Bayesian Network Management

Ayalvadi Ganesh*, Peter Green†,
Neil O'Connell, Susan Pitts‡
Basic Research Institute in the
Mathematical Sciences
HP Laboratories Bristol
HPL-BRIMS-97-06
May, 1997

We formulate some general network (and risk) management problems in a Bayesian context, and point out some of the essential features. We argue and demonstrate that, when one is interested in rare events, the Bayesian and frequentist approaches can lead to very different strategies: the former typically leads to strategies which are more conservative.

We also present an asymptotic formula for the predictive probability of ruin (for a random walk with positive drift) for large initial capital and large number of past observations. This is a preliminary investigation which raises many interesting questions for future research.

* Birkbeck College, London, United Kingdom
† University of Bristol, Bristol, United Kingdom
‡University of Cambridge, Cambridge, United Kingdom

1 Introduction

The current trend in telecommunications networks is towards systems which allow diverse traffic types like voice, video and computer data to share the same broadband network. These traffic types differ widely in the bandwidth and storage that they require from the network, in stark contrast to today's telephone networks, where user needs are homogeneous. The task of the network will be to integrate these services in a manner that ensures an adequate quality of service to all users, while at the same time exploiting the benefits of statistical resource sharing. This is not possible without some knowledge of the nature of the traffic generated by each user, a fact which has motivated much recent research on the possibility of estimating relevant characteristics of the traffic. The methods proposed so far typically separate the estimation problem from the network management problem. We propose to integrate the estimation problem with the goals of network management using a Bayesian framework: instead of producing point estimates of traffic characteristics, our aim is to produce a full description of our uncertainty, hence providing more information and coherence for the purpose of network management.

As a motivating example, consider a roulette player who repeatedly bets on black because she believes the wheel to be biased towards black. She starts with b pounds and bets a pound on each spin of the wheel: if it comes up black, she wins a pound; otherwise she loses the pound. Suppose she has observed n previous spins, k of which came up black, and would like to assess the probability of her ultimate ruin based on this observation. If the probability of black coming up, which we will denote by p, were known, then this ruin probability would be given by ρ^b, where $\rho = (1-p)/p$. (Here we are assuming that $p > 1/2$; otherwise ruin is certain.) From her past observations, she could estimate p by $\hat{p} = k/n$, say, and deduce an estimate for her ruin probability by plugging this into the formula. We will denote this estimate by $\hat{\rho}^b$. However, this estimate fails to account for the uncertainty in estimation of p, and, coupled by the extreme skewness of the transformation $p \rightarrow \rho^b$ for large b, gives very misleading inference. A Bayesian perspective leads to a predictive probability for her ultimate ruin, giving a complete measure of her uncertainty. This means putting a prior distribution on p, and computing the posterior distribution of p (given the data). It turns out that, for large values of b, the Bayesian and frequentist approaches lead to very different conclusions. It can be readily checked that, for any (proper) prior for p on $(1/2, 1]$,

$$\lim_{b\to\infty} \frac{1}{b} \log E[\rho^b|\text{data}] = \log r,$$

where r is the essential supremum of the induced prior distribution on ρ. The posterior expectation of the ruin probability is indeed the proper summary of the inference, under any cost structure that is linear in ruin probability; this is reasonable, at least for some telecommunications applications.

On the other hand,

$$\frac{1}{b} \log \hat{\rho}^b = \log \hat{\rho},$$

for all b. Note that for any sensible prior, $r > \hat{\rho}$. The Bayesian perspective is more conservative for larger values of b – the predictive probability of ruin is greater than the frequentist estimate – and the message is loud and clear: with a finite amount of data we cannot confidently extrapolate probabilities of rare events; the chance that the underlying probability of black (p) is much lower than the observed frequency, although it may be small, becomes very relevant when the initial capital (b) is large.

Although this example might seem trivial, it is in fact a canonical model for loss in queueing systems with finite resources; it is also a canonical risk process. What we have described is a qualitative phenomenon, and we expect it to be relevant in general network management situations, for arbitrary traffic types and network configurations. Such features will have implications for decision making in networks (buffer dimensioning, resource allocation, pricing, policing, etc.). They are also clearly relevant to the general problem of risk management, where there are similar decision problems.

In this paper we will formulate the network management problem in a very general context, and discuss the 'nature of the beast'. We will also present an asymptotic formula for studying predictive ruin probabilities. We expect this to be useful for comparing strategies based on the Bayesian paradigm with their frequentist counterparts.

2 Effective bandwidths

The performance of communications networks is analysed using queueing theory. Recent developments in telecommunications have stimulated significant advances in queueing theory involving, in particular, the use of large deviations theory. See Kelly [25], de Veciana and Walrand [9] and Duffield and O'Connell [11] for analyses of a single queue, and Chang [6], de Veciana *et al.* [10], O'Connell [29, 30, 31], Ganesh and Anantharam [19], Ganesh [20], Ganesh and O'Connell [22] and Bertsimas *et al.* [2] for networks. One of the important ideas to emerge from this work is the concept of the *effective bandwidth* of a source. This is a measure of its resource requirements and is directly related to the large deviation rate function associated with the source. In practice, we do not know the effective bandwidth of a source *a priori*, but have to estimate it from observations of the source output. One approach would be to hypothesize a source model, say an ARMA or Markov model, and use traffic data to estimate its parameters. These can be used to evaluate the limit in (1) below, and thus the effective bandwidth. The main disadvantage of this approach, from a practical point of view, is that different kinds of traffic will require different kinds of model: automating the modelling procedure appears impractical, given the diversity of telecommunication traffic sources. Also, the theory of large deviations tells us that for many purposes, there is more information in a complicated model than is needed for the purpose of network management. An alternative approach, based on directly estimating the expression in (1), has been studied by Crosby *et al.* [4, 5], Duffield *et al.* [12, 13] and Ganesh [21]. It has the advantage that it is model-independent, and fast. Estimation of a related quantity, using empirical queue-length distributions, had previously been studied by Courcoubetis *et al.* [3], where the idea of using a non-parametric approach was first introduced.

There have been similar recent developments [8, 14, 15, 16, 32, 33, 34] in the risk theory literature.

Consider a discrete-time queue fed by a single source and having a buffer with the capacity to hold b cells. Let A_n denote the number of cells generated by the source in the n^{th} time slot, with $\{A_n, -\infty < n < \infty\}$ being a (stationary) random process. Given a statistical model of the source, its effective bandwidth is defined, for each $\theta > 0$, by

$$\alpha(\theta) = \frac{1}{\theta} \lim_{n\to\infty} \frac{1}{n} \log E\left[\exp \theta \sum_{i=1}^{n} A_i\right], \tag{1}$$

assuming this limit exists. This quantity has the interpretation that if the queue is served at constant rate $\alpha(\theta)$ (cells per time slot), then the frequency of cell loss is approximately $e^{-\theta b}$. More precisely, if the service rate is c (cells per unit time) and $l(b)$ denotes the frequency of cell-loss in a buffer of size b, then

$$\lim_{b \to \infty} \frac{1}{b} \log l(b) = \delta := \sup\{\theta : \alpha(\theta) < c\}. \qquad (2)$$

In the risk theory context, the *risk reserve* plays the role of the buffer size, b, and the quantity δ is called the *risk adjustment coefficient.* It is therefore essential to estimate δ in order to achieve a desired bound on the frequency of cell loss, in the telecommunications context, or on the probability of ruin, in the insurance context. Effective bandwidths, or equivalently, δ, can be computed from a statistical source model. More importantly from a practical point of view, they can also be estimated non-parametrically from observed source statistics, without reference to a particular model.

3 A Bayesian framework

The task of a telecommunication network will be to provide a guaranteed quality of service (QoS) to users, typically expressed as a bound on the frequency of cell loss. A cell loss rate of about 10^{-8} is considered typical of what users may be prepared to tolerate in a real network. Around such low levels, cell loss rates can be very sensitive to the bandwidth provided. Therefore, *a small underestimate of the effective bandwidth can lead to drastic degradation of the QoS.* This is essentially a reiteration of the observation we discussed in the introduction, using the roulette example. Note that small overestimates are relatively benign, leading only to small losses of utilization efficiency in the network. We propose to take account of this asymmetry by posing the problem within a Bayesian framework. Gibbens *et al.* [24] is the only earlier study, to our knowledge, that considers the problem of network management in a Bayesian framework. It concentrates on the problem of call admission at a single buffer for specific traffic source models. There is also a significant recent literature on the Bayesian analysis of parametric queueing models (see, for example [1, 38, 40]), where many useful calculations are made. In this paper we formulate the problem in a very general framework, using large deviation theory and non-parametric Bayesian methods.

The canonical problem we will consider is that of computing the posterior distribution of δ, defined in (2), and, in particular, the expected frequency of cell loss (or predicive probability of ruin in the risk context). Without loss of generality, we can restrict ourselves to the *i.i.d* case: the case where the A_i's are weakly dependent can be reduced to the *i.i.d.* case by aggregation. Let μ denote the distribution of A_1. Then (1) becomes

$$\alpha(\theta) = \frac{1}{\theta} \log \int e^{\theta x} \mu(dx). \qquad (3)$$

A straightforward frequentist estimate of α (and hence of the frequency of overflow $l(b)$) can be constructed using the empirical distribution of the observations $\{A_1, \ldots, A_n\}$. This has formed the basis of earlier studies [4, 5, 12, 13, 21] (in a queueing context) and [8, 14, 15, 16, 32, 33, 34] (in a risk context). The Bayesian analogue involves putting a prior distribution on the (infinite-dimensional) parameter μ. The Dirichlet process priors introduced by Ferguson [17, 18] are a natural first choice, being a very rich class of conjugate priors with computable posteriors.

Let $n > 0$ and $a = (a_1, \ldots, a_n)$ be given. Suppose Z_i, $i = 1, \ldots, n$ are independent, with $Z_i \sim \mathcal{G}(a_i, 1)$, where $\mathcal{G}(a_i, 1)$ denotes the gamma distribution with shape parameter a_i and scale parameter 1, and $\sim$ denotes equality in distribution (if $a_i = 0$, we take $Z_i \equiv 0$). Let $Z = Z_1 + \ldots + Z_n$. The n-dimensional *Dirichlet distribution* with parameter $a = (a_1, \ldots, a_n)$, denoted $D(a)$, is defined to be the joint distribution of $(Z_1/Z, \ldots, Z_n/Z)$. This is a probability distribution on the n-simplex,

$$S^n = \{\mathbf{x} \in \mathbb{R}^n : x_i \geq 0, i = 1, \ldots, n, \sum_{i=1}^{n} x_i = 1\},$$

and if all the a_i are strictly positive, it can be expressed by the density

$$f(x_1, \ldots, x_{n-1}) = \frac{\Gamma(\sum_{i=1}^{n} a_i)}{\prod_{i=1}^{n} \Gamma(a_i)} \prod_{i=1}^{n-1} x_i^{a_i - 1} (1 - \sum_{i=1}^{n-1} x_i)^{a_n - 1}. \qquad (4)$$

Here $\Gamma(\cdot)$ denotes the gamma function: $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx$, $z > 0$.

Denote by $\mathcal{M}_+(\Omega)$ (respectively $\mathcal{M}_1(\Omega)$) the space of finite non-negative (respectively probability) measures on an arbitrary measurable space Ω. The "*Dirichlet process*" with parameter $\alpha \in \mathcal{M}_+(\Omega)$, which we denote by $\mathcal{D}(\alpha)$, is a probability distribution on $\mathcal{M}_1(\Omega)$, and is characterized as follows. A random probability measure, μ, on Ω has law $\mathcal{D}(\alpha)$ if, and only if, for each finite partition $A_1, \ldots, A_n$ of Ω, the vector $(\mu(A_1), \ldots, \mu(A_n))$ has the n-dimensional Dirichlet distribution $D(\alpha(A_1), \ldots, \alpha(A_n))$.

Suppose $\mathcal{P}$ is a prior distribution on the space Ω and assume $\mathcal{P}$ is a Dirichlet process with parameter α. Then, conditional on observing $\omega_1, \ldots, \omega_n$, it can be shown (see [17, 18]) that the posterior distribution is also a Dirichlet process, but with parameter $\alpha + \sum_{i=1}^{n} \delta_{\omega_i}$, where δ_x denotes the Dirac measure at x. (The Dirichlet processes $\mathcal{D}(\alpha)$, $\alpha \in \mathcal{M}_+(\Omega)$ are a *conjugate* family of priors.) This property facilitates computation of posterior distributions of quantities of interest and is very useful for analysis.

Now consider the problem of estimating the frequency of buffer overflow (or probability of ruin, depending on the context). Let A_i be *i.i.d.*, real-valued, as in Section 2. We fix $\alpha \in \mathcal{M}_+(\mathbb{R})$, and let $\mathcal{D}(\alpha)$ be our prior on the distribution, μ, of the A_i. So, $\mathcal{D}(\alpha)$ is a probability distribution on $\mathcal{M}_1(\mathbb{R})$. Now, conditional on the observations $A_1, \ldots, A_n$, we obtain the posterior distribution $\mathcal{D}(\alpha + \sum_{i=1}^{n} \delta_{A_i})$ for μ. We can use this to compute the posterior distribution of the effective bandwidth by substituting in (3), and the posterior expectated frequency of cell loss, using (2).

4 A useful asymptotic formula

One of the aims of research in this field is to develop network management strategies (tariffing, policing, call admission, resource allocation) based on observed traffic characteristics. It is to be expected that the Bayesian and frequentist approaches will lead to quite different strategies. This was indicated by some easy observations based on the roulette example of Section 1. We will now take a closer look at this example, and make some more refined observations. In particular, we will present an asymptotic formula for the predictive probability of ruin, which we expect to be useful for comparing strategies.

The roulette example is a special case of the general problem described in the last section, with $\Omega = \{\text{Red,Black}\}$. The $\mathcal{D}(\alpha)$ in this case corresponds to a beta distribution on p, the probability of black, with parameters $a_1 = \alpha(\text{Black})$ and $a_2 = \alpha(\text{Red})$, which we assume to be strictly positive. This distribution, denoted $\mathcal{B}(a_1, a_2)$, has density:

$$f_p(x) = \begin{cases} \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)} x^{a_1-1}(1-x)^{a_2-1}, & x \in [0,1], \\ 0, & \text{else}. \end{cases} \quad (5)$$

The posterior distribution of p, given that k of the last n spins have stopped at black, is $\mathcal{B}(a_1+k, a_2+n-k)$. Recall that the probability of eventual ruin is $((1-p)/p)^b \wedge 1$ if p is the probability of black and b denotes the initial capital of the gambler. Hence, the posterior probability of ruin is

$$P(\text{ruin}|\text{data}) =$$

$$\int_0^1 \frac{\Gamma(a_1+a_2+n)}{\Gamma(a_1+k)\Gamma(a_2+n-k)} x^{a_1+k-1}(1-x)^{a_2+n-k-1} \left[\left(\frac{1-x}{x}\right)^b \wedge 1\right] dx. \quad (6)$$

For comparison, the frequentist approach uses the estimate $\hat{p} = k/n$ for the probability of black turning up; the implied ruin probability is

$$P(\text{ruin}|\text{data}) = \left(\frac{1-\hat{p}}{\hat{p}} \wedge 1\right)^b. \quad (7)$$

The Bayesian estimates clearly depend on the choice of prior, as determined by a_1 and a_2. Roughly speaking, if a_1 and a_2 are small compared to n, then the prior is swamped by the data and the conclusion is not sensitive to the exact value of a_1 and a_2. For convenience, we shall simply substitute $a_1 = 0$ and $a_2 = 0$.

In Figure 1, we have plotted the logarithm of $P(\text{ruin}|\text{data})$ for $k/n = 0.6$ and a range of values of the initial capital, b. From top to bottom, the curves correspond to numbers of observations, n, equal to 100, 200, 500 and 1000. The straight line at the bottom corresponds to the frequentist estimate. As expected, the Bayesian estimates of the ruin probability are more conservative, and the effect is more pronounced for larger b. Also, as n increases, so does the range of b for which there is close agreement. These observations suggest that (a suitably scaled version of) the plot of log ruin probability versus b may have an asymptotic shape as b and n both go to infinity, with their ratio fixed. This is indeed true, as can be seen from Figure 2, where we have plotted $(1/n)\log P(\text{ruin}|\text{data})$ against b/n. From bottom to top, the plots correspond to $n = 50$, 100, 500 and 1000. The following asymptotic statement makes this precise; a proof is given in the appendix.

Proposition 1 *Fix $q > 0$ and let $P_{qn}(ruin|\hat{p}_n)$ denote the posterior ruin probability, when the gambler's initial capital is qn, conditional on $\hat{p}_n$ being the observed frequency of black in n spins of the roulette wheel. If $\hat{p}_n \to \hat{p}$ as $n \to \infty$, then*

$$\lim_{n\to\infty} \frac{1}{n} \log P_{qn}(ruin \mid \hat{p}_n) = \begin{cases} 0, & \text{if } \hat{p} < 1/2, \\ h(\hat{p}) - \log 2, & \text{if } \hat{p} - q < 1/2 \le \hat{p}, \\ h(\hat{p}) - h(\hat{p} - q), & \text{if } \hat{p} - q \ge 1/2, \end{cases}$$

where $h(x) = -x \log x - (1-x)\log(1-x)$.

Denote the above limit by $s_{\hat{p}}(q)$. It is easily seen from the above that, for fixed $\hat{p} > 1/2$, $s_{\hat{p}}$ is a convex, non-increasing function of q, and $s'_{\hat{p}}(0) = \log(1-\hat{p}) - \log \hat{p}$. Recall that the logarithm of the ruin probability estimated by the frequentist approach is $qn[\log(1-\hat{p}) - \log \hat{p}]$ and $\hat{p} > 1/2$. Hence, in this asymptotic regime, the Bayesian and frequentist estimates are close to each other for small values of q, but the Bayesian method grows progressively more conservative as q increases.

As one might expect, a similar result holds quite generally. In the setting of Section 3, it is possible to derive an almost sure large deviation principle (LDP) for the posterior distribution of μ (the law of the arrivals, $\{A_i\}$) under reasonable conditions on the prior distribution. Using this, we can deduce a general asymptotic formula for the predictive probability of ruin. Full details are given in [23]; what follows is a summary of the main observations.

Suppose the $\{A_i\}$ have true distribution μ. Under mild conditions on μ,

$$\lim_{n\to\infty} \frac{1}{n} \log P_{qn}(\text{ruin}) = \delta_\mu q,$$

where $P_{qn}(\text{ruin})$ denotes the true ruin probability when the initial capital is qn and δ_μ is some constant. Consider an asymptotic regime where the number of observations, n, and the initial capital, qn, go to infinity with q fixed. Then, almost surely, the logarithm of the predictive probability of ruin, divided by n, converges to a constant which we denote $s_\mu(q)$. The function s_μ is qualitatively similar to its gambler's ruin analogue, $s_{\hat{p}}$. In particular, it is convex and non-increasing, and $s'_\mu(0)$ is equal to δ_μ. Thus, as expected, the Bayesian estimate of the ruin probability is asymptotically guaranteed to be conservative, to a degree which becomes more pronounced as q increases.

5 Directions for future research

We have discussed the problem of managing networks in a manner that exploits the benefits of statistical resource sharing while at the same time guaranteeing a certain quality of service to users. Addressing this problem requires estimating user characteristics on the basis of observed traffic data; in particular, it requires inferences about probabilities of extremely rare events. We argue that the reliability of such inference based on a finite amount of a data is questionable and that therefore we need to be conservative in our estimates. This is particularly true because the consequences of network management decisions based on optimistic estimates lead to serious performance degradation, whereas the consequences of pessimistic estimates are not so serious. This asymmetry leads us to suggest that a Bayesian framework is a natural one to study the related decision problems. Implementing a Bayesian methodology requires a suitable choice of prior; we have argued that the Dirichlet processes provide a rich family of conjugate priors that are well suited to analytical work. We have derived simple asymptotic formulae that can be used to guide network management decisions.

A considerable amount of work remains to be done to establish whether the Bayesian approach offers real advantages in practice. This has to be done on the basis of economic criteria associated with network management strategies (tariffing, policing, call admission, resource allocation). As in the gambler's ruin example we expect the frequentist and Bayesian approaches to lead to very different strategies. An important aspect of future research will be to compare the performance of these strategies, and also of strategies (setting of premiums, etc) which arise naturally in the risk theory setting.

Practical implementation of the Bayesian approach will require fast computation of quantities like the posterior probability of cell loss. In principle, and for exploratory purposes, this can be done by numerical integration. However, we believe that Monte Carlo methods, coupled with importance sampling, will be quicker. There is a very well developed literature on computation of posterior quantities within these frameworks, mostly based on Markov chain Monte Carlo (MCMC) and bootstrap ideas (see, for example, [7, 27, 36, 37, 39]). For the finite mixture models, with weights, parameters and number of components unknown, we can adapt the MCMC approach in Richardson and Green [35]. (This methodology is generic, and is capable of being extended to the other classes of priors mentioned above; in particular, the Dirichlet process version has already been implemented.) The standard computational techniques will all need to be modified, using change of measure ideas from large deviation theory, to take account of the fact that we are interested in rare events. Computational issues will be important both for practical implementation of this approach and for simulation studies.

We have assumed so far that source characteristics are pre-determined. In fact, sources may be able to modify their behaviour in response to network congestion. For example, it may be possible to reduce the rate of a video source at the cost of a degradation in picture quality that is tolerable to the receiver. One way for the network to inform sources about congestion, while at the same time providing them with the incentive to modify their characteristics, is by the use of a *pricing* scheme. Pricing would also help to address another important issue, which is that, when network resources are shared, the misbehaviour of any user adversely affects the quality of service received by other users. One response to this problem has been the suggestion that sources be policed to ensure that they conform to certain agreed constraints on their output. A more efficient approach might be to charge users for violating these constraints, the penalty depending on the extent of violation. Low and Varaiya [28] suggest pricing buffer space and bandwidth separately and allowing users to choose the combination that best suits them. This has the disadvantage that it partitions resources between users, and therefore gives up the benefits of statistical multiplexing. Kelly [26] suggests pricing on the basis of effective bandwidths and illustrates this for sources with a known peak rate. A Bayesian framework is a very natural setting for studying decision problems like pricing and policing.

Figure 1: Plots of $\log P(\text{ruin}|\text{data})$ against the initial capital, b, for $n = 100$, 200, 500 and 1000 (from top to bottom) with $k/n = 0.6$. The straight line at the bottom corresponds to the frequentist estimate.

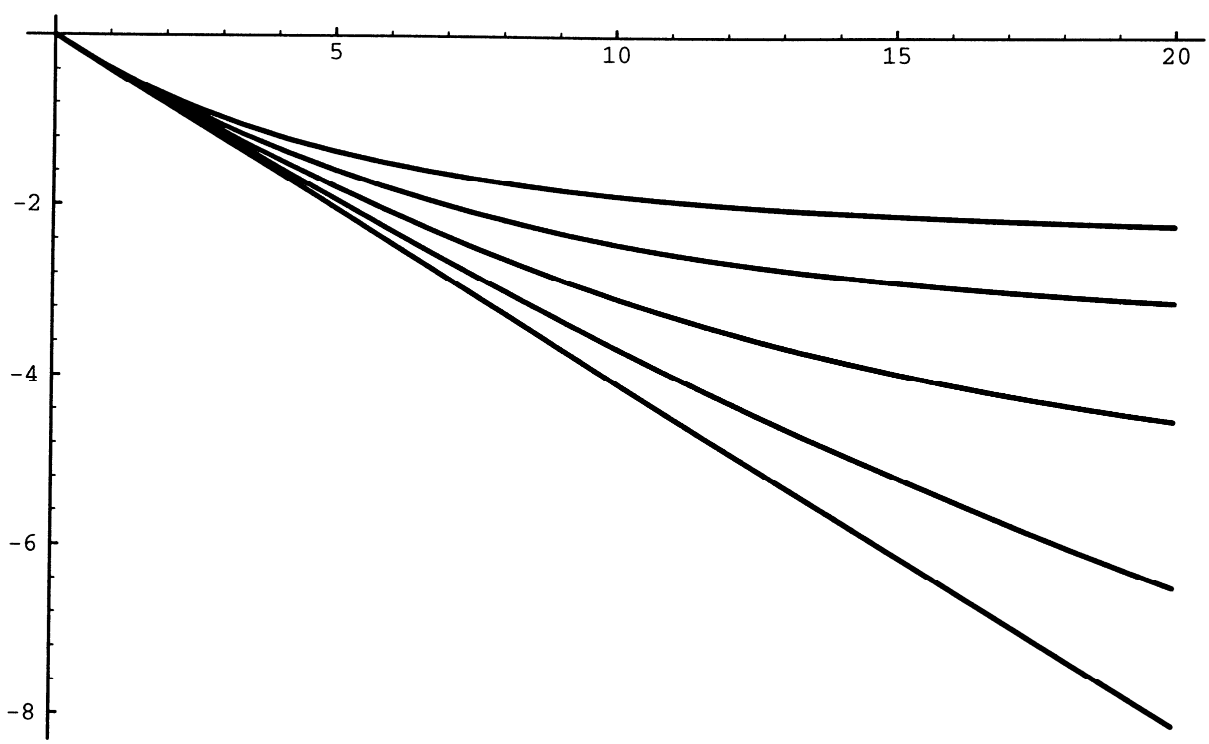
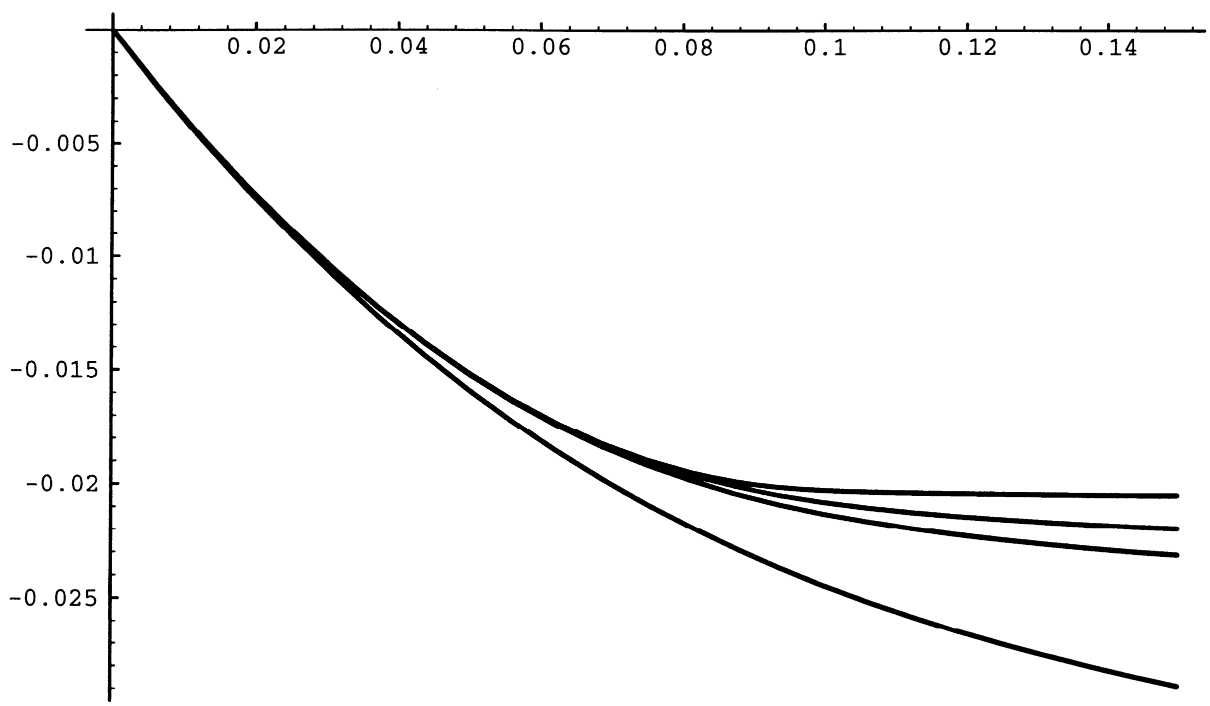


Figure 2: Plots of $(1/n) \log P(\text{ruin}|\text{data})$ against b/n, for $n = 50$, 100, 500 and 1000 (from bottom to top).



References

[1] C.Armero and M.J.Bayarri. Bayesian questions and answers in queues. (With discussion.) In: *Bayesian Statistics*, J.M.Bernardo, J.O.Berger, A.P.David and A.F.M.Smith (eds.). Oxford University Press, 1996.

[2] D. Bertsimas, I. Paschalidis and J. Tsitsiklis, On the large deviations behaviour of acyclic networks of G/G/1 queues. Submitted to *Ann. Appl. Probab.*

[3] C. Courcoubetis, G. Kesidis, A. Ridder, J. Walrand and R. R. Weber. Admission control and routing in ATM networks using inferences from measured buffer occupancy. *IEEE Trans. Comm.* 43 (1995) 1778-1784.

[4] S. Crosby, I. Leslie, J. T. Lewis, Neil O'Connell, R. Russell and F. Toomey. Bypassing modelling: an investigation of entropy as a traffic descriptor in the Fairisle ATM network. *Proc. 12th U.K. Teletraffic Symp.*, London 1995.

[5] S. Crosby, M. Huggard, I. Leslie, J. T. Lewis, F. Toomey and C. Walsh. Bypassing Modelling: Further Investigations of Entropy as a Traffic Descriptor in the Fairisle ATM network. *Proc. 1st Workshop on ATM Traffic Mgmt.*, Paris 1995.

[6] C. S. Chang. Stability, queue length and delay of deterministic and stochastic queueing networks. *IEEE Trans. Autom. Control* 39 (1994) 913-931.

[7] P. Damien, P. W. Laud and A. F. M. Smith. Random variate generation approximating infinitely divisible distributions with application to Bayesian inference. *J. Roy. Statist. Soc. B*, 57 (1995) 547-564.

[8] P. Deheuvals and J. Steinbach. On some alternative estimates of the adjustment coefficient in risk theory. *Scand. Actuarial J.* 1990: 135-159.

[9] G. de Veciana and J. Walrand. Effective bandwidths: call admission, traffic policing and filtering for ATM networks. *Queueing Systems*, 20 (1995) 37-59.

[10] G. de Veciana, C. Courcoubetis and J. Walrand. Decoupling bandwidths for networks: A decomposition approach to resource management for networks. *IEEE Infocom Proc., 1994.*

[11] N. Duffield and Neil O'Connell. Large deviations and overflow probabilities for the general single server queue, with applications. *Math. Proc. Cambridge Phil. Soc.* 118(1), 1995.

[12] N. G. Duffield, J. T. Lewis, Neil O'Connell, R. Russell and F. Toomey, Entropy of ATM traffic streams: A tool for estimating QoS parameters. *IEEE J. Sel. Areas in Comm.*, special issue on Advances in the Fundamentals of Networking, 13(6):981-990, 1995.

[13] N.G. Duffield, J.T. Lewis, Neil O'Connell, R. Russell and F. Toomey. The entropy of an arrivals process: a tool for estimating QoS parameters of ATM traffic. *Proc. 11th U.K. Teletraffic Symp.,* Cambridge 1994.

[14] P. Embrechts and H. Schmidli. Ruin estimation for a general insurance risk model. *Adv. Appl. Prob.* 26 (1994) 404–422.

[15] P. Embrechts, J. Grandell and H. Schmidli. Finite-time Lundberg inequalities in the Cox case. *Scand. Actuarial J.* 1 (1993) 17–41.

[16] P. Embrechts and T. Mikosch. A bootstrap procedure for estimating the adjustment coefficient. *Insurance: Mathematics and Economics* 10 (1991) 181–190.

[17] T. S. Ferguson. A Bayesian analysis of some non-parametric problems. *Ann. Statist.* 1 (1973) 209–230.

[18] T. S. Ferguson. Prior distributions on spaces of probability measures. *Ann. Statist.* 2 (1974) 615–629.

[19] A. Ganesh and V. Anantharam. Stationary tail probabilities in exponential server tandems with renewal arrivals, *Queueing Systems*, 22: 203-247, 1996.

[20] A. Ganesh. Large deviations of the sojourn time for queues in series. To appear in *Annals of Operations Research.*

[21] A. Ganesh . Bias correction in effective bandwidth estimation. *Perf. Eval. Review*, 27&28 (1996) 319–330.

[22] A. Ganesh and Neil O'Connell. The linear geodesic property is not generally preserved by a FIFO queue. To appear in *Ann. Appl. Probab.*.

[23] A. Ganesh and Neil O'Connell. Large deviations for Bayesian posteriors. In preparation.

[24] R. J. Gibbens, F. P. Kelly and P. B. Key. A decision-theoretic approach to call admission control in ATM networks. *IEEE J. Sel. Areas in Comm.*, 1995.

[25] F. P. Kelly. Effective bandwidths at multi-class queues. *Queueing Systems* 9 (1991) 5–15.

[26] F. P. Kelly. On tariffs, policing and admission control for multi-service networks. *Operations Research Letters* 15 (1994) 1–9.

[27] A. Y. Lo. A large sample study of the Bayesian bootstrap. *Ann. Statist.* 15 (1987) 360–375.

[28] S. H. Low and P. P. Varaiya. A new approach to service provisioning in ATM networks. *IEEE/ACM Trans. Networking* 1 (1993).

[29] Neil O'Connell. Large deviations for departures from a shared buffer. To appear in *J. Appl. Prob.*

[30] Neil O'Connell. Large deviations for queue lengths at a multi-buffered resource. DIAS Technical Report DIAS-APG-9434.

[31] Neil O'Connell. Queue lengths and departures at single-server resources. *Proc. RSS Stochastic Networks Workshop*, 1996.

[32] S. M. Pitts, R. Grübel and P. Embrechts. Confidence bounds for the adjustment coefficient. *Adv. Appl. Prob.* 28 (1996) 802–827.

[33] S. M. Pitts. Nonparametric estimation of the stationary waiting time distribution function in the $GI/GI/1$ queue. *Ann. Statist.* 22 (1994) 1428–1446.

[34] S. M. Pitts. Nonparametric estimation of compound distributions with applications in insurance. *Ann. Inst. Statist. Math.* 46 (1994) 537–555.

[35] S. T. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *University of Bristol Mathematics Research Report S-96-01. J. Roy. Statist. Soc. B*, 60 (1998). To appear.

[36] D. B. Rubin. The Bayesian bootstrap. *Ann. Statist.* 9 (1981) 130–134.

[37] A. F. M. Smith and G. O. Roberts. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. B*, 55 (1993) 3-23.

[38] Claudia Tebaldi and Mike West. Bayesian inference of network traffic using link count data. Preprint.

[39] S. Walker, Random variate generation from an infinitely divisible distribution via Gibbs sampling. Preprint.

[40] M.P.Wiper. Bayesian analysis of Er/M/1 and Er/M/c queues. IAMI Tech Report 96.6.

A Proof of Proposition 1

We assume in the following that the prior probability of black turning up in a spin of the roulette wheel has the beta distribution, $\mathcal{B}(a_1, a_2)$, for some constants $a_1 > 0$ and $a_2 > 0$. Observe from (6) that the posterior probability of ruin when the initial capital is qn, conditional on observing black turn up $n\hat{p}_n$ times in n spins of the roulette wheel, is given by

$$P_{qn}(\text{ruin}|\hat{p}_n) = \int_0^{1/2} f_{\hat{p}_n}(x)dx + \int_{1/2}^1 f_{\hat{p}_n}(x)\left(\frac{1-x}{x}\right)^{qn} dx, \tag{8}$$

where $f_{\hat{p}_n}$ is the density on $[0,1]$ given by

$$f_{\hat{p}_n}(x) = \frac{\Gamma(a_1+a_2+n)}{\Gamma(a_1+n\hat{p}_n)\Gamma(a_2+n(1-\hat{p}_n))} x^{a_1+n\hat{p}_n-1}(1-x)^{a_2+n(1-\hat{p}_n)-1}. \tag{9}$$

It is straightforward to approximate the above integrals using Laplace's method. For completeness, we state and prove a version of this method that is adequate for our purposes.

Lemma 1 *Let μ be a finite positive measure on a compact interval $[a,b] \subset \mathbb{R}$ such that $\mu(A) > 0$ for all Borel measurable $A \subset [a,b]$ whose Lebesgue measure is positive. Let $\phi_n : [a,b] \to \mathbb{R}$, $n = 1, 2, \ldots$, be a sequence of continuous functions converging uniformly to a (continuous) limit $\phi : [a,b] \to \mathbb{R}$.*

Then,

$$\lim_{n\to\infty}\frac{1}{n}\log\int_a^b e^{n\phi_n(x)}\mu(dx) = \sup_{x\in[a,b]}\phi(x).$$

Proof : Since ϕ is continuous on $[a,b]$, its maximum is achieved at some point $x^*\in[a,b]$. Since ϕ_n converge uniformly to ϕ, given $\epsilon>0$, there exists an N such that $\phi_n(x)\le\phi(x^*)+\epsilon$ for all $n>N$ and all $x\in[a,b]$. There also exists $\delta>0$ such that, if $x\in[a,b]$ and $|x-x^*|<\delta$, then $\phi_n(x)\ge\phi(x^*)-\epsilon$ for all $n>N$. Therefore,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\int_a^b e^{n\phi_n(x)}\mu(dx) &\le \limsup_{n\to\infty}\frac{1}{n}\log\int_a^b e^{n(\phi(x^*)+\epsilon)}\mu(dx)\\ &= \phi(x^*)+\epsilon, \end{aligned}\tag{10}$$

where the equality holds because $0<\mu([a,b])<\infty$. Likewise,

$$\begin{aligned}&\liminf_{n\to\infty}\frac{1}{n}\log\int_a^b e^{n\phi_n(x)}\mu(dx)\\ \ge\ &\liminf_{n\to\infty}\frac{1}{n}\log\int_{(x^*-\delta,x^*+\delta)\cap[a,b]} e^{n(\phi(x^*)-\epsilon)}\mu(dx)\\ =\ &\phi(x^*)-\epsilon,\end{aligned}\tag{11}$$

where the equality is because $0<\mu((x^*-\delta,x^*+\delta)\cap[a,b])<\infty$ by assumption. The claim of the lemma follows from (11) and (10) since $\epsilon>0$ is arbitrary.

We can rewrite (8) as

$$P_{qn}(\text{ruin}|\hat{p}_n) = c(n)\int_0^1 f(x)e^{n\phi_n(x)}dx \tag{12}$$

where

$$c(n) = \frac{\Gamma(a_1+a_2+n)\Gamma(a_1)\Gamma(a_2)}{\Gamma(a_1+n\hat{p}_n)\Gamma(a_2+n(1-\hat{p}_n))\Gamma(a_1+a_2)}, \tag{13}$$

$$f(x) = \frac{\Gamma(a_1+a_2)}{\Gamma(a_1)\Gamma(a_2)}x^{a_1-1}(1-x)^{a_2-1},\ \ x\in[0,1], \tag{14}$$

$$\phi_n(x) = \begin{cases}\hat{p}_n\log x+(1-\hat{p}_n)\log(1-x), & 0<x<1/2,\\ (\hat{p}_n-q)\log x+(1-\hat{p}_n+q)\log(1-x), & 1/2\le x<1.\end{cases} \tag{15}$$

For any a_1, $a_2 > 0$, we have by Stirling's formula that

$$\lim_{n\to\infty} \frac{1}{n} \log c(n) = -\hat{p}\log\hat{p} - (1-\hat{p})\log(1-\hat{p}) = h(\hat{p}), \tag{16}$$

For any $\delta > 0$, ϕ_n are continuous functions on $[\delta, 1-\delta]$, converging uniformly on this interval to the limit ϕ, defined on $(0,1)$ by

$$\phi(x) = \begin{cases} \hat{p}\log x + (1-\hat{p})\log(1-x), & 0 < x < 1/2, \\ (\hat{p}-q)\log x + (1-\hat{p}+q)\log(1-x), & 1/2 \le x < 1. \end{cases} \tag{17}$$

Also, f is the density of a probability measure μ on $[0,1]$ which satisfies the assumptions of the lemma (hence, so does its restriction to $[\delta, 1-\delta]$).

Suppose $0 < \hat{p} < 1$. Then it is straightforward to verify that the maximum of ϕ on $(0,1)$ is achieved at x^*, where

$$x^* = \begin{cases} \hat{p}, & \text{if } \hat{p} < 1/2, \\ 1/2, & \text{if } \hat{p} - q \le 1/2 \le \hat{p}, \\ \hat{p} - q, & \text{if } \hat{p} - q > 1/2. \end{cases} \tag{18}$$

Therefore, if $\delta > 0$ is small enough that $\hat{p} \in (\delta, 1-\delta)$, then Lemma 1, (17) and (18) imply that

$$\lim_{n\to\infty} \frac{1}{n} \log \int_{\delta}^{1-\delta} f(x) e^{n\phi_n(x)} dx = \begin{cases} -h(\hat{p}), & \text{if } \hat{p} < 1/2, \\ -\log 2, & \text{if } \hat{p} - q \le 1/2 \le \hat{p}, \\ -h(\hat{p}-q), & \text{if } \hat{p} - q \ge 1/2, \end{cases} \tag{19}$$

where $h(x) = -x\log x - (1-x)\log(1-x)$.

Since $\hat{p}$ lies in the interval $(\delta, 1-\delta)$, hence so does $\hat{p}_n$, for all n sufficiently large. For all such n, we see from (15) that $\phi_n'(x) > 0$ for $x \in (0,\delta)$ and $\phi_n'(x) < 0$ for $x \in (1-\delta, 1)$. Therefore, the maximum of ϕ_n over $[0,\delta]$ is achieved at δ and is equal to $\hat{p}_n \log\delta + (1-\hat{p}_n)\log(1-\delta)$, while its maximum over $[1-\delta, 1]$ is achieved at $1-\delta$ and is equal to $(\hat{p}_n - q)\log(1-\delta) + (1 - \hat{p}_n + q)\log\delta$. Therefore,

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_0^{\delta} f(x) e^{n\phi_n(x)} dx \le \hat{p}\log\delta + (1-\hat{p})\log(1-\delta), \tag{20}$$

$$\limsup_{n\to\infty} \frac{1}{n} \log \int_{1-\delta}^{1} f(x) e^{n\phi_n(x)} dx \le (\hat{p}-q)\log(1-\delta) + (1-\hat{p}+q)\log\delta. \tag{21}$$

We shall make use of the following fact, whose proof is easy and is left to the reader.

Fact: Let a_n and b_n be non-negative sequences, with

$$\lim_{n\to\infty}\frac{1}{n}\log a_n = A, \quad \limsup_{n\to\infty}\frac{1}{n}\log b_n \le B.$$

If $A > B$, then

$$\lim_{n\to\infty}\frac{1}{n}\log(a_n + b_n) = A.$$

As δ decreases to zero, the right hand sides of (20) and (21) decrease to $-\infty$. Therefore, by choosing $\delta > 0$ sufficiently small, they can be made smaller than the right hand side of (19). It then follows from the above fact that (19) holds with the limits of integration, δ and $1-\delta$, replaced with 0 and 1. The proof of Proposition 1 when $0 < \hat{p} < 1$ is now immediate from (16) and (19).

Suppose next that $\hat{p} = 0$. Then, for any $\delta > 0$, there exists N such that $\hat{p}_n < \delta$ for all $n > N$. Let $\delta < 1/2$. Then, by (15), we have for all $x \in (\delta/2, \delta)$ and all $n > N$ that

$$\phi_n(x) \ge \delta\log x + (1-\delta)\log(1-x) \ge \delta\log\frac{\delta}{2} + (1-\delta)\log\left(1-\frac{\delta}{2}\right).$$

Therefore, for all $n > N$,

$$\int_0^1 f(x)e^{n\phi_n(x)}dx \ge \int_{\delta/2}^{\delta} f(x)\exp\left\{n\Big[\delta\log\frac{\delta}{2} + (1-\delta)\log\left(1-\frac{\delta}{2}\right)\Big]\right\}.$$

Since f is strictly positive on $(0,1)$ by assumption,

$$\liminf_{n\to\infty}\frac{1}{n}\log\int_0^1 f(x)e^{n\phi_n(x)}dx \ge \delta\log\frac{\delta}{2} + (1-\delta)\log\left(1-\frac{\delta}{2}\right).$$

But $\delta > 0$ is arbitrary; as δ decreases to zero, so does the right hand side above. Also note that $\phi_n(x) \le 0$ for all $x \in (0,1)$, and so,

$$\limsup_{n\to\infty}\frac{1}{n}\log\int_0^1 f(x)e^{n\phi_n(x)}dx \le 0.$$

It follows from these two inequalities that

$$\lim_{n\to\infty}\frac{1}{n}\log\int_0^1 f(x)e^{n\phi_n(x)}dx = 0 = h(\hat{p}).$$

Together with (16), this establishes Proposition 1 when $\hat{p} = 0$. The case $\hat{p} = 1$ can be treated similarly. This concludes the proof of Proposition 1.

Remarks : Intuitively, if $\hat{p} < 1/2$, then we infer that ruin is certain, however large the initial capital of the gambler. For fixed $\hat{p} > 1/2$, the limit in the proposition is a decreasing function of q, as we would expect. However, it flattens out at $q = \hat{p} - 1/2$. The explanation for this is that, for larger q, the probability of ruin when the gambler's initial capital is qn is essentially the same as the predictive probability that eventual ruin is certain. This latter probability clearly does not depend on q. Finally, it can be verified by differentiation that the limit in the theorem is a decreasing function of $\hat{p}$ for fixed q, as expected.