Les Cahiers du GERAD

**Intelligent Simulation for the
Estimation of the Uplink Outage
Probabilities in CDMA Networks**

Felisa J. Vázquez-Abad
Irina Baltcheva

Intelligent Simulation for the Estimation of the Uplink Outage Probabilities in CDMA Networks

Felisa J. Vázquez-Abad and **Irina Baltcheva**

GERAD and Département d'informatique et recherche opérationnelle
Université de Montréal, Canada

ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN)
Department of Electrical and Electronic Engineering
University of Melbourne, Australia

{vazquez,baltchei}@iro.umontreal.ca, fva@ee.mu.oz.au

December, 2002

Abstract

In CDMA mobile networks, callers that are transmitting through a power station may cause interference at other power stations. When many users are already connected in the network, a new call may cause the signal to noise ratio to drop below a tolerance threshold. This phenomenon is called "outage" and it provides an important measure of performance, useful in the design and control of the system. Evaluating this probability analytically has proven unsuccessful and only approximations exist today. Direct simulation of such networks is at present very slow because outage occurs infrequently– it may take hours to simulate directly a realistic model if a reasonable precision is desired. Thus this approach is not useful for design problems where one wishes to evaluate and compare performance of many different network models. In this work we implement a change of measure to estimate the outage probability using Importance Sampling. We present a functional estimator and a stochastic approximation method that are capable of learning the best parameters for the change of measure.

Résumé

Dans les résaux CDMA de téléphones mobiles, les utilisateurs transmettant à travers une station de base peuvent créer de l'interférence aux autres stations. Lorsque plusieurs mobiles sont déja connectés au réseau, un nouvel appel peut faire en sorte que le rapport "signal-bruit" descende sous le seuil de tolérance. Ce phénomène est appelé "outage" (blocage) et c'est un critère de performance utile pour la conception et le contrôle du système. Évaluer cette probabilié de façon analytique s'étant avéré sans succès, il n'existe de nos jours que des approximations. Or, une simulation directe de tels réseaux est présentement très lente, car le blocage est un événement rare. Parfois, des heures de simulation sont nécessaires afin d'obtenir une précision raisonable. Ainsi, cette approche n'est pas très utile lorsqu'il s'agit de comparer la performance de différents modèles. Dans ce travail, nous développons un changement de mesure nous permettant d'estimer la probabilité de blocage. Nous présentons un estimateur fonctionnel et une méthode d'approximation stochastique par le gradient qui sont capables de trouver les paramètres optimaux du changement de mesure.

1   Introduction

CDMA is the next radio-frequency technology for future wireless networks, such as the Universal Mobile Telecommunications System (UMTS) or IMT-2000. The outage probability is an important performance measure of CDMA cellular mobile networks, and it is useful for planning and design. While mobiles are transmitting, a slow fading effect attenuates the signal, which is thus received at the various power stations at different intensities. The power station which is allocated for transmission for each mobile is the one that has the smallest attenuation, that is, the station that receives the highest intensity signal. Typically only neighboring power stations are included in the set of possible candidates for transmission (soft handover). In this work we consider the model for perfect power control, where each mobile can choose the exact transmission power so that its signal is received at unit power at the chosen station [2]. The total interference at a power station is the sum of all the interferences caused by mobiles not connected to that station. Outage at a given base station occurs whenever a new mobile in the network causes the signal to interference ratio (SIR) to decrease below a fixed threshold.

Analytical expressions for the outage probability are unavailable. Many studies have proposed approximations under limiting regimes such as increasing the number of users per cell, but most approximations are inaccurate for realistic values of the model parameters. In this work we use a simulation method to evaluate the outage probability. It is common to use simulation to validate or compare bounds and approximations, but most of the simulation studies use crude Monte Carlo methods to replicate the dynamic model and compute whether outage occurs or not at a given station. Because outage is typically a small quantity, the problem is akin to rare event estimation: many samples of the process must be simulated before an outage event occurs and this translates into highly inefficient algorithms, requiring prohibitive CPU times to obtain accurate estimates. Because it is necessary to simulate many scenarios when designing networks, it is therefore not viable to use crude Monte Carlo. Importance Sampling is a dramatically successful technique when dealing with rare events [9]. This method is based on a change of measure and it is possible to achieve a uniformly bounded relative error, when the parameters are well chosen. However, in this problem it is not straightforward to use the large deviation bounds that could point to the optimal change of measure. The difficulty is that the fading is modeled according to the usual power path loss, which has a lognormal distribution. Moreover, the distribution of the individual interferences is not known and thus it is impossible to solve analytically for the best change of measure. In this work we implement two simulation-based methods capable of learning the optimal parameters of a particular change of measure, in order to minimise the variance of the ensuing estimator. The functional estimation approach of [6] is implemented in Section 5, and in Section 6 , we apply the self-optimising method introduced in [11] and [10]. But first, we start with the presentation of the model in Section 2, then Section 3 describes the methods used in our simulation, especially the random number generation, while in Section 4, we present the change of measure.

2   Model and Problem Formulation

The model discribed here is based on the one used in [2]. There are K cells in the network, each with a *neighborhood* of "adjacent" cells $\mathcal{V}(k), k = 1, \ldots, K$. At each cell k the mobile callers are placed according to independent Poisson spacial processes with intensity function $\lambda_k(\cdot)$. In this

work we consider a homogeneous Poisson process per cell, so that each λ_k is a constant, but they may be different for different k. For example, later in Section 6.1, we consider a particular case where there are two different intensities: one for the neighborhood of the central cell and one for the rest of the cells.

Let $N_k \sim \text{Poisson}(\lambda_k)$ be the number of users in cell k. For each mobile $i = 1, \ldots, N_k$ in cell k denote by $d_{i,k}(l)$ the distance from caller i to power station $l \in \mathcal{V}(k)$. The signal transmitted by a mobile suffers degradation and some of the information is lost in the path due to the propagation attenuation and physical obstacles like mountains, high buildings, etc. The loss of intensity in the signal precision of the mobile i situated in $l \in \mathcal{V}(k)$ is called the *attenuation factor* at station l and is denoted by $\Gamma_{ik}(l)$. Set:

$$\Gamma_{ik}(l) = \frac{1}{d_{ik}(l)^4} 10^{Z_{ik}(l)}, \quad C_{i,k} = \max_{l \in \mathcal{V}(k)} \{\Gamma_{ik}(l)\},$$

$$\kappa(i) = \text{argmax}\{\Gamma_{ik}(l) : l \in \mathcal{V}(k)\}$$

where $\{Z_{ik}(l)\}$ are iid normal zero-mean variables $\mathcal{N}(0, \sigma^2)$. Obviously, $\Gamma_{ik}(l)$ depends on the distance from the mobile to the power station, but also on the term $10^{Z_{ik}(l)}$ which represents the *shadowing effect* that occurs randomly in the environment. Caller i is connected to the station $\kappa(i)$ which has the best transmission (the stronger signal from all the neighboring stations). If a mobile transmits at unit power, the fading factor makes that it is received at the cell-site with power $\Gamma_{ik}(l)^{-1}$. In our model we suppose perfect power control, meaning that each user can choose the exact transmission power so that he can be received at unit power at the receiving station. Because the attenuation factor with the chosen station is C_{ik}, the mobile adjusts his transmission power to $\frac{1}{C_{ik}}$ and this guarantees that all received signals are unity at $\kappa(i)$. The *interference* caused by mobile i in cell k to the base station 0 is $\Gamma_{i,k}(0)/C_{i,k}$. Notice that if $\{0\} \in \mathcal{V}(k)$ then all users of cell k will cause an interference bounded by one. However if $\{0\} \notin \mathcal{V}(k)$ it is possible that $\Gamma_{i0}/C_{i,k} > 1$, actually this random variable will have a heavy tail.
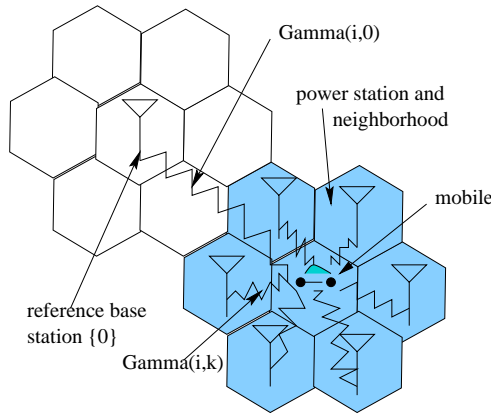

Figure 1: Mobile and base stations

In a network, the sum of all particular interference (the one of each user in the system) is called the *signal to noise ratio*. *Outage* is the event that the signal to noise ratio at the reference base station 0 is smaller than a threshold α. Because signals are always received at unit power,

this is equivalent to all the interference being larger than $1/\alpha$. If we sum the signal to noise ratio over all the users in each cell, we get the following event:

$$\mathcal{B} = \left\{ \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{\Gamma_{i,k}(0)}{C_{i,k}} > \alpha^{-1} \right\}. \tag{1}$$

Let

$$p = \mathsf{E}(\mathbf{1}\,(\mathcal{B})),$$

where $\mathbf{1}\,(\mathcal{B})$ is the indicator function of the event $\mathcal{B}$. The problem is to calculate the outage probability $p = \mathsf{E}(\mathbf{1}\,(\mathcal{B}))$.

3 Simulation

The "direct" way of simulating is very simple:

```
1. GENERATE the number of mobiles (LAMBDA per cell k);
2. PLACE mobiles in space;
3. CONNECT each mobile i:
       for all cells j in neighborhood(k) except 0 (reference cell)
           compute GAMMA{i,j} and choose the max;
       for the reference cell
           calculate GAMMA{i,0};
           S += (GAMMA{i,0} / C_i);
4. TEST outage:
       if (S > ALPHA)
           count outage;
```

The simulation program in itself is not very complicated, but the large number of random variables needed to be generated makes it very slow if not optimised in that direction. Here, we present an important aspect of our simulation: the methods we use to generate different random variables. In our model, generation of normal and Poisson variates is needed. The first is a continuous one, while the second is discrete. The best known method is the Inversion which is based on the knowledge of the distribution function $F(x)$. In the case of the normal distribution, this function is not easy to invert. We use for example the Box-Muller algorithm or the polar method. Regarding the Poisson variables, there are different methods whose performance depends on the mean λ. For example, there is a particular algorithm for the Poisson variates that works well with small mean. The Inversion method using a predefined array described in Section 3.2.2 is always very efficient, because after an initialization requiring a time $O(N)$, the cost of every access is constant for each variable. This method is useful when many variables must be generated with the same mean. Here are the algorithms found in our simulation with a few proofs of their efficiency.

3.1 Normal variables $X \sim \mathcal{N}(0,1)$

The normal variables are necessary to calculate the lognormal fading factor. Here is the algorithm for the polar method (see also [7]):

1. Repeat:

 – Generate U_1 and U_2 iid $U(-1, 1)$,
 – until $W = V_1^2 + V_2^2 < 1$;

2. Let $Y = \sqrt{(-2 \log W)/W}$.

3. Return $X_1 = YV_1$ and $X_2 = YV_2$. Then, X_1 and X_2 are iid $\mathcal{N}(0, 1)$.

The polar method is a modification of the Box-Muller one. Recall that the density function of the normal variable $\mathcal{N}(0, 1)$ is:

$$f(x) = \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2}.$$

The idea is to generate two independent variables $X, Y \sim \mathcal{N}(0, 1)$, using polar coordinates. The joint density function of X and Y is:

$$f(x, y) = \frac{1}{2\pi} \, e^{\frac{-(x^2 + y^2)}{2}}.$$

If we make a change of variables, letting $r = x^2 + y^2$, $\theta = \tan^{-1}(y/x)$, or

$$\begin{aligned} x &= \sqrt{r} \cos \theta \\ y &= \sqrt{r} \sin \theta, \end{aligned} \tag{2}$$

which Jacobian is equal to $1/2$, then the joint density function becomes

$$f(r, \theta) = \left(\frac{1}{2\pi} \right) \left(\frac{1}{2} e^{\frac{-r}{2}} \right), \quad 0 < r < \infty, \quad 0 < \theta < 2\pi.$$

The latter can be viewed as a product of two density functions: the uniform $U[0, 2\pi]$ and the exponential $\exp(2)$. Therefore, we can see r and θ as two independent random variables, with exponential and uniform densities: $r \sim \exp(1/2)$ and $\theta \sim U[0, 2\pi]$. This transformation (Box-Muller) implies calculation of trigonometric functions for every random variable generated (see Eq. (2)), which may take time. This is why an alternative way to generate normals was found, avoiding calculation of sinus and cosinus. In the polar method, instead of generating the radius and an angle, we generate the two sides of the triangle defining the coordinates of a point, via the acception/rejection method. This is what is done in step 1 of the algorithm above. The triangle's sizes V_i's are $U[-1, 1]$ and are accepted when inside the unit circle. Now, we can define

$$\begin{aligned} \sin \theta &= \frac{V_2}{\sqrt{V_1^2 + V_2^2}} = V_2/W \\ \cos \theta &= \frac{V_1}{\sqrt{V_1^2 + V_2^2}} = V_1/W, \end{aligned} \tag{3}$$

where $\theta \sim U[0, 2\pi]$. Remark that W and θ are uniform and independent. The radius r is still $exp(2)$ and will be generated by inversion, using W: $r = \sqrt{-2 \log W}$, W being an uniform $[0, 1]$ variable. Finally,

$$\begin{aligned} X_1 &= R \cos \theta = \sqrt{\frac{-2 \log W}{W}} \, V_1 \\ X_2 &= R \sin \theta = \sqrt{\frac{-2 \log W}{W}} \, V_2. \end{aligned} \tag{4}$$

The variables being generated by pairs, time can be gained when calling the procedure one time on two. Moreover, using W in the last step of the algorithm avoids generation of one extra random number, so the only time-consuming place is step 1.

3.2 Poisson variables $X \sim$ **Poisson**(λ)

For the Poisson(λ) variates, we use three methods: Inversion, Inversion with predefined array and the particular method for λ small. When $\lambda < 1$, we use the particular method, when $\lambda > 1$, we use Inversion with array or not, depending on the number of variates we need.

3.2.1 Particular method for λ small The following result is well known in simulation and presents a method for generating Poisson random numbers for small values of intensity. We reproduce it here for completeness of our presentation.

Lemma 1 *Given λ, let $U_1, U_2, \ldots \sim U(0,1)$ be a sequence of iid uniform random variates and set $p_n = U_n\, p_{n-1}$, with $p_0 = e^\lambda$. Then the random variable $M = \min\{n : p_n < 1\} - 1$ has Poisson distribution with mean λ.*

Proof: The idea of this algorithm is to simulate a Poisson process with mean λ and count the number of arrivals in the interval (0,1]. This number will be the returned value of the random variable M.

If $p_n = U_n\, p_{n-1}$, $p_0 = e^\lambda$ ($\{p_i\}$ is a decreasing sequence) and M is defined as $M = \min\limits_n\{p_n < 1\}$, then

$$
\begin{aligned}
M &= \min_n\{p_n < 1\} \\
&= \min_n\{p_n < 1 \le p_{n-1}\} \\
&= \min_n\{e^\lambda U_1 \ldots U_n < 1 \le e^\lambda U_1 \ldots U_{n-1}\}
\end{aligned}
\tag{5}
$$

After dividing by e^λ and taking the natural logarithm, we can equally write

$$
\begin{aligned}
\mathsf{P}\{M = n\} &= \mathsf{P}\left\{\sum_{i=1}^{n} \log U_i < -\lambda \le \sum_{i=1}^{n-1} \log U_i\right\} \\
&= \mathsf{P}\left\{\sum_{i=1}^{n-1}\left(-\frac{1}{\lambda}\log U_i\right) \le 1 < \sum_{i=1}^{n}\left(-\frac{1}{\lambda}\log U_i\right)\right\}
\end{aligned}
\tag{6}
$$

If we set the random variables A_i as iid $\exp(1/\lambda)$, (6) becomes

$$
\mathsf{P}\{M = n\} = \mathsf{P}\left\{\sum_{i=1}^{n-1} A_i \le 1 < \sum_{i=1}^{n} A_i\right\}
\tag{7}
$$

By definition of a Poisson process with rate λ, time intervals between arrivals are exponential independent random variables with mean $1/\lambda$. A_i, $i = 1, 2, \ldots$ can be viewed as interarrival times of a Poisson process with mean λ and the variable counting the number of arrivals in the interval (0,1], which is $M = \min\limits_n\{p_n < 1\}$, has then a Poisson distribution.

<div align="right">QED</div>

The algorithm for the particular method is the following:

1. Let $X = 0$ and $p = e^\lambda$;

2. Generate $U \sim U(0,1)$; $p \leftarrow p * U$; $X \leftarrow X + 1$;

3. If $p > 1$, go at 2;

4. Else return $X - 1$;

This method requires generation of many uniform variates, but it is very efficient when λ is small. If λ increases, e^{λ} increases also and many iterations are necessary before $p < 1$. The mean number of iterations before satisfying this criteria is λ, so to avoid big values of λ, convolution can be used (see next section for a proof).

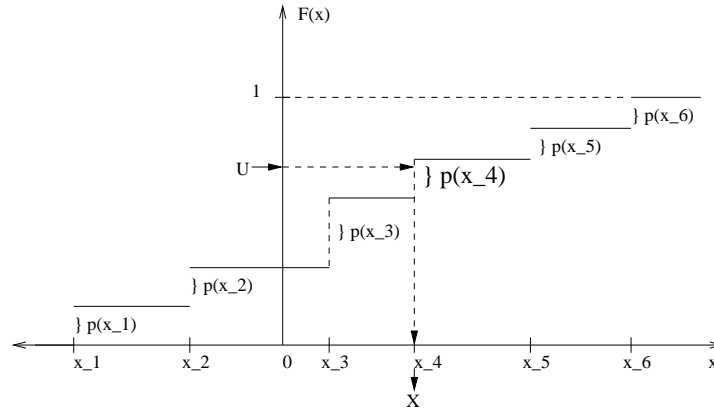### 3.2.2 Inversion method with predefined array

Figure 2: Distribution function

The idea here is to represent the probabilities $p(i)$ in an array. To understand better the way this method works, refer to Figure 2 which shows a distribution function $F(x)$. The inverse function on $[0,1]$ is defined as $F^{-1}(u) = inf\{x : F(x) \geq u\}$. The inverse method for discrete variables generates $U \sim U[0,1]$ and returns $X = \lfloor F^{-1}(U) \rfloor$. To accelerate the process, we use a fixed size array, initialized at the beginning and representing the distribution function (truncated). The problem is to calculate the size of the array t. We decide to ignore the X's for which $P[X < t] < 10^{-6}$. For the Poisson variable with mean $\lambda = 8$ (intensity used in our tests), this gives us an array of size $t = 1000$. One should be careful when the average λ increases, because then the size of the array may become too big and consume too much memory. Once initialized, the proportion of the array occupied by one value is the probability of that value. Figure 3 shows the contents of the array.
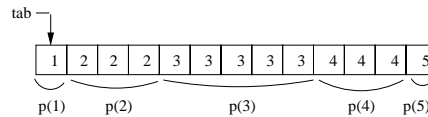

Figure 3: Array for the inversion method

To avoid the generation of variables with large mean, we use the following result on the convolution of Poisson variables:

Lemma 2 *If* $X \sim Poisson(K\lambda)$, *then* $Y = \sum_{i=1}^{K} X_i$, *with* X_i *iid* $Poisson(\lambda)$ *is also* $Poisson(K\lambda)$.

Proof: We show the result by induction, for the general case where $X_i \sim \text{Poisson}(\lambda_i)$, $i = 1, 2, \ldots, K$. The generating function of a Poisson(λ) variable X is:

$$\mathsf{E}[s^X] = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} s^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda s)^n}{n!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}.$$

When $K = 2$, let $Y = X_1 + X_2$ with $X_1 \sim \text{Poisson}(\lambda_1)$, $X_2 \sim \text{Poisson}(\lambda_2)$. Then

$$\mathsf{E}[s^Y] = \mathsf{E}[s^{(X_1+X_2)}] = \mathsf{E}[s^{X_1} s^{X_2}]$$
$$\stackrel{indep.^*}{=} [\mathsf{E}s^{X_1}][\mathsf{E}s^{X_2}] = e^{-\lambda_1(1-s)} e^{-\lambda_2(1-s)} = e^{-(\lambda_1+\lambda_2)(1-s)},$$

which means that Y has a Poisson distribution with mean $(\lambda_1 + \lambda_2)$. By induction, if $X_1, X_2, \ldots, X_K$ are independent Poisson variables with means $\lambda_1, \lambda_2, \ldots, \lambda_K$ respectively, then $Y = \sum_{i=1}^{K} X_i$ will be Poisson($\sum_i X_i$). In particular, when $\lambda_i = \lambda$, $\forall i$, we obtain the result.

<div align="right">QED</div>

Remark: The proof of $\mathsf{E}[s^{X_1} s^{X_2}] = [\mathsf{E}s^{X_1}][\mathsf{E}s^{X_2}]$ is in [4].

Here is the initialization algorithm:

1. Let: $U \sim U(0, 1)$; $p \leftarrow e^{-\lambda}$; $F \leftarrow 0$; $n \leftarrow 0$;
2. For $\forall m$, $0 \leq m < t$,
   while $m/t > F$ do:
   $n \leftarrow n + 1$;
   $p \leftarrow p * \frac{\lambda}{n}$;
   $F \leftarrow F + p$;
3. Array[m] = n;

After the initialization, we generate U, a uniform distributed number between 0 and t (the array size) and return $tab[\lfloor U \rfloor]$. This method is very convenient for all the Poisson numbers with average bigger than 1, because the access time in the array is $O(1)$.
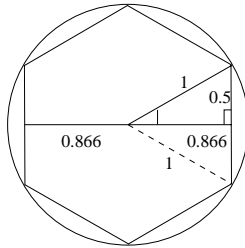

Figure 4: Dimensions of a cell

3.2.3   Generate the placing of the users in the hexagonal cells The points representing the mobiles in an hexagonal cell are generated using the acceptance-rejection method. We need
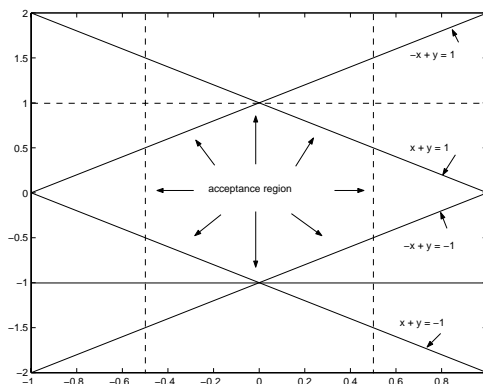
Figure 5: Acceptance-rejection method

here to generate uniformly distributed points in an hexagonal shape which represents the cell. Figure 4 shows the dimensions of a cell. A unit circle majorates the hexagon. The procedure is very simple and it is illustrated in Figure 5.

1. Generate a point uniformly distributed in the rectangle $[0.5, 0.5) \times [1, 1)$;

2. If this point lies outside of the "squashed hexagon" bounded by $\pm x + y = \pm 1$, it is rejected and a new one is chosen;

3. If the point is inside the "acceptance region", multiply the x coordinate of the point by $\sqrt{3}$ to normalize and return this point.

4 The Change of Measure

Estimating small probabilities with direct simulation may be a difficult problem. More precisely, if $X_i, i = 1, \ldots, N$ is an iid sample of Bernoulli random variables with parameter p, then the sample average has a relative error proportional to $\sqrt{p(1-p)}/(Np)$, which implies that the number of replications N has to be considerably large if the event in question has small probability. Estimation of probabilities of rare events has been the subject of numerous recent contributions ([9]). In some cases, using large deviation theory, it can be shown that a particular change of measure, called exponential tilting, can yield estimators with uniformly bounded relative error (in p) for fixed values of N, see [1] and Chapter 8 of [8], as we now explain.

To motivate the argument, recall that the outage probability is given by (1). Suppose that $X = (X_n; n = 1, 2, \ldots)$ are iid zero-mean random variables and let $S_n = \sum_{i=1}^{n} X_i$. The problem is estimating the probability:

$$p_n = \mathsf{P}\left(\frac{S_n}{n} \geq a\right) = \mathsf{E}(I_n(X)),$$

where $I_n(X) = \mathbf{1}(S_n \geq na)$, for $a > 0$. Clearly as $n \to \infty$ this event becomes rarer and the probability sought tends to zero. Actually, under the appropriate conditions it tends to zero exponentially fast: large deviations theory helps determining the decay rate of p_n. Re-write now the expectation as:

$$\mathsf{E}(I_n(X)) = \int_{\mathbb{R}^n} \mathbf{1}\left(\sum_{i=1}^{n} x_i \geq na\right) \prod_{i=1}^{n} f(x_i)\, dx$$

$$= \int_{\mathbb{R}^n} \mathbf{1}\left(\sum_{i=1}^{n} x_i \geq na\right) \left(\prod_{i=1}^{n} \frac{f(x_i)}{f_\theta(x)}\right) \prod_{i=1}^{n} f_\theta(x_i)\, dx$$

$$\equiv \mathsf{E}(L_n(\theta, \tilde{X}) I_n(\tilde{X})),$$

where $\tilde{X} = (\tilde{X}_n, n = 1, 2, \ldots)$ is now a sequence of iid variables with density f_θ. Clearly, the above is only true when the division by f_θ is well defined. The necessary condition is that for every null set $B \subset \mathbb{R}$ with $\int_B f_\theta(x)\, dx = 0$ it holds that $\int_B f(x)\, dx = 0$. This condition is in general known as absolute continuity: the probability measure defined through f must be *absolutely continuous* with respect to the new measure, defined through f_θ, and the random variable $L_n(\theta, X)$ is known as the *Radon Nikodym derivative* or the *likelihood ratio*.

Let us look at the variance of the estimation through such a change of measure. Because $\mathsf{E}[L_n(\theta, \tilde{X}) I_n(\tilde{X})] = p_n$, then:

$$\mathsf{Var}[L_n(\theta, \tilde{X}) I_n(\tilde{X})] = \mathsf{E}[L_n^2(\theta, \tilde{X}) I_n(\tilde{X})] - p_n^2 \geq 0,$$

which implies (variance is always non negative):

$$\frac{1}{n} \log\left(\mathsf{E}[L_n^2(\theta, \tilde{X}) I_n(\tilde{X})]\right) \geq \frac{2}{n} \log p_n.$$

The idea is to use Cramer's bound, namely that

$$\lim_{n\to\infty} \inf \frac{2}{n} \log p_n = -2\left(\sup_\theta \{a\theta - \log M(\theta)\}\right),$$

where $M(\theta) = \mathsf{E}[e^{\theta X}]$ is the moment generating function of X_i. The right hand side can be easily calculated, because the function is concave in θ and has only one critical point θ^*:

$$\frac{d}{d\theta}(a\theta - \log M(\theta))\bigg|_{\theta=\theta^*} = a - \frac{M'(\theta^*)}{M(\theta^*)} = 0.$$

Putting all this together, the following is a lower bound of the asymptotic expectation. $\mathsf{Var}[L_n(\theta, \tilde{X}) I_n(\tilde{X})] \geq 0 \implies$

$$\lim_{n\to\infty} \inf \frac{1}{n} \log\left(\mathsf{E}[L_n^2(\theta, \tilde{X}) I_n(\tilde{X})]\right) \geq -2(a\theta^* - \log M(\theta^*)). \tag{8}$$

Consider the new density $f_\theta(x) = e^{\theta x} f(x)/M(\theta)$, called the *exponential tilt* of f. The the Radon Nikodym derivative is simply the expression:

$$L_n(\theta, X) = \prod_{i=1}^{n} e^{-\theta X_i} M(\theta) = M^n(\theta)\, e^{-\theta S_n}.$$

Using the definition of $I_n(\tilde{X})$, $\mathsf{E}[L_n^2(\theta, \tilde{X})I_n(\tilde{X})] \le M^{2n}(\theta)e^{-2\theta na}$, so, diving by n and taking the logarithm yields:

$$\frac{1}{n}\log\left(\mathsf{E}[L_n^2(\theta, \tilde{X})I_n(\tilde{X})]\right) \le -2a\theta + 2\log M(\theta), \tag{9}$$

which gives the uper bound on the asymptotic rate of decay. Comparing the inequalities (8) and (9) it follows that the asymptotic variance attains its possible minimal value precisely at $\theta = \theta^*$.

Unfortunately, direct application of this theory would require tilting a heavy tailed random variable, for which the moment generating function does not exist. The change of measure that we implement uses a Poisson number of callers per cell with a different mean, together with an exponential tilt of the shadowing factors $Z_{i,k}(0)$ rather than transforming the heavy tailed path loss $\Gamma_{i,k}(l)$ or the ratio $\Gamma_{i,k}(l)/C_{i,k}$. Let $N_k \sim$ Poisson (λ_k), $N_k^* \sim$ Poisson (θ_k), $\forall k$.

Lemma 3 *Let $N = (N_1, \ldots, N_K)$, where $\{N_k\}$ are independent Poisson random variables with means λ_k, and let $Z_{i,k}(0), \ldots Z_{N_k,k}(0)$ be iid $\mathcal{N}(0, \sigma^2)$, conditionally independent of N. For any value of the parameter $(\theta_k, \mu_k; k = 1, \ldots, K) \in \mathbb{R}^{2K}$ and for any real valued function $\phi(N; Z)$:*

$$\mathsf{E}[\phi(N; Z)] = \mathsf{E}[L(N^*; Z^*)\phi(N^*; Z^*)], \tag{10}$$

where $N_k^* \sim Poisson(\theta_k)$, $Z_{i,k}^* \sim \mathcal{N}(\mu_k, \sigma^2)$ are independent normal random variables, and:

$$L(N; Z) = \exp\left\{\sum_{k=1}^{K}\left[(\theta_k - \lambda_k) + \frac{N_k\mu_k^2}{2\sigma^2} - \frac{\mu_k}{\sigma_k^2}B_{N_k,k}\right]\right\}\prod_{k=1}^{K}\left(\frac{\lambda_k}{\theta_k}\right)^{N_k} \tag{11}$$

with $B_{N_k,k} = \sum_{i=1}^{N_k} Z_{i,k}^*$.

Proof: Denote by $\bar{\phi}(N) = \mathsf{E}[\phi(N, Z) \mid N]$ the conditional expectation. Then $\mathsf{E}[\phi(N; Z)] = \mathsf{E}[\bar{\phi}(N)]$. If we multiply and divide by the probability of the star measure, we get:

$$
\begin{aligned}
\mathsf{E}[\bar{\phi}(N)] &= \sum_{k=0}^{K}\sum_{n_k=0}^{\infty} \bar{\phi}(n_1, \ldots, n_K) \times \prod_{k=0}^{K}\left(\frac{\mathsf{P}(N = n_k)}{\mathsf{P}(N^* = n_k)}\,\mathsf{P}(N^* = n_k)\right) \\
&= \sum_{k=0}^{K}\sum_{n_k=0}^{\infty} \bar{\phi}(n_1, \ldots, n_K) \times \prod_{k=0}^{K}\left(\frac{e^{-\lambda_k}\frac{\lambda_k^{n_k}}{n_k!}}{e^{-\theta_k}\frac{\theta_k^{n_k}}{n_k!}}\left[\frac{e^{-\theta_k}\theta_k^{n_k}}{n_k!}\right]\right) \\
&= \sum_{k=0}^{K}\sum_{n_k=0}^{\infty} \bar{\phi}(n_1, \ldots, n_K) \times \prod_{k=0}^{K}\left(e^{-(\lambda_k-\theta_k)}\left(\frac{\lambda_k}{\theta_k}\right)^{n_k}\left[\frac{e^{-\theta_k}\theta_k^{n_k}}{n_k!}\right]\right) \\
&= \sum_{k=0}^{K}\sum_{n_k=0}^{\infty} \bar{\phi}(n_1, \ldots, n_K) \times \prod_{k=0}^{K}\mathsf{P}(N^* = n_k) \times e^{-\sum_k(\lambda_k-\theta_k)} \times \prod_{k=0}^{K}\left(\frac{\lambda_k}{\theta_k}\right)^{n_k} \\
&= \mathsf{E}\left[\bar{\phi}(N^*)\,e^{-\sum_k(\lambda_k-\theta_k)}\prod_k\left(\frac{\lambda_k}{\theta_k}\right)^{N_k^*}\right], \tag{12}
\end{aligned}
$$

where $N_k^* \sim$ Poisson (θ_k). Similarly, by independence of the normal random variables, letting $d = \sum_{k=1}^{K} N_k$, $z = (z_{i,k}) \in \mathbb{R}^d$ it follows that:

$$
\begin{aligned}
\mathsf{E}[\phi(N;Z) \mid N] &= \int_{\mathbb{R}^d} \phi(N;z)\, \mathsf{P}(\phi(N;Z) \mid N)\, dz \\
&= \int_{\mathbb{R}^d} \phi(N;z)\, \frac{\mathsf{P}(\phi(N;Z) \mid N)}{\mathsf{P}(\phi(N;Z^*) \mid N)}\, \mathsf{P}(\phi(N;Z^*) \mid N)\, dz \\
&= \int_{\mathbb{R}^d} \phi(N;z) \prod_{k=1}^{K} \prod_{i=1}^{N_k} \left(\frac{e^{z_{i,k}^2/2\sigma^2}}{e^{(z_{i,k}-\mu_k)^2/2\sigma^2}} \times \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{(z_{i,k}-\mu_k)^2/2\sigma^2} \right) dz \\
&= \int_{\mathbb{R}^d} \phi(N;z) e^{\left\{ \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(\frac{\mu_k^2 - 2\mu_k z_{i,k}}{2\sigma^2} \right) \right\}} \prod_{k=1}^{K} \prod_{i=1}^{N_k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(z_{i,k}-\mu_k)^2/2\sigma^2}\, dz \\
&= \mathsf{E}\left[\exp\left\{ \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left(\frac{\mu_k^2 - 2\mu_k Z_{i,k}^*}{2\sigma^2} \right) \right\} \phi(N;Z^*) \mid N \right] = \bar{\phi}(N), \qquad (13)
\end{aligned}
$$

where for each k, Z_{ik}^* are iid normal $\mathcal{N}(\mu_k, \sigma^2)$. Replacing the expression (13) for $\bar{\phi}(N^*)$ in (12) above, we get:

$$
\begin{aligned}
\mathsf{E}[\bar{\phi}(N)] &= \mathsf{E}[\mathsf{E}[\phi(N,Z) \mid N]] \\
&= \mathsf{E}\left[\bar{\phi}(N^*)\, e^{-\sum_k (\lambda_k - \theta_k)} \prod_k \left(\frac{\lambda_k}{\theta_k} \right)^{N_k^*} \right] \\
&= \mathsf{E}\left[\mathsf{E}\left[e^{\sum_{k=1}^{K} \sum_{i=1}^{N_k^*} \left(\frac{\mu_k^2 - 2\mu_k Z_{i,k}^*}{2\sigma^2} \right)} \phi(N^*;Z^*) \mid N^* \right] e^{-\sum_k (\lambda_k - \theta_k)} \prod_k \left(\frac{\lambda_k}{\theta_k} \right)^{N_k^*} \right] \\
&= \left[e^{\sum_{k=1}^{K} \sum_{i=1}^{N_k^*} \left(\frac{\mu_k^2 - 2\mu_k Z_{i,k}^*}{2\sigma^2} \right)} e^{-\sum_k (\lambda_k - \theta_k)} \prod_k \left(\frac{\lambda_k}{\theta_k} \right)^{N_k^*} \right] \mathsf{E}\left[\mathsf{E}\left[\phi(N^*;Z^*) \mid N^* \right] \right] \\
&= \left[\exp\left\{ \sum_{k=1}^{K} \left[(\theta_k - \lambda_k) + \frac{N_k^* \mu_k^2}{2\sigma^2} - \frac{\mu_k}{\sigma_k^2} \sum_{i=1}^{N_k^*} Z_{i,k}^* \right] \right\} \prod_{k=1}^{K} \left(\frac{\lambda_k}{\theta_k} \right)^{N_k^*} \right] \mathsf{E}\left[\mathsf{E}\left[\phi(N^*;Z^*) \mid N^* \right] \right] \\
&= L(N^*;Z^*)\, \mathsf{E}[\phi(N^*;Z^*)] \\
&= \mathsf{E}[L(N^*;Z^*)\, \phi(N^*;Z^*)]
\end{aligned}
$$

which yields the result.

<div align="right">QED</div>

The above result leads to the following simulation procedure for estimating outage probability: K random variables $N_k^* \sim \text{Poisson}(\theta_k)$ are generated, and for each k, standard normal variables $\xi_{i,k}(l) \sim \mathcal{N}(0,1)$; $i = 1, \dots, N_k$; $l \in \mathcal{V}(k) \cup \{0\}$ are generated independently. The shadowing factors are set to $Z_{i,k}(l) = \sigma \xi_{i,k}(l)$. Then the attenuation terms are defined as:

$$
\Gamma_{i,k}^*(l) = \begin{cases} \Gamma_{i,k}(l) & l \neq 0, \\ 10_k^{\mu}\, \Gamma_{i,k}(l) & l = 0, \end{cases} \quad C_{ik}^* = \max_{l \in \mathcal{V}(k)} \left(\Gamma_{i,k}^*(l) \right).
$$

Using the transformation $Z^*_{i,k}(0) = \mu_k + Z_{i,k}(0)$, the Brownian motion in (11) can be written as $(\mu_k/\sigma^2)B^*_{N^*_k,k} = N^*_k(\mu^2_k/\sigma^2) + (\mu_k/\sigma)\sum_{i=1}^{N^*_k}\xi_{i,k}(0)$. Finally, an unbiased estimator of outage at the base station $\{0\}$ is:

$$\hat{p}(\theta,\mu) = e^{\sum_{k=1}^{K}\left[(\theta_k-\lambda_k)-\frac{N^*_k\mu^2_k}{2\sigma^2}-\frac{\mu_k}{\sigma_k}\sum_{i=1}^{N^*_k}\xi_{i,k}(0)\right]} \times \prod_{k=1}^{K}\left(\frac{\lambda_k}{\theta_k}\right)^{N^*_k}\mathbf{1}\left(\sum_{k=1}^{K}\sum_{i=1}^{N^*_k}\frac{10^{\mu_k}\Gamma^*_{i,k}(0)}{C^*_{i,k}} > \alpha^{-1}\right),$$

The efficiency of such estimation is measured in terms of the computational effort required to achieve a certain relative precision and in this case improving efficiency is equivalent to finding the values of θ_k, μ_k that *minimise* the variance $\mathsf{Var}[\hat{p}(\theta,\mu)]$. Because $\mathsf{Var}(\hat{p}(\theta,\mu)) = \mathsf{E}(\hat{p}^2(\theta,\mu)) - p^2$ and p is independent of $(\theta,\mu) \in \mathbb{R}^{2K}$, it suffices to find the values of θ and μ that minimise $\mathsf{E}[\hat{p}(\theta,\mu)^2]$.

5   Functional Estimation

To simplify the exposition, we consider the homogeneous model where $\lambda_k = \lambda$ is independent of the cell. The notation can be simplified by calling $M = \sum_{k=1}^{K} N_k$ and numbering all the mobiles in sequence $i = 1,\ldots,M$ regardless of the cell where they belong. Then M is a Poisson number with mean $K\lambda$. Consider the change of measure using also a homogeneous model: $\theta_k = \theta, \mu_k = \mu$. In this section we propose a method to estimate

$$F(\theta,\mu) \equiv \mathsf{E}[\hat{p}^2(\theta,\mu)] = \mathsf{E}[L^2\mathbf{1}(\mathcal{B})], \tag{14}$$

as a function of the simulation parameters. Using the simplified notation, where $M^* \sim \mathrm{Poisson}(K\theta)$, for $\theta > \lambda$,

$$L^2\mathbf{1}(\mathcal{B}) = \exp\left\{2K(\theta-\lambda) - M^*\frac{\mu^2}{\sigma^2} - \frac{2\mu}{\sigma}\sum_{i=1}^{M^*}\xi_i(0)\right\} \times \left(\frac{\lambda}{\theta}\right)^{2M^*}\mathbf{1}\left(\sum_{i=1}^{M^*}\frac{10^\mu\Gamma_i(0)}{C_i} > \alpha^{-1}\right).$$

The idea is to generate a plot of the function in the plane (θ,μ). One way of doing this is to generate a number of independent replications of the simulation for each value in a prespecified grid. By the nature of the problem, it is clear that for many such values, the variance of the estimator may be difficult to estimate (it may be as hard as estimating the outage probability itself). Functional estimation presents two problems. First with independent replications, the computational effort is proportional to the number of points, and for each point a large number of replications may be necessary to obtain a reasonable estimate. The second problem is that functional estimation deals with the estimation of a curve, and therefore the error in the interpolated curves may be considerably larger than the pointwise error estimated at each run. Most of the computational effort at each point goes into generation of random numbers and the calculation of the outage event itself. A more sensible approach for functional estimation is to use common random numbers (Chapter 4, [3]), which works well under stochastic monotonicity of the function in the parameters (see Proposition 1). Suppose that we want to calculate $F(\theta,\mu)$ by simulation, for a range of values $(\theta_1,\ldots,\theta_r) \times (\mu_1,\ldots,\mu_s)$, in order to produce the plot of the variance.

By "functional estimation with CRN" it is understood that the random variables $L, \mathbf{1}(\mathcal{B})$ are evaluated from a common simulation trajectory ω under different values for the pair (θ, μ). The direct way to achieve this is to simulate first $M_1 \sim \text{Poisson}(K\lambda)$ callers ($\theta_0 = \lambda$) and place them in the network, then for each value of μ_m, evaluate the corresponding attenuation *using the same variable $\xi_{i,k}$ per caller*. That is, under CRN, $\Gamma_{i,0}(\theta_0, \mu_m) = 10^{\mu_m}\Gamma_{i,0}(\theta_0, \mu_0)$, for $\mu_0 = 0$. As well, if $k \neq \{0\}$, then the change of measure does not affect the attenuation and $\Gamma_{i,k}(\theta, \mu) = \Gamma_{i,k}(\lambda, 0)$. Next, for $\theta_{n+1} = \theta_n + \Delta\theta_n$, generate and place *only the new callers*, using the fact that the sum of independent Poisson random numbers is Poisson, so that the new callers are just $\text{Poisson}(\Delta\theta_n)$, and for these new mobile users and each μ_m compute the new attenuation factors only.

Proposition 1 *Functional estimation using CRN implies that $\mathbf{1}(\mathcal{B})(\theta, \mu)$ is non-decreasing a.s. in both arguments.*

Proof: Denote the interference by $I(\theta, \mu)$,

$$I(\theta, \mu) = \sum_{i=1}^{M_n} \frac{\Gamma_{i,0}(\theta, \mu)}{C_i(\theta, \mu)}.$$

If outage occurs at a given value (θ_0, μ_0), then for all values $\theta \geq \theta_0$ and $\mu = \mu_0$, the system (under CRN) will have the same mobile callers' interference *plus* a random number of new users that can only contribute to $I(\theta, \mu_0)$ with positive terms.

To complete the proof of the claim, it suffices to show that if $I(\theta_0, \mu_0) > \alpha^{-1}$, then $I(\theta_0, \mu) > \alpha^{-1}$ for all $\mu \geq \mu_0$. Let M_0 be the Poisson number of mobiles for the given sample where outage occurs. For any $\mu \geq \mu_0$, the term $I(\theta_0, \mu)$ is affected only from the changes in the attenuation factors $\Gamma_{i,0}(\theta_0, \mu)$ and the corresponding $C_i(\theta_0, \mu)$. There are a number of possibilities for each user $i = 1, \ldots, M_0$:

1. If $\kappa(i; \theta_0, \mu_0) = 0$ then $C_i(\theta_0, \mu) = \Gamma_{i,0}(\theta_0, \mu) = 10^{\mu-\mu_0}\Gamma_{i,0}(\theta_0, \mu_0) > \Gamma_{i,0}(\theta_0, \mu_0) \geq \Gamma_{ik}(\theta_0, \mu)$, for all $k \in \mathcal{V}(k)$, because $\Gamma_{ik}(\theta_0, \mu)$ is independent of μ (only $\Gamma_{i,0}$ is accelerated with the change of measure). Thus the contribution of this user is unchanged: $\Gamma_{i,0}(\theta_0, \mu)/C_i(\theta_0, \mu) = 1$.

2. If $\kappa(i; \theta_0, \mu_0) = k \neq 0$ then $C_i(\theta_0, \mu_0) = \Gamma_{i,k}(n)$ and two situations may happen:

   (a) If $\kappa(i; \theta_0, \mu) = k$ in the new simulation, that is if $C_i(\theta_0, \mu) = C_i(\theta_0, \mu_0) = \Gamma_{i,k}(\theta_0, \mu_0)$, then its contribution to the interference will be larger in the new system:

   $$\frac{\Gamma_{i,0}(\theta_0, \mu)}{C_i(\theta_0, \mu)} > \frac{\Gamma_{i,0}(\theta_0, \mu_0)}{C_i(\theta_0, \mu_0)},$$

   and this happens even when $\{0\} \notin \mathcal{V}(k)$.

   (b) If the increase in $\Gamma_{i,0}(\theta_0, \mu) > \Gamma_{i,k}(\theta_0, \mu_0)$ causes a change in power station (which can only happen if $\{0\} \in \mathcal{V}(k)$), then $C_i(\theta_0, \mu) = \Gamma_{i,0}(\theta_0, \mu), \kappa(i, \theta_0, \mu) = 0$ and consequently its contribution becomes unity. Because $C_i(\theta_0, \mu_0) > \Gamma_{i,0}(\theta_0, \mu_0)$, the contribution of this caller to $I(\theta_0, \mu_0)$ was necessarily less than one:

   $$\frac{\Gamma_{i,0}(\theta_0, \mu_0)}{C_i(\theta_0, \mu_0)} < 1 = \frac{\Gamma_{i,0}(\theta_0, \mu)}{C_i(\theta_0, \mu)}$$

This shows that for each $i \leq M_0$ the individual contributions are no smaller for the new system, or:

$$\sum_{i=1}^{M_0} \frac{\Gamma_{i,0}(\theta_0, \mu)}{C_i(\theta_0, \mu)} \geq I(\theta_0, \mu_0); \quad \mu \geq \mu_0,$$

which establishes stochastic monotonicity.

<div align="right">QED</div>

The advantages of functional estimation with CRN are twofold: CPU time is saved by reusing random variables and computation, plus the shape of the function can be judged better, because relative variability will be minimal between samples.

A more sophisticated method can be used to dynamically adjust the grid sizes according to the observations themselves. The method is a generalisation of the simulation tree method of [6]. The idea is to produce a *single* simulation (with common random numbers), to generate the function $F(\cdot)$ by adjusting the breakpoints to those particular values of θ that effect jumps in the function: indeed with high probability (when $\Delta\theta_n$ is small) many simulations of the systems for θ_n and $\theta_n + \Delta\theta_n$ will have the same number of mobiles and will contribute nothing new to the estimation. The simulation tree evolves calculating such points in θ that will produce a discontinuous effect in L^2 and $\mathbf{1}(\mathcal{B})$. Assume that the Poisson random variables for each θ_n use the same random variates in the generator:

- Let $U_1, U_2, \ldots$ be iid uniform random variates,
- set $p_n = U_n p_{n-1}$, with $p_0 = e^\theta$
- $M = \min\{n : p_n < 1\} - 1$.

The algorithm is based on the following observation. Fix a value of μ. From (1), it is clear that the indicator of outage will be a piecewise constant function of θ *for each trajectory*, that is, for each sequence of iid uniform variates $\{U_n\}$ and each sequence of iid standard normal variables $\{\xi_i\}$, different values of θ may yield different values of M^*, but between jumps in the number of callers, all values of θ yield the same outage. Recall that:

$$L^2 = \exp\left\{ 2K(\theta - \lambda) - M^* \frac{\mu^2}{\sigma^2} - \frac{2\mu}{\sigma} \sum_{i=1}^{M^*} \xi_i \right\} \left(\frac{\lambda}{\theta}\right)^{2M^*},$$

where $\sum_{i=1}^{M^*} \xi_i$ will be piecewise constant a.s. with the same breakpoints as M^*. It is therefore possible to reconstruct the values of L^2 between breakpoints by keeping track of the corresponding values of M^*, B_{M^*} and the subinterval in θ corresponding to a constant M^*. This is the principle of "retrospective simulation" where a single sample (simulation trajectory) can be used to keep track of the corresponding value of the random variable L^2 for a whole range of values of θ: the simulation is performed without need for specifying the parameter of the Poisson variable *a priori*. Let us now describe how the algorithm proceeds.

1. Start with $n = 0, m = 0, \theta_0 = \lambda, \mu_0 = 0$ and generate $M_0 \sim \text{Poisson}(K\theta_0)$. Place the callers and generate $Z_{i,0}(l) \sim \mathcal{N}(0, \sigma^2), l \in \mathcal{V}(k(i))$, where now $k(i)$ is the cell where mobile i is located.

2. Given θ_n, for each value $\mu_{m+1} = \Delta\mu_n + \mu_m$ use the common random variable $Z_{i,0}(l)$ to calculate the attenuation recursively by: $\Gamma_{i,0}(\theta_n, m+1) = 10^{\Delta\mu_n}\Gamma_{i,0}(\theta_n, m)$. All other quantities $\Gamma_{i,l}(\theta_n, m), l \neq 0$ are independent of μ and remain unchanged. Keep the values M^*, B_{M^*} for the retrospective evaluation of L^2.

3. If $n = S$ end. Otherwise set $n = n + 1$. Generate $U_n \sim U(0, 1)$ and look for the value $\Delta\theta_n$ for which there is exactly *one more* mobile $M^*(n+1) = M^*(n)$. Simple algebra shows that $\Delta\theta_n = -\log(U_n)$. The current breakpoint is set at $\theta_n = \theta_{n-1} + \Delta\theta_n$. Notice that now θ_n is a random variable. Go to 2.

At the end of S simulations, for every value of $\mu \in \{\mu_1, \ldots, \mu_s\}$ one has a list with the $r + 1$ breakpoints of each simulation, and the corresponding values of the variables $M^*(s), B_{M^*}(s), \mathbf{1}(\mathcal{B})(s)$. The average curve can be generated by merging all the partitions and averaging the S different values of the estimator $L^2\mathbf{1}(\mathcal{B})$ evaluated at, say the midpoint of the interval.
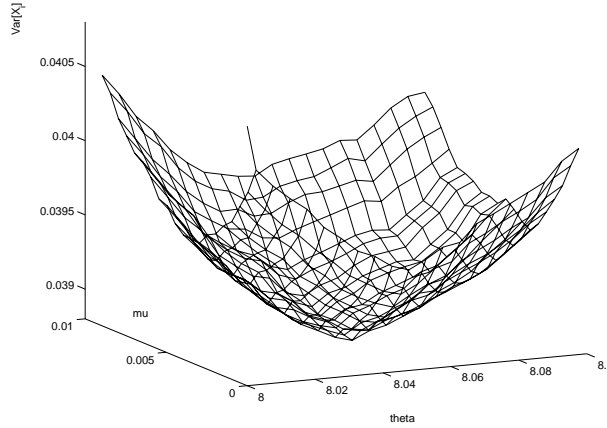

Figure 6: Functional Estimation

While the algorithm is harder to code than a prefixed grid, it is often the case that a reasonable range of values of θ is not known *a priori* and often many pilot simulations must be performed. In contrast, the simulation tree lets the system define the most significant breakpoints. Figure 6 shows the result of functional estimation on the plane, where convexity of the variance is clear.

6 Self-Optimised Importance Sampling

The idea of intelligent simulation is to use stochastic approximation as in [11] to adjust the values of θ and μ in the direction of decreasing variance. Let:

$$G(\theta, \mu) = \nabla_{\theta,\mu}\mathsf{E}[L^2(M^*; Z^*)\mathbf{1}(\mathcal{B}(M^*; Z^*)],$$

and suppose that $(\hat{G}_\theta(n), \hat{G}_\mu(n))$ is an unbiased estimator of the gradient, obtained with $S_n > 0$ independent simulations of the variables $(N; Z)$ at a value $(\theta(n), \mu(n))$. Consider the recursion:

$$
\begin{aligned}
\theta(n+1) &= \theta(n) - \epsilon_n \hat{G}_\theta(n) \\
\mu(n+1) &= \mu(n) - \epsilon_n \hat{G}_\mu(n)
\end{aligned}
\tag{15}
$$

Then $\theta(n) \to \theta^*$ and $\mu(n) \to \mu^*$ a.s. (Chapter 5, [5]).

If the function $L^2(N^*; Z^*)\mathbf{1}(\mathcal{B}^*)$ is a.s. Lipschitz continuous with integrable Lipschitz constant, then the derivative and expectation can be interchanged and the stochastic gradient is unbiased for $\nabla F(\cdot)$ [6]. For our problem, discontinuities arise because both the number of terms M^* and the outage event may jump as the values of θ and μ change. To overcome this difficulty and avoid gradient estimation methods that deal specifically with discontinuities, [10] propose to change back the measure, as we now do. From (11), we have for the homogeneous case:

$$
L(M^*; Z^*) = \exp\left\{ K(\theta - \lambda) + M^* \frac{\mu^2}{2\sigma^2} - \frac{\mu}{\sigma^2} \sum_{i=1}^{M^*} Z_i^*(0) \right\} \left(\frac{\lambda}{\theta} \right)^{M^*}.
\tag{16}
$$

The outage itself (under the star measure) is:

$$
\mathcal{B}(M^*; Z^*) = \mathbf{1}\left(\sum_{i=1}^{M^*} \frac{\Gamma_i^*(0)}{C_i^*} > \alpha^{-1} \right),
$$

where M^* is Poisson$(K\theta)$, and $Z_i^*(0), i = 1, \ldots, M^*$ are iid $\sim \mathcal{N}(\mu, \sigma^2)$. As before, let $B_{M^*}^* = \sum_{i=1}^{M^*} Z_i^*(0)$.

Lemma 4 *Under the change of measure, the estimators:*

$$
\hat{G}_\theta = \left(K - \frac{M^*}{\theta} \right) L^2(M^*; Z^*)\mathbf{1}(\mathcal{B}(M^*; Z^*))
\tag{17}
$$

$$
\hat{G}_\mu = \left(\frac{M^*\mu - B_{M^*}^*}{\sigma^2} \right) L^2(M^*; Z^*)\mathbf{1}(\mathcal{B}(M^*; Z^*))
\tag{18}
$$

are unbiased for the derivatives of $F(\theta, \mu)$ w.r.t. θ and μ, respectively.

Proof: Using Lemma 3 the expectation in $F(\theta, \mu)$ can be written in terms of the *original* measure, so that

$$
G(\theta, \mu) = \nabla_{\theta, \mu} \mathsf{E}[L(M; Z)\mathbf{1}(\mathcal{B}(M; Z))],
$$

where now $M \sim$ Poisson$(K\lambda)$ and $Z_i(0), i = 1, \ldots, M$ are iid $\sim \mathcal{N}(0, \sigma^2)$ are *independent* of θ and μ. For every $M; Z$ the function $L^2(M; Z)$ is differentiable in θ and μ, and this derivative has uniformly bounded expectation over compact sets (in θ, μ). Indeed simple algebra yields:

$$
\frac{\partial}{\partial \theta} L(M; Z) = \left(K - \frac{M}{\theta} \right) L(M; Z)
\tag{19}
$$

$$
\frac{\partial}{\partial \mu} L(M; Z) = \left(\frac{M\mu - B_M}{\sigma^2} \right) L(M; Z)
\tag{20}
$$

where $B_M = \sum_{i=1}^{M} Z_i(0)$. To show that these derivatives have uniformly bounded expectations over compact sets in (θ, μ) we use the result that a Poisson random number has a finite moment generating function, as shown below. For the sake of space only the result for (19) is shown: similar arguments apply for (20). Clearly $(K - M/\theta)L\mathbf{1}(\mathcal{B})$ is bounded a.s. by $(K - M/\theta)L$, so it suffices to show that this random variable has uniformly bounded expectation over any θ-compact set. Because now the measure is the original one, $M \sim \text{Poisson}(K\lambda)$ and $Y(M) = (\mu/\sigma^2)B_M \sim \mathcal{N}(0, M(\mu^2/\sigma^2))$. Conditioning on M, and using $\mathsf{E}(e^{-Y} \mid M) = e^{-M\mu^2/\sigma^2}$, it follows that $\mathsf{E}\left[\left(K - \frac{M}{\theta}\right)L(M; Z)\right] = \mathsf{E}\left[\left(K - \frac{M}{\theta}\right)e^{K(\theta-\lambda)}\left(\frac{\lambda}{\theta}\right)^M\right]$, where this latter expectation is w.r.t. a Poisson random number M. For such random variables the moment generating function $\phi(a) = \mathsf{E}[e^{aM}]$ exists and

$$\mathsf{E}\left[-\frac{M}{\theta}\left(\frac{\lambda}{\theta}\right)^M\right] = \sum_{m \geq 0} -\frac{m}{\theta}\left(\frac{\lambda}{\theta}\right)^m \frac{\lambda^m}{m!} = \frac{\partial}{\partial\theta}\sum_{m \geq 0}\left(\frac{\lambda}{\theta}\right)^m \frac{\lambda^m}{m!} = \frac{\partial}{\partial\theta}\phi[\ln(\lambda/\theta)] < \infty,$$

thus proving that the estimator has indeed a uniformly bounded expectation. This result establishes an unbiased estimator of the gradient to use in the stochastic approximation (15), the practical problem is that the simulation itself is carried out under the change of measure, namely M^* is generated as a $\text{Poisson}(K\theta)$ number. As proposed in [10], the measure is changed back, this time for the stochastic gradients (19) and (20), to establish the result:

$$G_\theta = \mathsf{E}\left[\left(K - \frac{M^*}{\theta}\right)L^2(M^*; Z^*)\mathbf{1}(\mathcal{B}(M^*; Z^*))\right]$$

$$G_\mu = \mathsf{E}\left[\left(\frac{M^*\mu - B^*_{M^*}}{\sigma^2}\right)L^2(M^*; Z^*)\mathbf{1}(\mathcal{B}(M^*; Z^*))\right]$$

<div align="right">QED</div>

6.1 Stratified Importance Sampling

By symmetry of our model, we use different control variables (θ_1, μ_1), (θ_2, μ_2) for those cells within and without the neighborhood $\mathcal{V}(0)$, respectively. However, our method could be easily adapted to other network models. The motivations are several: better performance can be achieved as the dimension of the control variables grows; generally, simulation time increases with the dimension (so more stratification is not useful); contrary to functional estimation, the computational overhead for intelligent simulation is independent of the dimension of the problem. For those reasons, we decided to stratify the network this way which gives us a more sensitive estimation of the optimal parameters.

Here are the details of the calculation of each derivate. The likelihood ratio can be rewritten in terms of the neighborhood as:

$$L(M^*, Z^*) = L_1(M_1^*, Z_1^*) \times L_2(M_2^*, Z_2^*)$$

$$L_1(M_1^*, Z_1^*) = \exp\left\{7(\theta_1 - \lambda) + \frac{M_1^* \mu_1^2}{2\sigma^2} - \frac{\mu_1}{\sigma^2}\sum_{i=1}^{M_1^*} Z_i(0)\right\}\left(\frac{\lambda}{\theta_1}\right)^{M_1^*}$$

$$L_2(M_2^*, Z_2^*) = \exp\left\{(K-7)(\theta_2 - \lambda) + \frac{M_2^* \mu_2^2}{2\sigma^2} - \frac{\mu_2}{\sigma^2}\sum_{i=1}^{M_2^*} Z_i(0)\right\}\left(\frac{\lambda}{\theta_2}\right)^{M_2^*}$$

where all the variables with index 1 are the ones associated with the neighborhood $\mathcal{V}(0)$, while those with index 2 are the others. Remark that because the cells are hexagonal, there are 7 cells in the neighborhood of the central cell and $(K-7)$ cells outside. Now, if we derivate each component with respect to each variable passing through the original measure (like done in the previous section), we obtain the following estimators:

$$G_{\theta_1} = \mathsf{E}\left[\left(7 - \frac{M_1^*}{\theta_1}\right) L_1^2(M^*; Z^*) L_2^2(M^*; Z^*) \quad \mathbf{1}\left(\mathcal{B}(M^*; Z^*)\right)\right] \tag{21}$$

$$G_{\theta_2} = \mathsf{E}\left[\left((K-7) - \frac{M_2^*}{\theta_2}\right) L_1^2(M^*; Z^*) L_2^2(M^*; Z^*) \quad \mathbf{1}\left(\mathcal{B}(M^*; Z^*)\right)\right] \tag{22}$$

$$G_{\mu_1} = \mathsf{E}\left[\left(\frac{M_1^* \mu_1 - B_{M_1^*}^*}{\sigma^2}\right) L_1^2(M^*; Z^*) L_2^2(M^*; Z^*) \quad \mathbf{1}\left(\mathcal{B}(M^*; Z^*)\right)\right] \tag{23}$$

$$G_{\mu_2} = \mathsf{E}\left[\left(\frac{M_2^* \mu_2 - B_{M_2^*}^*}{\sigma^2}\right) L_1^2(M^*; Z^*) L_2^2(M^*; Z^*) \quad \mathbf{1}\left(\mathcal{B}(M^*; Z^*)\right)\right] \tag{24}$$

where $B_{M_1^*}^* = \sum_{i=1}^{M_1^*} Z_i^*(0)$ and $B_{M_2^*}^* = \sum_{i=1}^{M_2^*} Z_i^*(0)$.

Intelligent simulation uses the observations of the process under Importance Sampling to estimate both the outage as well as the gradient of the variance, to adjust the simulation parameters as the estimation itself is being carried out. Contrary to functional estimation, the computational overhead of the self-optimised method is independent of the dimension of the problem, and better performance can be achieved with higher dimensional control variables. To compare, Figure 7 shows the trajectories of the self-optimised parameters with and without stratification. The dotted line is the optimal value as estimated from Figure 6. Not surprisingly, stratifying yields virtually no change of measure for those callers outside the neighborhood of the base station, while those inside have a higher intensity and tilted shadowing. The optimal variance was 15% better than without stratification. The intelligent simulations took a total of 5.1 minutes, while the graph in Figure 6 required a total of 25.75 minutes.
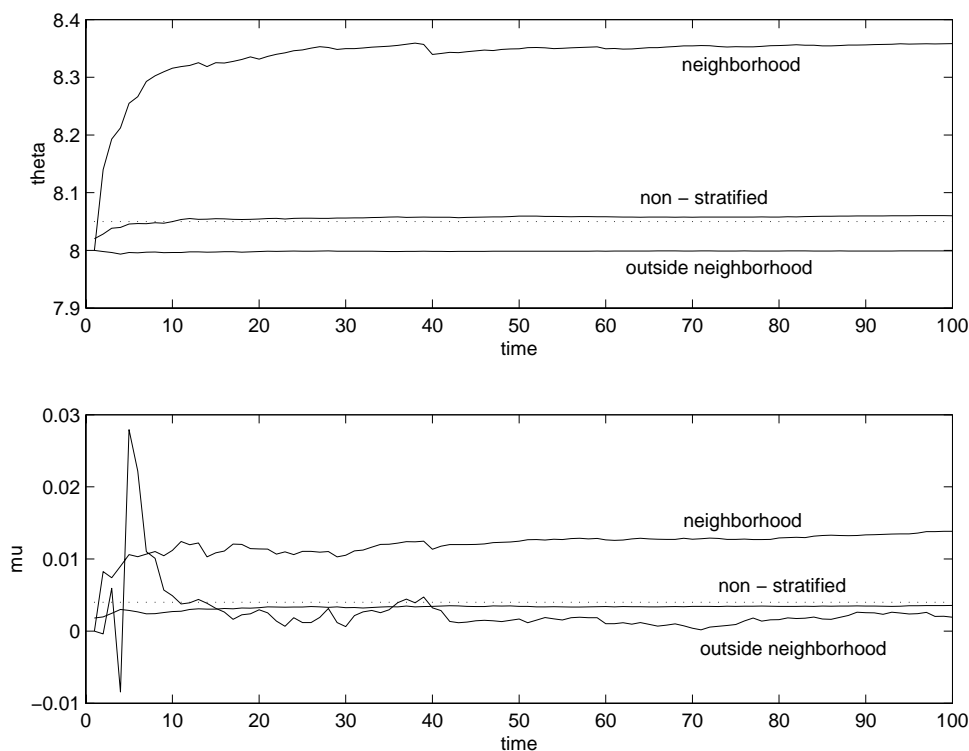
Figure 7: Intelligent Simulation

7   Conclusion

In this paper we have shown how two recent simulation-based methods can be applied to the important and yet unresolved problem of evaluating the performance of CDMA wireless networks. The algorithms are robust: they can easily be adapted to imperfect power control, inhomogeneous traffic models and non-constant traffic intensities. While this would require code preparation to set up the basic simulations, calculation of the change of measure and derivatives do not require more execution time. Our change of measure is not provably BRE (bounded relative error), mostly because of the difficulty in tilting the distribution of the interference itself. Other changes of measure could be used, in which case both the simulation tree as well as the accelerated simulation could be applied to learn the optimal parameters. It is well known that one of the major problems applying stochastic approximation is the choice of initial values and step sizes: the transient behaviour may be very sensitive to these. We are currently exploring the possibility of using a simulation tree with adaptive grid to start the method and seek a reasonable step size, as well as to determine an initial region for the parameters.

Because there are no other methods to solve this problem, simulation has become a benchmark against which to test approximate solutions. We propose to provide the researches with an intelligent simulation method that may give useful results in a reasonable amount of time. The gain in computational time can be substantial and the optimisation of the parameters is done concurrently with the simulation, requiring virtually no pilot simulations.

References

[1] J. A. Bucklew. *Large Deviations Techniques in Decision Simulation and Estimation*. Wiley, New York, 1990.

[2] C. Chan and S. Hanly. Calculating the outage probability in a cdma network with spatial poisson traffic. *IEEE Transactions on Vehicular Technology*, 50(1):183–204, 2001.

[3] G. Fishman. *Monte Carlo concepts, algorithms and applications*. Springer-Verlag, New York, 1996.

[4] D. Foata and A. Fuchs. *Calcul des probabilités, 2ème édition*. Dunod, Paris, 1998.

[5] H.J. Kushner and G. Yin. *Stochastic Approximation Algorithms and Applications*. Springer-Verlag, New York, 1997.

[6] B. Martin P. L'Ecuyer and F.J. Vázquez-Abad. Functional estimation for a multicomponent age replacement model(special issue on modern digital methodology). *American Journal of Mathematical and Management Sciences*, 19(1 and 2):135–156, December 1999.

[7] B. D. Ripley. *Stochastic Simulation*. John Wiley, USA, 1987.

[8] S. M. Ross. *Simulation, 2ème édition*. Academic Press, USA, 1997.

[9] P. Shahabuddin. Importance sampling for the simulation of highly reliable Markovian systems. *Management Science*, 40(3):333–352, 1994.

[10] Y. Su and M.Fu. Importance sampling in derivative securities pricing. In K. Kang J.A. Joines, R.R. Barton and P.A. Fishwick, editors, *Proceedings of the 2000 Winter Simulation Conference*, pages 587–596, 1994.

[11] F.J. Vázquez-Abad and D. Dufresne. Accelerated simulation for pricing asian options. In J.S. Carson D.J. Medeiros, E.F. Watson and M.S. Manivannan, editors, *Proceedings of the 1998 Winter Simulation Conference*, pages 1493–1500, 1998.