Project Assignment Bayesian Inference

Christel Faes

2024-2025

- This is a group project. You can find the group composition online.
- You are assigned to one of the assignments, so check which project you have to make!
- Answer all questions of the assignment that is assigned to your group. The answer to each of the questions should include:
  - A statistical explanation of your solution (e.g. an explanation of the likelihood and prior you used, the explanation of how you derive the solution from an MCMC chain, ...)
  - The answer to the problem (e.g. table with parameter estimates, a figure, ...)
  - The interpretation of the answer (e.g. how to interpret the results, what you can conclude from the analysis, ...)
- Report in question-answer style.
- The R script should be submitted as well. The code should reproduce all the results (up to numerical differences)!
- Report and R script should be submitted online (one submission per group is sufficient).
- Due by 01/06/2025.

1 Vaccination against Varicella

The Center for Disease Control (CDC) reports the vaccination coverage of Varicella among young children. Varicella, commonly known as chickenpox, is a highly contagious viral infection caused by the varicella-zoster virus (VZV). Vaccination against chickenpox has been highly effective in reducing the incidence and severity of the disease. In the United States, vaccination against varicella has been part of the routine childhood immunization schedule since the mid-1990s. Since the vaccine's introduction, there has been a dramatic decline in the number of chickenpox cases, hospitalizations, and deaths associated with the disease. The target for vaccination coverage of varicella (chickenpox) in the United States, as set by the Centers for Disease Control and Prevention (CDC), is typically around 90% or higher for children. This high coverage rate is aimed at achieving herd immunity and preventing outbreaks of chickenpox within communities.

2 Questions

Select the project that is assigned to your group. So only one of the following projects needs to be made.

In the report, only the BUGS language of your model should be included. All other R-code should go to the R script. Make sure that in each answer you explain what you did, and what you are calculating. E.g. make sure that you explain which priors you have assumed, how to interpret a result that you give, which convergence checks you have done, ... And, do the analysis in a Bayesian way!

You are allowed to make use of BUGS language of Jags, Winbugs, Openbugs or Nimble.

2.1 Project 1: Insurance

The next table summarizes, based on a survey, the number of children in the birth cohort 2014-2017 that had at least one dose of the Varicella vaccine. It gives the number of vaccinated children (Vaccinated) amongst the number of children in the survey (Sample Size). The information is provided for 5 regions of the US, and split according to insurance status (private insurance, uninsured or any Medicaid).

Geography	Insurance	Vaccinated	Sample Size
North Carolina	Any Medicaid	380	419
North Carolina	Private Insurance Only	632	673
North Carolina	Uninsured	28	34
Georgia	Any Medicaid	363	396
Georgia	Private Insurance Only	527	576
Georgia	Uninsured	36	50
Wisconsin	Any Medicaid	282	332
Wisconsin	Private Insurance Only	514	548
Wisconsin	Uninsured	16	34
Florida	Any Medicaid	446	490
Florida	Private Insurance Only	588	628
Florida	Uninsured	28	39
Mississippi	Private Insurance Only	400	441
Mississippi	Uninsured	27	32

1. Derive analytically the posterior of the vaccination coverage per geography and insurance group. Use a conjugate prior that (1) reflects no knowledge on the vaccination coverage, and (2) reflects that vaccination coverage is typically around 90% or higher. Give posterior summary measures of the vaccination coverage per geography and insurance group. Is the choice of the prior impacting your results?

2. Investigate whether the vaccination coverage is associated with the insurance status using a logistic regression model

   $$Y_{ij} \sim Binom(\pi_{ij}, N_{ij})$$

   with

   $$\text{logit}(\pi_{ij}) = \alpha_0 + \alpha_1 I_{AnyMedicaid,ij} + \alpha_2 I_{Uninsured,ij},$$

where i is the location, j is the insurance status, π_{ij} is the vaccination coverage and $I_.$ are dummy variables. Assume non-informative priors for the parameters to be estimated. Write and explain the code in BUGS language.

3. Run the MCMC method and check convergence of the MCMC chains. Give the details on how you checked convergence.

4. Make a plot of the posterior of the model parameters and give posterior summary measures. Interpret the results.

5. Give the posterior estimate of the vaccination coverage per region and insurance status. Compare with the analytical results you obtained in Question 1.

6. Based on the logistic regression model, what is the probability (a posteriori) that coverage amongst children that have private insurance is higher than amongst children that have any medicaid? And compared to children with no insurance?

7. Secondly, investigate whether the vaccination coverages are distinct at the different locations by adding a location-specific intercept.

$$\text{logit}(\pi_{ij}) = \alpha_{0i} + \alpha_1 I_{AnyMedicaid} + \alpha_2 I_{Uninsured},$$

   Assume non-informative priors for the parameters to be estimated. Write the code in BUGS language. Give a brief summary of the convergence checks you performed. Compare posteriors of vaccination coverages with results from Question 1.

8. Compare the vaccination coverage in each of the location with the vaccination coverage in North Carolina:

$$\theta_{ij} = \frac{\pi_{ij}}{\pi_{NorthCarolina,j}}.$$

   Interpret the results.

9. Make a caterpillar plot of the estimated coverage (per location and insurance status). Include also the observed vaccination proportion in the plot.

10. No data is given for children insured by Any medicaid in Mississippi. Predict the number of vaccinated individuals in Mississippi among 519 children with any medicaid.

2.2 Project 2: Birth year

The next table summarizes, based on a survey, the number of children that had at least one dose of the Varicella vaccine. It gives the number of vaccinated children (Vaccinated) among the number of children in the survey (Sample Size). The information is provided for 3 regions of the US, and split according to birth cohort (2011-2020).

1. Derive analytically the posterior of the vaccination coverage per birth year and region. Use a conjugate prior that (1) reflects no knowledge on the vaccination coverage, and (2) reflects that vaccination coverage is typically around 90% or higher. Give posterior summary measures of the vaccination coverage per birth year and region. Is the choice of the prior impacting your results?

2. Investigate whether there is a change in the vaccination coverage over the birth years 2011-2019 using a logistic regression model

$$Y_{ij} \sim Binom(\pi_{ij}, N_{ij})$$

   with

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 * BirthYear_j,$$

   where i is the location, j is birth cohort and π_{ij} is the vaccination coverage. Assume non-informative priors for the parameters to be estimated. Write and explain the code in BUGS language.

3. Run the MCMC method and check convergence of the MCMC chains. Give the details on how you checked convergence.

4. Make a plot of the posterior densities and give summary measures of the posterior distributions of the model parameters. Interpret the results.

5. Give the posterior estimate of the vaccination coverage per birth year. Compare with the analytical results you obtained in Question 1.

6. Secondly, investigate whether the vaccination coverage trends are distinct at the different locations by adding a location-specific intercept and slope:

$$\text{logit}(\pi_{ij}) = \beta_{0i} + \beta_{1i} * BirthYear_j.$$

Use data from the years 2011-2019. Assume non-informative priors for the parameters to be estimated. Write the code in BUGS language. Give a brief summary of the convergence checks you performed. Give the posterior estimates of this model.

7. What is the probability (a posteriori) that there is an increase in vaccination coverage (per location)?

8. Make a plot of the estimated vaccination coverage (per location and birth year), including the uncertainty on the estimates. Include also the observed vaccination proportion in the plot.

9. Investigate whether the observed number of vaccinated children in 2020 is in line with the expectations from earlier years. For this, compare the observed number of vaccinated children in 2020 with the prediction intervals for number of vaccinated children in 2020.

10. Make pairwise comparisons of the vaccination coverage in 2019 by estimating the ratio of the vaccination coverage in 2019 in two locations. Interpret the results.

Geography	Birth Year	Vaccinated	Sample Size
Georgia	2011	196	219
Georgia	2012	248	270
Georgia	2013	261	276
Georgia	2014	252	284
Georgia	2015	276	306
Georgia	2016	311	334
Georgia	2017	265	292
Georgia	2018	246	282
Georgia	2019	251	273
Georgia	2020	165	188
Wisconsin	2011	207	225
Wisconsin	2012	205	226
Wisconsin	2013	212	235
Wisconsin	2014	195	224
Wisconsin	2015	231	262
Wisconsin	2016	246	275
Wisconsin	2017	215	238
Wisconsin	2018	214	241
Wisconsin	2019	197	224
Wisconsin	2020	156	177
Mississippi	2011	171	198
Mississippi	2012	208	230
Mississippi	2013	190	217
Mississippi	2014	215	239
Mississippi	2015	243	272
Mississippi	2016	276	307
Mississippi	2017	290	321
Mississippi	2018	242	276
Mississippi	2019	304	324
Mississippi	2020	161	181

2.3 Project 3: Age

The next table summarizes, based on a survey, the number of young children that had at least one dose of the Varicella vaccine in 2020. It gives the number of vaccinated children (Vaccinated) among the number of children in the survey (Sample Size). The information is provided for 5 regions of the US, and split according to age (in months).

Geography	Age	Vaccinated	Sample Size
Mississippi	13 Months	123	181
Mississippi	19 Months	153	181
Mississippi	24 Months	161	181
Mississippi	35 Months	161	181
Florida	13 Months	143	212
Florida	19 Months	188	212
Florida	24 Months	195	212
Florida	35 Months	198	212
North Carolina	13 Months	201	280
North Carolina	19 Months	250	280
North Carolina	24 Months	246	280
North Carolina	35 Months	255	280
Georgia	13 Months	114	188
Georgia	19 Months	164	188
Georgia	24 Months	165	188
Georgia	35 Months	174	188
Wisconsin	13 Months	102	177
Wisconsin	19 Months	152	177
Wisconsin	24 Months	156	177
Wisconsin	35 Months	158	177

1. Derive analytically the posterior of the vaccination coverage per age and region. Use a conjugate prior that (1) reflects no knowledge on the vaccination coverage, and (2) reflects that vaccination coverage is typically around 90% or higher. Give posterior summary measures of the vaccination coverage per age and region. Is the choice of the prior impacting your results?

2. Investigate the trend in the vaccination coverage with age using the

logistic regression model

$$Y_{ij} \sim Binom(\pi_{ij}, N_{ij})$$

with

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 * Age_j,$$

where i is the location, j is age and π_{ij} is the vaccination coverage. Assume non-informative priors for the parameters to be estimated. Write and explain the code in BUGS language.

3. Run the MCMC method and check convergence of the MCMC chains. Give the details on how you checked convergence.

4. Make a plot of the posterior densities and give summary measures of the posterior distributions of the model parameters. Interpret the results.

5. The target vaccination coverage is 90%. Based on the logistic regression model, calculate, for each age group, the posterior probability that the target is reached (i.e. the posterior probability that the vaccination coverage is equal to or above 90%)?

6. Give the posterior estimate of the vaccination coverage per age. Compare with the analytical results you obtained in Question 1.

7. Compare the vaccination coverage at each age group with the vaccination coverage at 13 months of age. Do this by calculating the ratio of the vaccination coverages at the different age levels with the vaccination coverage at 13 months of age. Interpret the results.

8. Secondly, instead of modeling the number of vaccinated individuals using a binomial model, model the observed vaccination coverage $r_{ij} = Y_{ij}/N_{ij}$ directly using a non-linear normal regression model (exponential growth model):

$$r_{ij} \sim N(\mu_{ij}, \sigma^2)$$

with

$$\mu_{ij} = \alpha - (\alpha - \beta)e^{-k*Age_j}$$

In this model, α is the upper asymptote, β is the lower asymptote and k is the growth rate. Assume non-informative priors for the parameters to be estimated. Write the code in BUGS language. Check convergence of the MCMC chains. (In this case, give a summary of the results of your checks of convergence).

9. Make a plot of the estimated vaccination coverage as a function of age, including the uncertainty on the estimates. Include also the observed vaccination proportion in the plot.

10. Predict the number of vaccinated children at the age of 15 months (per location).

2.4 Project 4: Poverty

The next table summarizes, based on a survey, the number of children from the birth cohort 2016-2019 that had at least one dose of the Varicella vaccine. It gives the number of vaccinated children (Vaccinated) amongst the number of children in the survey (Sample Size). The information is provided for 9 regions of the US, and split according to poverty status. Poverty is measured based on the households income, and whether or not it falls within a specific range relative to the Federal Poverty Level (FPL). For example, $< 133\%$ FPL means that the household's income is less than 133% of the income threshold set by the federal government for poverty.

1. Derive analytically the posterior of the vaccination coverage per poverty level. Use a conjugate prior that (1) reflects no knowledge on the vaccination coverage, and (2) reflects that vaccination coverage is typically around 90% or higher. Give posterior summary measures of the vaccination coverage per age and region. Is the choice of the prior impacting your results?

2. Investigate the differences in vaccination coverage among poverty groups using the logistic regression model

$$Y_{ij} \sim Binom(\pi_{ij}, N_{ij})$$

with

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 * I_{133-400FPL,ij} + \beta_2 * I_{>400PFL,ij}$$

where i is the location, j is poverty level, π_{ij} is the vaccination coverage and $I_{.}$ are dummy variables. Assume non-informative priors for the parameters to be estimated. Write and explain the code in BUGS language.

3. Run the MCMC method and check convergence of the MCMC chains. Give the details on how you checked convergence.

4. Make a plot of the posterior densities and give summary measures of the posterior distributions of the parameters of interest. Interpret the results.

5. The target vaccination coverage is 90%. Calculate, for each poverty group, the posterior probability that the target is reached (i.e. the

posterior probability that the vaccination coverage is equal to or above 90%)?

6. Give the posterior estimate of the vaccination coverage per poverty level. Compare with the analytical results you obtained in Question 1.

7. Calculate how much the vaccination coverage is reduced due to poverty, by calculation the ratio of the vaccination coverages (with as reference > 400 PFL). Interpret the results.

8. Investigate whether the vaccination coverages are distinct at the different locations by adding a location-specific intercept.

$$\text{logit}(\pi_{ij}) = \beta_{0i} + \beta_1 * I_{133-400FPL} + \beta_2 * I_{>400PFL}$$

   Assume non-informative priors for the parameters to be estimated. Write the code in BUGS language. Check convergence of the MCMC chains. (In this case, give a summary of the results of your checks of convergence).

9. Make a caterpillar plot of the estimated vaccination coverage (per location and poverty level). Include also the observed vaccination proportion in the plot.

10. Predict the total number of vaccinated children amongst 300 children in Missisipi, of which 100 children are in the poverty group $< 133\%$ FPL, 100 in poverty group $133-400$ FPL and 100 in the poverty group > 400 PFL.

Geography	Poverty	Vaccinated	Sample Size
Pennsylvania	<133% FPL	539	584
Pennsylvania	133% to <400% FPL	633	684
Pennsylvania	>400% FPL	918	943
Missouri	<133% FPL	276	335
Missouri	133% to <400% FPL	503	550
Missouri	>400% FPL	371	401
Washington	<133% FPL	254	296
Washington	133% to <400% FPL	519	578
Washington	>400% FPL	612	670
Oklahoma	<133% FPL	258	312
Oklahoma	133% to <400% FPL	313	354
Oklahoma	>400% FPL	192	205
Mississippi	<133% FPL	314	354
Mississippi	133% to <400% FPL	436	485
Mississippi	>400% FPL	259	281
Florida	<133% FPL	298	335
Florida	133% to <400% FPL	463	490
Florida	>400% FPL	380	404
North Carolina	<133% FPL	229	256
North Carolina	133% to <400% FPL	394	427
North Carolina	>400% FPL	393	403
Georgia	<133% FPL	241	281
Georgia	133% to <400% FPL	376	419
Georgia	>400% FPL	357	367
Wisconsin	<133% FPL	160	193
Wisconsin	133% to <400% FPL	321	373
Wisconsin	>400% FPL	300	317