The EZK Report

Net Zhang

4/21/2021

Background & Definitions

The World Health Organization sponsored a small clinical trial run in 10 countries where the K9C9 virus in endemic. This project aims to propose a Bayesian hierarchical model for K9C9 status basing on the EZK test results. The model can later be applied to examine the diagnostic ability of the EZK test.

At each of the N^c countries, a highly accurate and expensive diagnostic test was given by $N^{total} = 100$ randomly selected participants. Let Y_{ic} be the indicator variable of whether the i^{th} participant at the c^{th} country is being infected by K9C9, for $i = 1, ..., N^{total}$. Meanwhile, let x_{ic} be the indicator variable that the i^{th} person's test result, where $x_{ic} = 1$ indicate that the individual is tested positive for K9C9.

$$Y = \{Y_{ic} : i = 1, ..., N^{total}, c = 1, ..., N^C\}$$

$$X = \{x_{ic} : i = 1, ..., N^{total}, c = 1, ..., N^C\}$$

First, we assume that the likelihood can be written as below:

$$p(y|\alpha, \beta) = \prod_{c=1}^{N^c} \prod_{i=1}^{N^{total}} p(y_{ic}|\alpha_c, \beta_c),$$

where for $i = 1, ..., N^{total}, c = 1, ..., N^C$

$$Y_{ic}|\alpha_c, \beta_c \sim Bern(\theta_{ic})$$

and

$$logit(\theta_{ic}) = log(\frac{\theta_{ic}}{1 - \theta_{ic}}) = \alpha_c + \beta_c x_{ic}$$

For $\alpha = (\alpha_1, ..., \alpha_{N^c})$ and $\beta = (\beta_1, ..., \beta_{N^c})$, we assume that:

$$p(\alpha, \beta|\mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2) = \prod_{c=1}^{N^c} p(\alpha_c|\mu_\alpha, \sigma_\alpha^2) p(\beta_c|\mu_\beta, \sigma_\beta^2)$$

where for all $c = 1, ..., N^c$,

$$\alpha_c|\mu_\alpha, \sigma_\alpha^2 \sim N(\mu_\alpha, \sigma_\alpha^2)$$

and

$$\beta_c|\mu_\beta, \sigma_\beta^2 \sim N(\mu_\beta, \sigma_\beta^2)$$

and lastly, we assume the parameters below are independent, that is to say,

$$p(\mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2) = p(\mu_\alpha)p(\sigma_\alpha^2)p(\mu_\beta)p(\sigma_\beta^2)$$

where the priors are specified as below:

$$\mu_\alpha \sim N(0, 100), \sigma_\alpha \sim Unif(0, 3)$$

$$\mu_\beta \sim N(0, 100), \sigma_\beta \sim Unif(0, 3)$$

Modeling

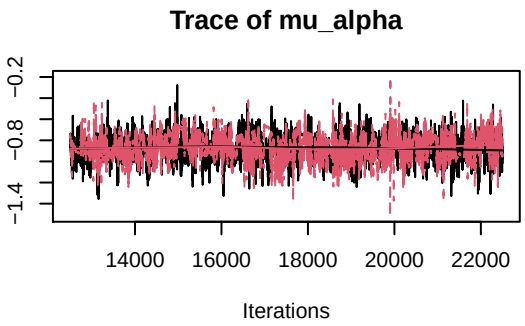
- **Initializing Values:**

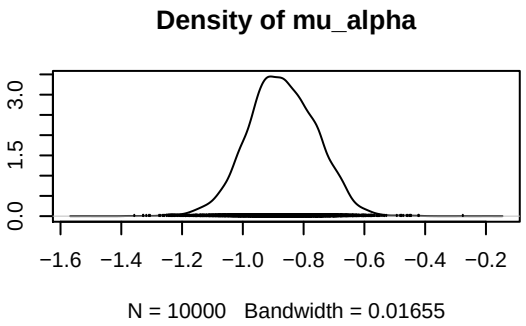
$$\mu_\alpha = 0, \sigma_\alpha = 1$$

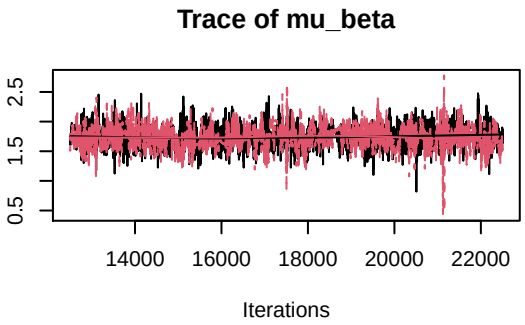
$$\mu_\beta = 0, \sigma_\beta = 1$$

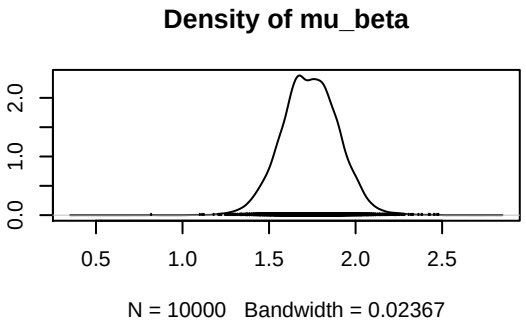
- **Iterations:**
  - number of total iteration = 10000
  - number of burn-in steps = 2500
  - number of adaptation steps = 2500
- **Convergence:**

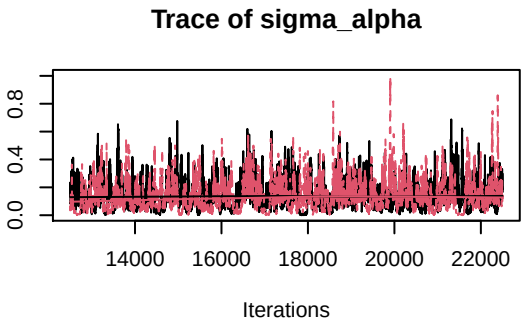
Here, since there are too many parameters, we only show the tracing plots for the first layer parameters that don't depend on other parameters. From the tracing plot, both two chains appear to be sampling from the same distribution. Therefore we have clear evidence that our algorithm has reached convergence.

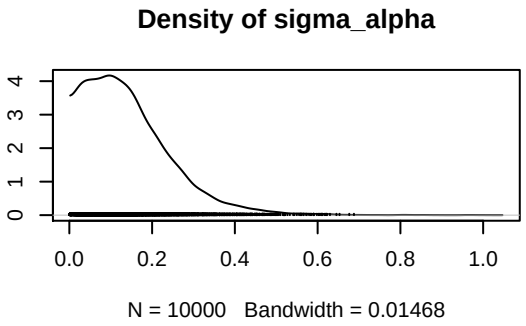


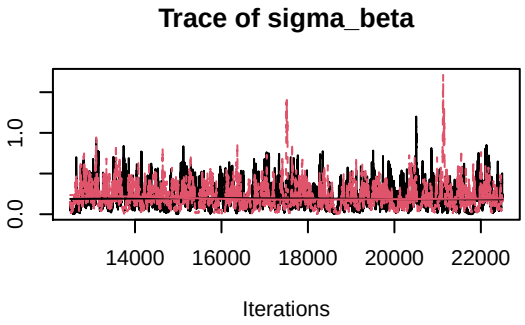


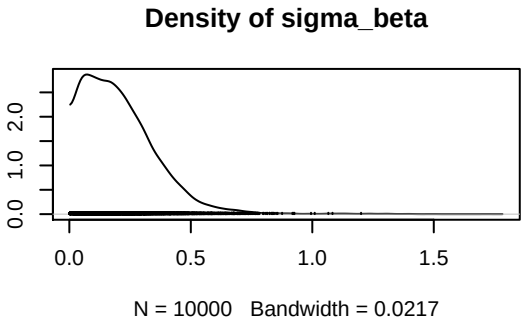












Results & Interpretation

Parameters Interpretation

- α_c:

α_c is the estimated **logit probability** of an individual being diagnosed with K9C9 while the EZK test came

out negative at country c. If the EZK test came out negative, then $expit(\alpha_c)$ is the probability that the person actually has the K9C9 virus.

- β_c

β_c could be interpreted as the **log-odds ratio** at the c country. At country c, a participant who has been tested positive for the K9C9 is expected to be $expit(\beta_c)$ times more (or less) likely to carry the K9C9 virus than the participant who has been tested negative for the virus.

- θ_{ic}

θ_{ic} is the expected probability for i^{th} participant at c country getting infected with the K9C9 virus.

- μ_α

μ_α is the mean of the **logit** infected probability with negative test EZK testing result across all ten countries involved in the trial.

- σ_α

σ_α captures the standard deviation of the **logit** infected probability with negative test EZK testing result across all ten countries involved in the trial.

- μ_β

μ_β is the mean of the **log odds ratio** of being infected with K9C9 virus across all of the ten countries participated in the trial.

- σ_β

σ_β captures the standard deviation of the **log odds ratio** of being infected with K9C9 virus across all of the ten countries participated in the trial.

Model Results Analysis

High-Level Analysis

First, we are interested in the accuracy of the EZK test under this expensive trial. Therefore, using our posterior sample we estimated how likely is an individual actually carried the K9C9 virus while the testing result came out negative:

$$E(\mu_\alpha|Y) = -0.8663$$
$$P(Y = 1|EZK = 0) = expit(E(\mu_\alpha|Y)) = 0.296$$

Hence as we can see, the estimated rate is nearly 0.30, which means that among all of the people who tested negative in the EZK test around the world, about 30% of the participants are actually carrying the K9C9 virus. This of course varies by countries, and with our estimation on σ_α, the 95% credible interval for μ_α is $(0.25, 0.35)$.

Next, we want to know what's the probability of people who don't really carry the virus got tested with positive result:

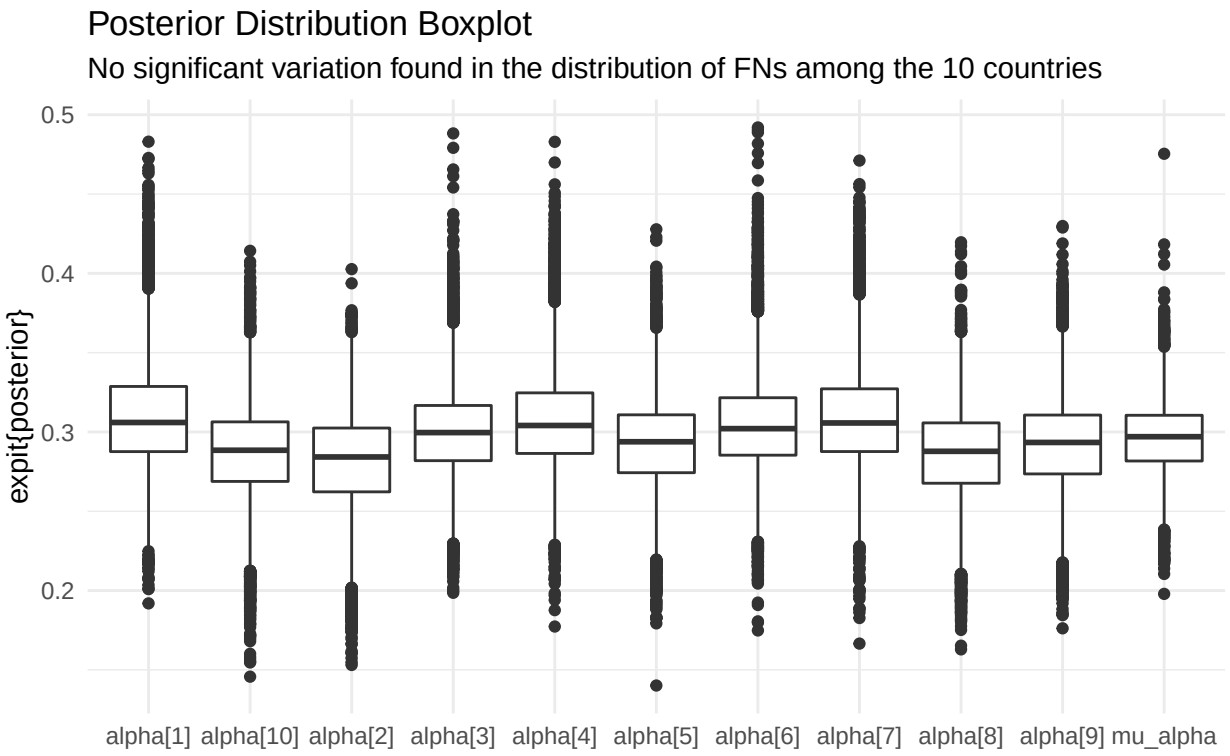
$$P(Y = 1|\hat{E}ZK = 1) = expit\{E(\mu_\alpha|Y) + E(\mu_\beta|Y)\} = 0.7034$$

$$P(Y = 0|\hat{E}ZK = 1) = 1 - P(Y = 1|\hat{E}ZK = 1) = 0.2966$$

Hence, generally speaking, the estimated probability of an individual with no K9C9 virus getting tested positive in EZK is about 0.30.

Country Level Analysis

In this section, we want to explore the distribution of parameters between different countries. From the plot below, we haven't found any significant variation in the distribution of α_c in each countries. Based on the posterior sample from α_2, we conclude that the estimated probability of being infected while tested negative is the lowest at Country B.



Conclusion

To conclude, the EZK test did a decent job achieving the overall false-negative rate at around 30%. The test performance is slightly different across counties, but this won't lead to any conclusion that some countries have hard-to-detect virus variants. The false-positive and false-negative rates are similar under this testing trial. Still, we think to mitigate the spread of the K9C9 virus and avoid a potential global pandemic, there's an urgent need to improve the performance of the EZK test to bring down the false-negative rate even if we have to make some sacrifice in bringing up the false positives.

Appendix

rjags model scripts:

```
model {
  # Prior ================================================================
  mu_alpha ~ dnorm(0, 1/100)
  sigma_alpha ~ dunif(0, 3)
  mu_beta ~ dnorm(0, 1/100)
  sigma_beta ~ dunif(0, 3)

  for (i in 1:nCountry) {
    alpha[i] ~ dnorm(mu_alpha, 1/(sigma_alpha)^2)
    beta[i] ~ dnorm(mu_beta, 1/(sigma_beta)^2)
  }

  # Likelihood ===========================================================
  for (i in 1:nTotal) {
    logit_theta[i] <- alpha[country[i]] + beta[country[i]] * x[i]
    theta[i] <- exp(logit_theta[i])/(1+exp(logit_theta[i]))
    y[i] ~ dbern(theta[i])
  }
}
```