# Model checking, hierarchical modeling and combined M-H and Gibbs

SISMID/July 18–20, 2022

Instructors: Vladimir Minin, Kari Auranen, Elizabeth Halloran

# Outline

- ▶ Chain binomial model for household outbreaks of measles
    - ▶ Bayesian analysis of incompletely observed data, using data augmentation (cf. the earlier lecture and computer lab)
    - ▶ Checking the model fit through comparison of predictive data with the observed data of the final number infected
- ▶ Model extension by allowing heterogeneity across households → a hierarchical model
- ▶ Implementation of posterior sampling in the hierarchical model through a combined Gibbs and Metropolis algorithm

# The observed outbreak sizes

Recall the observed data in the chain binomial model:

| Chain | Chain probability | Frequency | Observed frequency | Final number infected |
|-------|-------------------|-----------|--------------------|-----------------------|
| 1 | $q_j^2$ | $n_1$ | 34 | 1 |
| $1 \to 1$ | $2q_j^2 p_j$ | $n_{11}$ | 25 | 2 |
| $1 \to 1 \to 1$ | $2q_j p_j^2$ | $n_{111}$ | not observed | 3 |
| $1 \to 2$ | $p_j^2$ | $n_{12}$ | not observed | 3 |
| Total | 1 | $N$ | 334 | |

- ▶ If the final number infected is 1 or 2, the actual chain is observed
- ▶ If the final number infected is 3, the actual chain data are not observed
    - ▶ We still know that $N_3 \equiv n_{111} + n_{12} = 275$
- ▶ In the previous analysis, we assumed $q_j = q$ for $j = 1, \ldots, 334$, i.e., for all 334 households

# Prediction

- Recall that new (predictive) data $y^{pred}$ can be generated by drawing from their posterior predictive distribution $f(y^{pred}|y)$
- Posterior predictive distribution because
    - conditioning on the observed data $y$
    - predicting a future observable $y^{pred}$
- Predictive data can be compared with the observed data to assess the fit of the model
- In this example, we compare the predictive and observed frequencies of chains $1$ and $1 \rightarrow 1$

# Posterior predictive distribution

▶ Denote the model parameters by $\theta$. Then

$$f(y^{\text{pred}}|y) \quad = \int f(y^{\text{pred}}, \theta|y)d\theta = \int f(y^{\text{pred}}|\theta, y)f(\theta|y)d\theta$$

$$= \int f(y^{\text{pred}}|\theta)f(\theta|y)d\theta$$

▶ This means that samples from the posterior predictive distribution can be realised as follows:

[1 ] Draw an MCMC sample $\theta_k$ from the posterior $f(\theta|y)$ of the model parameters

[2 ] Given $\theta_k$, draw a sample $y_k^{\text{pred}}$ from $f(y^{\text{pred}}|\theta_k)$

[3 ] Repeat steps [1] and [2] $K$ times ($k = 1, \ldots, K$)

# Model checking

▶ The posterior predictive distribution of the chain frequencies $(n_1, n_{11}, n_{111}, n_{12})$ is now

$$f(n_1^{\text{pred}}, n_{11}^{\text{pred}}, n_{111}^{\text{pred}}, n_{12}^{\text{pred}}|n_1, n_{11}, N_3)$$
$$= \int_0^1 f(n_1^{\text{pred}}, n_{11}^{\text{pred}}, n_{111}^{\text{pred}}, n_{12}^{\text{pred}}|q)f(q|n_1, n_{11}, N_3)dq$$

▶ Samples from the posterior predictive distribution:
   [1 ] Draw an MCMC sample $q^{(k)}$ from the posterior $f(q|n_1, n_{11}, N_3)$
   [2 ] Draw a sample $(n_1^{(k)}, n_{11}^{(k)}, n_{111}^{(k)}, n_{12}^{(k)})$ from Multinomial(334,$(q^{(k)}, 2(q^{(k)})^2 p^{(k)}, 2q^{(k)}(p^{(k)})^2, p^{(k)})$)
   [3 ] Repeat steps [1] and [2] $K$ times ($k = 1, \ldots, K$)

# Model checking continues

- Comparison of a sample from the joint predictive posterior of $(n_1^{pred}, n_{11}^{pred})$ with the actually observed point (34,25) reveals a poor model fit (next page)

  - The prediction shows too few chains of size 1 and too many chains $1 \rightarrow 1$, meaning that estimated escape probability appears to be too small

- The model did not take into account possible heterogeneity across households in the escape probability

- Therefore, we'll next consider a model extension through allowing such heterogeneity

# Model checking continues
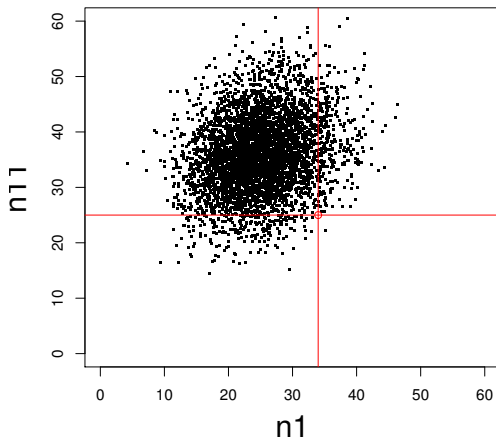
# A hierarchical model

- In household $j$, let $(n_1^{(j)}, n_{11}^{(j)}, n_{111}^{(j)}, n_{12}^{(j)})$ follow a multinomial distribution with size 1 and probability vector $(q_j^2, 2q_j^2 p_j, 2q_j p_j^2, p_j^2)$, $j = 1, \ldots, 334$

- Assume that the *household-specific* escape probabilities $q_j$ follow a Beta$(\alpha, \beta)$ distribution

- Assuming an uninformative joint prior* for $\alpha$ and $\beta$, the hierarchical model becomes fully defined:

$$
\begin{aligned}
(n_1^{(j)}, n_{11}^{(j)}, n_{111}^{(j)}, n_{12}^{(j)})|q_j &\sim \text{Multinomial}(1, (q_j^2, 2q_j^2 p_j, 2q_j p_j^2, p_j^2)) \\
q_j|\alpha, \beta &\sim \text{Beta}(\alpha, \beta) \\
(\alpha, \beta) &\propto (\alpha + \beta)^{-5/2}
\end{aligned}
$$

* The odd-looking joint prior distribution for $\alpha$ and $\beta$ follows from assuming independent uniform prior distibutions for $\alpha/(\alpha + \beta)$ (expectation of the Beta distribution) and $1/(\alpha + \beta)^{-1/2}$: see Chapter 5.3, Gelman et al.

# A hierarchical model continues
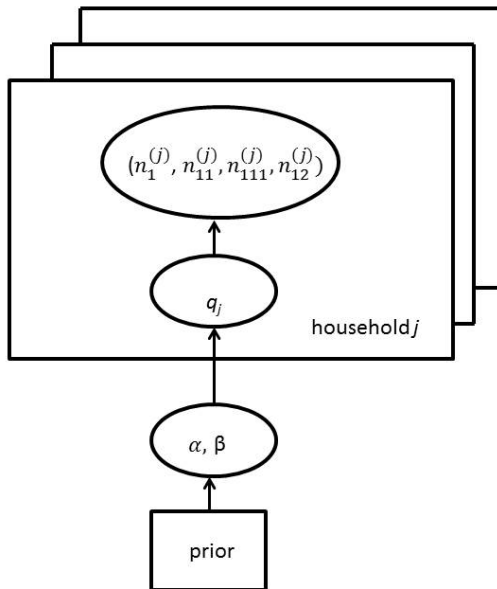
# The joint distribution

- The joint distribution of the parameters $\alpha$ and $\beta$, the household-specific escape probabilities $q_j$ $(j = 1, \ldots, 334)$ and the chain frequencies is

$$\prod_{j=1}^{334} \left( f(n_1^{(j)}, n_{11}^{(j)}, n_{111}^{(j)}, n_{12}^{(j)} | q_j) f(q_j | \alpha, \beta) \right) f(\alpha, \beta),$$

- The model unknowns are parameters $\alpha$ and $\beta$, frequencies (i.e. indicators) $n_{111}^{(j)}$ for all 275 households with outbreak size 3, as well as all 334 household-specific escape probabilities $q_j$

- This means that there are altogether 611 unknown quantities to be estimated in this hierarchical model

# Sampling from the posterior cont.

- In each household, the full conditional (Beta) distribution of $q_j^{(k)}$ depends on the current iterates of the numbers of escapes ($e_j^{(k-1)}$) and infections ($d_j^{(k-1)}$) *in that household* and the prior parameters $\alpha^{(k-1)}$ and $\beta^{(k-1)}$

- So, $q_j^{(k)} \sim \text{Beta}(e_j^{(k-1)} + \alpha^{(k-1)}, d_j^{(k-1)} + \beta^{(k-1)})$

- The numbers of escapes and infections are given in the table below

| Chain | Number of escapes $e_j^{(k-1)}$ | Number of infections $d_j^{(k-1)}$ |
|---|---|---|
| 1 | 2 | 0 |
| $1 \to 1$ | 2 | 1 |
| $1 \to 1 \to 1$ | $1 = n_{111}^{(j,k-1)}$ | 2 |
| $1 \to 2$ | $0 = n_{111}^{(j,k-1)}$ | 2 |

- Note that the unknown number of escapes in a household with outbreak size 3 (i.e. with either chain $1 \to 1 \to 1$ or $1 \to 2$) is equal to $n_{111}^{(j,k-1)}$; this is useful in implementing the sampling routine (see the computer lab)

## Sampling from the posterior

▶ A sketch of the steps in the $k$th iteration of the sampling algorithm:

$$q_j^{(k)}|\alpha^{(k-1)}, \beta^{(k-1)}, \overbrace{n_1^{(j)}}^{=1} \sim \text{Beta}(2 + \alpha^{(k-1)}, \beta^{(k-1)}), j=1,\dots, 34$$

$$q_j^{(k)}|\alpha^{(k-1)}, \beta^{(k-1)}, \overbrace{n_{11}^{(j)}}^{=1} \sim \text{Beta}(2 + \alpha^{(k-1)}, 1 + \beta^{(k-1)}), \ j=35,\dots, 59$$

$$q_j^{(k)}|\alpha^{(k-1)}, \beta^{(k-1)}, n_{111}^{(j,k-1)} \sim \text{Beta}(n_{111}^{(j,k-1)} + \alpha^{(k-1)}, 2 + \beta^{(k-1)}),$$
$$j=60,\dots, 334$$

$$n_{111}^{(j,k)}|q_j^{(k)} \sim \text{Binom}(1, 2q_j^{(k)}/(2q_j^{(k)} + 1)), j=60,\dots,334$$

$\alpha^{(k)}|\beta^{(k-1)}, q_1^{(k)}, \dots, q_{334}^{(k)}$ using a Metropolis-Hastings step

$\beta^{(k)}|\alpha^{(k)}, q_1^{(k)}, \dots, q_{334}^{(k)}$ using a Metropolis-Hastings step

# Sampling from the posterior

- ▶ Parameters $\alpha$ and $\beta$ require a Metropolis-Hastings step
- ▶ For example, for parameter $\alpha$, if the current iterate is $\alpha^{(k-1)}$, a new value $\bar{\alpha}$ is first proposed (e.g.) uniformly about the current iterate (this is a symmetric proposal)
- ▶ The proposal is then accepted, i.e., $\alpha^{(k)} := \bar{\alpha}$, with probability

$$
\min\{1, \frac{\displaystyle\prod_{j=1}^{334} f(q_j^{(k)}|\bar{\alpha}, \beta^{(k-1)})f(\bar{\alpha}, \beta^{(k-1)})}{\displaystyle\prod_{j=1}^{334} f(q_j^{(k)}|\alpha^{(k-1)}, \beta^{(k-1)})f(\alpha^{(k-1)}, \beta^{(k-1)})}\}
$$

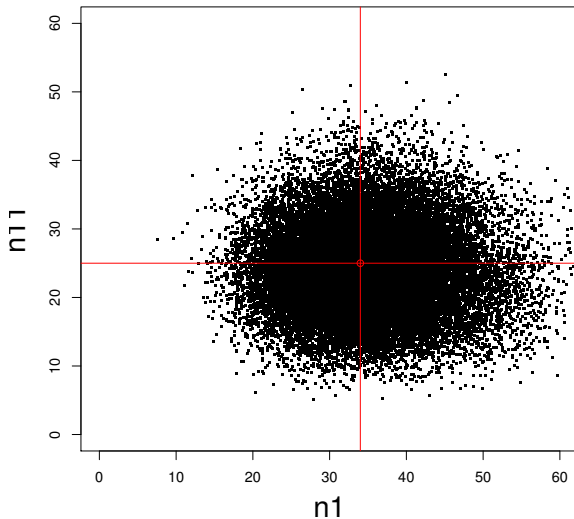# Posterior distributions of $\alpha$ and $\beta$
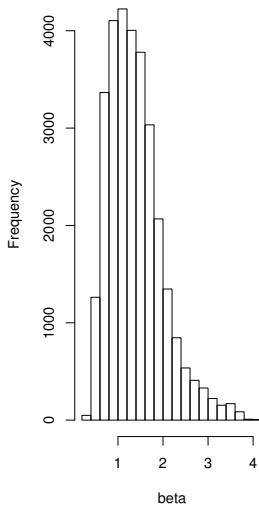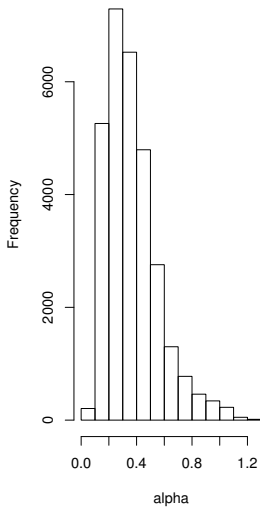
# Predictive distribution of the household-specific escape probability

► Posterior predictive distribution of the household-specific escape probability, $\int f(q_{335}|\alpha, \beta) f(\alpha, \beta|n_1, n_{11}, N_3) d\alpha d\beta$



posterior predictive distribution of the escape probability
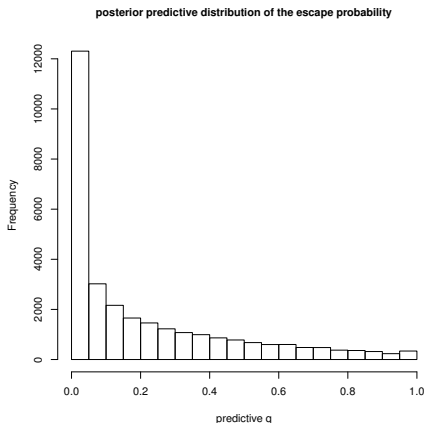
# Checking the hierarchical model

# An alternative approach

- In the model of this lecture, it is actually possible to marginalise $q_j$ over its prior distribution (see [3]), bypassing the need to build the computations according to the hierarhical model set-up

- This means calculating the chain probabilities as expectations of the probabilities given in the Table on page 3 with respect to $\text{Beta}(\tilde{q}/w, (1 - \tilde{q})/w)$

  - Note that the Beta distribution is now parameterised differently: $\tilde{q} = \alpha/(\alpha + \beta)$, $w = 1/(\alpha + \beta)$

| Chain | Chain probability | Frequency | Observed frequency | Final number infected |
|---|---|---|---|---|
| 1 | $\tilde{q}(\tilde{q} + w)/(1 + w)$ | $n_1$ | 34 | 1 |
| $1 \rightarrow 1$ | $2\tilde{p}\tilde{q}(\tilde{q} + w)/((1 + w)(1 + 2w))$ | $n_{11}$ | 25 | 2 |
| $1 \rightarrow 1 \rightarrow 1$ | $2\tilde{p}\tilde{q}(\tilde{p} + w)/((1 + w)(1 + 2w))$ | $n_{111}$ | missing | 3 |
| $1 \rightarrow 2$ | $\tilde{p}(\tilde{p} + w)/(1 + w)$ | $n_{12}$ | missing | 3 |

▶ The following identity helps to calculate the expectations:

$$\mathsf{E}(p_j^u q_j^v) = \frac{\tilde{q}(\tilde{q} + w)\dots(\tilde{q} + w(u-1))\tilde{p}(\tilde{p} + w)\dots(\tilde{p} + w(v-1))}{(1+w)\dots(1+w(u+v-1))}$$

▶ Using the probabilities as given in the table, it is straightforward to implement a Metropolis-Hastings algorithm to draw samples from the posterior of parameters $\tilde{q}$ and $w$

▶ A suitable prior for $\tilde{q}$ and $w$ is of course needed

# References

[1] Bailey T.J.N. The Mathematical Theory of Infectious Diseases. Charles Griffiths and Company, London 1975.

[2] O'Neill Ph. and Roberts G. Bayesian inference for partially observed stochastic processes. Journal of the Royal Statistical Society, Series A, 1999; 162: 121–129.

[3] Becker N. Analysis of infectious disease data. Chapman and Hall, New York 1989.

[4] O'Neill Ph. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. Mathematical Biosciences 2002; 180:103-114.

[5] Gelman, Carlin, Stern, Rubin: Bayesian Data Analysis, Chapman and Hall, 2004; see Chapter 5.3.