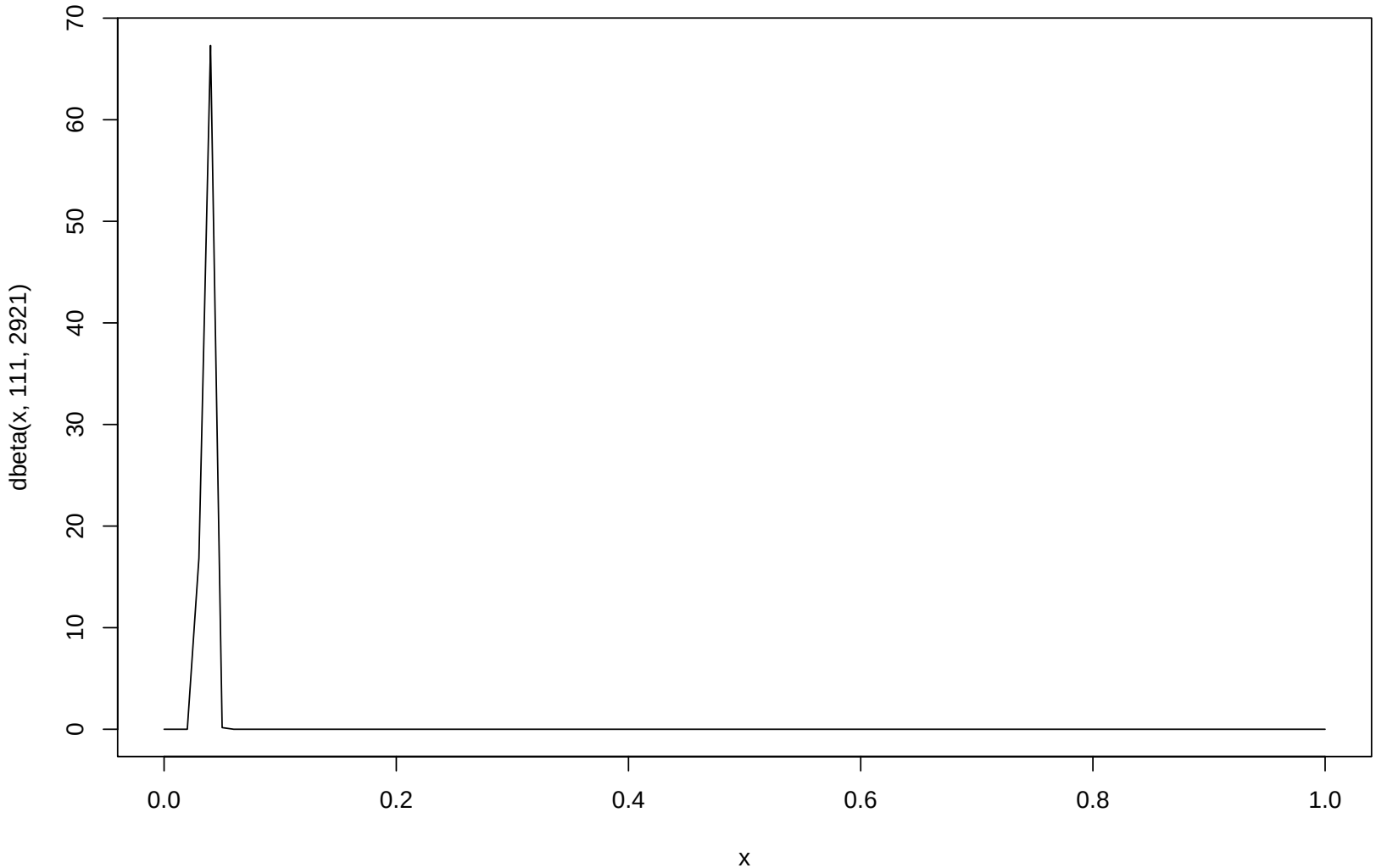
Which values for a prior beta distribution?

- 1. Based on a previous publication: e.g. if 11 out of 303 individuals tested positive:
  $P(\pi|data) = Beta(\pi|x + \alpha, n - x + \beta)$
  $P(\pi|data) = Beta(\pi|11 + 1, 303 - 11 + 1)$
  $P(\pi|data) = Beta(\pi|12, 293)$



Which values for a prior beta distribution?

- 1. Based on a previous publication: e.g. if 110 out of 3030 individuals positive: $P(\pi|data) = Beta(\pi|111, 2921)$



Which values for a prior beta distribution?

2. Based on expert opinion
   - Betabuster https://shiny.vet.unimelb.edu.au/epi/beta.buster/
   - R package PriorGen 'Based on the available literature the mean value for the sensitivity of a test is expected to be 0.90 and we can be 95% sure that it is higher than 0.80.'

```
library(PriorGen)
findbeta(themean=0.90, percentile=0.95,
         lower.v=FALSE, percentile.value=0.80)
```

Beta(27.79, 3.09)

Which values for a prior distribution

- data for prior should be independent of data for likelihood
- do a sensitivity analysis

Ex.5

- Have a look at the excel file 'CLIA_LFIA_ELISA_final.xlsx'
- Use the file 'runjags_multiple_populations.R' to run the models 'blcm_A.R','blcm_B.R','blcm_C.R'
- Describe the different models, in which respect do they differ?
- Do you have other suggestions for 'better' models?

Model selection: inclusion of conditional dependencies between se and sp of different tests

- Pragmatic approach:
  - Look at 95% credibility intervals and histograms of posterior covariances: do they include a zero?
  - Are the other posteriors affected when including a covariance?
  - If either se or sp equal 1 (is perfect), then it will always be conditionally independent of the se or sp of the other test(s)
- Analytical approach:
  - DIC: deviance information criterion (Spiegelhalter, 2002)

Ex. 6 check with a pragmatic approach which model is better (data from Ex.5)

Model selection DIC

Original Article | Full Access

The deviance information criterion: 12 years on

David J. Spiegelhalter, Nicola G. Best, Bradley P. Carlin, Angelika van der Linde

First published: 08 April 2014 | https://doi.org/10.1111/rssb.12062 | Citations: 205

Model selection criteria

- Question: if there are several possible models, which one is 'better'?
  - for example for nested models in generalised linear models, one could use deviance (likelihood ratio)
  - an alternative for non-nested models AIC Akaike information criterion $AIC = 2k - 2ln(\hat{L})$ with k the number of parameters and $ln(\hat{L})$ the maximum likelihood estimate.
  - AIC does not work in models with ninformative prior information
  - related BIC Bayesian information criterion $BIC = kln(n) - 2ln(\hat{L})$ with n the number of data points.
  - DIC: deviance information criterion $D(\theta) = -2log(p(y|\theta)) + C$

with θ unknown parameters, $p(y|\theta)$the likelihood and C a constant - WAIC might be an alternative?

Deviance information criterion (Spiegelhalter 2022, 2012)

- measure of 'effective number of parameters' p_D
  $DIC = D(\bar{\theta}) + 2p_D$ with $\bar{\theta}$ the expectation of θ
- some criticisms:
  - p_D is not invariant to reparametrization
  - lack of consistenc
  - not based on a proper predictive criterion
  - has a weak theorertical foundation

Ex. 7

- With the Cassaniti data set, try to obtain DIC values for model with different conditional dependencies between sensitivities.
- look at 'DIC.R' and 'runjags_version_deviance.R'
- (you might have a look at the WAIC.R file)

Covariates: Latent variable logistic regression (Lewis 2012)

- A binomial regression model with a logit link function between the latent true prevalence and covariates associated with disease occurrence can be definded as follows, for covariate pattern i $Pr(Y_i = y_i|n_i) = \binom{n_i}{k_i} q_i^{y_i}(1-q_i)^{n_i-y_i}$ where $q_i = Se\pi_i + (1-Sp)(1-\pi_i)$ and $log(\frac{\pi_i}{1-\pi_i}) = x_i^T\beta$
- When $Se = 1$ and $Sp = 1$ then the model reverts to the classical logistic regression model

Ex 8 Covariates

- Explore the data set 'echinococcus.xlsx' PCR for either E. multilocularis or E. granulosus, ELISA for both, eggs found by arecoline purgation, Tawnia co-infection, age and sex
- Run classical 'risk factor analysis': is sex, Taenia co-infection or age a risk factor for echinococcus (PCR-prevalence, seroprevalence or purges)? Obtain p-values and ORs with confidence intervals.

Ex 9 Covariates

- Prepare the data set in the correct format (dump, add ones) for BLCM
- Run a model for three tests (assume a very high sensitivity for arecoline purgation)
- Try different priors
- Evidence of conditional dependencies
- obtain DICs
- Is there evidence for a covariate effect on the prevalence?
- Compare your finding with Ex.8