A Proof of the Jensen-Dragomir Inequality

P.J. Fitzsimmons

March 29, 2012

A general (measure theoretic) version of Dragomir’s gloss on Jensen’s inequality has recently been published by Aldaz [1]. Below we state this theorem for random variables in \(d\)-dimensional Euclidean space, and provide what we feel is a simple proof. To state the theorem let \(\mu\) and \(\nu\) be Borel probability measures concentrated on some open set \(U \subset \mathbb{R}^d\) and let \(\varphi : U \to \mathbb{R}\) be convex. Assume that \(\int_U |x| \mu(dx) + \int_U |x| \nu(dx) < \infty\). Let \(V(\mu) := \int_U \varphi(x) \mu(dx) - \varphi(\int_U x \mu(dx))\) denote the \(\text{“Jensen functional”}\) associated with \(\varphi\). Observe that \(\int_U \varphi(x) \mu(dx)\) is well-defined, taking values in \((-\infty, +\infty]\).

**Theorem.** Suppose that \(\mu \ll \nu\) and let \(k \in (0, \infty]\) denote the \(L^\infty(\nu)\)-norm of the Radon-Nikodym derivative \(h := \frac{d\mu}{d\nu}\). Then

\[
V(\mu) \leq k \cdot V(\nu).
\]

**Proof.** Given \(y \in U\) let \(\gamma(y) \in \mathbb{R}^d\) be a subgradient of \(\varphi\) at \(y\). This means that

\[
f(x, y) := \varphi(x) - \varphi(y) - \gamma(y) \cdot (x - y) \geq 0
\]

for all \(x \in U\). We write \(m = \int_U x \mu(dx)\) and \(n = \int_U x \nu(dx)\), and compute

\[
V(\mu) = \int_U [\varphi(x) - \varphi(m)] \mu(dx) = \int_U f(x, m) \mu(dx)
\]

\[
= \int_U f(x, n) \mu(dx) + \int_U [f(x, m) - f(x, n)] \mu(dx)
\]

\[
\leq k \cdot \int_U f(x, n) \nu(dx) + \int_U [\varphi(n) - \varphi(m) - \gamma(m) \cdot (x - m) + \gamma(n) \cdot (x - n)] \mu(dx)
\]

\[
= k \cdot V(\nu) + \int_U [\varphi(n) - \varphi(m) + \gamma(n) \cdot (x - n)] \mu(dx)
\]

\[
= k \cdot V(\nu) - f(m, n)
\]

\[
\leq k \cdot V(\nu).
\]

(We adhere to the usual convention that \(\infty \cdot 0 = 0\); in particular, \(k \cdot \int_U f(x, n) \nu(dx) = 0\) if \(k = \infty\) and \(\int_U f(x, n) \nu(dx) = 0\).)

There is a companion lower bound \(k_* V(\nu) \leq V(\mu)\) in which \(k_*\) is the \(\nu\)-essential infimum of \(h\); this bound is obtained by reversing the roles of \(\mu\) and \(\nu\). See [1; Cor. 2.9] for details.
Henceforth we assume that \( \varphi \) is strictly convex. Suppose that the two sides of (1) are equal (and finite). If they are both equal to 0, then by the aforementioned convention, \( V(\mu) = 0 = V(\nu) \), so (by strictly convexity of \( \varphi \)) both \( \mu \) and \( \nu \) are unit point masses at \( m = n \).

If the two sides of (1) are equal and strictly positive (as well as finite) then \( k \) must be finite, both \( \int_U \varphi(x) \mu(dx) \) and \( \int_U \varphi(x) \nu(dx) \) must be finite, and

\[
(2) \quad f(m, n) = 0,
\]

and

\[
(3) \quad \int_U [k - h(x)] \cdot f(x, n) \nu(dx) = 0.
\]

Because \( \varphi \) is strictly convex, (2) forces \( m = n \), and then (3) implies that

\[
(4) \quad \nu\{x \in U : h(x) = k \text{ or } x = m\} = 1.
\]

Define \( p := \nu\{m\} \) and \( \tilde{\nu}(A) := (1 - p)^{-1} \nu(A \setminus \{m\}) \) for Borel \( A \subset U \). (If \( p = 1 \) then we take \( \tilde{\nu} \) to be \( \delta_m \), the unit point mass at \( m \).) Then \( \nu \) decomposes as

\[
\nu = p \cdot \delta_m + (1 - p) \tilde{\nu},
\]

and, in view of (4),

\[
\mu = ph(m) \cdot \delta_m + (1 - p) k \cdot \tilde{\nu}.
\]

(Notice that \( h(m) \) is uniquely determined as \( \mu\{m\}/p \) if \( p > 0 \).) Consequently, \( 1 - \mu\{m\} = k \cdot (1 - \nu\{m\}) \).

References
