Biometry Research Unit
Internal Report 2003-02

R0 estimation by the martingale method

Michael Höhle

Danish Institute of Agricultural Sciences
Research Centre Foulum

There are two publication series of Biometry Research Unit at the Danish Institute of Agricultural Science: Technical Reports and Internal Reports.

Technical Reports: This series aims to cover the publication of pre-prints and research papers in an advanced stage of elaboration. The material published should have some original content, but reviews are also suitable. There are, in principle, no size limitations and extended versions with full details that are referenced later in the final publication are acceptable. The authors assume entirely the responsibility regarding the quality and validity of the published material. There are no editorial checks. The publication in the series does not disqualify the material for later presentation to a journal that requires originality. The preferred language in this series is English. The reports are distributed through a mailing list. Copies can be obtained from the editor.

Internal Reports: This is an informal series intended to publish material of some interest that are not suitable for publishing in the technical report series. Typical articles are: documentation of useful programs or macros, pedagogic and divulgation articles in biostatistical methods, reports of interesting consultation cases, etc. The reports are distributed by the authors.

An updated list of the reports and a complete collection of both series will be kept at the departments library and the central library at the Research Centre Foulum. Copies of the papers of both series can be obtained from the editor. The reports may also be downloaded in pdf-format from the webpage:
http://www.jbs.agrsci.dk/Biometry/Reports/

Editor:
Ulrich Halekoh
Biometry Research Unit
Danish Institute of Agricultural Sciences
Research Centre Foulum

Internal Report 2003-02

R0 estimation by the martingale method

Michael Höhle

Ministry of Food, Agriculture and Fisheries
Danish Institute of Agricultural Sciences
Biometry Research Unit
Research Centre Foulum
P.O. Box 50 • DK-8830 Tjele
Denmark

R_0 estimation by the martingale method

Michael Höhle
hoehle@dina.kvl.dk

4th March 2003

Abstract

This is a short note about the martingale method of moment estimation of R_0 for the general epidemic model from final size data. A equation is derived in [Becker, 1989], but the "unlikely" case of all susceptible becoming infectious is not treated in much detail. In this case, a correction term has to be used due to infectivity wasted as nobody is susceptible anymore. But how this correction should be computed is somewhat unspecified. This note covers the missing derivations and obtains a similar expression as in [De Jong and Kinman, 1994], who do not show any derivations either.

1 Introduction

The martingale method of moment is a method to estimate R_0 of an epidemic, when only final size data is available. It is documented in [Becker, 1989], which although does not deal too much with the "unlikely" case that all susceptible turn infectious during the epidemic. Unfortunately this is often the case in disease transmission experiments performed in a veterinarian context. The standard reference for R_0 estimation in disease transmission experiments is [De Jong and Kinman, 1994], who apply a correction based on the fraction of infectivity lost. How this scheme can be derived from the underlying martingales is although unclear.

In the following the derivations necessary to obtain their result are shown by following the approach of [Becker, 1989]. Notation is a mix of [Andersson and Britton, 2000; Becker, 1989] and the article written by Höhle et. al.

2 The model

Consider a community of initially $S(0) = n$ susceptible and $I(0) = m$ infectives and let $N(t)$ be the number of individuals who have become infected during $(0, t]$. Set $N(0) = 0$ and introduce $R(t)$ to denote the number of removed in a classical SIR notion. Let $H(t)$ be the history of the process (N, I, R) up to time t. The

progress of the epidemic is described as follows.

$$\begin{aligned} P(dN(t) = 1,\, dR(t) = 0 \mid H(t)) &= \beta I(t)S(t)\,\mathrm{dt} \\ P(dN(t) = 0,\, dR(t) = 1 \mid H(t)) &= \gamma I(t)\,\mathrm{dt} \\ P(dN(t) = 0,\, dR(t) = 0 \mid H(t)) &= 1 - \beta I(t)S(t)\,\mathrm{dt} - \gamma I(t)\,\mathrm{dt} \end{aligned}$$

For the above SIR epidemic the basic reproduction ratio, also known as the threshold parameter or R_0, is given as β/γ.

3 Inference about R_0

Assume that the only information available is $S(0)$, $I(0)$, $S(T_N)$, and $I(T_N)$, where T_N is the time where the epidemic has ceased. The aim of this section is to estimate R_0 from this data.

To do this let M_1 and M_2 be the two processes

$$\begin{aligned} M_1(t) &= N(t) - \int_0^t \beta I(x)S(x)\,\mathrm{dx} \\ M_2(t) &= R(t) - \int_0^t \gamma I(x)\,\mathrm{dx}, \end{aligned}$$

which can be shown to be zero mean martingales with respect to $H(t)$. Because $S(t)$ and $I(t)$ are often not observable throughout the epidemic, it is necessary to construct martingales only involving observable quantities before a method of moment estimation can be applied.[1]

Let

$$B(x) = \frac{J(x-)}{S(x-)}, \text{where } J(x) = I(S(x) > 0),$$

and define $B(x) = 0$ when $J(x-) = 0$.

Now define the process $M_1^*(t)$ obtained by integrating the predictable function $B(x)$ wrt. M_1.

$$M_1^*(t) = \int_0^t B(x)dM_1(x) = \int_0^t B(x)dN(x) - \beta \int_0^t I(x)J(x)\,\mathrm{dx}$$

is a zero mean process.[2] Now, construct a martingale, such that R_0 becomes part of the equation. Observations in [Becker, 1989, p.149] lead to $M = M_1^* - R_0 M_2$

[1] In general, method of moments estimation is claimed to be worse than MLE due to bias etc. But we construct a score process, such that it evaluated at the true parameter values it becomes a zero mean martingale. [Andersson and Britton, 2000, p.89]. This also gives the connection to the likelihood, because similar results for β or γ can be derived using MLE.

[2] Integrating a predictable function wrt. a zero mean martingales yields zero mean martingale. Can be derived in some way from Theorem II.3.1 in [Andersen *et al.*, 1993]. Stated by Becker [1989, p.147]

as crafty proposal of a suitable zero mean martingale, i.e. [3]

$$M(t) = \int_0^t B(x)dN(x) - R_0 R(t) + \beta \int_0^t I(x)(1 - J(x))\,\mathrm{dx}\,. \qquad (1)$$

It can be shown that the first term is completely determined by $S(t-)$. If a suitable time t can be found, where the number of removed is known, the first two terms in (1) would therefore be deducible by exploiting knowledge about $S(0)$. I will assume that the final size of the epidemic is observable at the end of the outbreak. The only nuisance left is thus the third term in $M(t)$, which is dealt with by introducing a suitable stopping time.

Let T_N be the earliest time, where there are either no more susceptible left or the number of infectious has dropped to zero.

$$T_N = \inf\{t \geq 0 \mid (S(t) = 0) \vee (I(0) + N(t) - R(t) = 0)\,\}$$

Note that if $S(T_N) > 0$ then $\int_0^{T_N} I(x)(1 - J(x))\,\mathrm{dx} = 0$. By equating Equation 1 to its mean zero[4] and solve with respect to R_0 one obtains

$$\hat{R}_0 = \frac{\int_0^t B(x)dN(x)}{R(T_N)} = \frac{1}{R(T_N)} \sum_{i=S(T_N)+1}^{S(0)} \frac{1}{i}. \qquad (2)$$

The problem is if $S(T_N) = 0$. Here Becker [1989, p.150] writes about T_N not being identifiable, which I think is wrong. In my opinion, the problem is a mathematical one, because $I(x)(1-J(x))$ is not zero anymore, thus making it impossible to solve for R_0 in (1). Let T_R be the occurrence time of the last recovery,

$$T_R = \inf\{t \geq 0 \mid I(0) + N(t) - R(t) = 0\,\}.$$

Note, if $S(T_N) = 0$ then $T_R > T_N$, otherwise $T_N = T_R$. The assumptions about observability were such, that it was only at T_R and beyond that we could observe $R(t)$.

The intuition is now that after $S(t) = 0$, all excreted infectivity is wasted. It is therefore necessary to correct for this, which Becker [1989] does by simply replacing $R(T_N)$ with $R(T_R) - Z$ in Equation 2. The quantity Z is being set to "the size of the last generation of cases, a quantity which one possibly can determine". If $S(T_N) > 0$ then $Z = 0$. Looking at it strictly mathematically, Z could be the correction in the equation $R(T_R) = R(T_N) + Z$. That makes sense, but why all the trouble if, as in [De Jong and Kinman, 1994], it is possible to observe $R(T_N)$, and hence (2) can be used immediately?[5] Kroese and De Jong [2001] are

[3] Linear combination of martingales is still a martingale.

[4] This is why the method is called method of moments

[5] Well, there is one problem...$R(T_N) = 0$ in the experiment. The Figure showing $I(t)$ in [De Jong and Kinman, 1994] gives $I(t)$ as a function of time. With an assumption that recoveries and infection do not occur on the same day and a (questionable) assumption that the sampling frequency of one day exactly sufficiently covers the continuous time event occurrences, it is possible to deduce $S(t)$ from the graph. This is all we need to estimate $\hat{\beta}$ and $\hat{\gamma}$ separately to get a better estimate on $\hat{R}$.

a bit more explicit by saying that "a choice has to be made for Z, which will be debatable, as the moment an animal is infected or stops being infective cannot be determined exactly". They quote Becker for recommending a maximum likelihood approach if all turn infectious. This boils down to exploiting the final size equation, e.g. [Andersson and Britton, 2000, Theorem 2.2], to derive the probability of obtaining a specific final size as function of R_0. A maximum likelihood estimate can be obtained by numerically optimizing for R_0 in this expression, which will although only work numerically for relatively small populations. For more details see [Andersson and Britton, 2000; Kroese and De Jong, 2001].[6]

To treat the $S(T_N) = 0$ case formally, let

$$c = \frac{\int_0^{T_N} I(x)\,\mathrm{dx}}{\int_0^{T_R} I(x)\,\mathrm{dx}}$$

be the amount of infectivity excreted while there still are susceptibles. In analogue with (1) a zero mean martingale is constructed with the aim to cancel out the nuisance introduced by the 3rd term in (1) once $S(t) = 0$. Let $M' = M_1^* - c\theta M_2$, then

$$M(T_R) = 0 \Leftrightarrow \int_0^{T_R} B(x)dN(x) - c\theta R(T_R) + \beta \int_0^{T_R} I(x)(c - J(x))\,\mathrm{dx} = 0$$

Observe that

$$\int_0^{T_R} I(x)(c - J(x))\,\mathrm{dx} = c\int_0^{T_R} I(x)\,\mathrm{dx} - \int_0^{T_R} I(x)J(x) = 0,$$

due to the choice of c. As a result hereof,

$$\hat{R}_0 = \frac{\int_0^t B(x)dN(x)}{cR(T_N)} = \frac{1}{cR(T_N)} \sum_{i=S(T_N)+1}^{S(0)} \frac{1}{i}, \tag{3}$$

which corresponds to the equation used in [De Jong and Kinman, 1994][7] Recasting to the Z correction introduced by Becker, $Z = (1 - c)R(T_R)$, which is then used in conjunction with [Becker, 1989, Equation 7.4.5] to derive the standard deviation of $\hat{R}_0$ as

$$s.e.(\hat{R}_0) = \frac{1}{cR(T_R)} \left[\left(\sum_{i=S(T_N)+1}^{S(0)} \frac{1}{i^2} \right) + c\hat{R}_0^2 R(T_R) \right]^{\frac{1}{2}}. \tag{4}$$

[6] Another point is that the method of moment method is rather free of distributional assumptions, even though I have exploited the Markovian model in my derivations. A more rigorous treatment of this subject is although beyond the scope of this note.

[7] It is though never explicitly written in the article.

4 Implementation

The R file `dejong.R` handles the above estimation, which can also be combined to cover several experiments. Examples are given in the code to estimate R_0 for the data presented in [De Jong and Kinman, 1994; Dewulf *et al.*, 2001].

References

P.K. Andersen, Ø. Borgan, R.D. Gill, and N. Keiding. *Statistical Models Based on Counting Processes*. Springer-Verlag, 1993.

H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*. Number 151 in Lecture notes in statistics. Springer, 2000.

N. G. Becker. *Analysis of Infectious Disease Data*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1989.

M.C.M. De Jong and T. G. Kinman. Experimental quantification of vaccine-induced reduction in virus transmission. *Vaccine*, 12:761–766, 1994.

J. Dewulf, H. Laevens, F. Koenen, H. Vanderhallen, K. Mintiens, H. Deluyker, and A. de Kruif. An experimental infection with classical swine fever in E2 sub-unit marker-vaccine vaccinated and in non-vaccinated pigs. *Vaccine*, 19:475–482, 2001.

A.H. Kroese and M.C.M. De Jong. Design and analysis of transmission experiments. In F.D. Menzies and S.W.J. Reid, editors, *Proceedings of a meeting held at the Golden Tulip Conference Centre, Leeuwenhorst, Noordwijekerhout, The Netherlands*, pages 21–37. Society for Veterinary Epidemiology and Preventive Medicine, 2001.

Recent Publications in the Internal Report Series of the Biometry Research Unit:

2003-02) Höhle , M. (2003). R0 estimation by the martingale method

2003-01) Jørgensen, E. (2003). Foderforbrug pr kg tilvækst hos slagtesvin. Fordeling mellem forbrug til vedligehold og til produktion i besætninger under den Rullende Afprøvning.

2002-06) Højsgaard, S. (2002). mimR (Version 1.0)- An interface from R to MIM for graphical modelling in R

2002-05) Jørgensen, E. (2002). Evaluation of height measurement for automatic weight assessment of slaughter pigs

2002-04) Foldager, L. (2002). Kvantificering af Hydrauliske Egenskaber: Vandretention og Hydraulisk Ledningsevne

2002-03) Højsgaard, S. and Erik Jørgensen (2002). Mixed Models Applied in Agricultural Sciences II based on A Course in Mixed Models for Use in Research in Animal Nutrition and Physiology

2002-02) Foldager, L. (2002). Analyse af sammenhængen mellem hjertefrekvens og varmeproduktion hos drægtige søer - en anvendelse af mixed models

2002-01) Foldager, L., Labouriau, R. and Danfær, A. (2002). An emprirical study of a pig simulation model for feed evaluation.

2001-06) Højsgaard, S. and Jørgensen, E. (2001). Mixed Models applied in Agricultural Sciences Based on A Course in Mixed Models for use in Animal Health and Animal Welfare Research.

2001-05) Højsgaard, S. and Jørgensen, E. (2001). Note Concerning Analysis of Stratified Data – Objective Evaluation of Health Status in Danish Pig Herds.

2001-04) Højsgaard, S. and Jørgensen, E. (2001). Lecture Slides on Mixed Models Based on A Course in Mixed Models for Use in Animal Health and Animal Welfare Research.

2001-03) Kristensen, K. (2001). A collection of some statistical issues to consider when testing for GM seeds in conventional seed lots.

2001-02) Hansen, B. (2001). Approximate Standard Errors And the Use of Bootstrapping.

2001-01) Foldager, L. (2001) Usikkerhed på optimal kvælstofmængde – Bootstrap metoder.

2000-05) Foldager, L. (2000) Usikkerhed på optimal kvælstofmængde – variationsanalyse.

2000-04) Larsen, B., Nielsen, L.N., Labouriau, R. and Lundbye-Christensen, S. (2000). A binomial state-space model for studying the occurrence of respiratory disease in Danish diary calves.

2000-03) Labouriau, R. and Elmholt, S. (2000). Comparing fungal abundance in a multi-soil study: an application of generalized linear mixed models.

2000-02) Jørgensen, E. (2000). Elements of Bayesian network specification in an animal health economy research project.

2000-01) Labouriau, R., Schulin-Zeuthen, M. and Danfær, A (2000). Statistical analysis of pigs development: An application of Richards regression models.

1999-04) Hansen, B. (1999): On curve regristration and functional data analysis. pp. 141

1999-03) Laboriau, R., Andersen, J-O. and Nielsen, M. (1999): Statistical analysis of biocristallazation patterns of carrots extracts

Recent Publications in the Technical Report Series of the Biometry Research Unit:

2002-03) Foldager, L. and Pedersen, J. (2002). Spatial Distribution of Detected Single Fish. An Application of the Log Gaussian Cox Process.

2002-02) Tøgersen, F. A. and Waagepetersen, R. (2002). Statistical modelling and deconvolution of yield meter data.

2002-01) Jørgensen, E. (2002). Circadian variation. Estimation of a time maximum.

2000-04) Hansen, M.B., Møller, J. and Tøgersen, F. A. (2000). Bayesian contour detection in a time series of ultra-sound images through dynamic deformable template models.

2000-03) Waagepetersen, R. and Sørensen, D.(2000). A tutorial on reversible jump MCMC with a view towards applications in QTL-mapping.

2000-02) Jørgensen, B. and Labouriau, R. (2000). Exponential Families and Theoretical Inference.

2000-01) Højsgaard, S. (2000). Statistical inference in context specific interaction models for contingency tables.

1999-03) Kristensen, K. (1999). Statistical methods for analysing genotype environment interactions in variety trials.

1999-02) Labouriau, R (1999). Estimating functions for semiparametric models with moment restrictons.

1999-01) Nielsen, J., Studnitz, M. og Søegaard (1999) On the use of stereological methods in the agricultural sciences.

1998-04) Baddeley, A. J., Møller, J., and Waagepetersen, R. (1998): Non- and semi-parametric estimation of interaction in inhomogeneous point patterns.

1998-03) Pedersen, A. R. (1998). Measuring the nitrous oxide emission rate from the soil surface by means of the Cox, Ingersoll & Ross process.

1998-02) Labouriau, R & Høyer, L. (1998): A note on the temperature dependence of ethylene induced leaflet abscission in "Radamarchera sinica".

1998-01) Skjøth, F. & Værbak, S. (1998): A keeping quality study on "Primula vulgaris" based on a Markov model with aggregated data.

1997-05) Skjøth, F. & Lundbye-Christensen, S. (1997): An investigation of a State-space model for Longitudinal count data, with an application to the modelling of cucumber yield. (Second edition)

1997-04) Jørgensen, B. (1997): The Theory of Dispersion Models.

1997-03) Labouriau, R. (1997): Estimating functions for semiparametric models.

1997-02) Skjøth, F. & Lundbye-Christensen, S. (1997): An investigation of a State-space model for Longitudinal count data, with an application to the modelling of cucumber yield.

1997-01) Labouriau, R. (1997): The Laplace transform and polynomial approximation in L^2.

1996-04) Nielsen, J. (1996): The volume of isotropic random projections of simplices.

1996-03) Labouriau, R. (1996): Path and functional differentiability.

1996-02) Højsgaard, S. (1996): Split models for contingency tables.

1996-01) Labouriau, R. (1996): Semiparametric Models with L^2 Restrictions - part I.