Tail bound for sums of bounded random variables

Warren D. Smith*
wds@math.temple.edu

April 1, 2005

***Abstract* — We bound the tail probability for a sum $X_1 + X_2 + \cdots + X_N$ of N bounded random variables. The variables are not assumed to be independent, but the mean and upper and lower bounds on each X_n are assumed independent of all the other X_m.**

Keywords —

In 1963 Hoeffding [1] (as his "theorem 2" on page 16) gave an excellent bound on the tail probability of a sum of independent bounded random variables.

We now improve this situation by showing that the same bound still holds under a much weaker independence assumption.

Let $S = X_1 + X_2 + \cdots + X_N$ be the sum of N random variables.

We assume each random variable X_i is has a mean $\mu_i = E(X_i)$ and is bounded: $A_i \le X_i - E(X_i) \le B_i$. We shall *not* assume that the x_n are independent, but we do assume this amount of "independence": the mean μ_n of each X_n, and the lower and upper bounds A_n and B_n on it, are unaffected by (i.e. valid regardless of) what the other X_m do.

One way this scenario often arises is this. The X_n each are bounded by $|X_n| < B_n$ and have mean $\mu_n = 0$; and although the X_n are quite interdependent, the *sign* of X_n is gotten by a coin toss independent of all the other X_m.

Theorem 1 (Main result). *Let E denote expectation. For $t \ge 0$*

$$\text{Prob}\,(S - E(S) \ge t) \;\le\; \exp\left(\frac{-2t^2}{\sum_{i=1}^N (B_i - A_i)^2}\right). \tag{1}$$

Proof: We first prove the theorem under the assumption that the X_n are independent; our proof follows Hoeffding [1] page 22.

Consider the indicator variable $1_{S-E(S)\ge t}$. This indicator variable is dominated by $\exp\left((S - E(S) - t)h\right)$ for any constant $h > 0$. So

$$\begin{aligned} \text{Prob}\,(S - E(S) \ge t) &= E(1_{S-E(S)\ge t}) \;\le\; E\exp\left((S - E(S) - t)h\right) \;=\; \exp(-ht) E \exp\left(S - E(S)\right) & (2) \\ &= \exp(-ht) \prod_{n=1}^N E \exp\left(hX_n - hE(X_n)\right) \qquad \text{(due to independence)} & (3) \end{aligned}$$

We now upper-bound $E\exp\left(hX_n - hE(X_n)\right)$ by using the inequality (a special case of Jensen's inequality for convex-∪ functions arising when the function is e^x)

$$Ee^{hX} \;\le\; \frac{B - EX}{B - A} e^{hA} + \frac{EX - A}{B - A} e^{hB}. \tag{4}$$

Let $\mu_n = EX_n$. Then the proof continues

$$\ldots \;\le\; \exp(-ht) \prod_{n=1}^N E\left(\frac{B_n - \mu_n}{B_n - A_n} e^{hA_n} + \frac{\mu_n - A_n}{B_n - A_n} e^{hB_n}\right) \le\; \exp(-ht) \exp \sum_{n=1}^N L_n(h_n) \tag{5}$$

where

$$h_n = (B_n - A_n)h \quad \text{and} \quad L_n(h_n) = \ln\left(1 - p_n + p_n \exp(h_n)\right) - p_n h_n \quad \text{where} \;\; p_n = \frac{\mu_n - A_n}{B_n - A_n}. \tag{6}$$

Now

$$L'(u) = -p + \frac{pe^u}{1 - p + pe^u}, \qquad L''(u) = \frac{(1-p)pe^u}{(1 - p + pe^u)^2}. \tag{7}$$

*21 Shore Oaks Drive, Stony Brook NY 11790.

By considering maximizing $L''(u)$ with respect to p (which uniquely happens when $p = 1/(1+e^u)$) we find that $L''(u) \le 1/4$. Therefore by Taylor's formula $L(h_n) \le L(0) + L'(0)h_n + (1/8)h_n^2$. And evidently $L(0) = L'(0) = 0$. So the proof continues

$$\ldots \quad \le \quad \exp(-ht) \exp \sum_{n=1}^{N} (B_n - A_n)^2 h^2/8. \tag{8}$$

We now choose $h = 4t / \sum_{n=1}^{N} (B_n - A_n)^2$ to minimize this, yielding the theorem.

In the case of independent X_n this result is called "Hoeffding's inequality" and we have just re-proved it.

Now let us consider what happens if we allow dependence among the X_n. There was only one step in the above proof where independence was used: EQ 3.

We can justify that step (as a $\le$ ratehr than an $=$) *without* independence by instead relying on the following

Lemma 2 (Expectation of product). *Let X and Y be random variables and let X be a non-negative random variable. Then*

$$E(XY) \;\le\; E(X) \, \max_k E(Y|X=k) \tag{9}$$

where $E(Y|X=k)$ means the conditionally expected value of Y given that $X=k$.

Proof:

$$E(XY) = \int_{x \ge 0} x \, E(Y|X=x) \, \mathrm{prob}(x) \, \mathrm{d}x \le \int_{x \ge 0} x \left(\max_k E(Y|X=k) \right) \mathrm{prob}(x) \mathrm{d}x = E(X) \, \max_k E(Y|X=k). \tag{10}$$

Q.E.D.

Because $\exp(x) > 0$ for all x and because EQ 3 concerned the expectation of a product of exponentials, we now have a product of non-negative random variables and the lemma is applicable. Because our EQ 4 gives an upper bound on the expectation of any such exponential Y valid *regardless* of what the other X_m do, it is at least as large as $\max_k E(Y|\text{other } X_m)$ and therefore it is valid to use it in the lemma.

The theorem follows. Q.E.D.

References

[1] Wassily Hoeffding: Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58,301 (March 1963) 13-30.