Statistical Science
2022, Vol. 37, No. 3, 394–410
https://doi.org/10.1214/21-STS831

Modeling the Occurrence of Events Subject to a Reporting Delay via an EM Algorithm

Roel Verbelen, Katrien Antonio, Gerda Claeskens and Jonas Crevecoeur

Abstract. A delay between the occurrence and the reporting of events often has practical implications such as for the amount of capital to hold for insurance companies, or for taking preventive actions in case of infectious diseases. The accurate estimation of the number of incurred but not (yet) reported events forms an essential part of properly dealing with this phenomenon. We review the current practice for analysing such data and we present a flexible regression framework to jointly estimate the occurrence and reporting of events. By linking this setting to an incomplete data problem, estimation is performed via an expectation-maximization algorithm. The resulting method is elegant, easy to understand and implement, and provides refined insights in the nowcasts. The proposed methodology is applied to a European general liability portfolio in insurance.

Key words and phrases: EM algorithm, nowcasting, Poisson regression model, reporting delay.

1. INTRODUCTION

This paper reviews and extends the literature on statistical models for problems where individuals (or objects) under study experience two events. The first, also called the initiating or primary, event occurs at time x, and the second, the so-called secondary or consequent, event only occurs at a later time $s \geq x$. The presence of the delay $u = s - x$ between the two events leads to statistical challenges, because the currently observed number of primary events is right-censored while observation delays are right-truncated, see, for example, the early contributions by Lagakos, Barraj and De Gruttola (1988), Kalbfleisch and Lawless (1989), Harris (1990) and Kalbfleisch and Lawless (1991) for an introduction to the field. The term *back-calculation* (Brookmeyer and Gail, 1988, Bacchetti, Segal and Jewell, 1993) refers to the reconstruction of the past history of first events that must have occurred to give rise to the observed pattern of second event cases, under the assumption of a known delay distribution. Nowadays, *nowcasting* is often used for estimating the current number of first events using only the available partial information on the reported or registered secondary events (see Höhle and an der Heiden, 2014, van de Kassteele, Eilers and Wallinga, 2019, Bastos et al., 2019, for recent examples of nowcasting problems in epidemiology).

Such delays occur in different ways and in a variety of subject areas. In an insurance setting a claim is only reported some time after its occurrence, because the damage was not immediately noticed or the insured needed some time to file the claim to the insurance company. A proper estimation of these unreported claims is important, since financial regulations force insurance companies to hold sufficient capital reserves to be able to fulfill their future liabilities with respect to such claims. Jewell (1989) sketches first contributions to the modeling of these occurred- or incurred-but-not-reported (OBNR or IBNR) claims in an insurance context. In disease modeling at least two examples of these delays are studied in the literature. In the first, x represents the time of diagnosis of a case (or another relevant event, like hospital admission

Roel Verbelen is Post-doctoral Researcher, Faculty of Economics and Business, KU Leuven, Belgium, LStat, Leuven Statistics Research Centre, KU Leuven, Belgium, LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium (e-mail: roel.verbelen@kuleuven.be). Katrien Antonio is Professor, Faculty of Economics and Business, KU Leuven, Belgium, Faculty of Economics and Business, University of Amsterdam, The Netherlands, LStat, Leuven Statistics Research Centre, KU Leuven, Belgium, LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium (e-mail: katrien.antonio@kuleuven.be). Gerda Claeskens is Professor, Faculty of Economics and Business, KU Leuven, Belgium, LStat, Leuven Statistics Research Centre, KU Leuven, Belgium (e-mail: gerda.claeskens@kuleuven.be). Jonas Crevecoeur is Post-doctoral Researcher, Faculty of Economics and Business, KU Leuven, Belgium, LRisk, Leuven Research Center on Insurance and Financial Risk Analysis, KU Leuven, Belgium (e-mail: jonas.crevecoeur@kuleuven.be).