A LEMMA ON PROXIMITY OF VARIANCES AND EXPECTATIONS\footnote{I thank Doe Monderer, Haim Raizman and Ishy Weissman for comments and discussions.}

Doron Sonsino\footnote{Technion, Israel Institute of Technology, Haifa, Israel; e-mail: sonsino@ie.technion.ac.il}

Abstract. We define a notion of delta-variance maximization and show it implies epsilon-proximity in expectations.

AMS Subject Classification. 60A99.

Received March 19, 2000.

We present a lemma stipulating that when the variance of each element in a collection of random variables is maximal with respect to some larger class of random variables, then the corresponding expectations must be very close.

First, we formally define our notion of a large class of random variables.

\textbf{Definition.} A collection \( \mathcal{G} \) of random variables on a probability space \((\Omega, \mathcal{F}, P)\) is CPC (Closed under Piecewise Compositions) if for every \( y, y' \in \mathcal{G} \) and every measurable \( E \in \mathcal{F} \), \( y \cdot 1_E + y' \cdot 1_{E^c} \in \mathcal{G} \).

In words, the collection \( \mathcal{G} \) is closed under piecewise compositions if whenever \( y \) and \( y' \) are in \( \mathcal{G} \) and \( E \) is a measurable set, the random variable that takes the value \( y(\omega) \) when \( \omega \) belongs to \( E \) and takes the value \( y'(\omega) \) elsewhere, is in the collection \( \mathcal{G} \) as well.

We say that the collection of random variables \( \mathcal{G} \) is a CPC-extension of the collection \( \mathcal{D} \) if \( \mathcal{G} \) is a CPC-collection of random variables that contains \( \mathcal{D} \) (where all the random variables in both collections are defined on the same probability space \((\Omega, \mathcal{F}, P))\).

We now define the notion of \( \delta \)-variance maximization.

\textbf{Definition.} A collection \( \mathcal{D} \) of random variables in \( L^2(\Omega, \mathcal{F}, P) \) is \( \delta \)-variance maximizing if there is a CPC-extension of \( \mathcal{D} \), \( \mathcal{G} \), and a finite \( k \) such that:

\begin{align*}
(1) \quad \text{Var}(y) &\leq k \quad \forall y \in \mathcal{G}, \text{ and} \\
(2) \quad \text{Var}(y) &\geq k - \delta \quad \forall y \in \mathcal{D}.
\end{align*}

Put differently, the collection of random variables \( \mathcal{D} \) is \( \delta \)-variance maximizing if there is a large (in the sense of CPC) collection of random variables \( \mathcal{G} \) that contains \( \mathcal{D} \) such that the variance of each of the random variables in \( \mathcal{D} \) is within \( \delta \) from the supremum of variances over the larger collection \( \mathcal{G} \).

Note that if \( y \) and \( y' \) belong to a \( \delta \)-variance maximizing collection of random variables, then \( |\text{Var}(y) - \text{Var}(y')| \leq \delta \) and the two random variables have similar variances. Clearly, closeness in variances alone is not

\begin{flushleft}
Keywords and phrases: Variance, expectation.\end{flushleft}
strong enough to induce similarity in expectations. The MVSE lemma\(^3\) however says that the stronger concept of \(\delta\)-variance maximization is sufficient for this purpose.

**The MVSE lemma.** Let \(\mathcal{D}\) be a collection of random variables on a non atomic probability space \((\Omega, \mathcal{F}, P)\). If \(\mathcal{D}\) is \(\delta\)-variance maximizing then \(|E[y] - E[y']| \leq 2\sqrt{\delta}\) for every \(y, y' \in \mathcal{D}\).

**Proof.** Let \(\mathcal{D}\) be a \(\delta\)-variance maximizing collection of random variables on a non atomic probability space \((\Omega, \mathcal{F}, P)\). Fix \(y, y'\) in \(\mathcal{D}\) and assume w.l.g. that \(E[y] > E[y']\). Assume by way of contradiction that \(E[y] - E[y'] = 2\sqrt{\tau} > 2\sqrt{\delta}\).

Let

\[
A = \{ \omega \in \Omega \mid (y(\omega) - E[y] + \sqrt{\tau})^2 \geq (y'(\omega) - E[y'] - \sqrt{\tau})^2 \}
\]

\[
B = A^C = \{ \omega \in \Omega \mid (y(\omega) - E[y] + \sqrt{\tau})^2 < (y'(\omega) - E[y'] - \sqrt{\tau})^2 \}
\]

Observe that (by standard arguments) \(A\) and \(B\) are measurable w.r.t \(\mathcal{F}\).

Note that since \(E[y] - E[y'] = \int (y - y') \, dP = 2\sqrt{\tau}\), it is either the case that

\[
\int_A (y - y') \, dP \geq \sqrt{\tau}, \quad \text{or} \quad E[y] - E[y'] = 2\sqrt{\tau}
\]

\[
\int_B (y - y') \, dP \geq \sqrt{\tau}.
\]

In case 1, let \(E\) be a measurable subset of \(A\) such that

\[
\int_E (y - y') \, dP = \sqrt{\tau}.
\]

The existence of such a measurable \(E\) follows from the assumption that \(P\) is non atomic (see Billingsley [1], 2.17, p. 31).

Set \(z = y \cdot I_E + y' \cdot I_{E^C}\) and observe that \(z\) must belong to any CPC-extension of \(\mathcal{D}\). Thus, the assumptions that \(\mathcal{D}\) is \(\delta\)-variance maximizing implies that

\[
\text{Var}(z) \leq \text{Var}(y) + \delta \quad \text{and} \quad \text{Var}(z) \leq \text{Var}(y') + \delta.
\]

But note that by definitions of \(z\) and \(E\),

\[
E[z] = \int z \, dP = \int_E y \, dP + \int_{E^C} y' \, dP = \int_{\Omega} y' \, dP + \int_E (y - y') \, dP = E[y] + \sqrt{\tau},
\]

and similarly

\[
E[z] = \int_{\Omega} y \, dP - \int_{E^C} (y - y') \, dP = \int_{\Omega} y \, dP - \int_{\Omega} (y - y') \, dP + \int_E (y - y') \, dP = E[y] - \sqrt{\tau}.
\]

Thus,

\[
\text{Var}[z] = \int (z - E[z])^2 \, dP = \int_E (y - E[y] + \sqrt{\tau})^2 \, dP + \int_{E^C} (y' - E[y'] - \sqrt{\tau})^2 \, dP,
\]

\(^3\)MVSE stands for Maximal Variance Similar Expectations.
and since \((y - E[y] + \sqrt{\epsilon})^2 \geq (y' - E[y']) - \sqrt{\epsilon})^2\) on \(E \subseteq A\),

\[
\text{Var}[z] \geq \int_{\Omega} (y' - E[y']) - \sqrt{\epsilon})^2 dP = \int_{\Omega} (y' - E[y'])^2 dP + \epsilon = \text{Var}[y'] + \epsilon \geq \text{Var}[y'] + \delta,
\]

which contradicts (4) and proves that case (1) is impossible.

In a very similar way we may argue that case (2) leads to a contradiction as well: assume by way of contradiction that the condition in (2) holds. Let \(E\) be a measurable subset of \(B\) such that

\[
\int_{E} (y' - y') dP = \sqrt{\epsilon}.
\]

(Existence follows again from the assumption that \(P\) is non atomic.)

Let \(z = y' \cdot I_{E} + y \cdot I_{E^C}\).

Note (as above) that

\[
E[z] = E[y] - \sqrt{\epsilon} = E[y'] + \sqrt{\epsilon}.
\]

Thus,

\[
\text{Var}[z] = \int (z - E[z])^2 dP = \int_{E} (y' - E[y']) - \sqrt{\epsilon})^2 dP + \int_{E^C} (y - E[y]) + \sqrt{\epsilon})^2 dP
\]

\[
\geq \int_{\Omega} (y - E[y]) + \sqrt{\epsilon})^2 dP = \int_{\Omega} (y - E[y])^2 dP + \epsilon = \text{Var}[y] + \epsilon \geq \text{Var}[y] + \delta,
\]

which again contradicts (4) and proves that case (2) is impossible as well.

\[\square\]

**Reference**