Estimating Variance

Andreas Maurer

The purpose of this note is to give an upper bound on the variance of a bounded vector- or real valued random variable in terms of an iid sample. Let $X, X', X_1, ..., X_n$ be iid random variables with values in a ball B of diameter 1 in some Hilbert space (which could be $\mathbb{R}$). We write

$$V = \frac{1}{2}\mathbb{E}\left\|X - X'\right\|^2 \text{ and } \hat{V} = \frac{1}{2n(n-1)}\sum_{i,j}\left\|X_i - X_j\right\|^2.$$

Theorem 1 $\Pr\left\{V - \hat{V} > t\right\} \leq \exp\left(-nt^2/4V\right)$.

The proof relies on the following concentration inequality (Theorem 13, 2nd conclusion in [4], similar results from [2] or [3] could also be used):

Theorem 2 *Let $\mathbf{X} = (X_1, ..., X_n)$ be a vector of independent random variables with values in some set Ω. For $1 \leq k \leq n$ and $y \in \Omega$, we use $\mathbf{X}_{y,k}$ to denote the vector obtained from $\mathbf{X}$ by replacing X_k by y. Suppose that $a > 1$ and that $Z = Z(\mathbf{X})$ satisfies the inequalities*

$$Z(\mathbf{X}) - \inf_{y\in\Omega} Z(\mathbf{X}_{y,k}) \leq 1, \forall k \tag{1}$$

$$\sum_{k=1}^{n}\left(Z(\mathbf{X}) - \inf_{y\in\Omega} Z(\mathbf{X}_{y,k})\right)^2 \leq aZ(\mathbf{X}) \tag{2}$$

almost surely. Then, for $t > 0$,

$$\Pr\{\mathbb{E}Z - Z > t\} \leq \exp\left(\frac{-t^2}{2a\mathbb{E}Z}\right).$$

Proof of Theorem 1. Write $Z = n\hat{V}$. Fix some k and choose any $y \in B$. Then

$$Z(\mathbf{X}) - Z(\mathbf{X}_{y,k}) = \frac{1}{n-1}\sum_j\left(\left\|X_k - X_j\right\|^2 - \left\|y - X_j\right\|^2\right) \leq \frac{1}{n-1}\sum_j\left\|X_k - X_j\right\|^2.$$

It follows that $Z(\mathbf{X}) - \inf_{y\in\Omega} Z(\mathbf{X}_{y,k}) \leq 1$. We also get

$$\begin{aligned}\sum_k\left(Z(\mathbf{X}) - \inf_{y\in\Omega} Z(\mathbf{X}_{y,k})\right)^2 &\leq \sum_k\left(\frac{1}{n-1}\sum_j\left\|X_k - X_j\right\|^2\right)^2 \\ &\leq \frac{1}{n-1}\sum_k\sum_j\left\|X_k - X_j\right\|^2 = 2Z(\mathbf{X}),\end{aligned}$$

so that Z satisfies (1) and (2) with $a = 2$. Since $\mathbb{E}Z = n\mathbb{E}\hat{V} = nV$, Theorem 2 gives the conclusion. ■

In the real valued case the exponent can be improved. This also furnishes an opportunity to show the advantages of Theorem 2 over the results in [2] and [3]. We need a technical lemma.

Lemma 3 *Let X, Y be iid random variables with values in an interval $[a, a+1]$. Then*

$$\mathbb{E}_X\left[\mathbb{E}_Y(X-Y)^2\right]^2 \leq (1/2)\,\mathbb{E}(X-Y)^2.$$

Proof. The right side above is of course the variance $\mathbb{E}\left[X^2 - XY\right]$. One computes

$$\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}_Y(X-Y)^2\right]^2 &= \mathbb{E}\left(X^2 - 2X\mathbb{E}(Y) + \mathbb{E}\left(Y^2\right)\right)^2 \\
&= \mathbb{E}\left(X^4\right) + 3\mathbb{E}\left(X^2\right)^2 - 4\mathbb{E}\left(X^3\right)\mathbb{E}(X) \\
&= \mathbb{E}\left[X^4 + 3X^2Y^2 - 4X^3Y\right]
\end{aligned}$$

We therefore have to show that $\mathbb{E}\left[g(X,Y)\right] \geq 0$ where

$$g(X,Y) = X^2 - XY - X^4 - 3X^2Y^2 + 4X^3Y$$

A rather tedious computation gives

$$\begin{aligned}
g(X,Y) + g(Y,X) &= X^2 - XY - X^4 - 3X^2Y^2 + 4X^3Y \\
&\quad + Y^2 - XY - Y^4 - 3X^2Y^2 + 4Y^3X \\
&= (X - Y + 1)(Y - X + 1)(Y - X)^2.
\end{aligned}$$

The latter expression is clearly nonnegative, so

$$2\left[\mathbb{E}g(X,Y)\right] = \mathbb{E}\left[g(X,Y) + g(Y,X)\right] \geq 0,$$

which completes the proof. ■

Theorem 4 *If $B = [0,1]$ then* $\Pr\left\{V - \hat{V} > t\right\} \leq \exp\left(-(n-1)\,t^2/2V\right)$.

Proof. We apply Lemma 3 to the empirical measure uniform on $(X_1, ..., X_n)$, multiply with n^3 and divide by $(n-1)^2$ to obtain obtain

$$\sum_k \left(\frac{1}{n-1}\sum_j (X_k - X_j)^2\right)^2 \leq \frac{n}{2(n-1)^2}\sum_{kj}(X_k - X_j)^2 = \frac{n}{n-1}\hat{V}.$$

The conditions of Theorem 1 are therefore satisfied with $a = n/(n-1)$. Proceed as above. ■

Corollary 5 *(i) For $\delta > 0$ we have with probability at least δ that*

$$V \leq \hat{V} + \sqrt{\frac{4\hat{V}\ln 1/\delta}{n}} + \frac{4\ln 1/\delta}{n}.$$

(ii) If $B = [0,1]$ we have with probability at least δ that

$$V \leq \hat{V} + \sqrt{\frac{2\hat{V}\ln 1/\delta}{n-1}} + \frac{2\ln 1/\delta}{n-1}.$$

(iii) If $B = [0,1]$ and $\hat{X} = (1/n)\sum X_i$ then, for $\delta > 0$, we have with probability at least δ that

$$\mathbb{E}X \leq \hat{X} + \sqrt{\frac{2\hat{V}\ln 2/\delta}{n}} + \frac{7\ln 2/\delta}{3\,(n-1)}.$$

Proof. Equating the right side of the bound in Theorem 1 to δ and solving for t gives, with probability at least δ,

$$\begin{aligned} V &\leq \hat{V} + \sqrt{\frac{4V\ln 1/\delta}{n}} \implies \\ \sqrt{V} &\leq \sqrt{\hat{V} + \frac{\ln 1/\delta}{n}} + \sqrt{\frac{\ln 1/\delta}{n}}. \end{aligned}$$

Squaring and the estimate $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ then give conclusion (i). For (ii) we use Theorem 4 in an analogous way to obtain

$$\sqrt{V} \leq \sqrt{\hat{V} + \frac{\ln 1/\delta}{2\,(n-1)}} + \sqrt{\frac{\ln 1/\delta}{2\,(n-1)}} \tag{3}$$

and the conclusion. Finally recall that Bernstein's inequality [5] implies

$$\mathbb{E}X \leq \hat{X} + \sqrt{V}\sqrt{\frac{2\ln 1/\delta}{n}} + \frac{\ln 1/\delta}{3n},$$

so that the last conclusion follows from combining this with (3) in a union bound and some simple estimates. ■

Part (iii) above is a version of an *empirical Bernstein bound* (see Audibert et al [1]). In [1] the result is obtained in a triple application of Bernsteins inequality, resulting in a slightly larger constant in the last term.

References

[1] J. Y. Audibert, R. Munos, C. Szepesvári. Exploration-exploitation trade-off using variance estimates in multi-armed bandits, Preprint.

[2] S. Boucheron, G. Lugosi, P. Massart, *A sharp concentration inequality with applications in random combinatorics and learning*, Random Structures and Algorithms, (2000) 16:277-292.

[3] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities using the entropy method*, Annals of Probability (2003) 31:1583-1614.

[4] Maurer, A. (2006). Concentration inequalities for functions of independent variables. *Random Structures Algorithms* 29 121–138.

[5] C. McDiarmid, *Concentration*, in *Probabilistic Methods of Algorithmic Discrete Mathematics*, (1998) 195-248. Springer, Berlin