Automated and Early Detection of Disease Outbreaks

Kasper Schou Telkamp

Section for Dynamical Systems

DTU Compute
Department of Applied Mathematics and Computer Science

Outline

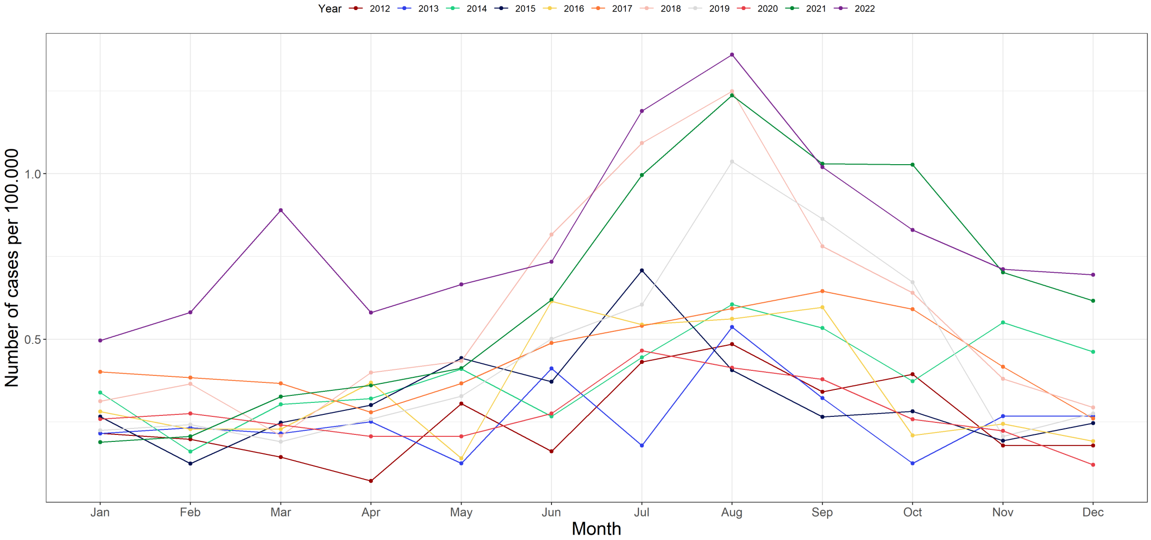
- Data exploration
  - Shiga- and verotoxin producing E. coli.
- State-of-the-art
  - Farrington
  - Noufaily
- Hierarchical models
  - Hierachical Poisson-Normal model
  - Hierachical Poisson-Gamma model
- Comparison of methods
- References
- Appendix

Shiga- and verotoxin producing E. coli.

Date	ageGroup	y_{it}	x_{it}
2008-01-01	$<$1 year	2	64137
2008-01-01	1-4 years	2	259910
2008-01-01	5-14 years	2	680529
2008-01-01	15-24 years	1	635838
...	...	...	...
2022-12-01	5-14 years	5	634139
2022-12-01	15-24 years	1	721286
2022-12-01	25-64 years	10	3031374
2022-12-01	65+ years	12	1204892

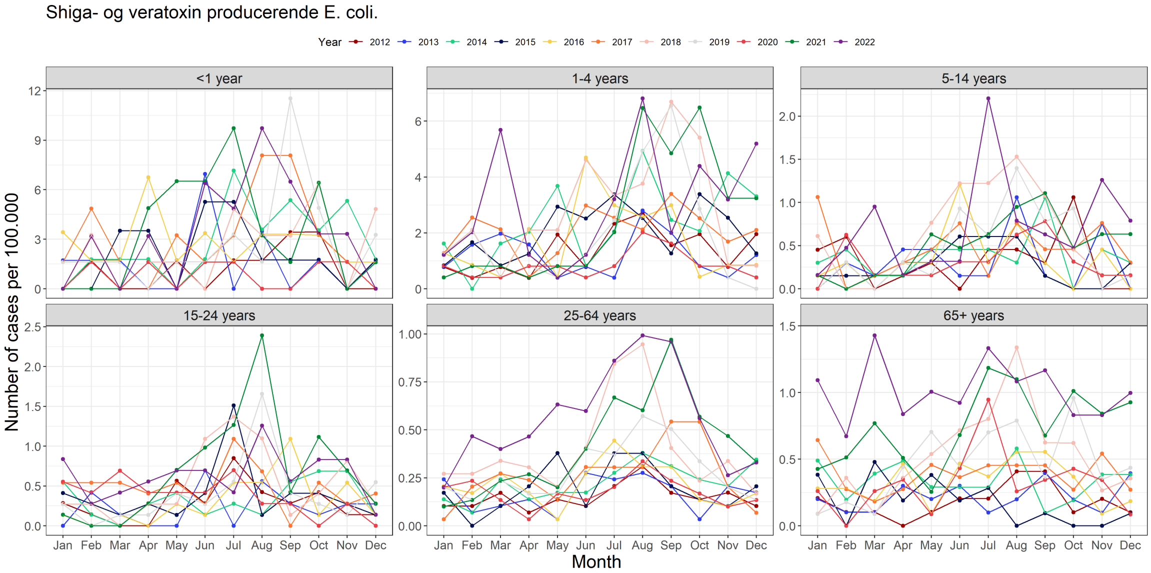
Shiga- and verotoxin producing E. coli.

Shiga- og veratoxin producerende E. coli.



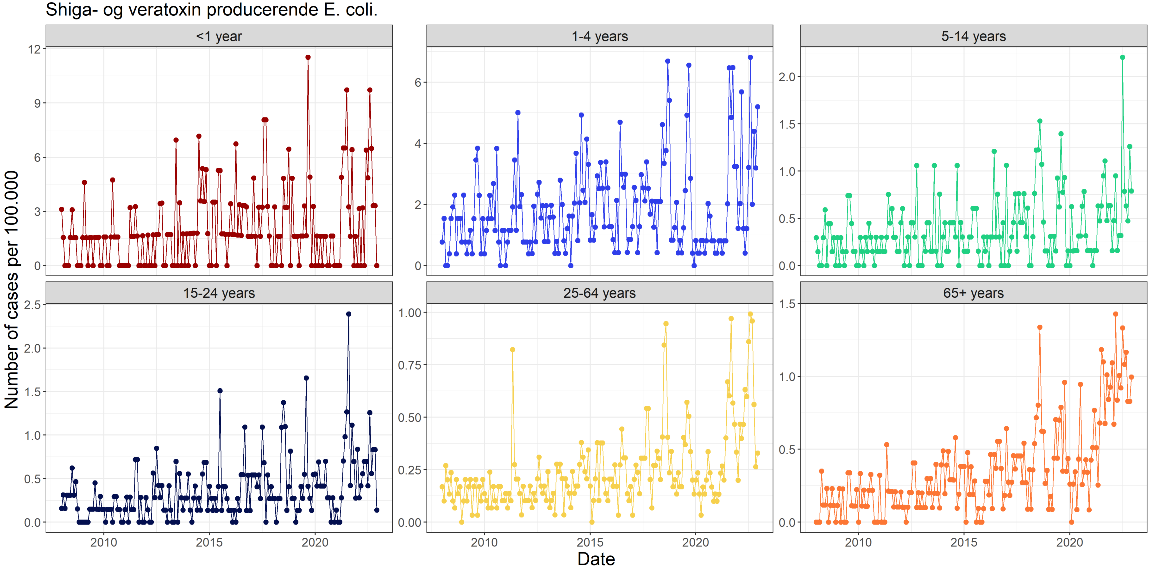
Data exploration

Shiga- and verotoxin producing E. coli.



Data exploration

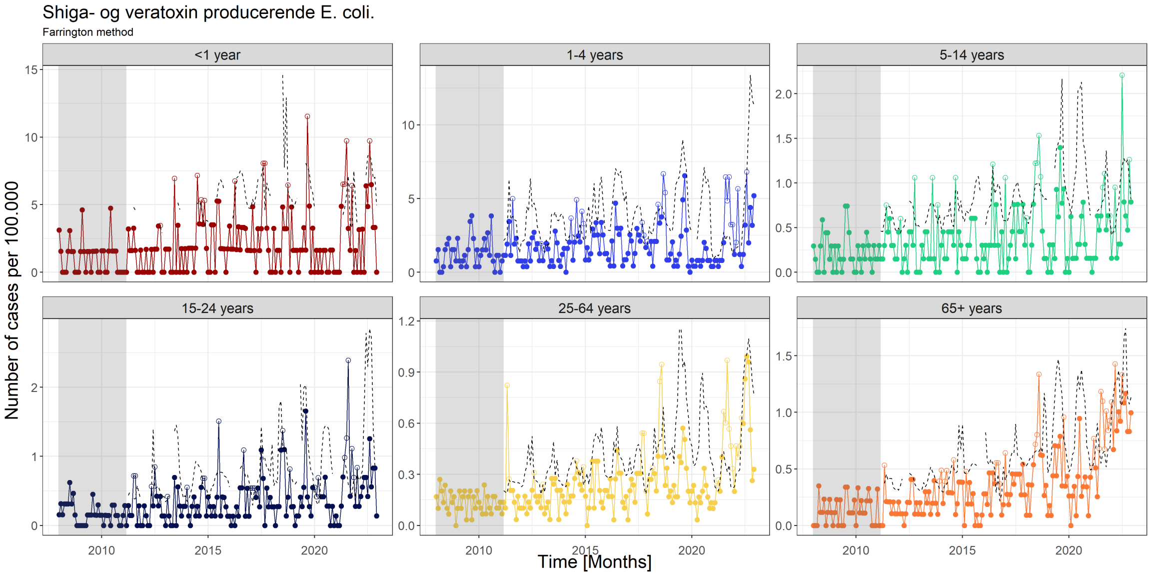
Shiga- and verotoxin producing E. coli.



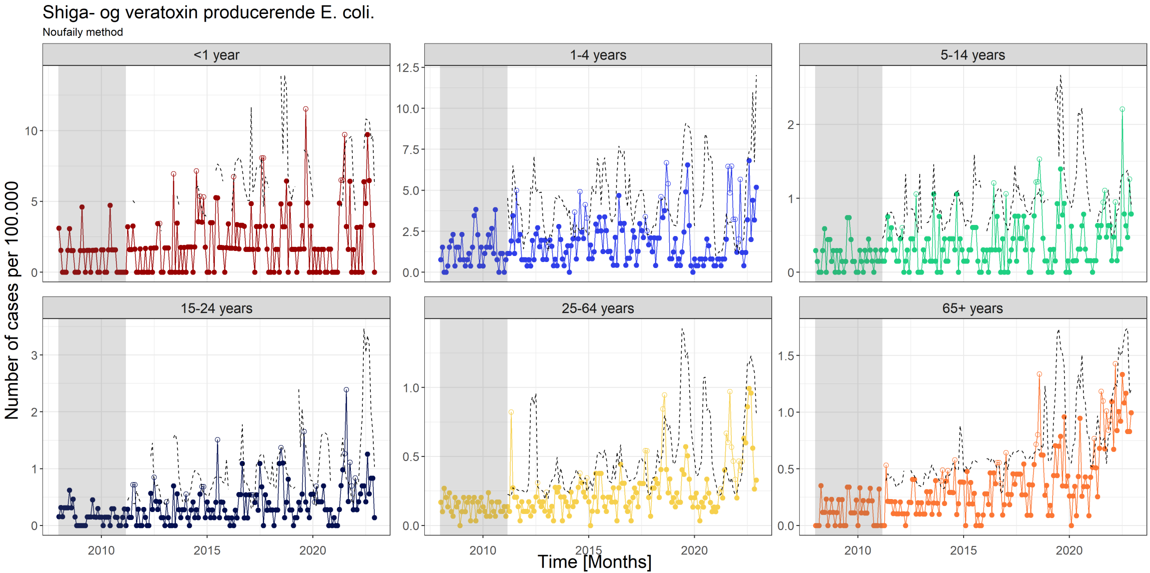
State-of-the-art

State-of-the-art methods for aberration detection is presented in Salmon, Schumacher, and Höhle 2016 and implemented in the R package **surveillance**. 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.

Farrington


Shiga- og veratoxin producerende E. coli.
Farrington method
<1 year
1-4 years
5-14 years
15-24 years
25-64 years
65+ years
Number of cases per 100.000
Time [Months]
2010
2015
2020

Noufaily


Shiga- og veratoxin producerende E. coli.
Noufaily method
<1 year
1-4 years
5-14 years
15-24 years
25-64 years
65+ years
Number of cases per 100.000
Time [Months]
2010
2015
2020

Hierachical Poisson-Normal model

The conditional distribution, $Y|u$, of the count observations are assumed to be a Poisson distribution with intensities λ_{it}. Also, we shall assume that the count is proportional to the population size, x_{it}, within each age group, i, at a given time point, t. Hence, in terms of the canonical link for the Poisson distribution the model for the fixed effect is

$$\log(\lambda_{it}) = \mathbf{X}_i^T \beta_{it} + \log(x_{it}) \tag{1}$$

Here $\mathbf{X}_i$ is $T \times 6$-dimensional, and β_{it} contains the corresponding fixed effect parameter. The random effects u_{it} are assumed to be Gaussian.

$$u_{it} = \epsilon_{it} \tag{2}$$

where $\epsilon_{it} \sim \mathrm{N}(0, \sigma^2)$ is a white noise process, and σ is a model parameter.

Formulation

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

$$Y_{it}|u_{it} \sim \mathrm{Pois}(\lambda_{it} e^{u_{it}}) \tag{3a}$$

$$u_{it} \sim \mathrm{N}(0, \sigma^2) \tag{3b}$$

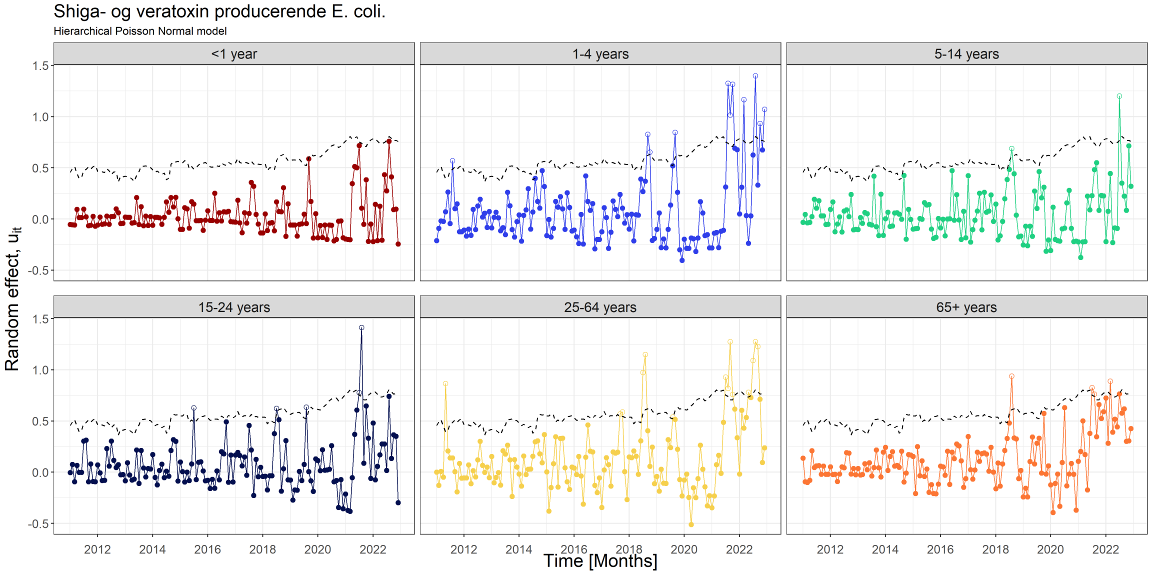
Threshold calculation

Here should be something about threshold calculation

Results

Parameter	Estimate	Std. Error
$\beta_{<1year}$	10.83	0.11
$\beta_{1-4years}$	13.64	0.09
$\beta_{5-14years}$	13.94	0.10
$\beta_{15-24years}$	14.05	0.09
$\beta_{25-64years}$	16.61	0.08
$\beta_{65+years}$	14.83	0.09
σ	1.01	0.04

Results


Shiga- og veratoxin producerende E. coli.
Hierarchical Poisson Normal model
<1 year
1-4 years
5-14 years
15-24 years
25-64 years
65+ years
Random effect, u_{it}
1.5
1.0
0.5
0.0
-0.5
2012
2014
2016
2018
2020
2022
Time [Months]

Hierachical 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

$$\log(\lambda_{it}) = \mathbf{X}_i^T \beta_{it} + \log(x_{it}) \tag{4}$$

Here $\mathbf{X}_i$ is $T \times 6$-dimensional, and β_{it} contains the corresponding fixed effect parameter. Additionally, the random effects u_{it} are assumed to be Gamma distributed.

Formulation

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

$$Y_{it}|u_{it} \sim \mathrm{Pois}(\lambda_{it} u_{it}) \tag{5a}$$

$$u_{it} \sim \mathrm{G}(1/\phi_i, \phi_i) \tag{5b}$$

Probability function for Y

$$
\begin{aligned}
P[Y = y_i] &= g_Y(y;\lambda,\phi) \\
&= \frac{\lambda^y}{y!\Gamma(1/\phi)\phi^{1/\phi}} \frac{\phi^{y+1/\phi}\Gamma(y+1/\phi)}{(\lambda\phi+1)^{y+1/\phi}} \\
&= \frac{\Gamma(y+1/\phi)}{\Gamma(1/\phi)y!} \frac{1}{(\lambda\phi+1)^{1/\phi}} \left(\frac{\lambda\phi}{\lambda\phi+1}\right)^y \\
&= \binom{y+1/\phi-1}{y} \frac{1}{(\lambda\phi+1)^{1/\phi}} \left(\frac{\lambda\phi}{\lambda\phi+1}\right)^y, \; for\ y = 0,1,2,\ldots
\end{aligned}
\tag{6}
$$

where we have used the convention

$$
\binom{z}{y} = \frac{\Gamma(z+1)}{\Gamma(z+1-y)y!} \tag{7}
$$

The marginal distribution of Y is a negative binomial distribution, $Y \sim \mathrm{NB}\left(1/\phi, 1/(\lambda\phi+1)\right)$

Inference on individual group means

Consider the hierarchical Poisson-Gamma model in (5), and assume that a value $Y = y$ has been observed. Then the conditional distribution of u for given $Y = y$ is a Gamma distribution,

$$u|Y = y \sim \mathrm{G}\left(y + 1/\phi, \phi/(\lambda\phi + 1)\right) \tag{8}$$

with mean

$$\mathrm{E}[u|Y = y] = \frac{y\phi + 1}{\lambda\phi + 1} \tag{9}$$

and variance

$$\mathrm{V}[u|Y = y] = \frac{(y\phi^2 + \phi)}{(\lambda\phi + 1)^2} \tag{10}$$

Threshold calculation

Here should be something about threshold calculation

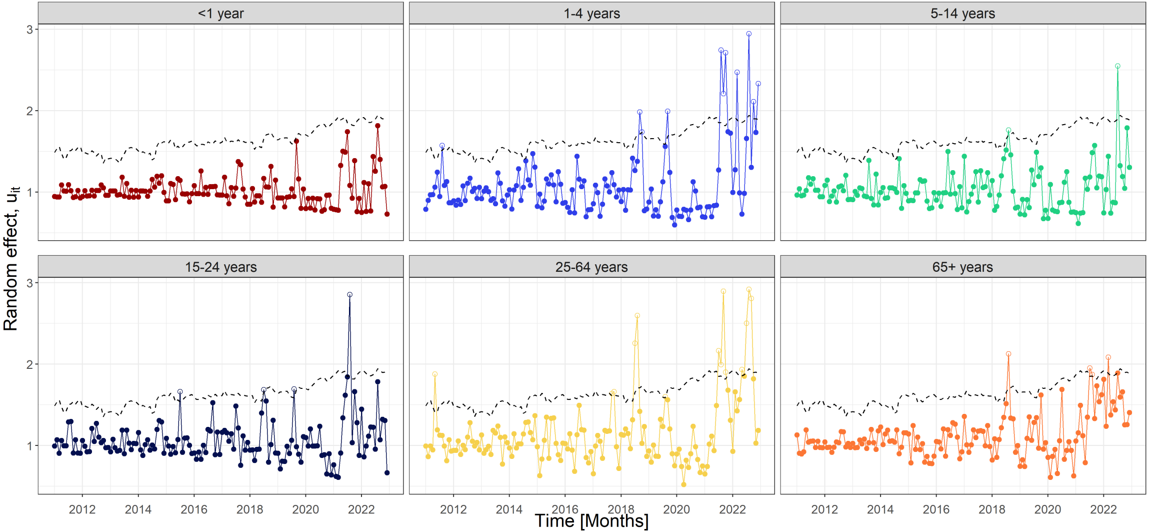
Results

Parameter	Estimate	Std. Error
$\beta_{<1 year}$	11.23	0.08
$\beta_{1-4 years}$	13.94	0.06
$\beta_{5-14 years}$	14.30	0.07
$\beta_{15-24 years}$	14.42	0.07
$\beta_{25-64 years}$	16.87	0.06
$\beta_{65+ years}$	15.24	0.06
ϕ	0.45	0.03

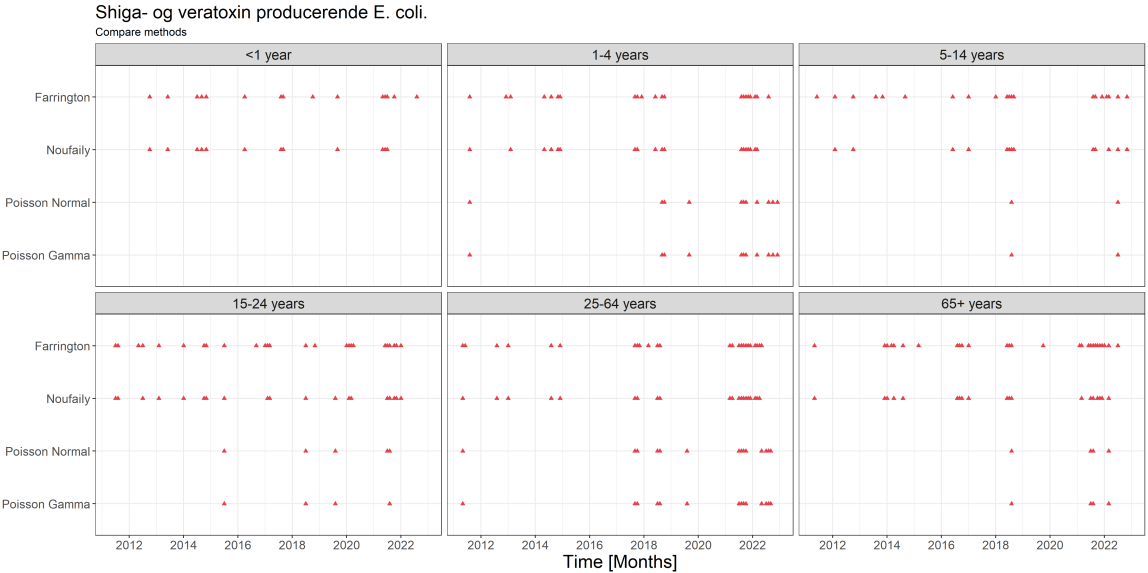
Results

Shiga- og veratoxin producerende E. coli.

Compound Poisson Gamma model



Comparison of methods



References

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. (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.

Proof - Probability function for Y

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

$$f_{Y|u}(y; \lambda, u) = \frac{(\lambda u)^y}{y!} \exp(-\lambda u) \tag{11}$$

and the probability density function for the distribution of u is

$$f_u(u; \phi) = \frac{1}{\phi \Gamma(1/\phi)} \left(\frac{u}{\phi} \right)^{1/\phi - 1} \exp(-u/\phi) \tag{12}$$

Proof - Probability function for Y

Given (11) and (12), the probability function for the marginal distribution of Y is determined from

$$\begin{aligned} g_Y(y;\lambda,\phi) &= \int_{u=0}^{\infty} f_{Y|u}(y;\lambda,u) f_u(u;\phi)\,du \\ &= \int_{u=0}^{\infty} \frac{(\lambda u)^y}{y!}\exp(-\lambda u)\frac{1}{\phi\Gamma(1/\phi)}\left(\frac{u}{\phi}\right)^{1/\phi-1}\exp(-u/\phi)\,du \\ &= \frac{\lambda^y}{y!\Gamma(1/\phi)\phi^{1/\phi}}\int_{u=0}^{\infty} u^{y+1/\phi-1}\exp\big(-u(\lambda\phi+1)/\phi\big)\,du \end{aligned} \tag{13}$$

Proof - Probability function for Y

In (13) it is noted that the integrand is the *kernel* in the probability density function for a Gamma distribution, $\mathrm{G}\left(y+1/\phi, \phi/(\lambda\phi+1)\right)$. As the integral of the density shall equal one, we find by adjusting the norming constant that

$$\int_{u=0}^{\infty} u^{y+1/\phi-1} \exp\left(-u/\Big(\phi/(\lambda\phi+1)\Big)\right) du = \frac{\phi^{y+1/\phi}\Gamma(y+1/\phi)}{(\lambda\phi+1)^{y+1/\phi}} \tag{14}$$

and then (6) follows

Proof - Conditional distribution of Y

The conditional distribution is found using Bayes Theorem

$$
\begin{aligned}
g_u(u|Y=y) &= \frac{f_{y,u}(y,u)}{g_Y(y;\lambda,\phi)} \\
&= \frac{f_{y|u}(y;u)g_u(u)}{g_Y(y;\lambda,\phi)} \\
&= \frac{1}{g_Y(y;\lambda,\phi)}\left(\frac{(\lambda u)^y}{y!}\exp(-\lambda u)\frac{1}{\phi\Gamma(1/\phi)}\left(\frac{u}{\phi}\right)^{1/\phi-1}\exp(-u/\phi)\right) \\
&\propto u^{y+1/\phi-1}\exp\big(-u(\lambda\phi+1)/\phi\big)
\end{aligned}
\tag{15}
$$

We identify the *kernel* of the probability density function

$$
u^{y+1/\phi-1}\exp(-u(\lambda\phi+1)/\phi) \tag{16}
$$

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