Evaluation numérique de l'information de Fisher pour des observations irrégulières de l'état d'une file d'attente

Olivier Cappé, François Roueff
CNRS LTCI / GET - Télécom Paris,
46 rue Barrault, 75634 Paris cedex 13, France
cappe -@- tsi.enst.fr, roueff -@- tsi.enst.fr

Résumé – Cette contribution est consacrée à l'évaluation des performances d'estimation (bornes d'estimation ainsi que performances d'estimateurs particuliers) pour des modèles à temps continu de type "file d'attente" observés à des instants discrets et de manière irrégulière. Il s'agit d'un problème important dans le domaine de la métrologie des réseaux. Les solutions que nous apportons sont de nature algorithmique et permettent l'évaluation numériques des performances statistiques. Ces méthodes apportent de nouveaux résultats concernant l'efficacité relative de certains schémas d'échantillonnage utilisés en métrologie des réseaux (réguliers, en paires, etc.)

Abstract – In this contribution, we study the problem of performance evaluation for the estimation of the parameters of a queuing model observed at, possibly irregular, discrete instants. This is a crucial problem for evaluating the parameters of a communication network from sampled measurements. To our knowledge, effective computations of the information for such models has not been addressed so far. We provide algorithms which allow for the numerical evaluation of statistical performances. These methods are applied for comparing different sampling schemes proposed in the field of active network measurements (random Poisson sampling, regular sampling, packet pairs...)

1 Contexte

Une campagne de mesure active consiste à envoyer des sondes d'un point à un autre du réseau et à recueillir les données auxquelles celles-ci permettent d'accéder : délais, pertes, informations de routage, etc. L'objectif de ces mesures est de permettre d'estimer des paramètres physiques du réseau ainsi que les paramètres du trafic empruntant la même route que les sondes. Les travaux fondateurs étudiant les possibilités offertes par ce type de mesures remontent à un peu moins d'une décennie [3], [4], [8]. A partir de modèles probabilistes de files d'attente (qui rendent compte de la dynamique des délais dans le réseau), l'estimation des paramètres de réseau par mesure active se ramène à un problème d'estimation statistique des paramètres d'une file d'attente échantillonnée à des instants discrets $\{C(t_k),\, t_1 < \cdots < t_n\}$. Dans ce cadre, une solution simple [1] consiste à utiliser des estimateurs de moments (de la forme $f(1/n \sum_{l=1}^{n} g(C_{t_l})))$ basés sur la loi stationnaire du modèle considéré. Dans le cas de la file d'attente M/M/1 considéré ci-dessous, $1/n \sum_{l=1}^{n} \mathbb{1}(C(t_l) > 0)$ permet (par exemple) d'estimer le taux de charge de la file. Toutefois même pour ces estimateurs de forme très simple, la question des performances d'estimation et, en particulier, de leur dépendance vis à vis du choix des instants d'observations $t_1, \ldots, t_n$ n'a pas de réponse simple. Cette problématique, souvent évoquée dans ce type de problème, concerne le choix des instants d'envoi des sondes. Dans un cadre plus général, nous nous demanderons comment choisir la répartition temporelle des observations d'une file d'attente pour optimiser l'estimation de ses paramètres.

Dans cet article, nous ne détaillerons que le cas, plus simple, où les observations portent sur le nombre de clients (ou tâches) présentes dans la file d'attente à des instants donnés. Le cas où les observations correspondent à des mesures du temps de traitement, aussi appelé "charge" du serveur (dans le cas des réseaux, celui-ci est lié la taille totale des données présentes dans la file d'attente), est plus complexe mais peut être traité avec une extension de l'approche proposée. Dans la suite nous considérerons une file d'attente du type M/M/1. Soit $(C(t))_{t\in\mathbb{R}^+}$ la chaîne de Markov à temps continu qui représente l'évolution du nombre de clients dans la fille d'attente. On suppose que l'on dispose de mesures (non bruitées) de l'état du processus à des temps $t_1, \ldots, t_n$. La principale difficulté posée par ce modèle statistique est que les observations ne fournissent qu'une observation très partielle du système. A contrario, si l'on observait toute la trajectoire (à temps continu) entre les dates t_1 et t_n, l'estimation au sens du maximum de vraisemblance se ramènerait à problème bien maîtrisé. Ce problème est concrètement posé par le fait que les matrices de transition de la chaîne échantillonnée sont plus difficiles à manier pour le calcul des quantités habituellement utilisées en théorie de l'estimation. Ce type de question a été étudié dans le contexte de l'estimateur par maximum de vraisemblance des processus Markovien de sauts (voir [5]).

Une hypothèse importante de ce travail consiste à considérer que le processus de mesure utilisé ne perturbe pas l'état du système. En pratique, il est clair que l'envoi de paquets sondes sur le réseau est susceptible de modifier, au moins marginalement, son état [1]. Dans un premier temps, il nous semble toutefois plus informatif d'investiguer les performances d'une technique de mesure supposée non intrusive.

2 Rappels sur la file M/M/1

Nous faisons un rappel très succinct sur la file M/M/1. Le lecteur pourra avec profit se reporter aux nombreux ouvrages existants (citons [2]). Une file d'attente modélise un système de traitement de tâches. Celles-ci arrivent à des temps modélisés par un processus d'arrivée et à chacune d'elles est assigné un temps de traitement. La stratégie de traitement considérée est PAPS (premier arrivé, premier servi). Pour une file M/M/1 les arrivées, disons à partir de $t = 0$ (l'état de la file étant donné à l'instant $t = 0$ par des conditions initiales), sont modélisées par un processus ponctuel de Poisson homogène d'intensité λ marqué par des temps de traitements i.i.d. de loi exponentielle de paramètre μ et indépendants des temps d'arrivées. Le processus du nombre de clients $(C(t))_{t\in\mathbb{R}^+}$ est un processus de vie et de mort, les clients arrivant aux instants définis par le processus de Poisson des arrivées déjà mentionné, et sortant (tant que la file est non-vide) à des instants correspondant à un processus de Poisson indépendant d'intensité μ. Nous considérons le cas où la file est stable c'est-à-dire, $\lambda < \mu$. Dans ces conditions, la file converge géométriquement vers sa loi stationnaire et repasse régulièrement vers l'état vide (en un temps d'espérance finie).

3 Information de Fisher

Considérons une observation $\mathbf{X}_n = (X_0 \ \cdots \ X_n) \in \mathcal{X}^{(n+1)}$ associée à un modèle statistique paramétrique dominé $\{\mathbb{P}_\theta = p_\theta(\cdot)\, d\gamma(\cdot)\,,\, \theta \in \Theta\}$ où $\Theta \subset \mathbb{R}^d$. On définit la matrice d'information de Fisher par

$$\mathbf{I}_n(\theta) = \mathbb{E}_\theta\left[\nabla_\theta L_\theta(\mathbf{X}) \nabla_\theta^T L_\theta(\mathbf{X})\right] = \mathrm{cov}_\theta\left(\nabla_\theta L_\theta(\mathbf{X})\right), \quad (1)$$

où $L_\theta \overset{\text{def}}{=} \log p_\theta$. Sous des hypothèses supplémentaires d'intégrabilité uniforme des dérivées, on a de plus $\mathbf{I}_n(\theta) = -\mathbb{E}_\theta\left[\nabla^2_{\theta_\theta T} L_\theta(X)\right]$. L'inverse de la matrice d'information de Fisher permet de minorer la variance des estimateurs sans biais de θ ("borne de Cramer-Rao") et fournit, dans les modèles réguliers, un équivalent asymptotique de la variance de l'estimateur du maximum de vraisemblance quand le nombre d'observations n tend vers l'infini. Dans la suite, nous nous ramènerons toujours (éventuellement par discrétisation et troncature) au cas où $\mathcal{X}$ est un ensemble fini (de cardinal noté r) et l'on ne fait donc pas figurer la mesure de domination γ dans les expressions ci-dessous.

On s'intéresse ici au cas particulier où $\mathbf{X}_n$ correspond à l'observation d'un nombre fini de valeurs successives d'une chaîne de Markov inhomogène, à valeur dans un ensemble $\mathcal{X}$ fini. Plus précisément,

$$\mathbb{P}_\theta(x_0, x_1, \ldots, x_n) = \nu_\theta(x_0) \prod_{k=0}^{n-1} P_{\theta,k}(x_k, x_{k+1}),$$

$\{P_{\theta,k} : \mathcal{X}^2 \to [0,1],\, k \in \mathbb{N}\}$ étant une séquence de matrices de transition paramétrées par θ, ν_θ la distribution initiale (elle aussi paramétrée par θ). Pour une mesure ν et une matrice de transition P, la notation νP désigne classiquement la mesure définie par $\sum_x \nu(x) P(x, \cdot)$ (les mesures sont donc assimilées à des vecteurs lignes). L'élément d'indices x, x' de la matrice de transition $P_{\theta,k}$ correspond à $\mathbb{P}_\theta(X_{k+1} = x' | X_k = x)$. Soient $I_\nu(\theta)$ la matrice d'information de Fisher associée à ν_θ et $I_k : \mathcal{X} \times \Theta \to \mathbb{R}^{2d}$ la matrice d'information conditionnelle associée à la loi $P_{\theta,k}(x, \cdot)$, c'est-à-dire,

$$I_k(x, \theta) = \mathrm{cov}_\theta\left(\nabla_\theta \log P_{\theta,k}(x, X_{k+1})\right).$$

Par conditionnements successifs, il est aisé de vérifier que la matrice d'observation de Fisher a une structure additive (en le nombre d'observations) :

$$\mathbf{I}_n(\theta) = (I_\nu(\theta) +) \sum_{k=0}^{n-1} \mathbb{E}_\theta\left[I_k(X_k, \theta)\right]. \quad (2)$$

Dans la suite, la partie entre parenthèses (liée à la condition initiale) ne nous intéressera pas dans la mesure où, lorsque le nombre d'observations n augmente, elle est clairement négligeable face au second terme.

On considère dorénavant le cas où l'on échantillonne un processus de Markov homogène à temps continu $(X(t), t \geq 0)$, de loi paramétrée par θ, à des instants $t_0, \ldots, t_n$. De par la propriété de Markov, on est ramené dans le contexte étudié précédemment à condition de considérer les noyaux de transition

$$P_{\theta,k}(x, x') = \mathbb{P}_\theta(X(t_{k+1}) = x' | X(t_k) = x).$$

L'inhomogénéité est ici causée uniquement par le fait que le processus à temps continu n'est pas (nécessairement) échantillonné de manière régulière. Si Q_θ désigne le générateur infinitésimal associé au processus à temps continu [2], on dispose de l'expression explicite suivante

$$P_{\theta,k} = \exp\left[(t_{k+1} - t_k) Q_\theta\right], \quad (3)$$

où exp désigne l'exponentielle matricielle. Une autre conséquence importante, est que dans ce cas particulier, les matrices de transition $P_{\theta,k}$ bien qu'elles soient inhomogènes, possèdent toutes la même loi stationnaire qui est celle du processus à temps continu $X(t)$ (que nous noterons π_θ). En particulier, si la loi initiale ν_θ coïncide avec π_θ, l'espérance qui figure dans l'équation (2), se fait toujours par rapport à la loi π_θ puisque $X_k = X(t_k)$ suit la loi π_θ (quel que soit l'indice k). Nous nous placerons dans ce cas par la suite. Il est alors clair que l'ordre des termes figurant dans la somme de l'équation (2) peut être modifié sans que la matrice d'information de Fisher ne soit modifiée. Par exemple, le schéma d'échantillonnage "en paires" ($t_{2i} = i, t_{2i+1} = i + \epsilon$) pour $i = 0, \ldots, \ell$ (avec $\epsilon < 1$) conduit à la même matrice d'information que $(t_i = i\epsilon)_{0\leq i\leq \ell}$ suivi de $(t_i = \ell\epsilon + (i+1)(1-\epsilon))_{0\leq i\leq \ell}$ (échantillonnage à une cadence rapide, puis lente). Rappelons une fois encore que l'on ne considère ici que les performances statistiques en ignorant les perturbations apportées au trafic par la mesure (aspect sous lequel les deux schémas évoqués ci-dessus sont probablement assez distincts).

4 Méthodes de calcul numérique

On considère maintenant diverses approches permettant d'évaluer numériquement la matrice d'information de Fisher donnée par (2) dans le cas de l'échantillonnage d'un processus de Markov à temps continu stationnaire. D'après la discussion qui précède, il suffit alors de parvenir à évaluer le terme générique de (2) qui s'écrit en utilisant (1) et (3)

$$I_{(h)}(\theta) = \sum_{x,x' \in \mathcal{X}} \pi_\theta(x) P_{\theta,(h)}(x,x') \nabla_\theta L_{\theta,(h)}(x,x') \nabla_\theta^T L_{\theta,(h)}(x,x'),$$

où $P_{\theta,(h)} \stackrel{\text{def}}{=} \exp(hQ_\theta)$ et $L_{\theta,(h)} \stackrel{\text{def}}{=} \log P_{\theta,(h)}$, $h \in \mathbb{R}^+$ désignant l'écart générique entre deux instants d'échantillonnage successifs.

Nous avons mis en œuvre trois méthodes différentes permettant d'évaluer numériquement cette information de Fisher :

Evaluation numérique directe Tous les environnements de calcul numériques comportent en général une fonction permettant de calculer directement l'exponentielle matricielle. La méthode réputée la plus fiable numériquement utilise une approximation de type Padé [7]. On peut donc calculer directement $P_{\theta,(h)}$ et évaluer approximativement $\nabla_\theta \log P_{\theta,(h)}$ par différences finies (en évaluant l'exponentielle matricielle en d points "proches" de θ). Cette approche a le mérite d'être facilement implémentable et de ne requérir aucune connaissance sur la dépendance en θ de Q_θ (autre que d'être capable de l'évaluer numériquement). Elle est toutefois lourde à mettre en œuvre quand $\mathcal{X}$ est un ensemble de cardinal r élevé et fournit une évaluation du gradient dont la précision est impossible à quantifier précisément (dépend de la dérivée seconde dans la direction des perturbations).

Décomposition propre et calcul opérateur Une autre solution consiste à calculer numériquement une décomposition propre de Q_θ sous la forme $Q_\theta = V_\theta D_\theta V_\theta^{-1}$ où D_θ est une matrice diagonale contenant les valeurs propres $(d_{\theta,k})_{1 \leq k \leq r}$. Dès lors, on peut évaluer $\exp(hQ_\theta)$ sous la forme $V_\theta D_{\theta,(h)} V_\theta^{-1}$ où $D_{\theta,(h)}$ est une matrice diagonale dont les éléments valent $\mathrm{e}^{hd_{\theta,k}}$. Par ailleurs, l'exponentielle matricielle admet également une représentation spectrale sous la forme

$$\exp(hQ_\theta) = \frac{1}{2i\pi} \int_\Gamma \mathrm{e}^s (sI - Q_\theta)^{-1} ds,$$

où Γ est un contour de $\mathbb{C}$ englobant les valeurs propres d_k [6]. En utilisant la formule des résidus, il est possible de montrer que (ici pour un paramètre θ scalaire)

$$\frac{\partial}{\partial\theta} \exp(hQ_\theta) = V_\theta \left\{ \left[V_\theta^{-1} \left(\frac{\partial}{\partial\theta} Q_\theta \right) V_\theta \right] \odot G_{(h)} \right\} V_\theta^{-1}$$

où $G_{(h),kk} = h\,\mathrm{e}^{hd_k}$ et $G_{(h),kl} = (\mathrm{e}^{hd_k} - \mathrm{e}^{hd_l})/(d_k - d_l)$ pour $k \neq l$ (le symbole $\odot$ désigne le produit dit de Hadamard ou terme à terme). Cette méthode, qui suppose de savoir dériver (analytiquement) Q_θ par rapport à θ fournit une forme relativement explicite de la dérivée de Q_θ. Elle est également moins coûteuse en temps de calcul en cas d'évaluations répétées pour des valeurs de h différentes puisque la décomposition propre n'est calculée qu'une fois. Elle est malheureusement moins stable numériquement [7], en particulier pour des grandes valeurs de r.

Discrétisation temporelle Lorsque h tends vers 0, on sait que $P_{\theta,(h)}$ est approximativement égal à $\widehat{P}_{\theta,(h)}$ où $\widehat{P}_{\theta,(h)}$ est une matrice de transition dont le terme d'indices k, l (pour $k \neq l$) vaut $hQ_{\theta,kl}$ [2]. En fixant un pas de discrétisation temporelle $1/l$ (avec $l \gg 1$), on peut donc approximer $P_{\theta,(h)}$ par

$$\widehat{P}_{\theta,(h),l} = \prod_{i=1}^{\lfloor h/l \rfloor} \widehat{P}_{\theta,(1/l)}. \tag{4}$$

Cette méthode est liée à celle (plus générique) qui consiste à tronquer le développement en série entière de $\exp(hQ_\theta)$ mais elle garantit que l'approximation de $P_{\theta,(h)}$ obtenue est bien une matrice de transition. On peut montrer que cette méthode est stable dans le sens où pour tout h, et toute loi initiale ν,

$$\|\nu \widehat{P}_{\theta,(h),l} - \nu P_{\theta,(h)}\|_1 \leq C_\theta / l,$$

C_θ étant une constante qui ne dépend que de r et de $\max_{x \in \mathcal{X}}\{-Q_\theta(x,x)\}$.Pour évaluer les dérivées, on utilise simplement une relation de récurrence obtenue en dérivant (4) par rapport à θ. Cette dernière approche a le défaut d'être sensible à la valeur absolue des taux $Q_\theta(x,x')$, une discrétisation plus fine devenant nécessaire lorsque les taux augmentent. Elle a toutefois le mérite de montrer qu'il est possible d'approcher la vraisemblance (et ses dérivées) par un calcul récursif, ce qui suggère une possible implémentation (approchée) adaptative de l'estimateur du maximum de vraisemblance.

5 Simulations et résultats

A titre d'exemple, nous considérons ici le processus du nombre de clients $(C(t), t \in \mathbb{R}^+)$ associé à la file d'attente M/M/1 dans laquelle le processus d'arrivée des clients est de taux λ et le processus de traitement de taux μ.

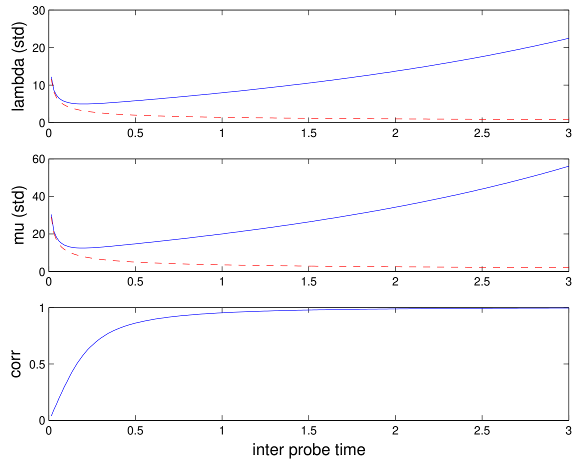

FIG. 1: *De haut en bas: Ecarts-types pour λ, μ et coefficient de corrélation en fonction de h. Superposés en tirets, les comportement théorique des écarts-types pour les faibles valeurs de h.*

Les figures 1 et 2 montrent les résultats obtenus pour $\lambda = 2, \mu = 5$ en utilisant la troisième méthode d'évaluation numérique mentionnée au paragraphe précédent, avec $l =$

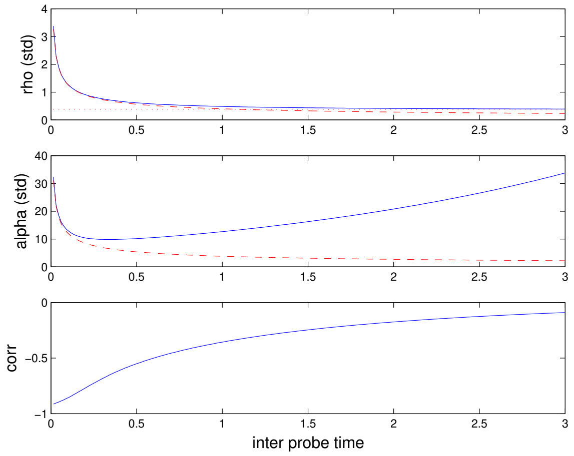

FIG. 2: *De haut en bas: Ecarts-types pour ρ, α et coefficient de corrélation en fonction de h. Superposés, en tirets, les comportement théorique des écarts-types pour les faibles valeurs de h et, en traits pointillés, le comportement de l'écart-type sur ρ pour les fortes valeurs de h.*

500. Notons que pour l'implémentation, il est également nécessaire de tronquer l'espace d'état du système, les figures 1 et 2 correspondent en fait à un modèle M/M/1/20 (ou le nombre maximal de clients est 20), mais cette troncature est sans effet sur les résultats présentés ici. Sur ces figures, on représente les éléments de $I_{(h)}^{-1}(\theta)$ plutôt que ceux de $I_{(h)}(\theta)$ en prenant la racine carré pour les termes diagonaux pour obtenir des grandeurs homogènes à des écarts-types et la forme normalisée (coefficient de corrélation) pour le terme hors diagonale.

La figure 1 concerne la paramétrisation originale $\theta = (\lambda, \mu)$. Sur ces courbes, on constate que les performances d'estimation sont optimales pour des valeurs de h de l'ordre du temps moyens entre deux événements (arrivée ou départ) dans la file étudiée. Les performances se dégradent très rapidement lorsque h augmente avec des erreurs fortement corrélées sur λ et μ. En effet, la file étant stable, $P_{\theta,(h)}$ tend assez rapidement vers la loi stationnaire du système, or cette dernière (loi géométrique de paramètre λ/μ) ne dépend que du rapport λ/μ. Pour les faibles valeurs de h, il est possible de faire une analyse simplifiée en considérant que les trois seuls événements de probabilités significatives susceptibles de se produire pendant la durée h sont l'arrivée d'un client (probabilité λh), le départ d'un client (probabilité μh, sauf si la file est déjà vide), où le fait que l'état de la file ne change pas. Sur la base de ce raisonnement, on obtient

$$I_{(h)}(\theta) = \begin{pmatrix} h/\lambda & 0 \\ 0 & (\lambda/\mu)h/\mu \end{pmatrix} + O(h^2),$$

qui correspond aux courbes représentées en traits pointillés.

Compte tenu de ces résultats, il est plus sain de considérer la paramétrisation $\theta = (\rho, \alpha) = (\lambda/\mu, \mu - \alpha)$ comme sur la figure 2. Le plus frappant ici est que les performances d'estimation pour ρ s'améliorent de manière monotone avec h pour atteindre (pour les grandes valeurs de h) l'information correspondant à la loi stationnaire $I_\pi(\rho) = 1/(\rho(1-\rho)^2)$. La figure 2 suggère que pour l'estimation de ρ, l'écart temporel entre les observations importe peu, du moment qu'il est suffisant, et que la situation la plus favorable correspond au cas de quasi indépendance entre les observations successives (grandes valeurs de h). Pour la différence des taux α, les performances d'estimation sont faibles et se dégradent assez rapidement lorsque h augmente. Il est à noter que dans le contexte de la mesure active de réseau où l'observation se fait par émission d'un paquet sur le réseau, la plage de h qui semble la plus favorable sur la figure 2 (temps de l'ordre du temps moyen entre deux événements) correspond en fait à des régimes tous à fait irréalistes dans lequel le trafic généré par les paquets sondes n'est plus du tout négligeable face au trafic réel de données. Par conséquent, la partie la plus pertinente de la courbe est la partie droite qui montre une sévère dégradation pour les performances d'estimation de α (lorsque h augmente). Qualitativement, on obtient les mêmes conclusions pour toutes les files M/M/1 nettement stables (telles que ρ soit nettement inférieur à 1) ce qui correspond au cas le plus représentatif pour les mesures de trafic de bout en bout.

6 Conclusions

L'approche présentée ci-dessus fournit une évaluation numérique des performances statistiques qui s'avère fructueuse en ce qu'elle permet effectivement de comparer divers schémas d'échantillonnage proposés dans la littérature, dans le cadre d'un modèle (Markovien) de file d'attente. Il est également possible d'étendre directement cette approche au cas du processus de charge de la file M/M/1 stable grâce à un résultat d'approximation du générateur.

Références

[1] S. Alouf, P. Nain, and D. Towsley. Inferring network characteristics via moment-based estimators. In *IEEE Infocom 01*, 2001.

[2] Søren Asmussen. *Applied probability and queues.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1987.

[3] J. C. Bolot. Characterizing end-to-end packet delay and loss in the Internet. *Jounal of high-speed Networks*, 2(3):305–323, 1993.

[4] J. C. Bolot. End-to-end packet delay and loss behavior in the Internet. In *SIGCOMM'93*, 1993.

[5] D. Dehay and J.-F. Yao. On likelihood estimation for a discretely observed markov jump process. submitted.

[6] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.

[7] C. B. Moler and C. F. Van Loan. Nineteen dubious ways to compute the exponential of a matrix. *SIAM Review*, 20:801–836, 1979.

[8] V. Paxson. *Measurements and analysis of End-to-end Internet Dynamics.* PhD thesis, U.C. Berkeley, 1997.