Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 76 (2006) 1514–1521
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Compensator and exponential inequalities for some suprema of counting processes

P. Reynaud-Bouret

DMA, École Normale Supérieure de Paris, 45 rue d'Ulm, 75230 Paris Cedex 05, France

Received 24 March 2004; received in revised form 20 February 2006; accepted 1 March 2006
Available online 18 April 2006

Abstract

Talagrand [1996. New concentration inequalities in product spaces. Invent. Math. 126 (3), 505–563], Ledoux [1996. On Talagrand deviation inequalities for product measures. ESAIM: Probab. Statist. 1, 63–87], Massart [2000a. About the constants in Talagrand's concentration inequalities for empirical processes. Ann. Probab. 2 (28), 863–884], Rio [2002. Une inégalité de Bennett pour les maxima de processus empiriques. Ann. Inst. H. Poincaré Probab. Statist. 38 (6), 1053–1057. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov] and Bousquet [2002. A Bennett concentration inequality and its application to suprema of empirical processes. C. R. Math. Acad. Sci. Paris 334 (6), 495–500] have obtained exponential inequalities for suprema of empirical processes. These inequalities are sharp enough to build adaptive estimation procedures Massart [2000b. Some applications of concentration inequalities. Ann. Fac. Sci. Toulouse Math. (6) 9 (2), 245–303]. The aim of this paper is to produce these kinds of inequalities when the empirical measure is replaced by a counting process. To achieve this goal, we first compute the compensator of a suprema of integrals with respect to the counting measure. We can then apply the classical inequalities which are already available for martingales Van de Geer [1995. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. Ann. Statist. 23 (5), 1779–1801].

MSC: 60E15; 60G55

Keywords: Exponential inequalities; Counting process; Supremum of centered integrals

1. Introduction

Counting processes can model a large number of biomedical situations, (see Andersen et al., 1993). In all these problems, the intensity of the process has to be estimated. If we want to use the penalized model selection method of estimation developed by Birgé and Massart (see Birgé and Massart, 2001; Massart, 2000b for instance), some very sharp exponential inequalities have to be available.

Birgé's and Massart's framework is usually the white noise model or the i.i.d. n-sample framework. There is therefore a certain structure that produces concentration inequalities. More precisely, the inequalities

E-mail address: Patricia.Reynaud-Bouret@ens.fr.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2006.03.012

developed by Talagrand (1996), Ledoux (1996), Massart (2000a), Rio (2002) and Bousquet (2002) consist of exponential inequalities for suprema of countably many empirical processes.

For counting processes, we extensively use the martingale properties. There are already a lot of exponential inequalities for martingales using the exponential semi-martingales approach (see Van de Geer, 1995; Kallenberg, 1997, Theorem 23.17 for instance). We are hence left with the precise computation of the compensator of a supremum of countably many integrals with respect to a counting process to obtain the desired exponential inequalities.

In Section 2, we explain how to build the compensator of a supremum. We can then apply classical exponential inequalities to this supremum, which is done in Section 3. In Section 4, we derive a more suitable version for the statistical applications.

2. Compensator of the supremum

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability triple and $(\mathscr{F}_t, t \geqslant 0)$ be a filtration satisfying the usual conditions (see Kallenberg, 1997, p. 124 for the definition and use of the usual augmentation of a classical filtration). Let $(N_t)_{t\geqslant 0}$ be a counting process adapted to $(\mathscr{F}_t, t \geqslant 0)$ and let $(\Lambda_t)_{t\geqslant 0}$ be its compensator, i.e. the nondecreasing predictable function such that $(M_t = N_t - \Lambda_t)_{t\geqslant 0}$ is a martingale with respect to $(\mathscr{F}_t, t\geqslant 0)$ (for precise definitions, see for instance, Brémaud, 1981). Let T be some positive fixed time, eventually infinite. We suppose in the whole paper the following assumption.

Assumption 1. The compensator $(\Lambda_t)_{t\geqslant 0}$ is absolutely continuous and almost surely finite on $[0, T]$.

Let $\{(H_{a,t})_{t\geqslant 0}, a \in \mathscr{A}\}$ be a countable family of predictable processes. We suppose them to be locally bounded in t and uniformly bounded in a. Let $(Z_t)_{t\geqslant 0}$ be the process defined by

$$\forall t \geqslant 0, \quad Z_t = \sup_{a\in\mathscr{A}} \left[\int_0^t H_{a,s}\, \mathrm{d}M_s \right]. \tag{1}$$

The process $(Z_t)_{t\geqslant 0}$ is therefore an adapted process with bounded variations. For all $t \leqslant T$, let $(T_i, 1 \leqslant i \leqslant N_t)$ be the ordered jumps of N before t, for there is almost surely a finite number of these jumps. Indeed, it is a consequence of the finiteness of the compensator (Assumption 1) and of Theorem II-8 (α) of Brémaud (1981). Since the compensator is continuous (Assumption 1), the jumps of Z only happen when N jumps. We can consequently write:

$$\forall t \leqslant T, \quad Z_t = \sum_{T_i \leqslant t} [Z_{T_i} - Z_{T_i -}] + Z_{t-} - Z_{T_{n_t}} + \sum_{T_i \leqslant t} [Z_{T_i -} - Z_{T_{i-1}}] \text{ a.e.}, \tag{2}$$

where $Z_{T_0} = Z_0 = 0$. (Z_{s-} denotes the left limit of the process Z at time s).

Reasoning by induction, it is straightforward to prove the following result, using the absolute continuity and Corollary 6.18 of Lieb and Loss (1997).

Lemma 1. *Assume $\mathscr{A} = \{1, \ldots, k\}$ to be finite and ordered. Let i be a positive integer. Let v be a real number in $]T_{i-1}; T_i[$. Then under Assumption 1, $Z_v - Z_{T_{i-1}} = -\int_{T_{i-1}}^{v} H_{\hat{a}_{s-},s}\, \mathrm{d}\Lambda_s$, where $\hat{a}_{s-}$ is the first index where Z_{s-} is attained.*

We have deliberately taken the left limit to prove the forthcoming proposition. Of course, we could have taken $H_{\hat{a}_s,s}$, for this is equal to $H_{\hat{a}_{s-},s}$ on the intervals between the jumps of N.

Using this lemma we get the following result.

Proposition 1. *Let T be a fixed positive number. Let $(Z_t)_{t\geqslant 0}$ be defined by (1). Under Assumption 1, if $\mathscr{A}$ is finite, then*

$$\forall 0 \leqslant t \leqslant T, \quad Z_t = \int_0^t \Delta Z(s)\, \mathrm{d}N_s - \int_0^t H_{\hat{a}_{s-},s}\, \mathrm{d}\Lambda_s \ \textit{a.s.},$$

where

$$\Delta Z(s) = \sup_{a\in\mathscr{A}}\left[H_{a,s} + \int_0^{s-} H_{a,u}\,\mathrm{d}M_u\right] - \sup_{a\in\mathscr{A}}\left[\int_0^{s-} H_{a,u}\,\mathrm{d}M_u\right].$$

The compensator of $(Z_{t\wedge T})_{t\geqslant 0}$ *is then defined by*

$$\forall t\geqslant 0,\quad A_t = \int_0^{t\wedge T} [\Delta Z(s) - H_{\hat{a}_{s-},s}]\,\mathrm{d}\Lambda_s.$$

If $\mathscr{A}$ *is countable, the compensator of* $(Z_{t\wedge T})_{t\geqslant 0}$, $(A_t)_{t\geqslant 0}$ *exists, is nonnegative, nondecreasing and*

$$\forall 0\leqslant t\leqslant T,\quad Z_t - A_t = \int_0^t \Delta Z(s)\,\mathrm{d}M_s.$$

Proof. Let us first assume that $\mathscr{A}$ is finite. The first integral in Z_t is exactly the first sum on the right-hand side of (2). In the second sum on the right-hand side of (2), all the differences are between two consecutive jumps and we can use the previous lemma. Moreover, $\Delta Z(s)$ introduced in the proposition is predictable. The compensator is then obvious. As $\Delta Z - H_{\hat{a}_{s-},s}$ is nonnegative and Λ nondecreasing, A is nonnegative nondecreasing.

If $\mathscr{A}$ is just countable, $\mathscr{A}$ is an increasing union of finite sets B_n. Let us denote by Z^n the supremum over B_n instead of $\mathscr{A}$. As, for all n, B_n is finite, Z^n satisfies the first part of the proposition. But, for all $t\leqslant T$, the predictable process $Z_t^n - \int_0^t \Delta Z^n(s)\,\mathrm{d}N_s$ converges almost surely to $X_t = Z_t - \int_0^t \Delta Z(s)\,\mathrm{d}N_s$. The process $(X_t)_{t\geqslant 0}$ is consequently predictable. As a result, the process $(A_t)_{t\geqslant 0}$, defined by $A_t = \int_0^t \Delta Z(s)\,\mathrm{d}\Lambda_s + X_t$, is the compensator of $(Z_{t\wedge T})_{t\geqslant 0}$ and it stays nonnegative and nondecreasing as a limit of nonnegative nondecreasing functions. □

3. Exponential inequalities for supremum

We first present some well-known facts about exponential inequalities for martingales in our particular framework.

Let $(H_t)_{t\geqslant 0}$ be a locally bounded predictable process and $(Z_t)_{t\geqslant 0}$ be defined by $Z_t = \int_0^t H_s\,\mathrm{d}M_s$ for all $t\geqslant 0$. Let ϕ be defined by $\phi(u) = e^u - u - 1$ for all u. Let

$$\forall t\geqslant 0,\quad E_t = \exp\left(\lambda Z_t - \int_0^t \phi(\lambda H_s)\,\mathrm{d}\Lambda_s\right).$$

Let T be a positive real number. Let I be an interval such that for all λ in I, $\int_0^T e^{\lambda H_s}\,\mathrm{d}\Lambda_s$ is almost surely finite. Under Assumption 1, a consequence of Theorem VI-2 of Brémaud (1981) is that $(E_{t\wedge T})_{t\geqslant 0}$ is a supermartingale and that for all stopping time τ $(\tau\leqslant T)$, $\mathbb{E}(E_\tau)$ is less than 1. This implies that for all λ in I,

$$\forall \varepsilon > 0,\quad \mathbb{P}\left(\sup_{[0,T]} Z_t \geqslant \varepsilon\right) \leqslant e^{-\lambda\varepsilon} \exp\left\| \int_0^T \phi(\lambda H_s)\,\mathrm{d}\Lambda_s \right\|_\infty.$$

The next result is a consequence of the previous inequality. It could also be seen as an application to this special framework of Van de Geer (1995) except that the absolute values are inside the integral in (3) when applying van de Geer's results.

Proposition 2. *Let* $(Z_t)_{t\geqslant 0}$ *be the process*:

$$\forall t\geqslant 0,\quad Z_t = \int_0^t H_s\,\mathrm{d}M_s,$$

where $(H_t)_{t\geqslant 0}$ *is a predictable process. Let* T *be a positive real number. If there exist* c *and* v *positive constants, such that*

$$\forall k\geqslant 2, \left|\int_0^T H_s^k\,\mathrm{d}\Lambda_s\right| \leqslant c^{k-2} v \frac{k!}{2}, \tag{3}$$

then under Assumption 1,

$$\forall u \geqslant 0, \quad \mathbb{P}\left(\sup_{[0,T]} Z_t \geqslant \sqrt{2vu} + cu\right) \leqslant \exp(-u).$$

There exists a simpler form when $(H_t)_{t\geqslant 0}$ is bounded. This form is well known; it can be found in Shorack and Wellner (1986), for instance.

Corollary 1. *With the notations of Proposition* 2, *if there exist b and v positive constants such that*, *on* $[0,T]$, $(H_t)_{t\geqslant 0}$ *has values in* $[-b;b]$ *and*

$$\left|\int_0^T H_s^2 \, \mathrm{d}\Lambda_s\right| \leqslant v$$

then under Assumption 1,

$$\forall u \geqslant 0, \quad \mathbb{P}\left(\sup_{[0,T]} Z_t \geqslant \sqrt{2vu} + \frac{b}{3}\, u\right) \leqslant \exp(-u).$$

In Section 2, we have written the centered supremum, $(Z_t - A_t)_{0\leqslant t\leqslant T}$, as an integral of a predictable process with respect to the centered counting measure. To find concentration inequalities for the supremum, it is now sufficient to apply the previous results to $H_s = \Delta Z(s)$.

Theorem 1. *Let* $(N_t)_{t\geqslant 0}$ *be a counting process satisfying Assumption* 1. *Let* $\{(H_{a,t})_{t\geqslant 0}, a \in \mathscr{A}\}$ *be a countable family of predictable processes. Let*

$$\forall t \geqslant 0, \quad Z_t = \sup_{a\in\mathscr{A}} \left[\int_0^t H_{a,s} \, \mathrm{d}M_s\right].$$

Let T be a positive real number. Let $(A_t)_{t\geqslant 0}$ *be the compensator of the process* $(Z_{t\wedge T})_{t\geqslant 0}$, *defined by Proposition* 1.

(a) *If there exist positive constants b and v such that the* H_a*'s have values in* $[-b;b]$ *on* $[0,T]$ *and such that* $\int_0^T \sup_{a\in\mathscr{A}} [H_{a,s}^2] \, \mathrm{d}\Lambda_s \leqslant v$, *then*

$$\forall u \geqslant 0, \quad \mathbb{P}\left(\sup_{[0,T]} (Z_t - A_t) \geqslant \sqrt{2vu} + \frac{1}{3}\, bu\right) \leqslant \exp(-u).$$

(b) *If there exist positive constants c and v*, *such that*

$$\forall k \geqslant 2, \quad \left(\int_0^T \sup_{a\in\mathscr{A}} |H_{a,s}|^k \, \mathrm{d}\Lambda_s\right) \leqslant c^{k-2} v \frac{k!}{2}$$

then

$$\forall u \geqslant 0, \quad \mathbb{P}\left(\sup_{[0,T]} (Z_t - A_t) \geqslant \sqrt{2vu} + cu\right) \leqslant \exp(-u).$$

Let us compare this result to the inequalities successively obtained by Talagrand (1996), Ledoux (1996), Massart (2000a), Rio (2002) and Bousquet (2002) by looking at the counting process as an empirical measure and at the compensator as an expectation. At first glance, it seems that this new inequality is in some sense stronger: we can manage random (predictable) functions and one has also a "moment" version (see (b)), which does not assume an absolute bound on the family of functions to integrate (or to sum in the i.i.d. framework). One can remark that a "moment" version of Talagrand's inequality in the i.i.d framework has recently been proved (Massart, 2005). The presence of the $\sup_{[0,T]}$ is just a refinement due to the martingale structure but this does not affect the orders of magnitude.

However, we lose something important with respect to Talagrand's inequality. Let us compare in each case what is usually called the "variance term", v. In Talagrand's inequality, v can be seen as

$$v = \sup_{a \in \mathscr{A}} \mathrm{Var}\left(\int H_a \, \mathrm{d}\mathbb{P}_n \right),$$

where the H_a's are here deterministic functions and where $\mathrm{d}\mathbb{P}_n$ is the empirical measure. The supremum is outside the integral. But in Theorem 1(a), it lies inside the integral and it is consequently of bigger order.

This phenomenon has already been underlined by Samson (2000). He recovers Talagrand's inequality for Φ-mixing, apart from this exchange between the supremum and the sum. For the Poisson processes, Wu (2000) and Houdré and Privault (2002) use martingales approach to derive exponential inequalities for very general functionals of the process. When we apply these inequalities to the supremum, this exchange also appears in the variance term. To have supremum on the left-hand side in the Poisson case (Reynaud-Bouret, 2003), we need some techniques using the infinitely divisible property of the Poisson process. Consequently, it seems that the exchange between supremum and sum (or integral) can only be done when there exists an independence property in the framework. For general counting processes, we have not been able to prove such results.

4. Statistical applications

Let us now describe the statistical framework for which these inequalities are made and let us give another corollary which is ready to be used in practice.

4.1. Statistical background

These exponential inequalities are useful to provide exponential deviations for the following χ^2-type statistics. Let T be a fixed positive real number and let $\{h_\lambda, \lambda \in m\}$ be a finite family of predictable processes. We set

$$\chi_T^2 = \sum_{\lambda \in m} \left(\int_0^T h_\lambda(t) \, \mathrm{d}M_t \right)^2. \tag{4}$$

This quantity naturally appears if we estimate the signal s by penalized model selection in the white noise framework (see Birgé and Massart, 2001). One has a model i.e. a finite dimensional linear subspace with orthonormal basis $\{\varphi_\lambda, \lambda \in m\}$ for the classical scalar product on $[0, T]$. The classical projection estimator on this subspace satisfies that the $\mathbb{L}_2$-distance between the least-square estimator and the true orthogonal projection of s, $\|s_m - \hat{s}_m\|^2$, is a χ_T^2 given by (4), with φ_λ instead of h_λ (i.e. deterministic functions) and with $\mathrm{d}W$, the white noise, instead of $\mathrm{d}M$. In this case, this quantity obeys a real χ^2-distribution. The deviations of this quantity have to be controlled to prove the adaptive properties of the model selection procedure. In the white noise framework, they use the exponential inequalities available for χ^2-distributions.

If we estimate the density s from a n-sample by penalized model selection, we can still consider the same model as before. As previously, the least-square estimator, $\hat{s}_m$, is an unbias estimation of the classical projection, s_m. The distance $\|s_m - \hat{s}_m\|^2$ is also a χ^2-type statistics where $h_\lambda = \varphi_\lambda$, i.e. an orthonormal deterministic basis of the model for the classical scalar product. In this case, $\mathrm{d}M$ is replaced by the centered empirical measure. In this context, Birgé and Massart use Talagrand's inequality to provide control on the χ^2-type statistics (Birgé and Massart, 1997).

If we estimate the intensity s of a Poisson process N by penalized model selection, we can keep the same procedure and notations. The distance $\|s_m - \hat{s}_m\|^2$ is still a χ^2-type statistics where $\mathrm{d}M$ is the centered Poisson process, and where $\{h_\lambda = \varphi_\lambda, \lambda \in m\}$ represents an orthonormal deterministic family of $\mathbb{L}_2([0, T], \mathrm{d}t)$. In this case, we can use the concentration inequality of Reynaud-Bouret (2003) (see Proposition 3) to control these distances, which gives the same order of magnitude as Talagrand's inequality in the n-sample framework.

Generalizing the Poisson process, we can be interested by the Aalen multiplicative intensity model where the compensator of $(N_t)_{t \geqslant 0}$ satisfies $\mathrm{d}\Lambda_t = Y_t s(t) \, \mathrm{d}t$, with a predictable and known process $(Y_t)_{t \geqslant 0}$. For instance the right-censoring model (see Andersen et al., 1993, for a complete description) has an Aalen multiplicative intensity. We can estimate the deterministic function s using the observations of the processes N and Y by penalized model selection (see Reynaud-Bouret, 2002). In this case, we are using a random scalar product $\int_0^t f(s)g(s)Y_s\,\mathrm{d}s$ instead of the classical one. In this context, we cannot exactly use the same model as before but we can use predictable h_λ's. Indeed, if $\{\varphi_\lambda, \lambda \in m\}$ is an orthonormal family of $\mathbb{L}_2([0,1], \mathrm{d}t)$ (typically histograms or Fourier basis), $\{h_\lambda = \varphi_\lambda/\sqrt{Y}, \lambda \in m\}$ becomes an orthonormal predictable family for the random product (when Y_t is positive). The subspace generated by the h_λ's is the used model.

Our aim is now to provide for χ_T an exponential inequality which is ready to be used for the statistical applications.

4.2. An inequality which is ready for immediate application

In order to provide a concentration inequality for χ_T^2, we can remark that

$$\forall t \geqslant 0, \quad \chi_t = \sup_{\sum_{\lambda\in m} a_\lambda^2 = 1} \int_0^t \left(\sum_{\lambda \in m} a_\lambda h_\lambda(s) \right) \mathrm{d}M_s \tag{5}$$

is the square root of χ_t^2. We can consequently use Theorem 1 on a countable dense subset of the unit ball of $\mathbb{R}^m$. But as we do not know in practice the compensator of χ_t, we may prefer comparing it to $\sqrt{C_t}$ where

$$\forall t \geqslant 0, \quad C_t = \sum_{\lambda \in m} \int_0^t h_\lambda(s)^2\,\mathrm{d}\Lambda_s, \tag{6}$$

is the compensator of χ_t^2. This leads to the forthcoming result.

Corollary 2. *Let T be a fixed positive real number. Let χ_T be defined by* (4). *Then, for any positive number u, with probability larger than $1 - 2e^{-u}$,*

$$\chi_T - \sqrt{C_T} \leqslant 3\sqrt{2vu} + bu,$$

where

- C_T *is defined by* (6),
- $v = \|C_T\|_\infty$, *and*
- $\forall s \leqslant T, \sum_{\lambda\in m} h_\lambda^2(s) \leqslant b^2$.

Proof. Let $u \geqslant 0$. First, we can interpret χ_t as a supremum (see (5)). Moreover, let B be a countable dense subset of the unit ball of $\mathbb{R}^m$. We can say that

$$\chi_t = \sup_{a \in B} \int_0^t \left(\sum_{\lambda \in m} a_\lambda h_\lambda(s) \right) \mathrm{d}M_s.$$

We can therefore apply Proposition 1(a) with $H_a = \sum_{\lambda\in m} a_\lambda h_\lambda$. We obtain that $(\chi_{t\wedge T})_{t\geqslant 0}$ has a compensator $(A_t)_{t\geqslant 0}$ and that

$$\mathbb{P}\left(\sup_{[0,T]} (\chi_t - A_t) \geqslant \sqrt{2vu} + \frac{b}{3}u \right) \leqslant e^{-u}.$$

We can replace the H_a by $-H_a$ to obtain that

$$\mathbb{P}\left(\sup_{[0,T]} |\chi_t - A_t| \geqslant \sqrt{2vu} + \frac{b}{3}u \right) \leqslant 2e^{-u}.$$

Let $B_T = \sup_{[0,T]}|\chi_t - A_t|$. Now we must compare $(A_t)_{t\geqslant 0}$ and $(C_t)_{t\geqslant 0}$. One has for all $t\leqslant T$:

$$\begin{aligned}\chi_t^2 - A_t^2 &= (\chi_t - A_t)^2 + 2A_t(\chi_t - A_t)\\ &= (\chi_t - A_t)^2 + 2\int_0^t (\chi_{s-} - A_{s-})\,\mathrm{d}A_s + 2\int_0^t A_s\,\mathrm{d}(\chi_s - A_s).\end{aligned}$$

But the compensator of the first term is $\int_0^t (\Delta\chi)^2(s)\,\mathrm{d}\Lambda_s$ and the last term is a martingale. Moreover, $(A_t)_{t\geqslant 0}$ is predictable. We can therefore take the compensator of the previous expression to obtain:

$$C_t - A_t^2 = \int_0^t (\Delta\chi)^2(s)\,\mathrm{d}\Lambda_s + 2\int_0^t (\chi_{s-} - A_{s-})\,\mathrm{d}A_s.$$

Consequently, we have:

$$\begin{aligned}\chi_t - \sqrt{C_t} &= \chi_t - A_t + A_t - \sqrt{C_t}\\ &= \chi_t - A_t - \frac{C_t - A_t^2}{A_t + \sqrt{C_t}}\\ &= \chi_t - A_t - \frac{\int_0^t (\Delta\chi)^2(s)\,\mathrm{d}\Lambda_s + 2\int_0^t (\chi_{s-} - A_{s-})\,\mathrm{d}A_s}{A_t + \sqrt{C_t}}.\end{aligned}$$

As $(A_t)_{t\geqslant 0}$ is nonnegative and nondecreasing, we obtain that $\chi_t - \sqrt{C_t} \leqslant 3B_T$, for all $t\leqslant T$, which implies the result. □

4.3. Orders of magnitude

Let us now emphasize the difference between Talagrand's inequality and ours. To do so, let us rewrite Corollary 2 of Reynaud-Bouret (2003), which is the complete equivalent of Talagrand's inequality in the Poisson framework, and which is consequently closer to the general framework of counting processes:

Proposition 3. *Let N be a Poisson process on $(\mathbb{X},\mathscr{X})$ with finite mean measure v. Let $\{\psi_a, a\in A\}$ be a countable family of functions with values in $[-b,b]$. One considers*

$$Z = \sup_{a\in A}\left|\int_{\mathbb{X}} \psi_a(x)(\mathrm{d}N_x - \mathrm{d}v_x)\right| \quad \textit{and} \quad v_0 = \sup_{a\in A}\int_{\mathbb{X}} \psi_a^2(x)\,\mathrm{d}v_x.$$

Then

$$\forall\varepsilon>0,\quad \forall u\geqslant 0,\quad \mathbb{P}\Big(Z\geqslant (1+\varepsilon)\mathbb{E}(Z) + \sqrt{2\kappa v_0 u} + \kappa(\varepsilon)bu\Big)\leqslant \exp(-u),$$

where $\kappa = 6$ and $\kappa(\varepsilon) = 1.25 + 32/\varepsilon$.

This inequality has exactly the same orders of magnitude as Talagrand's inequality in the i.i.d. framework: the supremum lays outside the integral in the variance term v_0.

Let us compare this result to Proposition 2 applied to Poisson processes. Actually one can see Poisson processes as a special case of counting processes with Aalen multiplicative intensity. Let us look at the case where s is a constant equal to 1, Y is a constant equal to n and T equals 1 (i.e. $\mathrm{d}v = n\,\mathrm{d}s$ is the mean measure of the Poisson process). This is the case for the aggregated process built from the sum of n independent identically distributed homogeneous Poisson processes on $[0,1]$ with intensity 1.

Suppose the model (see Section 4.1) is the set of histograms constructed on a regular partition m of $[0,1]$. The basis is therefore deterministic and of the form $\sqrt{(D/n)}\mathbb{1}_I$ where D is the number of intervals in m.

If we apply Corollary 2 of Reynaud-Bouret (2003) on $[0,1]$, to

$$\chi = \sup_{\sum_{I\in m} a_I^2 = 1}\int_0^1 \sum_{I\in m} a_I\sqrt{\frac{D}{n}}\mathbb{1}_I(\mathrm{d}N_s - n\,\mathrm{d}s) = \sqrt{\sum_{I\in m}\frac{D}{n}\Big(N_I - \frac{n}{D}\Big)^2},$$

where N_I is the number of jumps of the process in the interval I, we obtain that

$$\forall \varepsilon > 0, \quad \forall u \geqslant 0, \quad \mathbb{P}\left(\chi \geqslant (1+\varepsilon)\sqrt{D} + \sqrt{2\kappa u} + \kappa(\varepsilon)\sqrt{\frac{D}{n}u}\right) \leqslant \exp -u. \tag{7}$$

But, if we apply Proposition 2, we obtain that

$$\forall u \geqslant 0, \quad \mathbb{P}\left(\chi \geqslant \sqrt{D} + 3\sqrt{2Du} + \sqrt{\frac{D}{n}u}\right) \leqslant 2\exp -u. \tag{8}$$

The variance term (the factor of $\sqrt{u}$) in (8) is bigger than the corresponding term in (7). It has the same order as the expectation ($\sqrt{D}$). The parameter D is indeed potentially as large as n for proper models, n tending to infinity. Consequently, (7) gives an order of magnitude of $\sqrt{D}$ when (8) gives an order of magnitude of $\sqrt{Du}$. In this sense, inequality (7) is better than (8).

But for more general Aalen multiplicative intensity processes, Y is no longer constant: it often decreases and it can become very small. In the right-censoring model, Y_t is equal to 1 when t tends to 1, but Y_0 equals n, the number of observations when $t = 0$. In this case, the third linear term becomes of order $u\sqrt{D}$. Hence we would not change the order of magnitude given by this inequality for processes with Aalen multiplicative intensity, even if we were able to improve the behavior of the quadratic term in the general case.

5. Conclusion

We have proved an inequality that generalizes Talagrand's inequality in the n-sample framework, apart from the increase of the quadratic term. This is especially useful to prove the adaptive properties of the penalized model selection procedure when we are dealing with counting processes which are more intricate than Poisson processes, like the processes with Aalen multiplicative intensity.

References

Andersen, P., Borgan, O., Gill, R., Keiding, N., 1993. Statistical Models Based on Counting Processes. Springer Series in Statistics, Springer, Berlin.

Birgé, L., Massart, P., 1997. From model selection to adaptive estimation. In: Festschrift for Lucien Le Cam. Springer, New York, pp. 55–87.

Birgé, L., Massart, P., 2001. Gaussian model selection. J. Eur. Math. Soc. (JEMS) 3 (3), 203–268.

Bousquet, O., 2002. A Bennett concentration inequality and its application to suprema of empirical processes. C. R. Math. Acad. Sci. Paris 334 (6), 495–500.

Brémaud, P., 1981. Point Processes and Queues. Springer, Berlin.

Houdré, C., Privault, N., 2002. Concentration and deviation inequalities in infinite dimensions via covariance representations. Bernoulli 8 (6), 697–720.

Kallenberg, O., 1997. Foundations of Modern Probability. Springer, Berlin.

Ledoux, M., 1996. On Talagrand deviation inequalities for product measures. ESAIM: Probab. Statist. 1, 63–87.

Lieb, E.H., Loss, M., 1997. Analysis. In: Graduate Studies in Mathematics, vol. 14. American Mathematical Society, Providence, RI.

Massart, P., 2000a. About the constants in Talagrand's concentration inequalities for empirical processes. Ann. Probab. 2 (28), 863–884.

Massart, P., 2000b. Some applications of concentration inequalities. Ann. Fac. Sci. Toulouse Math. (6) 9 (2), 245–303 Probability theory.

Massart, P., 2005. Lectures on Probability Theory and Statistics, Lecture Notes in Mathematics. Springer, Berlin, http://www.math.u-psud.fr/~massart/. In: J. Picard (Ed.), Lectures from the 33rd Probability Summer School held in Saint-Flour, July 7–24, 2003, to appear.

Reynaud-Bouret, P., 2002. Estimation adaptative de l'intensité de certains processus ponctuels par sélection de modèle. Ph.D. thesis, Université Paris XI, UFR Orsay.

Reynaud-Bouret, P., 2003. Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. P.T.R.F. 126 (1), 103–153.

Rio, E., 2002. Une inégalité de Bennett pour les maxima de processus empiriques. Ann. Inst. H. Poincaré Probab. Statist. 38 (6), 1053–1057 En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

Samson, P.-M., 2000. Concentration of measure inequalities for Markov chains and ϕ-mixing processes. Ann. Probab. 28 (1), 416–461.

Shorack, G., Wellner, J., 1986. Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics.

Talagrand, M., 1996. New concentration inequalities in product spaces. Invent. Math. 126 (3), 505–563.

Van de Geer, S., 1995. Exponential inequalities for martingales, with application to maximum likelihood estimation for counting processes. Ann. Statist. 23 (5), 1779–1801.

Wu, L., 2000. A new modified logarithmic Sobolev inequality for Poisson point process and several applications. P.T.R.F. 118 (3), 427–438.