

4 Methods

In this chapter, the current state-of-the-art methods for disease outbreak detection will be outlined. Furthermore, the novel outbreak detection algorithm will be introduced, along with the theory related to generalized mixed effects models and hierarchical generalized linear models. These modeling frameworks are utilized in this master's thesis to analyze the count observations denoted as y , but more importantly they play a crucial role in assessing the unobserved random variables or random effects represented by u , which are directly employed in the detection algorithm for characterizing outbreaks.

Due to the complexity of generalized mixed effects models, obtaining closed-form solutions is generally not feasible. Therefore, an overview of the Laplace approximation technique will be provided in this chapter, which allows for approximating the likelihood function in these models. Additionally, the implementation of these models in the R programming language will be presented.

4.1 State-of-the-art outbreak detection algorithm

In this section, the Farrington method introduced by Farrington et al. (1996) and the subsequent improvements proposed by Noufaily et al. (2013) will be outlined. These methods are recognized as the current state-of-the-art for disease outbreak detection and will be used as benchmarks to evaluate the performance of the novel outbreak detection algorithm proposed in this master's thesis. Both of the aforementioned methods have been implemented in the R package called **surveillance** by Salmon, Schumacher, and Höhle (2016), which can be accessed from the Comprehensive R Archive Network (CRAN) at <https://cran.r-project.org/web/packages/surveillance/index.html>. The presentation of these methods are strongly inspired by Salmon, Schumacher, and Höhle (2016).

Both methods follow the same steps in the algorithm. The first step involves fitting an over-dispersed Poisson generalized linear model (GLM) with a log link to the reference data. In this model, the baseline count y_t corresponding to the baseline time point t is assumed to have an expected value λ_t and a variance $\phi\lambda_t$, where $\phi \geq 1$ is ensured to account for over-dispersion. The systematic component of the model includes only a linear time trend in the frequency of reports. Therefore, the systematic component can be expressed as

$$\log(\lambda_t) = \alpha + \beta t \quad (4.1)$$

The original method incorporated seasonal effects by considering counts from comparable periods in past years for threshold calculation. This approach is similar to the one used by Stroup et al. (1989). The baseline weeks, which are used as reference, are determined by two integers: b represents the number of years back, and w represents the window half-width. For a given current week x of year h , only data from weeks $x - w$ to $x + w$ of years $h - b$ to $h - 1$ are considered, resulting in a total of $n = b(2w + 1)$ baseline weeks. The default values are $b = 5$ and $w = 3$, resulting in a total of $n = 35$ baseline values.

However, Noufaily et al. (2013) demonstrated that the algorithm performs better when utilizing more historical data without disregarding seasonality. To achieve this, the author introduced a 10-level factor with a 7-week reference period and nine additional 5-week periods in each year. As a result, the systematic component of the model is modified as follows

$$\log(\lambda_t) = \alpha + \beta t + \delta_{j(t)} \quad (4.2)$$

In this equation, $j(t)$ represents the seasonal factor level corresponding to time point t . The reference week t_0 is always associated with the reference seasonal level, denoted by $j(t_0) = 0$ and $\delta_0 = 0$.

The idea of incorporating more data while preserving seasonality has been further expanded in the implementation of the method in the **surveillance** R package. The package allows the user to choose an arbitrary number of periods in each year. Consequently, the systematic component is adjusted as follows

$$\log(\lambda_t) = \alpha + \beta t + \delta_{c(t)} \quad (4.3)$$

In this equation, $c(t)$ represents the coefficients of a zero-order spline with `noPeriods` + 1 knots. It can be conveniently represented as a `noPeriods`-level factor that captures seasonality. The function $c(t)$ indicates which season or period of the year t belongs to.

Furthermore, Noufaily et al. (2013) demonstrated that it is beneficial to exclude the last 26 weeks before t_0 from the baseline calculation. This exclusion helps prevent a reduction in sensitivity when an outbreak has recently started before t_0 .

In the second step, the algorithm predicts the expected number of counts λ_{t_0} for the current time point t_0 using the fitted generalized linear model. Both methods differ in their assumptions for calculating the upper bound U_{t_0} .

The original method assumes that a transformation of the prediction error, denoted as $g(y_{t_0} - \hat{\lambda}_{t_0})$, follows a normal distribution. For example, when using the identity transformation $g(x) = x$, the assumption becomes

$$y_{t_0} - \hat{\lambda}_{t_0} \sim \mathbf{N}(0, \mathbf{V}(y_{t_0} - \hat{\lambda}_{t_0})) \quad (4.4)$$

The upper bound of the prediction interval is then calculated based on this distribution. The variance of the prediction error is given by

$$\mathbf{V}(y_{t_0} - \hat{\lambda}_{t_0}) = \mathbf{V}(\hat{y}_{t_0}) + \mathbf{V}(\hat{\lambda}_{t_0}) = \phi \lambda_{t_0} \quad (4.5)$$

Here, $\mathbf{V}(\hat{y}_{t_0})$ represents the variance of an observation, and $\mathbf{V}(\hat{\lambda}_{t_0})$ represents the variance of the estimate. The threshold, defined as the upper bound of a one-sided $(1 - \alpha) \cdot 100\%$ prediction interval, is calculated as

$$U_{t_0} = \hat{\lambda}_0 + z_{1-\alpha} \hat{V}(y_{t_0} - \hat{\lambda}_{t_0}) \quad (4.6)$$

However, this method's weakness lies in the assumption of normality itself. Therefore, an alternative assumption was presented in Noufaily et al. (2013). This approach assumes that y_{t_0} follows a negative binomial distribution, denoted as $\text{NB}(\lambda_{t_0}, \nu)$, where λ_{t_0} represents the mean and $\nu = \frac{\lambda_{t_0}}{\phi - 1}$ represents the over-dispersion parameter. In this parameterization, the expected value of y_t remains λ_t , and the variance of y_t is $\phi \lambda_t$, with $\phi > 1$. If $\phi \leq 1$, a Poisson distribution is assumed for the observed count. The threshold is defined as a quantile of the negative binomial distribution using the plug-in estimates $\hat{\lambda}_{t_0}$ and $\hat{\phi}$.

In the final step, the observed count y_{t_0} is compared to the upper bound U_{t_0} , and an alarm is raised if $y_{t_0} > U_{t_0}$. The fitting of the GLM in both methods involves three important steps.

First, the algorithm optionally performs a power transformation to correct for skewness and stabilize the variance of the data.

Next, the significance of the time trend is checked. The time trend is included in the model only if it is statistically significant at a chosen significance level, there are more than three years of reference data, and there is no over-extrapolation due to the time trend.

Finally, past outbreaks are reweighted based on their Anscombe residuals. If the Anscombe residual of a count exceeds a certain weight threshold, it is reweighted in a second fitting of the GLM. In the original method by Farrington et al. (1996), a reweighting threshold of 1 was used. However, Noufaily et al. (2013) suggests using a value of 2.56 for the weight threshold to make the reweighting procedure less drastic, as it also reduces the variance of the observations.

4.2 Novel outbreak detection algorithm

In this section, the novel algorithm for the prospective detection of disease outbreaks proposed in this master's thesis is outlined. The algorithm utilizes a generalized mixed effects model or a hierarchical generalized linear model as a modeling framework to model the count observations y and assess the unobserved random effects u . These random effects are used directly in the detection algorithm to characterize an outbreak. The theoretical foundations of these models will be further discussed in Section 4.3 and Section 4.4.

The first step involves fitting either a hierarchical Poisson Normal or Poisson Gamma model with a log link to the reference data. Here, it is possible for the user to include an arbitrary number of covariates by supplying a model formula. In order to account for structural changes in the time series, e.g. an improved and more sensitive diagnostic method or a new screening strategy at hospitals, a rolling window with width k is used to estimate the time-varying model parameters. Also, it is assumed that the count is proportional to the population size n . Hence, in terms of the canonical link the model for the fixed effects is

$$\log(\lambda_{it}) = x_{it}\beta + \log(n_{it}), \quad i = 1, \dots, m, \quad t = 1, \dots, T \quad (4.7)$$

Here x_{it} and β are p -dimensional vectors of covariates and fixed effects parameters respectively, where p denotes the number of covariates or fixed effects parameters, m denotes the number of groups, and T denotes the length of the period.

Then, in the second step, as a new observation becomes available, the algorithm infers the one-step ahead random effect u_{it_1} for each group using the obtained model estimates $\hat{\theta}_{t_0}$. The threshold U_{t_0} for detecting outbreaks is defined as a quantile of the distribution of the random effects in the second stage model. This can be either a Gaussian distribution using the plug-in estimate $\hat{\sigma}$ or a Gamma distribution using the plug-in estimate $\hat{\phi}$.

In the final step, the inferred random effect \hat{u}_{it_1} is compared to the upper bound U_{t_0} , and an alarm is raised if $\hat{u}_{it_1} > U_{t_0}$. If an outbreak is detected, the related observation y_{i1} is omitted from the parameter estimation in the future. Thus, resulting in a smaller sample size for the rolling window until that specific observation is discarded.

4.3 General mixed effects models

In this section selected theory related to generalized mixed effects models is presented. The presentation of Section 4.3 and Section 4.4 is mostly inspired by Madsen and Thyregod (2011).

The general mixed effects model can be represented by its likelihood function

$$L_M(\theta; \mathbf{y}) = \int_{\mathbb{R}^q} L(\theta; \mathbf{u}, \mathbf{y}) d\mathbf{u} \quad (4.8)$$

where \mathbf{y} is the observed random variable, θ is the model parameters to be estimated and \mathbf{U} is the q unobserved random variables. The likelihood function L is the joint likelihood of both the observed and the unobserved random variables. The likelihood function for estimating θ is the marginal likelihood L_M obtained by integrating out the unobserved random variables. In general it is difficult to solve the integral in (4.8) if the number of unobserved random variables is more than a few and hence numerical methods must be used. Thus, an outline of the Laplace approximation is included in this section.

4.3.1 Hierarchical models

It is useful to formulate the model as a hierarchical model containing a *first stage model*

$$f_{Y|u}(\mathbf{y}; \mathbf{u}, \beta) \quad (4.9)$$

which is a model for the observed random variables given the unobserved random variables, and a *second stage model*

$$f_U(\mathbf{u}; \Psi) \quad (4.10)$$

which is a model for the unobserved random variables. Here β represent the fixed effects parameters and Ψ is a model parameter. The total set of parameters is $\theta = (\beta, \Psi)$. Hence the joint likelihood is given as

$$L(\beta, \Psi; \mathbf{u}, \mathbf{y}) = f_{Y|u}(\mathbf{y}; \mathbf{u}, \beta) f_U(\mathbf{u}; \Psi) \quad (4.11)$$

To obtain the likelihood for the model parameters (β, Ψ) the unobserved random variables are integrated out. The likelihood function for estimating (β, Ψ) is as in (4.8) the marginal likelihood

$$L_M(\beta, \Psi; \mathbf{y}) = \int_{\mathbb{R}^q} L(\beta, \Psi; \mathbf{u}, \mathbf{y}) d\mathbf{u} \quad (4.12)$$

where q is the number of unobserved random variables, and β and Ψ are the parameters to be estimated.

4.3.2 Laplace Approximation

The Laplace approximation will be outlined in the following. A thorough description of the Laplace approximation in nonlinear mixed effects models is found in Wolfinger and Lin (1997).

For a given set of model parameters θ the joint log-likelihood $\ell(\theta, \mathbf{u}, \mathbf{y}) = \log (L(\theta, \mathbf{u}, \mathbf{y}))$ is approximated using a second order Taylor approximation around the optimum $\tilde{\mathbf{u}} = \hat{\mathbf{u}}_\theta$ of the log-likelihood function w.r.t. the unobserved random variables \mathbf{u} , i.e.,

$$\ell(\theta, \mathbf{u}, \mathbf{y}) \approx \ell(\theta, \tilde{\mathbf{u}}, \mathbf{y}) - \frac{1}{2}(\mathbf{u} - \tilde{\mathbf{u}})^T \mathbf{H}(\tilde{\mathbf{u}})(\mathbf{u} - \tilde{\mathbf{u}}) \quad (4.13)$$

where the first-order term of the Taylor expansion disappears since the expansion is done around the optimum $\tilde{\mathbf{u}}$ and $\mathbf{H}(\tilde{\mathbf{u}}) = -\ell''_{uu}(\theta, \mathbf{u}, \mathbf{y})|_{\mathbf{u}=\tilde{\mathbf{u}}}$ is the negative Hessian of the joint log-likelihood evaluated at $\tilde{\mathbf{u}}$.

It is readily seen that the joint log-likelihood for the hierarchical model specified in (4.9) and (4.10) is

$$\ell(\theta, \mathbf{u}, \mathbf{y}) = \ell(\beta, \Psi, \mathbf{u}, \mathbf{y}) = \log f_{Y|u}(\mathbf{y}; \mathbf{u}, \beta) + \log f_U(\mathbf{u}; \Psi) \quad (4.14)$$

which implies that the Laplace approximation becomes

$$\ell_{M,LA}(\theta, \mathbf{y}) = \log f_{Y|u}(\mathbf{y}; \tilde{\mathbf{u}}, \beta) + \log f_U(\tilde{\mathbf{u}}, \Psi) - \frac{1}{2} \log \left| \frac{\mathbf{H}(\tilde{\mathbf{u}})}{2\pi} \right| \quad (4.15)$$

4.3.3 Formulation of the generalized mixed effects model

The generalized mixed effects model utilized in this master's thesis to model the count observations \mathbf{y} and assess the random effects \mathbf{u} is presented in Theorem 4.1, along with the joint likelihood function for the first and second stage models.

Theorem 4.1 (Hierarchical Poisson Normal model). *In order to simplify the notation, the probability density functions are presented for a specific observation and hence the subscripts indicating the group and time are omitted. The conditional distribution of the count observations is assumed to be a Poisson distribution with intensities λ*

$$f_{Y|u}(y; u, \beta) = \frac{\lambda \exp(u)^y}{y!} \exp(-\lambda \exp(u)) \quad (4.16)$$

Also, it is assumed that the count is proportional to the population size n . Hence, in terms of the canonical link for the Poisson distribution the model for the fixed effects is

$$\log(\lambda_{it}) = \mathbf{x}_{it}\beta + \log(n_{it}), \quad i = 1, \dots, m, \quad t = 1, \dots, T \quad (4.17)$$

The probability density function for the distribution of the random effects is assumed to follow a zero mean Gaussian distribution, $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, I\sigma^2)$, i.e.

$$f_U(\mathbf{u}; \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad (4.18)$$

where σ is a model parameter.

Henceforth, the total set of parameters are $\theta = (\beta, \sigma)$ and the model can be formulated as a two-level hierarchical model

$$Y|u \sim \text{Pois}(\lambda \exp(u)) \quad (4.19a)$$

$$u \sim N(0, I\sigma^2) \quad (4.19b)$$

The joint likelihood for the count observations y and the random effects u becomes

$$L(\beta, \sigma; u_{it}, y_{it}) = \prod_{t=1}^T \prod_{i=1}^m \frac{(\lambda_{it} \exp(u_{it}))^{y_{it}}}{y_{it}!} \exp(-\lambda_{it} \exp(u_{it})) \prod_{t=1}^T \prod_{i=1}^m \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{u_{it}^2}{2\sigma^2}\right) \quad (4.20)$$

4.4 Hierarchical generalized linear models

In this section selected theory related to hierarchical generalized linear models is presented. The model class was initially formulated by Lee and Nelder (1996) as a natural generalization of the generalized linear models to also incorporate random effects. A starting point in hierarchical modelling is an assumption that the distribution of random effects may be modeled by an exponential dispersion family. This family of models were first introduced by Fisher and Russell (1922), and has proven to play an important role in mathematical statistics because of their simple inferential properties. The exponential dispersion family considers a family of distributions, which can be written on the form

$$f_Y(y; \theta) = c(y, \lambda) \exp(\lambda\{\theta y - \kappa(\theta)\}) \quad (4.21)$$

Here the parameter $\lambda > 0$ is called the *precision parameter*, which in some cases represents a known shape parameter as for the Gamma distribution. In other cases the precision parameter represents an over-dispersion that is not related to the mean. These distributions combine with the so-called *standard conjugate distributions* in a simple way, and lead to marginal distributions that may be expressed in a closed form suited for likelihood calculations. For an introduction to the concept of *standard conjugate distributions* and the definition of a hierarchical generalized linear model, refer to Section 6.3 and Section 6.5 of Madsen and Thyregod (2011), respectively.

In general, when the conditional distribution of $Y|u$ is a Poisson distribution, and the conjugate distribution is assumed to be a Gamma distribution with mean value 1, it follows that the distribution of Y is a negative binomial distribution.

To further motivate this choice of distribution for the second stage model, an illustrative example is presented.

Example 4.1 (Variation between observed cases of LIST). Usually, it is reasonable to assume a Poisson distribution for count data, where the expected value and variance are both equal to λ . However, it should be noted that this assumption may not always hold true when modeling count data.

Table 4.1: The distribution of cases with 1, 2, \dots , 10+ cases of LIST.

Number of LIST cases	Number of times observed	Poisson expected	Negative binomial expected
0	4	6.26	14.44
1	5	15.52	21.41
2	28	25.66	24.75
3	25	31.83	24.58
4	26	31.58	22.01
5	34	26.11	18.27
6	19	18.51	14.31
7	14	11.48	10.72
8	10	6.33	7.73
9	3	3.14	5.41
10+	12	2.34	10.46

Take, for example, the monthly cases of LIST analyzed in this master's thesis. Table 4.1 presents the distribution of these cases from 2008 to 2022, along with the expected numbers from both the Poisson and Negative binomial distributions. The distribution of cases over time is visualized in Figure 4.1.

As observed, the actual distribution has significantly heavier tails compared to the Poisson distribution. Additionally, the mean of the Poisson distribution for the monthly LIST cases is $\hat{\lambda} = \bar{y} = 4.96$, while the empirical variance is $s^2 = 10$, which is considerably larger than the mean. These findings indicate overdispersion, suggesting that a Negative binomial model would be more appropriate for modeling the data.

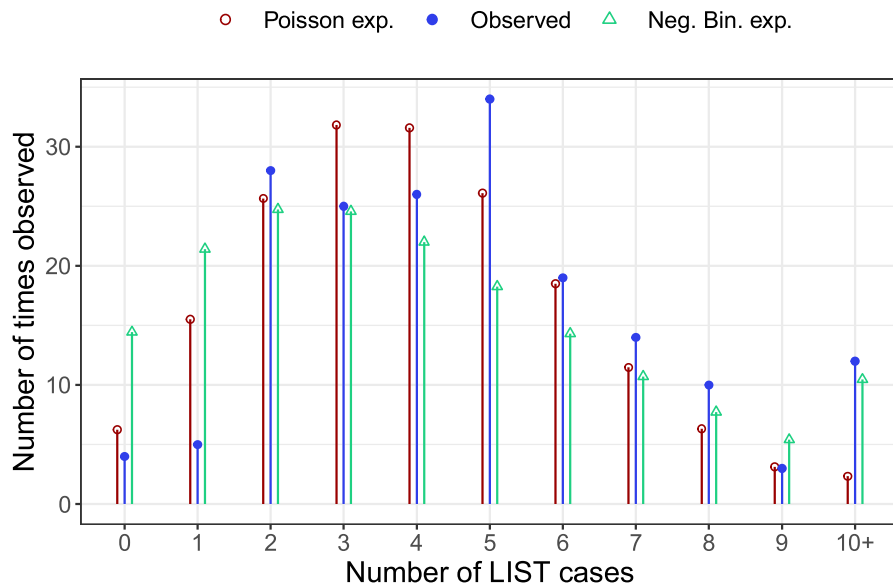


Figure 4.1: Distribution of monthly cases of LIST together with the expected numbers from both the Poisson and Negative binomial distributions

4.4.1 Formulation of the hierarchical model

The hierarchical model used in this master's thesis to model the count observations y and assess the random effects u is presented in Theorem 4.2, along with the derivation of the marginal distribution of Y and the joint likelihood function.

Theorem 4.2 (Compound Poisson Gamma model). *In the compound Poisson Gamma model the conditional distribution of the count observations are assumed to be a Poisson distribution with intensities λ*

$$f_{Y|u}(y; u, \beta) = \frac{(\lambda u)^y}{y!} \exp(-\lambda u) \quad (4.22)$$

The probability density function for the random effects u are assumed to follow a reparametrized Gamma distribution with mean 1, $u \sim G(1/\phi, \phi)$ that is

$$f_u(u; \phi) = \frac{1}{\phi \Gamma(1/\phi)} \left(\frac{u}{\phi}\right)^{1/\phi-1} \exp(-u/\phi) \quad (4.23)$$

Subsequently, the model can be formulated as a two-level hierarchical model

$$Y|u \sim \text{Pois}(\lambda u) \quad (4.24a)$$

$$u \sim G(1/\phi, \phi) \quad (4.24b)$$

Given (4.22) and (4.23), the probability function for the marginal distribution of Y is determined from

$$\begin{aligned} g_Y(y; \beta, \phi) &= \int_{u=0}^{\infty} f_{Y|u}(y; u, \beta) f_u(u; \phi) du \\ &= \int_{u=0}^{\infty} \frac{(\lambda u)^y}{y!} \exp(-\lambda u) \frac{1}{\phi \Gamma(1/\phi)} \left(\frac{u}{\phi}\right)^{1/\phi-1} \exp(-u/\phi) du \\ &= \frac{\lambda^y}{y! \Gamma(1/\phi) \phi^{1/\phi}} \int_{u=0}^{\infty} u^{y+1/\phi-1} \exp(-u(\lambda\phi + 1)/\phi) du \end{aligned} \quad (4.25)$$

In (4.25) it is noted that the integrand is the kernel in the probability density function for a Gamma distribution, $G(y + 1/\phi, \phi/(\lambda\phi + 1))$. As the integral of the density shall equal one, we find by adjusting the norming constant that

$$\int_{u=0}^{\infty} u^{y+1/\phi-1} \exp\left(-u/\left(\phi/(\lambda\phi + 1)\right)\right) du = \frac{\phi^{y+1/\phi} \Gamma(y + 1/\phi)}{(\lambda\phi + 1)^{y+1/\phi}} \quad (4.26)$$

Therefore, it can be shown that the marginal distribution of Y is a negative binomial distribution, $Y \sim \text{NB}(1/\phi, 1/(\lambda\phi + 1))$. The probability function for Y is

$$\begin{aligned}
P[Y = y] &= g_Y(y; \beta, \phi) \\
&= \frac{\lambda^y}{y! \Gamma(1/\phi) \phi^{1/\phi}} \frac{\phi^{y+1/\phi} \Gamma(y + 1/\phi)}{(\lambda\phi + 1)^{y+1/\phi}} \\
&= \frac{\Gamma(y + 1/\phi)}{\Gamma(1/\phi) y!} \frac{1}{(\lambda\phi + 1)^{1/\phi}} \left(\frac{\lambda\phi}{\lambda\phi + 1} \right)^y \\
&= \binom{y + 1/\phi - 1}{y} \frac{1}{(\lambda\phi + 1)^{1/\phi}} \left(\frac{\lambda\phi}{\lambda\phi + 1} \right)^y, \quad \text{for } y = 0, 1, 2, \dots
\end{aligned} \tag{4.27}$$

where we have used the convention

$$\binom{z}{y} = \frac{\Gamma(z + 1)}{\Gamma(z + 1 - y) y!} \tag{4.28}$$

for z real and y integer values. Consequently, the mean and variance of Y are given by

$$\mathbb{E}[Y] = \lambda \quad \mathbb{V}[Y] = \lambda(\lambda\phi + 1) \tag{4.29}$$

The joint likelihood function for estimating (β, ϕ) is

$$L(\beta, \phi; y_{it}) = \prod_{t=1}^T \prod_{i=1}^m \binom{y_{it} + 1/\phi - 1}{y_{it}} \frac{1}{(\lambda_{it}\phi + 1)^{1/\phi}} \left(\frac{\lambda_{it}\phi}{\lambda_{it}\phi + 1} \right)^{y_{it}} \tag{4.30}$$

Inference on individual groups

Consider the compound Poisson Gamma model in (4.24), and assume that a value $Y = y$ has been observed.

Then the conditional distribution of u for given $Y = y$ is found using Bayes Theorem. In order to simplify the notation, the subscript indicating the group and time are omitted.

$$\begin{aligned}
g_u(u|Y = y) &= \frac{f_{y,u}(y, u)}{g_Y(y; \lambda, \phi)} \\
&= \frac{f_{y|u}(y; u) g_u(u)}{g_Y(y; \lambda, \phi)} \\
&= \frac{1}{g_Y(y; \lambda, \phi)} \left(\frac{(\lambda u)^y}{y!} \exp(-\lambda u) \frac{1}{\phi \Gamma(1/\phi)} \left(\frac{u}{\phi} \right)^{1/\phi-1} \exp(-u/\phi) \right) \\
&\propto u^{y+1/\phi-1} \exp(-u(\lambda\phi + 1)/\phi)
\end{aligned} \tag{4.31}$$

We identify the *kernel* of the probability density function

$$u^{y+1/\phi-1} \exp(-u(\lambda\phi + 1)/\phi) \tag{4.32}$$

as the kernel of a Gamma distribution, $G(y + 1/\phi, \phi/(\lambda\phi + 1))$, i.e. the conditional distribution of u for given $Y = y$ can be written as

$$u|Y = y \sim G(y + 1/\phi, \phi/(\lambda\phi + 1)) \tag{4.33}$$

The mean of the conditional distribution is given by:

$$\mathbb{E}[u|Y = y] = \frac{y\phi + 1}{\lambda\phi + 1} \quad (4.34)$$

And the variance of the conditional distribution is:

$$\mathbb{V}[u|Y = y] = \frac{(\phi^2 + \phi)}{(\lambda\phi + 1)^2} \quad (4.35)$$

These formulas provide the mean and variance of the conditional distribution of u given the observed value $Y = y$.

Why is the Gamma distribution chosen to represent the variation between months?

The choice of the Gamma distribution for modeling the random effects has been motivated by several reasons. Firstly, the support of the Gamma distribution, which ranges from 0 to infinity, aligns with the mean-value space, denoted as \mathcal{M} , for the Poisson distribution. This ensures that the random effects are constrained within a meaningful range for the underlying Poisson process.

Secondly, the two-parameter family of Gamma distributions offers considerable flexibility, encompassing a wide range of shapes and distributions that can span from exponential-like distributions to fairly symmetrical distributions on the positive real line. This flexibility allows the model to capture various patterns and characteristics observed in the data.

Additionally, the choice of the Gamma distribution has benefits in terms of the derivation of the marginal distribution of the response variable Y . The kernel $u^{\alpha-1} \exp(-u/\beta)$ of the Gamma distribution used for modeling the random effects exhibits a similar structure to the kernel $u^y \exp(-u)$ of the likelihood function corresponding to the sampling distribution of Y . This similarity facilitates the analytical computation of the integral involved in deriving the marginal distribution, as it can be expressed in terms of known functions.

Overall, the Gamma distribution is selected due to its alignment with the mean-value space of the Poisson distribution, its flexibility in capturing diverse distributions, and its analytical convenience in computing the marginal distribution of the response variable.

4.5 Parameter estimation

In this section, the parameter estimation and implementation of the models used in the novel outbreak detection algorithm are presented. These model are implemented in R using the open-source R package **TMB** (Template Model Builder) developed by Kristensen et al. (2016). This package facilitates efficient maximum likelihood estimation and uncertainty calculations for the parameter set $\theta = (\beta, \Psi)$ and random effects u . The presentation of the parameter estimation conducted in **TMB** is strongly inspired by Chapter 2 in Kristensen et al. (2016) and Section 5.10 in Madsen and Thyregod (2011).

TMB maximizes a user-provided objective function in the form of a C++ template, to estimate the maximum likelihood for the parameter set $\theta = (\beta, \Psi)$. Refer to B to access the C++ template files used in this master's thesis. The objective function maximizes the marginal log-likelihood function, which integrates out the random effects u

$$\ell_M(\theta; \mathbf{y}) = \int_{\mathbb{R}^q} \ell(\theta; \mathbf{u}, \mathbf{y}) d\mathbf{u} \quad (4.36)$$

where $\ell(\theta, u, y)$ is the joint log-likelihood function of the data given the parameters and random effects. The maximizer \hat{u}_θ of the joint log-likelihood $\ell(\theta; u, y)$ with respect to the random effects u is defined as:

$$\hat{u}_\theta = \arg \max_u \ell(\theta; u, y) \quad (4.37)$$

Using $H(\hat{u}_\theta)$ to denote the negative Hessian of the joint log-likelihood evaluated at \hat{u}_θ ; i.e.,

$$H(\hat{u}_\theta) = -\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta} \quad (4.38)$$

The Laplace approximation for the marginal log-likelihood $\ell_M(\theta)$ is

$$\ell_{M,LA}(\theta, y) = \ell(\theta, u, y) - \frac{1}{2} \log \left| \frac{H(\hat{u}_\theta)}{2\pi} \right| \quad (4.39)$$

Our estimate of θ minimizes the negative log of the Laplace approximation, i.e.,

$$-\ell_{M,LA}(\theta, y) = -\ell(\theta, u, y) + \frac{1}{2} \log \left| \frac{H(\hat{u}_\theta)}{2\pi} \right| \quad (4.40)$$

The maximization of the Laplace approximation for the marginal likelihood is then performed using conventional R optimization routines (e.g., BFGS) to optimize the objective and obtain our estimate $\hat{\theta}$. Uncertainty of the estimate $\hat{\theta}$, or any differentiable function of the estimate $\phi(\hat{\theta})$, is obtained by the δ -method:

$$\mathbf{V}(\phi(\hat{\theta})) = -\phi'_{\theta}(\hat{\theta}) \left(\Delta^2 \ell_{M,LA}(\hat{\theta}, y) \right)^{-1} \phi'_{\theta}(\hat{\theta})^T \quad (4.41)$$

Additionally, **TMB** utilizes Automatic Differentiation (AD) techniques (Griewank and Walther 2008) to evaluate first, second, and potentially third-order derivatives. This approach enhances the computational efficiency and accuracy of the parameter estimation process in the implemented models. Therefore, even though the random effects are analytically integrated out in (4.30), and the Laplace approximation is not needed, implementing the joint log-likelihood function in TMB can still result in more efficient computations.

For a comprehensive introduction to the concept of AD, it is recommended to read Section 2.1 and Section 2.2 of Fournier et al. (2012).

4.6 Scoring rule

In this section, the scoring rule used to evaluate the overall score of the models is outlined. The approach is inspired by Bjerregård, Møller, and Madsen (2021). For a given time series $y_t = y_1, \dots, y_N$, each forecast and its corresponding realized observation pair (G_t, y_t) is evaluated. The overall score of the model is then reported as the average score:

$$\bar{S}(G, y) = \frac{1}{N} \sum_{t=1}^N S(G_t, y_t) \quad (4.42)$$

One commonly used scoring rule is the *logarithmic score* derived from likelihood theory, which is defined as $S(G, y) = -\log(f(y))$ (Good 1992; Gneiting and Raftery 2007). This scoring rule is based on the probability density function and is equivalent to the log-likelihood of the forecast model. It has desirable properties as it captures all possible information about the observed data in relation to the model. However, it has a potential drawback in that it heavily penalizes unlikely observations. Consequently, even small changes in the tails of a density forecast can lead to significant changes in the *logarithmic score*, even when the overall shape of the density remains unchanged.

The calculation of the logarithmic score is shown in Example 4.2, which is adapted from Bjerregård, Møller, and Madsen (2021).

Example 4.2 (Calculation of the logarithmic score). The Gamma distribution is used to represent the probabilistic forecast in this example. The Gamma distribution is parametrized by two parameters, shape (α) and rate (β), and its probability density function (PDF) is given by:

$$f(y) = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{y}{\beta}\right)^{\alpha-1} \exp\left(-\frac{y}{\beta}\right) \quad (4.43)$$

In this example, the parameters of the true model, denoted as f , is chosen to be $(\alpha, \beta) = (3, 3)$. We simulate 10 observations, denoted as y_1, y_2, \dots, y_{10} , which are shown in Table 4.2. The true model f is compared to a competing model, denoted as g , which is a Gamma distribution with parameters $(\alpha, \beta) = (3, 8)$. The true model f , the competing model g , and the observations y are illustrated in Figure 4.2.

Table 4.2: 10 simulated observations following a G(3,3)-distribution.

i	1	2	3	4	5	6	7	8	9	10
y_i	1.278	2.233	0.657	0.831	1.281	0.287	1.363	1.612	0.790	0.161

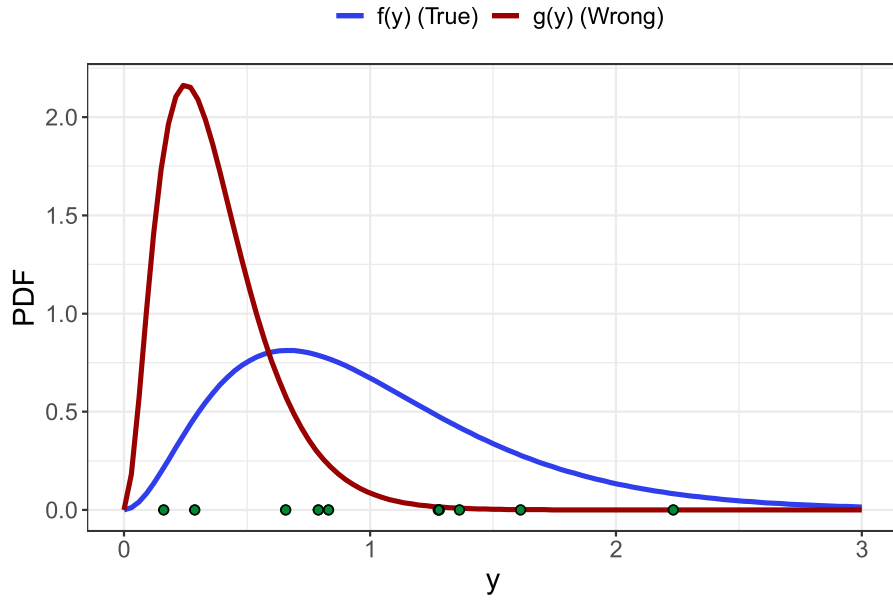


Figure 4.2: Observations (green dots) were simulated from a $G(3, 3)$ -distribution. The true model f (blue) is shown, along with the PDF for the competing model g , which follows a $G(3, 8)$ -distribution.

The logarithmic score of the true model f for the first observation, $y_1 = 1.278$, is calculated as follows

$$\begin{aligned}
 -\log(f(y_1)) &= -\log(f(1.278)) \\
 &= -\log\left[\frac{1}{\Gamma(3)4}\left(\frac{1.278}{4}\right)^{3-1}\exp(-1.278/4)\right] \\
 &= -1.156
 \end{aligned} \tag{4.44}$$

Similarly, the logarithmic scores for the other observations can be calculated using the same formula. The individual logarithmic scores for all 10 observations are presented in Table 4.3. Among the 10 observations, 8 of them are more likely to occur under the true model f compared to the competing model g . The final key quantity, the average logarithmic score, $\bar{S}(G, \mathbf{y})$, can be calculated using Equation (4.42).

$$\begin{aligned}
 \bar{S}(f, \mathbf{y}) &= 1.1 \\
 \bar{S}(g, \mathbf{y}) &= 3.22
 \end{aligned} \tag{4.45}$$

Table 4.3: Logarithmic scores of the two different Gamma models w.r.t. the 10 individual observations.

i	1	2	3	4	5	6	7	8	9	10
$\bar{S}(f, y_i)$	1.16	3.86	0.00	0.23	1.16	0.18	1.37	2.03	0.17	0.83
$\bar{S}(g, y_i)$	4.19	10.71	0.55	1.47	4.21	-0.75	4.74	6.40	1.25	-0.60

