Optimal Bounds on Tail Probabilities - A Simplified Approach

Aviad Cohen[1] Yuri Rabinovich[2] Assaf Schuster[3] and Hadas Shachnai[4]

[1] Israel Software Lab, Intel, Haifa, Israel. ‡
[2] Dept. of Mathematic and Computer Science, Ben-Gurion University of the Negev Beer-Sheva, Israel 84105. §
[3] Dept. of Computer Science, Technion IIT, Haifa, Israel 32000. ¶
[4] Dept. of Computer Science, Technion IIT, Haifa, Israel 32000. ‖

Abstract. Let $\{X_i\}_{i=1}^{\infty}$ be independent random variables, assuming values in $[0,1]$, having a common mean μ, and variances bounded by σ^2. Let $S_n = \sum_{i=1}^{n} X_i$. We give a general and simple method for obtaining asymptotically optimal upper bounds on probabilities of events of the form $\{S_n - E[S_n] \geq na\}$ with explicit dependence on μ and σ^2. For general bounded random variables the method yields the Bennett inequality, with a simplified proof. For specific classes of distributions the method can be used to derive bounds that are tighter than those achieved by the Bennett inequality. We demonstrate the power of the method by applying it to the case of symmetric three-point distributions, thus improving previous results for the List Update Problem.

1 Introduction

Let $\{X_i\}_{i=1}^{\infty}$ be independent random variables assuming values in a bounded interval $[m, M]$ and let $S_n = \sum_{i=1}^{n} X_i$. Suppose that X_i has mean μ and variance bounded by σ^2 and let $0 < a < M - \mu$. In computer science, discrete mathematics and statistics, the need often arises to give upper bounds on tail probabilities i.e. probabilities of the form: [9]

$$\Pr\left[S_n - E[S_n] \geq na\right] \quad .$$

In particular it is often desired that the upper bounds decrease exponentially with n (for a fixed a).

‡ e-mail: aviadc@iil.intel.com
§ e-mail: uri@cs.bgu.ac.il
¶ e-mail: assaf@cs.technion.ac.il
‖ e-mail: hadas@cs.technion.ac.il

[9] Probabilities of the form $\Pr[S_n - E[S_n] \leq -na]$ can be handled by defining $Y_i = C - X_i$ (C a constant), $S'_n = \sum_{i=1}^{n} Y_i$ and using

$$\Pr\left[S_n - E[S_n] \leq -na\right] = \Pr\left[S'_n - E[S'_n] \geq na\right]$$

The best known inequalities yielding exponentially decreasing upper bounds, are the Chernoff, Hoeffding, and Bennett inequalities. The well known Hoeffding inequality [9] asserts that if X_i assumes values in the interval $[0,1]$, then:

$$\Pr\left[S_n - E[S_n] \geq na\right] \leq \left\{ \left(\frac{\alpha}{\beta}\right)^{\beta} \left(\frac{1-\alpha}{1-\beta}\right)^{1-\beta} \right\}^n \tag{1}$$

where

$$\alpha = \mu \quad ; \quad \beta = \mu + a$$

The Hoeffding inequality generalizes the Chernoff inequality [4, 8], which makes the same statement about $\{0,1\}$ random variables. Thus in a certain sense, sums of random variables supported on $[0,1]$, are at least as concentrated around their mean, as sums of $\{0,1\}$ random variables with the same mean.

The Hoeffding inequality in the above form holds for any sequence of random variables with mean μ, no matter what their variance. However, it seems reasonable that sums of random variables having a variance smaller than the maximal value (being $\mu - \mu^2$) should be better concentrated around their mean. Thus it is desirable to give upper bounds that incorporate the variance and enable us to give estimates sharper than Hoeffding's. Bennett [1] solved this by showing that (1) holds with α and β depending on the variance, as follows:[10]

$$\alpha = \frac{\sigma^2}{\sigma^2 + (1-\mu)^2} \quad ; \quad \beta = \frac{\sigma^2 + a(1-\mu)}{\sigma^2 + (1-\mu)^2} \tag{2}$$

Indeed, when the variance assumes its maximal value ($\sigma^2 = \mu - \mu^2$), the Bennett bound as given in (2) reduces to the Hoeffding inequality. In all other cases the Bennett bound is strictly stronger. In particular, when the variance is zero, the Bennett bound asserts that tail probabilities equal to zero, as should be.

Although the Bennett bound has numerous applications, it is rarely used in the literature. A major reason is the long presentation of the bound and its tedious proof [3, 1]. Furthermore, while its generality ensures that the bound holds for any distribution, it is often the case that better bounds exist for specific distribution classes. In this paper we present a simple technique for proving Bennett-style upper bounds for any distribution class. Our optimization is relative to the specific parameters of the distributions, hence the resulting bounds are stronger than those achieved by using the general Bennett bound.

We first use the technique to give a new and simpler derivation of the Bennett inequality. (We comment that in some asymptotic sense the bound cannot be improved. We elaborate on this in [5]). Then we show how to use the new technique to achieve better bounds for specific distributions.

The rest of this paper is organized as follows. In Section 2 we give the outline of the general technique. In Section 3 we provide a new proof for the Bennett

[10] In fact the Bennett bound holds also for variables that are unbounded from one side [1]. Here we restrict the discussion to bounded variables.

bound using our simple approach. Section 4 presents an exponential version of the Bennett bound, which enables to determine its rate of convergence. In Section 5 we show how to use the method for achieving earlier stopping point to the dynamic reorganization of a linear list (known as the List Update Problem [2, 10, 12]). Due to lack of space we omit some of the proofs. These proofs can be found in [5].

2 Bounding Tail Probabilities with the Laplace Transform.

Many known derivations of exponentially decreasing upper bounds on tail probabilities share a common strategy. Let Ψ be a class of probability distributions and suppose we wish to give upper bounds on tail probabilities, that hold for **any** sequence of independent random variables distributed according to members of Ψ. The strategy consists of the following steps.

(1) Let χ be the indicator function of the event $(S_n - E[S_n]) \geq na$. Observe that for any $t > 0$:

$$\chi \leq e^{(S_n - E[S_n] - na)t} \tag{3}$$

Therefore for any $t > 0$:

$$\begin{aligned}\Pr[S_n - E[S_n] \geq na] &= E[\chi] \leq E[e^{(S_n - E[S_n] - na)t}] \\ &= e^{-nt(\mu+a)} \prod_{i=1}^{n} E[e^{X_i t}] = \left(e^{-t(\mu+a)}\phi_n(t)\right)^n\end{aligned} \tag{4}$$

where

$$\phi_n(t) = \left(\prod_{i=1}^{n} E[e^{X_i t}]\right)^{\frac{1}{n}}$$

$(t > 0)$ is the geometric mean of the one-sided Laplace transforms of the random variables $\{X_i\}_{i=1}^n$.

(2) Define $Z(t)$ by:

$$Z(t) = \sup_{Y \in \Psi} E[e^{Yt}]$$

Clearly $\phi_n(t) \leq Z(t)$ for every positive n. Therefore (4) yields for any $t > 0$:

$$\Pr[S_n - E[S_n] \geq na] \leq \left(e^{-t(\mu+a)} Z(t)\right)^n \tag{5}$$

Inequality (5) is the basic inequality of the strategy. Note that it continues to hold if $Z(t)$ is replaced by any function which majorizes it.

(3) The next step is to attempt to express $Z(t)$ as an explicit function of t and the parameters defining the class Ψ, and to plug this function in inequality (5). For example, in our case, we wish to find upper bounds for the class of random variables having a mean μ and a variance bounded by σ^2, and therefore we seek to express $Z(t)$ as an explicit function of μ, σ^2 and t. If we cannot find such an explicit form for $Z(t)$ we can instead substitute any other function which majorizes $Z(t)$ in the right side of (5).

(4) The final step is to minimize the right hand side of the inequality obtained in step (3) with respect to t.

Step (3) is the crucial step of the strategy. Here we need to express $Z(t)$ explicitly as a function of t and the parameters defining the class Ψ (or in the event that we fail to do so, to find a sufficiently good upper bound for it, in terms of the relevant parameters). There is one case in which this can be achieved. Suppose one can find a random variable $X \in \Psi$ whose Laplace transform $E[e^{Xt}]$ attains a maximal value (among the members of Ψ) for every $t > 0$. In this case $Z(t) = E[e^{Xt}]$, and if we can express $E[e^{Xt}]$ explicitly as a function of the relevant parameters we are done. Note that in such a case we get a bound which is optimal as far as the above strategy is concerned.

In fact the Hoeffding bound is obtained in this way, by finding a p.d. that maximizes the Laplace transform among all p.d. concentrated on $[0,1]$ and having mean μ, and expressing its Laplace transform as a function of μ and t. This is the random variable that assumes the value 0 with probability $1-\mu$, and the value 1 with probability μ. Using Cramer's Theorem, it can be shown that the above strategy is asymptotically optimal (see in [5]).

3 The Bennett Bound: A New Proof

In this section we give a bound on tail probabilities, which is optimal as far as the strategy of section 2 goes, namely: *the Bennett bound.* To this end we proceed as follows:

1. We find a probability distribution that maximizes the Laplace transform among all the probability distributions that are supported on $[0,1]$ and have a first moment μ and a second moment $\leq \nu$.
2. We plug the maximal Laplace transform in the rightmost part of (4) and optimize the resulting expression with respect to t.

3.1 Maximizing the Laplace transform

Let $\Psi(\mu,\nu)$ denote the collection of all probability distributions on $[0,1]$ with first moment μ and second moment $\leq \nu$. $\Psi(\mu,\nu)$ contains a unique member $\rho(\mu,\nu)$ satisfying the conditions that its second moment is precisely ν and that it is concentrated on two points, one of which is 1. If $\rho(\mu,\nu)$ assumes the value λ with probability p and the value 1 with probability q, then p,q,λ are determined by

$$p+q=1 \quad ; \quad p\lambda+q=\mu \quad ; \quad p\lambda^2+q=\nu$$

The solution of these equations can be expressed in the form:

$$\lambda = \frac{\mu-\nu}{1-\mu} \quad ; \quad p = \frac{1-\mu}{1-\lambda}\,; \quad q = \frac{\mu-\lambda}{1-\lambda} \tag{6}$$

Lemma 1. *$\rho(\mu,\nu)$ has maximal moment (of any order) among the members of $\Psi(\mu,\nu)$.*

The proof is given in [5]. $\blacksquare$

Corollary 2. *For any $t \geq 0$, $\rho(\mu, \nu)$ maximizes $E[e^{tX}]$ among $X \in \Psi(\mu, \nu)$.*

Proof: follows from the expansion $E[e^{tX}] = \sum_{i=0}^{\infty} \frac{t^i E[x^i]}{i!} = \sum_{i=0}^{\infty} \frac{t^i m_i}{i!}$, and the fact that each of the m_i is maximized by $\rho(\mu, \nu)$. $\blacksquare$

3.2 Obtaining the Inequality

Theorem 3. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent random variables assuming values in $[0,1]$. Suppose for each X_i: $E[X_i] = \mu$, $E[X_i^2] \leq \nu$, and let $a > 0$, then*

$$\Pr[S_n - E[S_n] \geq na] \leq \left\{ \left(\frac{\alpha}{\beta}\right)^{\beta} \left(\frac{1-\alpha}{1-\beta}\right)^{1-\beta} \right\}^n \tag{7}$$

where

$$\alpha = \frac{\sigma^2}{\sigma^2 + (1-\mu)^2} \quad ; \quad \beta = \frac{\sigma^2 + a(1-\mu)}{\sigma^2 + (1-\mu)^2}$$

Proof: Set $\delta = a + \mu$. From (4) one has for any $t > 0$:

$$\Pr[S_n - n\mu \geq na] \leq \left(e^{-t\delta} E[e^{tX}]\right)^n$$

By Corollary 2, $\rho(\mu, \nu)$ maximizes $E[e^{tX}]$. Therefore:

$$\Pr[S_n - n\mu \geq na] \leq \left(e^{-t\delta} \cdot (pe^{t\lambda} + qe^t)\right)^n = \left(e^{t(1-\delta)} \cdot (pe^{t(\lambda-1)} + q)\right)^n$$

where p, q, λ are as in (6).
Differentiating $e^{t(1-\delta)} \cdot (pe^{t(\lambda-1)} + q)$ with respect to t we find that this expression is minimized for $t = \tau$ that satisfies:

$$e^{\tau(\lambda-1)} = \frac{q}{p} \cdot \frac{1-\delta}{\delta-\lambda}$$

Substituting $t = \tau$ and the solution from (6) we obtain:

$$\begin{aligned}
\Pr[S_n - n\mu \geq na] &\leq \left\{ \left(\frac{q}{p} \cdot \frac{1-\delta}{\delta-\lambda}\right)^{\frac{1-\delta}{\lambda-1}} \frac{q(1-\lambda)}{\delta-\lambda} \right\}^n \\
&= \left\{ \left(\frac{\mu-\lambda}{\delta-\lambda}\right)^{1-\frac{1-\delta}{1-\lambda}} \left(\frac{1-\mu}{1-\delta}\right)^{\frac{1-\delta}{1-\lambda}} \right\}^n \\
&= \left\{ \left(\frac{\alpha}{\beta}\right)^{\beta} \left(\frac{1-\alpha}{1-\beta}\right)^{1-\beta} \right\}^n
\end{aligned}$$

with

$$\alpha = \frac{\mu-\lambda}{1-\lambda} = \frac{\sigma^2}{\sigma^2 + (1-\mu)^2} \quad ; \quad \beta = \frac{\delta-\lambda}{1-\lambda} = \frac{\sigma^2 + a(1-\mu)}{\sigma^2 + (1-\mu)^2}$$

$\blacksquare$

4 An Alternative Formulation of the Upper Bound

The upper bound given by theorem 3 contains the expression $F(\alpha, \beta)$ defined by:

$$F(\alpha, \beta) = \left(\frac{\alpha}{\beta}\right)^{\beta} \left(\frac{1-\alpha}{1-\beta}\right)^{1-\beta}$$

This expression is estimated in [9]. We will give here a different estimation of this quantity, which may prove useful under certain circumstances.

Consider the Entropy function $H(x)$ and its three first derivatives:

$$H(x) = -x \log x - (1-x) \log(1-x)$$

$$H'(x) = \log \frac{1-x}{x} \quad ; \quad H''(x) = -\frac{1}{x(1-x)} \quad ; \quad H'''(x) = \frac{1}{x^2} - \frac{1}{(1-x)^2}$$

Taking the logarithm of $F(\alpha, \beta)$ and putting $\Delta = \beta - \alpha$ we get

$$\begin{aligned} \log F(\alpha, \beta) &= -\beta \log \beta - (1-\beta) \log(1-\beta) + \beta \log \alpha + (1-\beta) \log(1-\alpha) \\ &= H(\beta) + (\alpha + \Delta) \log \alpha + ((1-\alpha) - \Delta) \log(1-\alpha) \\ &= H(\beta) - H(\alpha) - \Delta H'(\alpha) \end{aligned}$$

Thus one obtains from theorem 3 the following upper bound:

$$\Pr[S_n - E[S_n] \geq na] \leq e^{\left(H(\beta) - H(\alpha) - \Delta H'(\alpha)\right)n} \tag{8}$$

The Taylor expansion of $H(\beta)$ at α gives:

$$H(\beta) = H(\alpha) + \Delta H'(\alpha) + \frac{\Delta^2}{2} H''(\alpha) + R_2$$

Where R_2 is the remainder term of order 2 in the Taylor expansion of $H(\beta)$ at α. Upon substituting the definitions of α, β in terms of a, μ, σ^2 one obtains

$$\frac{\Delta^2}{2} H''(\alpha) = \frac{a^2}{2\sigma^2}$$

Thus we have arrived at an equivalent version of theorem 3:

Theorem 4. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of independent random variables assuming values in $[0, 1]$. Suppose for each X_i: $E[X_i] = \mu$, $E[X_i^2] \leq \nu$, and let $a > 0$, then*

$$\Pr[S_n - E[S_n] \geq na] \leq \exp\left(-\frac{na^2}{2\sigma^2} + R_2\right) \tag{9}$$

As an example where theorem 4 may be useful consider the case in which $\alpha \geq \frac{1}{2}$ or equivalently $\sigma^2 \geq (1-\mu)^2$. Since

$$R_2 = \frac{1}{2}\int_\alpha^\beta (\beta - t)^2 H'''(t)dt = \frac{1}{2}\int_\alpha^\beta (\beta - t)^2 \left(\frac{1}{t^2} - \frac{1}{(1-t)^2}\right) dt$$

it is easily seen that $R_2 \leq 0$ and therefore:

$$\Pr[S_n - E[S_n] \geq na] \leq \exp\left(-\frac{na^2}{2\sigma^2}\right) \tag{10}$$

The right side of inequality (10) is reminiscent of a normal distribution. In fact, suppose the random variable $\frac{S_n - E[S_n]}{\sqrt{n}}$ has a normal distribution with mean 0 and variance σ^2. Then, for a sufficiently large n we have

$$\Pr[S_n - E[S_n] \geq na] = \int_{\sqrt{n}a/\sigma}^{\infty} e^{-\frac{x^2}{2}} dx \leq \int_{\sqrt{n}a/\sigma}^{\infty} x e^{-\frac{x^2}{2}} dx = \exp\left(-\frac{na^2}{2\sigma^2}\right)$$

By the Central Limit theorem, the sequence of random variables $Y_n = \frac{S_n - E[S_n]}{\sqrt{n}}$ converges weakly to a random variable with a normal distribution of mean 0 and variance σ^2. However, this does not imply that one can assume a normal distribution for the random variable $\frac{S_n - E[S_n]}{\sqrt{n}}$ when computing an upper bound for $\Pr[S_n - E[S_n] \geq na]$. Thus, one can view Theorem 4 and inequality (10) as setting mathematical conditions under which one can obtain the same result as one would get by assuming that the random variable $\frac{S_n - E[S_n]}{\sqrt{n}}$ "has already converged to a normal distribution".

5 Application to the List Update Problem

Using the general strategy introduced in Section 2 one can obtain tighter bounds for specific classes of distributions. As noted in Section 2, when the strategy can be used to obtain a closed form of the bound for a specific class of distributions, the bound is asymptotically optimal with respect to this class. We demonstrate this below by applying the strategy to symmetric three-point distributions. As a result we obtain a stopping point to the dynamic reordering of a data structure, which improves previous results.

The List Update Problem has been widely studied (see in [12, 10]): A set of n items held as a linear list is accessed randomly. Each request involves a search for specific item (identified uniquely by its key). The probability of accessing the ith element, denoted by R_i, is $p_i \ \ \forall \ \ 1 \leq i \leq n$, $\sum_i p_i = 1$. The p_i's are fixed but initially *unknown*, therefore the list is dynamically reorganized along a reference sequence, so as to improve the relative ordering of the items. The sequential search implemented in each request, starting at the header, implies that the optimal static arrangement of the items is by decreasing order of the access probabilities.

The Counter Scheme (CS), which maintains a reference count for each element, and keeps the list by decreasing order of the counters, is known to converge to the optimal ordering [10]. However, the strong law of large numbers, which guarantees this asymptotic optimality, does not provide a measure to the rate of convergence.

Hofri and Shachnai present in [10] a stopping point for the reorganization process, in case the distribution of the accesses is known up to the mapping of the probabilities to keys. Therefore, while the access the probability vector is known to be $(p_1, \ldots, p_n)$, we cannot tell the optimal order of the items. As shown in [10], a direct computation of the average access cost to the list after m accesses requires $O(m^2n^2)$ steps, while using a tight bound on the cost reduces this complexity to $O(n^2)$.

In Theorem 5 we improve the bound in [10] (which is based on Chebyshev's bound). We first introduce a simpler formulation of the cost function. Denote by $C_m(CS|\bar{p}), C(OPT \mid \bar{p})$ the average access cost to the list after m references and the minimal expected access cost respectively, then by [10]

$$C_m(CS|\bar{p}) = C(OPT \mid \bar{p}) + \sum_{1 \le i < j \le n} (p_i - p_j) \cdot \mathrm{Prob}_{CS}(\sigma_m(j) < \sigma_m(i)) \ , \quad (11)$$

where $\mathrm{Prob}_{CS}(\sigma_m(j) < \sigma_m(i))$ is the probability that R_j precedes R_i, after m references to the list, when the reorganization rule is CS. Without loss of generality we assume a renumbering of the items, such that $p_1 \ge p_2, \ldots \ge p_n$.

Let $X_i^{(m)}, X_j^{(m)}$ denote the number of accesses to R_i, R_j respectively, within the first m references to the list, and

$$S_{ji}^{(m)} = X_j^{(m)} - X_i^{(m)} \ , \quad (12)$$

then

$$S_{ji}^{(m)} = \sum_{k=1}^{m} Y_k \ , \quad (13)$$

such that

$$Y_k = \begin{cases} -1 & p_i \\ 0 & 1 - p_i - p_j \\ 1 & p_j \end{cases} \quad (14)$$

and the Y_k's are mutually independent.

Note, that

$$\mathrm{Prob}_{CS}(\sigma_m(j) < \sigma_m(i)) = \mathrm{Prob}(S_{ji}^{(m)} \ge 0) \ . \quad (15)$$

We can bound the right-hand side of (15) by substituting into (5) the Laplace transform of Y_k, $1 \le k \le m$. Thus, we have

n\ ϵ	0.0001	0.001	0.01	0.05	0.1	0.15	0.2
10	22.06	5.77	3.25	2.77	2.41	2.36	2.24
20	11.66	4.63	3.61	2.96	2.6	2.47	2.38
25	9.85	4.48	3.78	3.00	2.66	2.51	2.46
50	6.78	4.45	4.13	3.04	2.72	2.52	2.52
100	5.56	4.78	4.14	3.10	2.73	2.71	2.58

Table 1. The required length for the reorganization process under CS to approach the optimum within $1+\epsilon$: The ratio between Chebyshev's bound and the stopping point in Theorem 5.

Theorem 5. (A Stopping Point for the CS) *For a list of n items with the probability vector $(p_1, \ldots, p_n)$, and any $0 < \epsilon < 1$, the CS approaches within a factor of $(1+\epsilon)$ to the minimal expected cost after $m*$ references, where*

$$m* = \min_m \left\{ \sum_{(i,j)\in S} (p_i - p_j)(1 - (\sqrt{p_i} - \sqrt{p_j})^2)^m \leq \sum_{(i,j)\in S} \epsilon(p_j + 2/n(n-1)) \right\} , \tag{16}$$

and

$$S = \{(i,j) \ : \ 1 \leq i < j \leq n \ ; \quad p_i - p_j > 2 \cdot \epsilon(p_j + 2/n(n-1))\} \quad . \tag{17}$$

The proof is given in [5]. ▮

Tables 1 and 2 quantify the improvement on the previous bound for the Zipf distribution, where $p_i = 1/iH_n$. The Zipf distribution often governs the reference process, especially when keys are drawn from a text file (as in Lisp [2]). In Tables 1 and 2 we present the ratio between the stopping points achieved by the Chebyshev and Hoeffding bounds and $m*$ as computed in (16), respectively. As shown in Table 2, for any ϵ, the bound presented in Theorem 5 improves the value obtained by the Hoeffding bound by at least a factor of ϵ. In comparison with Chebyshev (Table 1), which also uses the variance for computing the bound, the improvement increases as ϵ becomes smaller.

Acknowledgment: We would like to thank Micha Hofri, Kurt Mehlhorn and Ofer Zeitouni for helpful comments and suggestions.

References

1. Bennett G., *Probability Inequalities for the Sum of Independent Random Variables,*
2. Bentley J. L., McGeogh, *Amortized Analyses of Self-Organizing Sequential Search Heuristics*, *Comm. ACM*, **28**, pp. 404-411, 1985.

$n \setminus \epsilon$	0.05	0.1	0.15	0.2
10	7.73	6.53	6.06	5.37
20	18.21	15.75	14.25	13.0
25	23.84	20.81	18.84	17.48
50	52.13	48.24	34.58	34.28
100	87.03	79.83	65.24	52.33

Table 2. The required length for the reorganization process under CS to approach the optimum within $1+\epsilon$: The ratio between the Hoeffding bound and the stopping point in (16).

3. Bernstein S., *Sur une modification de l'inequalite de Tchebichef*, Annals Science Institute Sav. Ukraine, Sect. Math. I, 1924 (Russian, French summary).
4. Chernoff H., *A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations*, Annals of Math. Stat., **23**, 1952, 493–507.
5. Cohen A., Rabinovich Y., Schuster A., Shachnai H., "Optimal Bounds on Tail Probabilities - A Simplified Approach", TR #911, Technion IIT, CS Dept., May 1997.
6. Dembo A. and Zeitouni O., *Large Deviations and Applications: the finite dimensional case*, a book in preparation, 1992.
7. Feller W., *An Introduction to Probability Theory and its Applications*, John Wiley and Sons, 1957.
8. Hagerup T. and Rüb C., *A Guided Tour of Chernoff Bounds*, Inf. Proc. Lett., **33**, 1990, 305–308.
9. Hoeffding W., *Probability Inequalities for Sums of Bounded Random Variables*, J. Am. Stat. Ass., **58**, 1963, 13–30.
10. Hofri M., Shachnai H., *Self-Organizing Lists and Independent References - a Statistical Synergy*, *Jour. of Alg.*, **12**, pp. 533-555, 1991.
11. Krein M. G. and Nudelman A. A., *The Markov Moment Problem and Extremal Problems*, Translations of Math. Monographs (From Russian), Vol. 50, 1977, American Mathematical Society.
12. McCabe J., *On Serial Files with Relocatable Records*, *Operations Res.*, **13**, pp. 609-618, 1965.

This article was processed using the LaTeX macro package with LLNCS style