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

This chapter is an overview of the methods that we propose to solve an **important problem**.

instruction	expectation	rating
good	high	25
good	high	0
good	high	-16
good	high	5
good	high	11
good	high	-6

2 Literature

Here is a review of existing methods.

Farrington et al. (1996) was improved in Noufaily et al. (2013)

and Salmon, Schumacher, and Höhle (2016)

Yang, Santillana, and Kou (2015) and Ning, Yang, and Kou (2019) uses Google Search terms to predict influenza-like ilness (ILI).

Vega, Jose Eugenio Lozano, et al. (2013) introduces the *Moving Epidemic Method* (MEM), which is used to compare the intensity level of ILI's in Vega, José E. Lozano, et al. (2015).

An analysis of ILI spread in Sweden showed that, rates in some large countries could vary considerably from one region to another Skog et al. (2014).

Different methods for monitoring influenza surveillance using only recent data is presented in Cowling et al. (2006)

Hutwagner et al. (1997) applied cumulative sums (CUSUM), to detect Salmonella outbreaks in US.

Costagliola et al. (1991) implement a simple regression model and calculates a 95% CI for a non-epidemic curve and use this threshold to alert when an epidemic begins.

Stern and Lightfoot (1999) discusses an automated algorithm to detect disease outbreaks with salmonella in Australia

3 Dataset

Test

4 Methods

This chapter present the methods…

4.1 Hierarchical models

In this section the hierarchical model and the generalized mixed effect model that is used to model the count observation y_{it} for the specific age groups, $i = 1, \ldots, 11$, and in the monthly period ranging from 2008 to 2022, $t = 1, \ldots, 180$, is presented. For an introduction to the concept of hierarchical models see Madsen and Thyregod (2011).

4.1.1 Hierachical Poisson Normal model

Consider a hierarchical model for Y specified by

$$\begin{aligned} Y_{it}|u_{it} &\sim \text{Pois}\left(w_{it}\lambda_i \exp(u_{it})\right) && (4.1a) \\ u_{it} &\sim \text{N}(0, \sigma^2) && (4.1b) \end{aligned}$$

4.1.2 Hierarchical Poisson Gamma model

The count observations y_{it} is assumed to follow a Poisson distribution with age group specific intensities, λ_i. Furthermore, a variation between months, u_{it}, is represented with a Gamma distribution.

Formulation of the Compound Poisson Gamma model
Consider a hierarchical model for Y specified by

$$\begin{aligned} Y_{it}|u_{it} &\sim \text{Pois}(\lambda_i u_{it}) && (4.2a) \\ u_{it} &\sim \text{G}(1/\beta, \beta) && (4.2b) \end{aligned}$$

In the first stage a random value u_{it} is selected according to a reparameterized Gamma distribution with shape, $1/\beta$, and scale, β. Hence the mean value of the Gamma distribution is 1. Moreover, a fixed effect parameter, λ_i, 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, $Y \sim \text{NB}\left(1/\beta, 1//\lambda\beta + 1)\right)$. The probability function for Y is

$$\begin{aligned} P[Y = y] &= g_Y(y; \lambda, \beta) \\ &= \frac{\lambda^y}{y!\Gamma(1/\beta)\beta^{1/\beta}} \frac{\beta^{y+1/\beta}\Gamma(y + 1/\beta)}{(\lambda\beta + 1)^{y+1/\beta}} \\ &= \frac{\Gamma(y + 1/\beta)}{\Gamma(1/\beta)y!} \frac{1}{(\lambda\beta + 1)^{1/\beta}} \left(\frac{\lambda\beta}{\lambda\beta + 1}\right)^y \\ &= \binom{y + 1/\beta - 1}{y} \frac{1}{(\lambda\beta + 1)^{1/\beta}} \left(\frac{\lambda\beta}{\lambda\beta + 1}\right)^y, \; for \; y = 0, 1, 2, \ldots \end{aligned} \tag{4.3}$$

where we have used the convention

$$\binom{z}{y} = \frac{\Gamma(z+1)}{\Gamma(z+1-y)y!} \tag{4.4}$$

for z real and y integer values.

Inference on individual groups
Consider the compound Poisson Gamma model in (4.2), and assume that a value $Y = y$ has been observed. Then the conditional distribution of u for given $Y = y$ is a Gamma distribution

$$u|Y = y \sim \mathsf{G}\left(y + 1/\beta, \beta/(\lambda\beta + 1)\right) \tag{4.5}$$

with mean

$$\mathsf{E}[u|Y = y] = \frac{y\beta + 1}{\lambda\beta + 1} \tag{4.6}$$

and variance

$$\mathsf{V}[u|Y = y] = \frac{(y\beta^2 + \beta)}{(\lambda\beta + 1)^2} \tag{4.7}$$

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 reasons 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.

Bibliography

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.

Yang, Shihao, Mauricio Santillana, and S. C. Kou (2015). “Accurate estimation of influenza epidemics using Google search data via ARGO”. In: *Proceedings of the National Academy of Sciences* 112.47, pp. 14473–14478. DOI: 10.1073/pnas.1515373112. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1515373112. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1515373112.

Ning, Shaoyang, Shihao Yang, and S. C. Kou (Mar. 2019). “Accurate regional influenza epidemics tracking using Internet search data”. In: *Scientific Reports* 9.1. DOI: 10.1038/s41598-019-41559-6. URL: https://doi.org/10.1038/s41598-019-41559-6.

Vega, Tomás, Jose Eugenio Lozano, et al. (2013). “Influenza surveillance in Europe: establishing epidemic thresholds by the Moving Epidemic Method”. In: *Influenza and Other Respiratory Viruses* 7.4, pp. 546–558. DOI: https://doi.org/10.1111/j.1750-2659.2012.00422.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1750-2659.2012.00422.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1750-2659.2012.00422.x.

Vega, Tomás, José E. Lozano, et al. (2015). “Influenza surveillance in Europe: comparing intensity levels calculated using the moving epidemic method”. In: *Influenza and Other Respiratory Viruses* 9.5, pp. 234–246. DOI: https://doi.org/10.1111/irv.12330. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/irv.12330. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/irv.12330.

Skog, Lars et al. (July 2014). “Spatiotemporal characteristics of pandemic influenza”. In: *BMC Infectious Diseases* 14.1. DOI: 10.1186/1471-2334-14-378. URL: https://doi.org/10.1186/1471-2334-14-378.

Cowling, Benjamin J et al. (Aug. 2006). “Methods for monitoring influenza surveillance data”. In: *International Journal of Epidemiology* 35.5, pp. 1314–1321. ISSN: 0300-5771. DOI: 10.1093/ije/dyl162. eprint: https://academic.oup.com/ije/article-pdf/35/5/1314/14112661/dyl162.pdf. URL: https://doi.org/10.1093/ije/dyl162.

Hutwagner, L.C. et al. (1997). “Using Laboratory-Based Surveillance Data for Prevention: An Algorithm for Detecting Salmonella Outbreaks”. In: *Emerging Infectious Diseases* 3.3. Cited by: 131; All Open Access, Gold Open Access, Green Open Access, pp. 395–400. DOI: 10.3201/eid0303.970322. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0031181478&doi=10.3201%2feid0303.970322&partnerID=40&md5=0b5a2fea57368381526824f978f96ee3.

Costagliola, D et al. (Jan. 1991). “A routine tool for detection and assessment of epidemics of influenza-like syndromes in France.” In: *American Journal of Public Health* 81.1, pp. 97–99. DOI: 10.2105/ajph.81.1.97. URL: https://doi.org/10.2105/ajph.81.1.97.

Stern, L. and D. Lightfoot (1999). “Automated outbreak detection: A quantitative retrospective analysis”. In: *Epidemiology and Infection* 122.1. Cited by: 55; All Open Access, Bronze Open Access, Green Open Access, pp. 103–110. DOI: 10.1017/S0950268898001939. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0033065916&doi=10.1017%2fS0950268898001939&partnerID=40&md5=cc79bbef077747d0cfe003ef5ee3b1c5.

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

A Proofs

A.1 Hierarchical Poisson Gamma model

This section is a collection of proofs for the derivation of the compound Poisson Gamma model in (4.2).

A.1.1 Probability function for Y

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

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

and the probability density function for the distribution of u is

$$f_u(u;\beta) = \frac{1}{\beta\Gamma(1/\beta)}\left(\frac{u}{\beta}\right)^{1/\beta-1}\mathrm{exp}(-u/\beta) \tag{A.2}$$

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

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

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

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

and then (4.3) follows.

A.1.2 Conditional distribution of Y

The conditional distribution of Y is found using Bayes Theorem

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

We identify the *kernel* of the probability density function

$$u^{y+1/\beta-1} \, \mathrm{exp}(-u(\lambda\beta+1)/\beta) \tag{A.6}$$

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

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