Insurance: Mathematics and Economics 53 (2013) 237–251

Contents lists available at SciVerse ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A heavy traffic approach to modeling large life insurance portfolios

Jose Blanchet [a], Henry Lam [b,*]

[a] Department of Industrial Engineering and Operations Research, Columbia University, 120th Street, New York, NY 10027, USA

[b] Department of Mathematics and Statistics, Boston University, 111 Cummington Street, Boston, MA 02215, USA

HIGHLIGHTS

- We propose a bottom-up approach to model large life insurance portfolios.
- We use heavy-traffic approximation to derive and justify the structure of the risk processes.
- The risk processes are shown to depend on mortality and insurance contract structure in a tractable manner.
- We formulate and compute ruin probability that takes actuarial reserve into account.
- We identify explicitly the temporal and cross-sectional correlation structure of the derived risk processes.

ARTICLE INFO

Article history:
Received February 2011
Received in revised form
April 2013
Accepted 30 April 2013

ABSTRACT

We explore a new framework to approximate life insurance risk processes in the scenario of plentiful policyholders, via a bottom-up approach. Given the insurance contract structure, we aggregate the balance of individual policy accounts, and derive an approximating Gaussian process with computable correlation structure. The methodology is borrowed from heavy traffic theory in the literature of many-server queues, and involves the so-called fluid and diffusion approximations. Our framework is different from the individual risk model in that it takes into account the time dimension and the specific policy structure including the premium payments. It is also different from classical risk theory in that it builds the risk process from micro-level contracts and parameters instead of assuming aggregated claim and premium processes outright. As a result, our approximating process behaves differently depending on the issued contract structure. We also illustrate the flexibility of our approach by formulating a finite-horizon ruin problem that incorporates actuarial reserve in the consideration.

The study of risk processes is a central topic in actuarial science. Most of the literature focuses on the calculation of ruin probability and deficits (or overshoots) at the time of ruin, as well as the optimal control of premiums, reinsurance levels, and investment allocation. These questions have been studied under a variety of stochastic settings, from the classical Cramer–Lundberg approximation to diffusion processes. The central theme is that random-walk-type models, with a negatively drifted premium process and a jump process of claims, provide a rich framework to allow plenty of extensions, modifications and problem formulations (see, for example, Asmussen and Albrecher, 2010 for the survey on ruin probability calculations, and Schmidli, 2008 for the counterpart in stochastic control problems).

In this paper, we take a different view from the existing literature. Rather than focusing on the computation of risk-related quantities, we explore the question of the construction of risk process itself. The approach we use is bottom-up: given the structure and parameters of the individual insurance contracts, how does the risk process of the insurer look like on an aggregate scale?

Naturally, the risk process under this framework is the sum of all the individual accounts i.e. the balances of policyholders who entered contract with the insurer over time. For actuaries, this points to the standard one-period individual and collective risk models. However, these standard models do not consider the time dimension. This in turn also restrains the power of such models to capture the specific contract structure involved e.g. the premium payments.

In this regard, our work can be seen as a generalization of the standard risk models to a process-level approximation. Of course, mere summation of all individual accounts might end up getting an unpleasant process that is hardly computable. To tackle this issue, we borrow techniques in so-called heavy traffic theory in the queueing literature. The basic idea is that under the assumption of large number of customers or policyholders, one can approximate the functionals of these policyholders' statuses using fluid and diffusion approximations. In the statistics literature, these correspond

* Corresponding author.
E-mail addresses: jose.blanchet@columbia.edu (J. Blanchet), khhlam@gmail.com, khlam@bu.edu (H. Lam).

0167-6687/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.insmatheco.2013.04.011

to stochastic-process versions of Law of Large Numbers and Central Limit Theorems. With the sheer scale of major insurance companies, the assumption of plentiful policyholders is sensible, and so these approximation techniques can be used. As we will see, these heavy traffic approximation would then lead to a Gaussian process that is as analyzable as many standard processes used in the current risk theory literature. In particular, the correlation structure of this Gaussian process is explicitly computable given the contract structure (see Section 4). To illustrate our argument on tractability, we formulate a finite-horizon ruin problem based on our Gaussian approximation (see Section 3).

We distinguish our contribution from the classical risk theory and standard actuarial risk models in a few ways. First, our model explains how individual insurance policies lead to certain features of the aggregate risk process. The construction of our risk process depends intricately on the premium and benefit structure of single policies. This means that different types of insurance, such as whole life insurance, term life, endowment etc. would lead to different correlation structure of our resulting Gaussian process. This is in sharp contrast to the current model in risk theory, where premium and claim processes are modeled separately, each as a drifted random walk (or its variants) and marked point process. This feature can potentially provide a framework to analyze the effect of contract structure on the firm-wide risk level. Second, our model allows naturally the incorporation of actuarial reserve in our approximation. Indeed, the finite-horizon ruin problem that we formulate in Section 3 will involve the calculation of prospective reserve. Third, since serial correlation is explicitly computable, this provides a way to capture the fluctuation of our approximating process over time, which can be potentially applicable to dynamically monitoring mismatch on the insurer's balance sheet with regard to statistical error.

In a more organized fashion, we summarize our contributions as follows:

(A) Under the assumption of large number of policyholders, we construct the fluid limit and diffusion limit for the aggregate risk processes. (As we mentioned, these correspond to functional Law of Large Numbers and Central Limit Theorem respectively in the statistics community; throughout the paper we mostly use the former terminology to align with the queueing literature, but will also use the latter interchangeably when necessary.) The risk processes that we are interested in include the insurer's cash level, liabilities, and per basis reserve level. These will be discussed in Section 2. We prove and numerically demonstrate that these risk processes can be approximated by Gaussian processes with certain correlation structures.

(B) Using the theory of Gaussian processes, we illustrate how our result can be used to approximate the ruin probabilities. We model ruin as the situation in which the liabilities surpass the assets (plus the initial capital) within a given time horizon (see Section 3.1). This highlights the flexibility of our methodology in incorporating reserve calculation, and also the dependency on the underlying insurance contracts. In particular, we apply our results to several common types of insurance.

(C) Our diffusion approximation shows how, under the Equivalence Principle, the benefit reserve arises as the fluid limit of the empirical cash level per basis at any point in time (see Section 2). These results, we believe, provide a useful perspective into the basic concepts underlying the definition of benefit reserve; see the discussion following Theorem 1.

(D) We compute the correlation structures of our limiting processes, thereby showing their tractability. In particular, we illustrate how our approach allows to evaluate and compare the autocorrelation (as a function of time) of risk processes with different insurance types; see Section 4.

Let us emphasize that our purpose in applications such as (B) and (C) is to illustrate the concepts behind our ideas, and hence the models we are using in this paper are basic. There are certainly many practical considerations to make the model more realistic. We shall list out these generalizations and more realistic extensions that we believe are worth pursuing in Section 5.

In terms of methodology, as aforementioned, we will invoke primarily the machinery in heavy traffic theory i.e. fluid and diffusion approximations in the queueing literature. The ideas date back to Kingman (1961, 1962) for single-server queues, and they still constitute an active research area among the queueing theorists (see the standard surveys of Whitt, 2002 and Billingsley, 1999 for instance). Under fairly mild assumptions, the tools significantly simplify and single out the important elements of the system dynamics of interest, and provide approximate solutions to many important performance measures (in our context, the ruin probability mentioned in (B) constitutes one such example). More precisely, the results in this paper relate to the analysis of so-called many-server queues, which have been substantially studied in recent years. In these queueing systems, customers arrive and elicit service for a random amount of time, as long as there are available servers. When the number of servers is infinite, every customer can start service right at arrival. Connecting to our work, policyholders can be thought of as customers in the queueing system. While the feature of arrivals is not our focus in this paper, the death time of policyholders is analogous to the end of service, and hence the approximation technique is translatable. Some relevant references on the topic include Pang and Whitt (2010) and Decreusefond and Moyal (2008), which focus on infinite-server models, Halfin and Whitt (1981), Kaspi and Ramanan (2010) and Reed (2009), which study finite but large number of servers in different proportion (or so-called regime) to the number of customers, Puhalskii and Reiman (2000) that study queues with multiclass customers, and Dai et al. (2010) on queues with reneging. The common theme of all these work is the heavy traffic technique being applicable to various features of the queues.

Finally, we discuss two papers that use similar approach and highlight our difference. One is a recent working paper by Bensusan and El Karoui (2009), who propose a microstructural approach to model population dynamics to capture mortality/longevity risk. Their motivation is different from ours: instead of building our mortality distribution microstructurally, we make common assumptions on mortality; instead, our focus is on how this mortality assumption, under the interaction with the contract structure, benefit level and premium calculation, leads to a macroscopic fluctuation of total assets, liabilities and other actuarial quantities. Secondly, we note that diffusion approximation has been invoked by Iglehart (1969) in arguing the use of Brownian motion in modeling insurance risk process. However, he maintained a Cramer–Lundberg framework by assuming compound Poisson claims and constantly drifted premium, and showed that under certain scaling their difference converges to a diffusion process. Contract structure, relation between premium and benefit, and actuarial reserve etc. were not considered in his work.

The organization of this paper is as follows. In Section 1 we lay out our model assumptions and define the key quantities that we approximate. Section 2 is devoted to the statement of our main result and its discussion. Section 3 relates to applications in ruin probability computations and shows some examples. Section 4 identifies the autocorrelation structure of our approximating Gaussian processes. Section 5 discusses some extensions. Appendix constitutes an appendix, which is divided into two parts. The first part discusses basic facts about heavy traffic limit theorems and gives the proof of our main result; the second part contains a discussion on the simulation methodology that is used to generate various examples in this paper.