A STUDENTIZED CHEBYSHEV INEQUALITY

TAMÁS F. MÓRI

Department of Probability Theory and Statistics, Eötvös Loránd University, H–1088 Budapest, Múzeum krt. 6–8.

(Received January 2, 1984)

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample of X and let $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ denote the sample mean. If X has expectation μ and variance $\sigma^2 < \infty$ then Chebyshev's inequality gives

$$P(|\bar{X}-\mu| > \lambda\sigma) \leqq (n\lambda^2)^{-1} \text{ for every } \lambda > 0.$$

The upper bound appearing on the right-hand side has the following advantageous properties: it is valid for a large class of distributions and it tends to 0 as $\lambda \to \infty$ or $n \to \infty$.

If the scale parameter σ is not known, it is often replaced by its estimate

$$s_n = \left(\sum_{i=1}^{n} (X_i - \bar{X})^2/(n-1) \right)^{1/2}.$$

This substitution, called studentization, involves the problem of giving uniform upper bounds on the probability of the event $\{|\bar{X}-\mu| > \lambda s_n\}$ for general classes of distributions of X. In the wording of Birnbaum [2]:

"For many years statisticians have from time to time given some thought to a question which may be called the 'problem of a studentized Chebyshev inequality', and which can be stated as follows: is there a sequence of functions $\Psi_n(\lambda)$, decreasing to 0 as $\lambda \to +\infty$, such that

$$P(|\bar{X}-\mu| > \lambda s_n) \leqq \Psi_n(\lambda), \tag{1}$$

no matter what probability distribution X may have? To the author's knowledge, no answer has been given to this question."

Several papers are devoted to Chebyshev type inequalities in case of estimated parameters. Most of them deal with estimates for μ and σ other than $\bar{X}$ and s_n, evading thus the original problem. In [1]–[4] order statistics are used and nice inequalities are obtained in the case where μ and σ are estimated by the sample median and the interquantile range between two sample quantiles, resp. The aim of the present paper is to characterize wide classes of distributions for which a studentized Chebyshev inequality of type (1) can be given. Specializing our results we obtain a uniform bound for all unimodal distributions with mode μ, not supposing the existence of any moment.

2. Upper and lower bounds

First we give two sided estimates for the probability

$$P(|\bar{X}|>\lambda s_n/\sqrt{n}), \tag{2}$$

roughly speaking, in terms of the logarithmic concentration function. Since we do not wish to require moment conditions, the location parameter μ cannot be identified with expectation. Without loss of generality we may assume $\mu=0$, due to the invariance of the problem under translation. Further, use of (2) can be made in particular when estimating the tail probabilities of the distribution of Student's statistics $t=\sqrt{n}\,\bar{X}/s_n$ for a non-normal sample. This is why the norming factor $\sqrt{n}$ appears in (2).

Let us define

$$K^+(h)=\sup_x P(X>0,\quad x<\log X\leqq x+h),$$

$$K^-(h)=\sup_x P(X<0,\quad x<\log(-X)\leqq x+h)$$

and

$$K(h)=\max\{K^-(h),\quad K^+(h)\}.$$

This latter can be considered as the logarithmic concentration function of X. If $P(X=0)=0$, thus $\log|X|$ can be interpreted with probability 1, the concentration function of $\log|X|$ lies between $K(h)$ and $2K(h)$, for every $h>0$. K^+, K^- and K are increasing functions of h. Put

$$\lim_{h\to\infty} K^+(h)=P(X>0)=q^+,$$

$$\lim_{h\to\infty} K^-(h)=P(X<0)=q^-,$$

and

$$\lim_{h\to\infty} K(h)=\max\{q^+,q^-\}=q.$$

Theorem 1. *For* $\lambda>n-1$ and $\tau>1$

$$K^n\left[\left(1+\frac{\lambda^2}{2(n-1)}\right)^{-1/2}\right]\leqq P(|\bar{X}|>\lambda s_n/\sqrt{n})\leqq \tau K^{n-1}\left(\tau^*\frac{2(n-1)}{\lambda-(n-1)}\right) \tag{3}$$

where τ^* *is the number conjugated to* τ, *and defined by* $\dfrac{1}{\tau}+\dfrac{1}{\tau^*}=1$.

This theorem asserts that a lumpy distribution of X causes a heavy tail of Student's t.

Next we describe those classes of distributions, denoted by $\mathcal{F}$, for which

$$\Psi_n(\lambda) = \sup_{X\in\mathcal{F}} P(|\bar{X}| > \lambda s_n/\sqrt{n})$$

tends to 0 as $\lambda\to\infty$. First, let $\mathcal{F}$ consist of a single probability distribution. Since $\lim_{h\to+0} K(h) = \max_{x\neq 0} P(X = x)$, it is sufficient to require the continuity of the distribution of X everywhere but at 0. (This is obvious without referring to Theorem 1, because if X takes on a non-zero value with positive probability, then it is possible that $s_n = 0$ and at the same time $\bar{X} \neq 0$.) For larger families of distributions we obtain the following assertion as an immediate consequence of Theorem 1.

Corollary 1. *For any class $\mathcal{F}$ of distributions $\lim_{\lambda\to\infty} \Psi_n(\lambda) = 0$ holds if and only if*

$$\lim_{h\to+0}\sup_{X\in\mathcal{F}} K(h) = 0.$$

In words, necessary and sufficient condition for a class of distributions to satisfy a studentized Chebyshev inequality is that it consists of uniformly smooth probability measures. As an example one can set the notable class $\mathcal{F}_0$ of all unimodal distributions with mode 0 (a probability law belongs to $\mathcal{F}_0$ if the corresponding distribution function $F(x)$ is convex for $x<0$ and concave for $x>0$).

Corollary 2. *For $X\in\mathcal{F}_0$, $\lambda > n-1$*

$$P(|\bar{X}| > \lambda s_n/\sqrt{n}) \leqq 2^{n-1} n^n (\lambda-(n-1))^{-(n-1)}. \tag{4}$$

3. Further problems

Returning now to the properties listed under the formulation of the classical Chebyshev inequality one can see that our estimation (3) becomes meaningless as $n\to\infty$ and λ remains fixed. This raises the problem of giving estimates which are uniform in n. In general it is clearly impossible, because $\sqrt{n}(\bar{X}-\mu)/s_n$ is asymptotically Gaussian if X has a finite variance, therefore $P(|\bar{X}| > \lambda_n s_n/\sqrt{n})$ keeps away from 1 as $n\to\infty$ if and only if $\lambda_n - \sqrt{n}|\mu|/\sigma$ remains bounded from below.

The following extension of Theorem 1 is sharp in a certain sense.

Theorem 2. Let $\lambda_k = \left(\frac{k(n-1)}{n-k}\right)^{1/2}$ $(k = 1, 2, \ldots, n-1)$ *and* $\lambda_n = +\infty$. *Suppose $\lambda_{k-1} < \lambda \leqq \lambda_k$. Then for every $\tau > 1$*

$$P(\bar{X} > \lambda s_n/\sqrt{n}) \leqq \sum_{j>k} \binom{n}{j} (1-q^+)^{n-j} (q^+)^j + \\ + \tau \binom{n}{k} (1-q^+)^{n-k} q^+ \left[K^+\left(\tau^* \frac{2\lambda_{k-1}}{\lambda-\lambda_{k-1}}\right)\right]^{k-1}. \tag{5}$$

From this theorem it follows that $\lim_{n\to\infty} P(|\bar{X}|>\lambda s_n/\sqrt{n})=0$ if $\lim_{n\to\infty}\left(\lambda-\sqrt{n}\frac{q}{1-q}\right)=+\infty$. One can easily see that $\sup|\mu|/\sigma=\frac{q}{1-q}$ where the supremum is taken for all probability distributions having finite variance and satisfying $\max\{P(X>0),\ P(X<0)\}=q$, thus Theorem 2 cannot be improved significantly. However, our estimates seem rather rough e.g. for symmetric distributions. It would be of interest to describe those classes of distributions for which $\lim_{\lambda\to\infty}\sup_n \Psi_n(\lambda)=0$ holds. The class $\mathcal{F}_B$ of all symmetric unimodal distributions (also called bell-shaped) is expected to have this property. Moreover, we conjecture that $\sup_{X\in\mathcal{F}_B} P(|\bar{X}|>\lambda s_n/\sqrt{n})$ is attained for a symmetric uniform distribution. The assumption of symmetry seems relevant: the family $\mathcal{F}_{00}$ of unimodal distributions with mode 0, expectation 0 and median 0 is still too large for this stronger version of the studentized Chebyshev inequality (let the distribution of X be a mixture of uniform distributions $U\left(-\frac{1-\varepsilon}{2},\ -\varepsilon\right)$, $U(-\varepsilon,\ 0)$ and $U(0,\ 1)$ with weights ε, $\frac{1}{2}-\varepsilon$ and $\frac{1}{2}$, resp., for a small positive ε, and then let $\varepsilon\to 0$).

4. Proofs

Proof of Theorem 1. First we give a geometric description of the subset C of the sample space $\mathbf{R}^n$ which corresponds to the event $\{|\bar{X}|>\lambda s_n/\sqrt{n}\}$. Replacing X with $-X$ C remains unchanged, thus we can confine our attention to $C^+=\{\bar{X}>\lambda s_n/\sqrt{n}\}$.

Write $\mathbf{1}(\in\mathbf{R}^n)$ for the vector with coordinates all equal to 1 and let φ $(0\leqq\varphi<\pi/2)$ denote the angle of $\mathbf{X}=(X_1,\ldots,X_n)$ and $\mathbf{1}$. Then

$$\cos\varphi=\left(n\sum_{i=1}^n X_i^2\right)^{-1/2}\left(\sum_{i=1}^n X_i\right)=\left(1+\frac{n-1}{n}s_n^2/\bar{X}^2\right)^{-1/2}$$

i.e. C^+ is described by $\{\cos\varphi>\lambda(n-1+\lambda^2)^{-1/2}\}$ or equivalently, by $\{\tan\varphi<\sqrt{n-1}/\lambda\}$. Hence C is a (double) cone of revolution with vertex $\mathbf{0}$ and axis parallel to $\mathbf{1}$.

In order to estimate the probability of C^+ we cover it by a sequence of hypercubes $\{\beta\alpha^j<X_i\leqq\beta\alpha^{j+\tau},\ 1\leqq i\leqq n\}$, $j\in\mathbf{Z}$, where $\alpha>1$, $\beta>0$ are appropriate constants. This coverage is possible if and only if $C^+\subset\{X_i>0,\ 1\leqq i\leqq n\}$, i.e. if $\tan\varphi<(n-1)^{-1/2}$ in C^+. Concerning λ this means $\lambda>n-1$. Suppose $\mathbf{X}\in C^+$ and define j by $\beta\alpha^j<\min_{1\leqq i\leqq n} X_i\leqq\beta\alpha^{j+1}$, then $\max_{1\leqq i\leqq n} X_i\leqq\beta\alpha^{j+\tau}$ must hold, i.e. $\max X_i/\min X_i\leqq\alpha^{\tau-1}$ in C^+. We want to choose α as small as possible, therefore we need to find the supremum of $\max X_i/\min X_i$ in C^+. To do this let us fix $\max X_i=M$, $\min X_i=m$ and try to minimize φ, or equivalently, maximize $(\Sigma X_i)^2/\Sigma X_i^2$. Let $S = \Sigma X_i-(m+M)$ denote the sum of the "free" coordinates, then

$$\Sigma X_i^2 \geqq m^2+M^2+S^2/(n-2),$$

hence

$$(\sum X_i)^2/(\sum X_i^2) \leqq \frac{(m+M+S)^2}{m^2+M^2+S^2/(n-2)} \leqq n-2+\frac{(m+M)^2}{m^2+M^2}$$

and this upper bound is attained when the free coordinates are all equal to $(m^2+M^2)/(m+M)$. It follows by a simple computation that

$$\frac{\min X_i}{\max X_i} \leqq 1+[n(n-1)+(2n(n-1)\lambda^2-n(n-1)^2(n-2))^{1/2}]/[\lambda^2-(n-1)^2] \leqq$$

$$\leqq 1+\frac{2(n-1)}{\lambda-(n-1)}.$$

Let $\alpha^{\tau-1} = 1+\dfrac{2(n-1)}{\lambda-(n-1)},$ then

$$P(\mathbf{X}\in C^+) \leqq \sum_{j=-\infty}^{+\infty} P^n(\beta\alpha^j<X_1\leqq\beta\alpha^{j+\tau}) \leqq$$

$$\leqq \sum_{j=-\infty}^{+\infty} P^n(\log\beta+j\log\alpha<\log X_1\leqq\log\beta+j\log\alpha+\log\alpha^\tau, X_1>0) \leqq$$

$$\leqq [K^+(\log\alpha^\tau)]^{n-1}H(\log\beta, \log\alpha),$$

where

$$H(u,v) = \sum_{j=-\infty}^{+\infty} P(u+jv<\log X_1\leqq u+(j+\tau)v, X_1>0)$$

$$(-\infty<u<+\infty, 0<v<+\infty).$$

One can readily verify that $\int_0^v H(u,v)du = \tau v q^+$, thus for every $v>0$ u can be chosen such that $H(u,v)\leqq\tau q^+$. Since

$$\log\alpha^\tau = \frac{\tau}{\tau-1}\log\left(1+\frac{2(n-1)}{\lambda-(n-1)}\right) \leqq \tau^*\frac{2(n-1)}{\lambda-(n-1)},$$

we obtain

$$P(\mathbf{X}\in C^+) \leqq \tau q^+\left[K^+\left(\tau^*\frac{2(n-1)}{\lambda-(n-1)}\right)\right]^{n-1}, \tag{6}$$

from which the upper bound of (3) immediately follows.

In order to get a lower estimate of $P(X\in C^+)$ let us inscribe a hypercube $\{m<X_i\leqq M,\ 1\leqq i\leqq n\}$ into the cone C^+. First we determine the supremum of φ over the closure of the cube. Consider $(\Sigma X_i)^2/(\Sigma X_i^2)$ again. One can easily see that this expression is minimized at one of the vertices of the cube, i.e. the minimum is equal to

$$\frac{(lm+(n-l)M)^2}{lm^2+(n-l)M^2} \tag{7}$$

for a certain integer l, $0<l<n$. Letting l take on real values and minimizing (7) in l we obtain that

$$(\Sigma X_i)^2/(\Sigma X_i^2)\geqq n\frac{4mM}{(m+M)^2},$$

hence if

$$\frac{M}{m}\leqq 1+2(n-1)\lambda^{-2}[1+(1+\lambda^2/(n-1))^{1/2}] = \varrho \tag{8}$$

holds, the cube is contained entirely in C^+. Thus

$$P(\mathbf{X}\in C^+)\geqq [K^+(\log\varrho)]^n\geqq\left[K^+\left(\frac{\varrho-1}{\varrho}\right)\right]^n\geqq\left\{K^+\left[\left(1+\frac{\lambda^2}{2(n-1)}\right)^{-1/2}\right]\right\}^n. \ \square$$

Proof of Corollary 2. (4) will follow from (3) with $\tau = n$, $\tau^* = \dfrac{n}{n-1}$ if we show that $K(h)\leqq h$ for every $X\in\mathcal{F}_0$. If $0<x<y$, $P(x<X\leqq y)\leqq q^+\left(1-\dfrac{x}{y}\right)$ because of the unimodality at 0. Hence $K^+(h)\leqq q^+(1-e^{-h})\leqq q^+h$ and $K(h)\leqq h$. $\square$

Proof of Theorem 2. Suppose that C^+ contains a point $\mathbf{X}$ having exactly k positive coordinates, say $X_1, X_2, \ldots, X_k$ $(k<n)$, Then

$$\frac{n\lambda^2}{n-1+\lambda^2}<\left(\sum_{i=1}^n X_i\right)^2\Big/\left(\sum_{i=1}^n X_i^2\right)\leqq\left(\sum_{i=1}^k X_i\right)^2\Big/\left(\sum_{i=1}^k X_i^2\right)\leqq k,$$

thus

$$\lambda\leqq\left(\frac{k(n-1)}{(n-k)}\right)^{1/2} = \lambda_k.$$

Hence if $\lambda_{k-1}<\lambda\leqq\lambda_k$, every point of C^+ has at least k positive coordinates. Parts of C^+ consisting of points with more than k positive coordinates are estimated simply by

$$\sum_{j>k}\binom{n}{j}(1-q^+)^{n-j}(q^+)^j$$

the remaining parts of C^+ can be treated by projecting them onto the subspace spanned by the positive coordinates. The image thus obtained is a cone of lower dimensions. Suppose $X_1, X_2, \ldots, X_k > 0$ and let

$$\overline{X}_k = \sum_{i=1}^{k} X_i/k, \quad s_k^2 = \sum_{i=1}^{k} (X_i - \overline{X}_k)^2/(k-1).$$

Then $\mathbf{X} \in C^+$ implies $\overline{X}_k > \eta s_k/\sqrt{k}$, where $\dfrac{n\lambda^2}{n-1+\lambda^2} = \dfrac{k\eta^2}{k-1+\eta^2}$, i.e.

$$\eta = \lambda\left(\frac{\lambda_k^2 - 1}{\lambda_k^2 - \lambda^2}\right)^{1/2} \geqq \lambda\left(\frac{\lambda_k^2 - 1}{\lambda_k^2 - \lambda_{k-1}^2}\right)^{1/2} = \frac{\lambda}{\lambda_{k-1}}(k-1).$$

From Theorem 1 it follows that

$$P(\overline{X}_k > \eta s_k/\sqrt{k}) \leqq \tau\left[K^+\left(\tau^* \frac{2(k-1)}{\eta-(k-1)}\right)\right]^{k-1} \leqq \tau\left[K^+\left(\tau^* \frac{2\lambda_{k-1}}{\lambda - \lambda_{k-1}}\right)\right]^{k-1}$$

for every $\tau > 1$.

Now the proof can easily be completed. □

REFERENCES

[1] *Ahmad I. A.:* A class of asymptotically distribution-free statistics similar to Student's *t*. *Comm. Statist. A – Theory and Methods* 4 (1975), 863–871.

[2] *Birnbaum Z. W.:* On a statistics similar to Student's *t*. In: M. L. Puri, Ed., Nonparametric Techniques in Statistical Inference, Cambridge University Press. 1970, pp. 427–433.

[3] *Kuan K. S.:* The tail probability of a statistics similar to Student's *t*. *SEA Bull. Math.* 1 (1977), 46–49.

[4] *Sathe Y. S.* and *Shenoy R. G.:* On inequalities for statistics similar to Student's *t*. *Comm. Statist. A – Theory and Methods* 8 (1979), 575–579.

[5] *Shane H. D.:* On an inequality for order statistics. *Ann. Math. Statist.* 42 (1971), 1748–1751.