DTU Compute
Department of Applied Mathematics and Computer Science

Automated and Early Detection of Disease Outbreaks
AEDDO

Master Thesis

Automated and Early Detection of Disease Outbreaks
AEDDO

Master Thesis
August, 2023

By
Kasper Schou Telkamp

Cover photo: Vibeke Hempler, 2012
Published by: DTU, Department of Applied Mathematics and Computer Science, Richard Petersens Plads, Building 324, 2800 Kgs. Lyngby Denmark
www.compute.dtu.dk
ISSN: [0000-0000] (electronic version)
ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)
ISBN: [000-00-0000-000-0] (printed version)

Approval

This thesis has been prepared over six months at the Section for Dynamical Systems, Department of Applied Mathematics and Computer Science, at the Technical University of Denmark, DTU, in collaboration with Epidemiologisk Forskning / Modelgruppen at Statens Serum Institut, SSI, in partial fulfilment for the degree Master of Science in Engineering, MSc Eng., Quantitative Biology and Disease Modelling.

It is assumed that the reader has a basic knowledge in the areas of statistics.

Kasper Schou Telkamp - s170397

.....

Signature

.....

Date

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgements

Lasse Engbo Christiansen, Senior Researcher, Statens Serum Institut
Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Jan Kloppenborg Møller, Associate Professor, Technical University of Denmark
Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Contents

1 Introduction

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

2 Literature

In recent years there has been a surge in interest for statistical methods for automated and early detection of infectious disease outbreaks. The methodologies ranges the spectrum of statistical methods and includes regression techniques, time series methodology, methods inspired by statistical process control, methods incorporating spatial information, and multivariate outbreaks detection. A review of the aforementioned methods can be found in Buckeridge (2007) and Unkel et al. (2012).

Write a paragraph that focus on Farrington et al. (1996) **and** Noufaily et al. (2013)!

Moreover, state-of-the-art methods for aberration detection is presented in Salmon, Schumacher, and Höhle (2016) and implemented in the R package **surveillance**, which is available from the Comprehensive R Archive Network at https://cran.r-project.org/web/packages/surveillance/index.html. The R package includes methods such as the Farrington method introduced by Farrington et al. (1996) together with the improvements proposed by Noufaily et al. (2013). As a Bayesian counterpart to these methods the BODA method presented by Manitz and Höhle (2013) allows for easy integration of covariates. Common for these methods are that no accumulation of evidence takes place and they detect aberrations only when the count of the currently monitored timepoint is above the threshold. Another method, which is also implemented in **surveillance** is the negative binomial cumulative sum of Höhle and Paul (2008) that allows for the detection of sustained shifts by accumulating evidence over several timepoints.

3 Dataset

In Denmark the surveillance of infectious diseases is carried out by Statens Serum Institut (SSI). The surveillance is a central part of the national and international disease preparedness. Thanks to The National Board of Health Statutory Order on Physicians' Notification of Infectious Diseases, the quality of the Danish surveillance registers is very high. In the Statutory Order it is stated that a number of diseases[1] are individually notifiable for physicians and general practitioners. The notifications include relevant information on the patient, and are notified on a paper form to the Ministry of Health and to SSI, respectively. A modern surveillance system comprises not only collection and registration of disease data, but also timely and continuous communication of knowledge to authorities responsible for treatment, prevention, and control. The national surveillance system comprises only diseases of serious character, diseases that are particularly infectious, and most of the vaccine-preventable diseases.

For the purpose of this master thesis, only a subset of the individually notifiable diseases are considered. …(**WRITE SOMETHING ABOUT WHAT CHARACTERISES THE CHOSEN SUBSET I.E. SEASONALITY, INCIDENCE, ETC.**)… and the subset includes: Shiga toxin producing Escherichia coli (STEC), **X, X, X (CHOOSE 4)**.

Include a section with plots and tables of the data.

[1] For a full list of individually notifiable diseases see https://www.ssi.dk/sygdomme-beredskab-og-forskning/anmeldelse-af-sygdomme/lovpligtige-meldesystemer/individ_anmeldelses_sygdomme

4 Methods

In this chapter the outbreak detection algorithm together with theory related to generalized mixed effects models and hierarchical generalized linear models is presented. These models are used in this master thesis to model the count observations $\boldsymbol{y}$, but more importantly they are used to assess the unobserved random variables or random effects $\boldsymbol{u}$, which are used directly in the detection algorithm to characterize an outbreak. In general it is impossible to obtain closed form solutions for generalized mixed effects models and hence an outline of the Laplace approximation is included in this chapter. Moreover, the implementation of aforementioned models in R is presented. The presentation of this chapter is mostly inspired by Madsen and Thyregod (2011).

4.1 General mixed effects models

In this section selected theory related to generalized mixed effects models is presented along with a novel method for the prospective detection of disease outbreaks. The novel method utilize a hierarchical Poisson Normal model to model the count observations $\boldsymbol{y}$ and to assess the unobserved random effects $\boldsymbol{u}$. These random effects are in turn used directly in the detection algorithm to characterize an outbreak. Moreover, an implementation in R of the aforementioned model is presented.

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

$$L_M(\boldsymbol{\theta}; \boldsymbol{y}) = \int_{\mathbb{R}^q} L(\boldsymbol{\theta}; \boldsymbol{u}, \boldsymbol{y}) d\boldsymbol{u} \tag{4.1}$$

where $\boldsymbol{y}$ is the observed random variable, $\boldsymbol{\theta}$ is the model parameters to be estimated and $\boldsymbol{U}$ 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 $\boldsymbol{\theta}$ is the marginal likelihood L_M obtained by integrating out the unobserved random variables. In general it is difficult to solve the integral in (4.1) if the number of unobserved random variables is more than a few and hence numerical methods must be used.

4.1.1 Hierarchical models

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

$$f_{Y|u}(\boldsymbol{y}; \boldsymbol{u}, \boldsymbol{\beta}) \tag{4.2}$$

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

$$f_U(\boldsymbol{u}; \boldsymbol{\Psi}) \tag{4.3}$$

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

$$L(\boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{u}, \boldsymbol{y}) = f_{Y|u}(\boldsymbol{y}; \boldsymbol{u}, \boldsymbol{\beta}) f_U(\boldsymbol{u}; \boldsymbol{\Psi}) \tag{4.4}$$

To obtain the likelihood for the model parameters $(\boldsymbol{\beta}, \boldsymbol{\Psi})$ the unobserved random variables are integrated out. The likelihood function for estimating $(\boldsymbol{\beta}, \boldsymbol{\Psi})$ is as in (4.1) the marginal likelihood

$$L_M(\boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{y}) = \int_{\mathbb{R}^q} L(\boldsymbol{\beta}, \boldsymbol{\Psi}; \boldsymbol{u}, \boldsymbol{y}) d\boldsymbol{u} \tag{4.5}$$

where q is the number of unobserved random variables, and $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ are the parameters to be estimated.

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

$$\ell(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) \sim \ell(\boldsymbol{\theta}, \tilde{\boldsymbol{u}}, \boldsymbol{y}) - \frac{1}{2}(\boldsymbol{u} - \tilde{\boldsymbol{u}})^T \boldsymbol{H}(\tilde{\boldsymbol{u}})(\boldsymbol{u} - \tilde{\boldsymbol{u}}) \tag{4.6}$$

where the first-order term of the Taylor expansion disappears since the expansion is done around the optimum $\tilde{\boldsymbol{u}}$ and $\boldsymbol{H}(\tilde{\boldsymbol{u}}) = -\ell''(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y})|_{\boldsymbol{u}=\tilde{\boldsymbol{u}}}$ is the negative Hessian of the joint log-likelihood evaluated at $\tilde{\boldsymbol{u}}$ which will simply be referred to as “the Hessian”.

It is readily seen that the joint log-likelihood for the hierarchical model specified in Section 4.1.1 is

$$\ell(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{y}) = \ell(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{u}, \boldsymbol{y}) = \log f_{Y|u}(\boldsymbol{y}; \boldsymbol{u}, \boldsymbol{\beta}) + \log f_U(\boldsymbol{u}; \boldsymbol{\Psi}) \tag{4.7}$$

which implies that the Laplace approximation becomes

$$\ell_{M,LA}(\boldsymbol{\theta}, \boldsymbol{y}) = \log f_{Y|u}(\boldsymbol{y}; \tilde{\boldsymbol{u}}, \boldsymbol{\beta}) + \log f_U(\tilde{\boldsymbol{u}}, \boldsymbol{\Psi}) - \frac{1}{2} \log \left| \frac{\boldsymbol{H}(\tilde{\boldsymbol{u}})}{2\pi} \right| \tag{4.8}$$

4.1.3 Formulation of the hierarchical Poisson Normal model

The conditional distribution of the count observations, $f_{Y|u}(\boldsymbol{y}; \boldsymbol{u}, \boldsymbol{\beta})$, is assumed to be a Poisson distribution with intensities $\boldsymbol{\lambda}$. Also, it is assumed that the count is proportional to the population size $\boldsymbol{x}$. Hence, in terms of the canonical link for the Poisson distribution the model for the fixed effects is

$$\log(\lambda_i) = \boldsymbol{X}_i^T \boldsymbol{\beta} + \log(x_{it}) \tag{4.9}$$

Here $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$ are p-dimensional vectors of covariates and fixed effect parameters respectively where p denotes the number of fixed effects parameters. Moreover, the distribution of the random effects, $f_U(\boldsymbol{u}, \boldsymbol{\sigma})$, are assumed to follow a Gaussian distribution

$$u_{it} = \epsilon_{it} \tag{4.10}$$

where $\epsilon_{it} \sim \mathsf{N}(\mathbf{0}, \sigma^2)$ is a white noise process, and σ is a model parameter. Henceforth, the total set of parameters are $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)$ and the model can be formulated as a two-level hierarchical model

$$\boldsymbol{Y}|\boldsymbol{u} \sim \mathsf{Pois}\left(\boldsymbol{\lambda}\exp(\boldsymbol{u})\right) \tag{4.11a}$$
$$\boldsymbol{u} \sim \mathsf{N}(\mathbf{0}, \sigma^2) \tag{4.11b}$$

The joint likelihood for the count observations $\boldsymbol{y}$ and the random effects $\boldsymbol{u}$ becomes

$$L(\boldsymbol{\beta}, \sigma; \boldsymbol{u}, \boldsymbol{y}) = \tag{4.12}$$

4.1.4 Implementation in R using TMB

The hierarchical Poisson Normal model specified in (4.11) is implemented in R using the open source R package **TMB** (Template Model builder) by Kristensen et al. (2016). The joint likelihood specified in (4.12) is defined as a C++ template function. The package evaluates and maximizes the Laplace approximation of the marginal likelihood where the random effects are automatically integrated out. This approximation and its derivaties are obtained using automatic differentiation of the joint likelihood.

4.1.5 Novel outbreak detection algorithm

4.2 Hierarchical generalized linear models

4.2.1 Hierarchical Poisson Gamma model

Likewise, in the compound Poisson-Gamma model the conditional distribution, $Y|u$, of the count observations are assumed to be a Poisson distribution, but this time the intensities, λ_{it}, are defined as

$$\mathsf{log}(\lambda_{it} = \mathbf{X}_i^T \beta_{it} + \mathsf{log}(x_{it})) \tag{4.13}$$

Here $\mathbf{X}_i$ is $T \times 6-$dimensional, and β_{it} contains the corresponding fixed effect parameters. Additionally, the random effects $u_i t$ are assumed to be Gamma distributed. Subsequently, the model can be formulated as a two-level hierarchical model

$$Y_{it}|u_{it} \sim \mathsf{Pois}(\lambda_i u_{it}) \tag{4.14a}$$
$$u_{it} \sim \mathsf{G}(1/\phi, \phi) \tag{4.14b}$$

In the first stage a random value u_{it} is selected according to a reparameterized Gamma distribution with shape, $1/\phi$, and scale, ϕ. Hence the mean value of the Gamma distribution is 1. Moreover, a fixed effect parameter, λ_{it}, is found for each age group, $i = 1, \ldots, 11$. The Y is generated according to a Poisson distribution with $\lambda_i u_{it}$ as mean value. The the marginal distribution of Y is a negative binomial distribution, $Y \sim \mathsf{NB}\left(1/\phi, 1//\lambda\phi + 1)\right)$. The probability function for Y is

$$
\begin{aligned}
P[Y = y_i] &= g_Y(y; \lambda, \phi) \\
&= \frac{\lambda^y}{y!\Gamma(1/\phi)\phi^{1/\phi}} \frac{\phi^{y+1/\phi}\Gamma(y + 1/\phi)}{(\lambda\phi + 1)^{y+1/\phi}} \\
&= \frac{\Gamma(y + 1/\phi)}{\Gamma(1/\phi)y!} \frac{1}{(\lambda\phi + 1)^{1/\phi}} \left(\frac{\lambda\phi}{\lambda\phi + 1}\right)^y \\
&= \binom{y + 1/\phi - 1}{y} \frac{1}{(\lambda\phi + 1)^{1/\phi}} \left(\frac{\lambda\phi}{\lambda\phi + 1}\right)^y , \ for\ y = 0, 1, 2, \ldots
\end{aligned}
\tag{4.15}
$$

where we have used the convention

$$
\binom{z}{y} = \frac{\Gamma(z + 1)}{\Gamma(z + 1 - y)y!} \tag{4.16}
$$

for z real and y integer values. The marginal distribution of Y is a negative binomial distribution, $Y \sim \text{NB}\left(1/\phi, 1/(\lambda\phi + 1)\right)$. See proof in B.1.1.

Inference on individual groups
Consider the compound Poisson Gamma model in (4.14), and assume that a value $Y = y$ has been observed. Then the conditional distribution of u for given $Y = y$ is a Gamma distribution

Consider the hierarchical Poisson-Gamma model in (4.14), and assume that a value $Y = y$ has been observed. Then the conditional distribution of u for given $Y = y$ is a Gamma distribution,

$$
u|Y = y \sim \text{G}\left(y + 1/\phi, \phi/(\lambda\phi + 1)\right) \tag{4.17}
$$

with mean

$$
\text{E}[u|Y = y] = \frac{y\phi + 1}{\lambda\phi + 1} \tag{4.18}
$$

and variance

$$
\text{V}[u|Y = y] = \frac{(y\phi^2 + \phi)}{(\lambda\phi + 1)^2} \tag{4.19}
$$

Why do we choose the Gamma distribution to represent the variation between days?
The Gamma distribution is chosen for three simple reasons. First of all, the support of the Gamma distribution, $0 < u_{it} < \infty$ conforms to the mean-value space, $\mathcal{M}$ for the Poisson distribution. Secondly, the two-parameter family of Gamma distributions is a rather flexible class of unimodal distribution, ranging from an exponential distribution to a fairly symmetrical distribution on the positive real line. A third reason may be observed in the derivation of the marginal distribution of Y. The fact that the kernel $u^{\alpha-1}\exp(-u/\beta)$ of the mixing distribution have the same structure as the kernel $u^y\exp(-u)$ of the likelihood function corresponding to the sampling distribution of Y. This feature have the consequence that the integral has a closed form representation in terms of known functions.

4.2.2 Parameter estimation

5 Results

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.1 Case studies

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

5.2 Simulation study

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

6 Discussion

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

7 Conclusion

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Bibliography

Buckeridge, David L. (2007). “Outbreak detection through automated surveillance: A review of the determinants of detection”. In: *Journal of Biomedical Informatics* 40.4. Public Health Informatics, pp. 370–379. ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2006.09.003. URL: https://www.sciencedirect.com/science/article/pii/S1532046406000980.

Unkel, Steffen et al. (2012). “Statistical methods for the prospective detection of infectious disease outbreaks: a review”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 175.1, pp. 49–82. ISSN: 09641998, 1467985X. URL: http://www.jstor.org/stable/41409708 (visited on 02/15/2023).

Farrington, C. P. et al. (1996). “A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease”. In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159.3, pp. 547–563. ISSN: 09641998, 1467985X. URL: http://www.jstor.org/stable/2983331 (visited on 01/27/2023).

Noufaily, Angela et al. (Mar. 2013). “An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems”. en. In: *Online Journal of Public Health Informatics* 32.7, pp. 1206–1222.

Salmon, Maëlle, Dirk Schumacher, and Michael Höhle (2016). “Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance”. In: *Journal of Statistical Software* 70.10, pp. 1–35. DOI: 10.18637/jss.v070.i10. URL: https://www.jstatsoft.org/index.php/jss/article/view/v070i10.

Manitz, Juliane and Michael Höhle (2013). “Bayesian outbreak detection algorithm for monitoring reported cases of campylobacteriosis in Germany”. In: *Biometrical Journal* 55.4. Cited by: 18, pp. 509–526. DOI: 10.1002/bimj.201200141. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84880042358&doi=10.1002%5C%2fbimj.201200141&partnerID=40&md5=4658480d9c1cd84c23335bdb9269f0bc.

Höhle, Michael and Michaela Paul (2008). “Count data regression charts for the monitoring of surveillance time series”. In: *Computational Statistics and Data Analysis* 52.9. Cited by: 56, pp. 4357–4368. DOI: 10.1016/j.csda.2008.02.015. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-42749087852&doi=10.1016%5C%2fj.csda.2008.02.015&partnerID=40&md5=8d18fdad2a28d99ecbda5441a8fbeae3.

Madsen, Henrik and Poul Thyregod (2011). *Introduction to general and generalized linear models*. English. Texts in statistical science. CRC Press. ISBN: 9781420091557.

Wolfinger, Russell D. and Xihong Lin (1997). “Two Taylor-series approximation methods for nonlinear mixed models”. In: *Computational Statistics and Data Analysis* 25.4. Cited by: 87, pp. 465–490. DOI: 10.1016/S0167-9473(97)00012-1. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031222254&doi=10.1016%2fS0167-9473%2897%2900012-1&partnerID=40&md5=dbf7ca1e47d836029af4044ff9880463.

Kristensen, Kasper et al. (2016). “TMB: Automatic Differentiation and Laplace Approximation”. In: *Journal of Statistical Software* 70.5, pp. 1–21. DOI: 10.18637/jss.v070.i05.

A Some probability functions

This chapter serves as a reference, specifying notation, properties, and moments related to the various distributions used in this master thesis.

Name	Support	Density	E$[Y]$	V$[Y]$
Poisson Pois(λ)	$0, 1, 2, \ldots$ $\lambda \in \mathbb{R}_+$	$\frac{\lambda^y}{y!}\exp(-\lambda)$	λ	λ
Gamma G(α, β)	$\mathbb{R}_+$ $\alpha \in \mathbb{R}_+, \beta \in \mathbb{R}_+$	$\frac{1}{\Gamma(\alpha)\beta}\left(\frac{y}{\beta}\right)^{\alpha-1}\exp(-y/\beta)$	$\alpha\beta$	$\alpha\beta^2$
Neg. Bin. NB(r, p)	$0, 1, 2, \ldots$ $r \in \mathbb{R}_+, p \in]0, 1]$	$\binom{r+y-1}{y} p^r(1-p)^y$	$\frac{r(1+p)}{p}$	$\frac{r(1-p)}{p^2}$

Table A.1: Density, support, mean value, and variance for a number of distributions used in this master thesis.

A.1 The Poisson distribution model

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

A.2 The Gamma distribution model

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

A.3 The Negative Binomial distribution model

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

B Proofs

B.1 Hierarchical Poisson Gamma model

This section is a collection of proofs for the derivation of the compound Poisson Gamma model in (4.14).

B.1.1 Probability function for Y

The probability function for the conditional distribution of Y for given u

$$f_{Y|u}(y;\lambda,u) = \frac{(\lambda u)^y}{y!}\,\mathsf{exp}(-\lambda u) \tag{B.1}$$

and the probability density function for the distribution of u is

$$f_u(u;\phi) = \frac{1}{\phi\Gamma(1/\phi)}\left(\frac{u}{\phi}\right)^{1/\phi-1}\mathsf{exp}(-u/\phi) \tag{B.2}$$

Given (B.1) and (B.2), the probability function for the marginal distribution of Y is determined from

$$\begin{aligned} g_Y(y;\lambda,\phi) &= \int_{u=0}^{\infty} f_{Y|u}(y;\lambda,u) f_u(u;\phi)\,du \\ &= \int_{u=0}^{\infty} \frac{(\lambda u)^y}{y!}\,\mathsf{exp}(-\lambda u)\frac{1}{\phi\Gamma(1/\phi)}\left(\frac{u}{\phi}\right)^{1/\phi-1}\mathsf{exp}(-u/\phi)\,du \\ &= \frac{\lambda^y}{y!\Gamma(1/\phi)\phi^{1/\phi}}\int_{u=0}^{\infty} u^{y+1/\phi-1}\,\mathsf{exp}\big(-u(\lambda\phi+1)/\phi\big)\,du \end{aligned} \tag{B.3}$$

In (B.3) it is noted that the integrand is the *kernel* in the probability density function for a Gamma distribution, $\mathsf{G}\big(y+1/\phi, \phi/(\lambda\phi+1)\big)$. As the integral of the density shall equal one, we find by adjusting the norming constant that

$$\int_{u=0}^{\infty} u^{y+1/\phi-1}\,\mathsf{exp}\left(-u/\Big(\phi/(\lambda\phi+1)\Big)\right)du = \frac{\phi^{y+1/\phi}\Gamma(y+1/\phi)}{(\lambda\phi+1)^{y+1/\phi}} \tag{B.4}$$

and then (4.15) follows

B.1.2 Conditional distribution of Y

The conditional distribution is found using Bayes Theorem

$$\begin{aligned} g_u(u|Y=y) &= \frac{f_{y,u}(y,u)}{g_Y(y;\lambda,\phi)} \\ &= \frac{f_{y|u}(y;u)g_u(u)}{g_Y(y;\lambda,\phi)} \\ &= \frac{1}{g_Y(y;\lambda,\phi)}\left(\frac{(\lambda u)^y}{y!}\,\mathsf{exp}(-\lambda u)\frac{1}{\phi\Gamma(1/\phi)}\left(\frac{u}{\phi}\right)^{1/\phi-1}\mathsf{exp}(-u/\phi)\right) \\ &\propto u^{y+1/\phi-1}\,\mathsf{exp}\big(-u(\lambda\phi+1)/\phi\big) \end{aligned} \tag{B.5}$$

We identify the *kernel* of the probability density function

$$u^{y+1/\phi-1} \exp(-u(\lambda\phi+1)/\phi) \tag{B.6}$$

as the kernel of a Gamma distribution, $G(y+1/\phi, \phi/(\lambda\phi+1))$

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Technical
University of
Denmark

Richard Petersens Plads, Building 324
2800 Kgs. Lyngby
Tlf. 4525 1700

www.compute.dtu.dk