Jensen's inequality in $\mathbb{R}$

Matija Vidmar

November 4, 2013

Proposition 1 (Jensen's inequality)**.** *Let X be a random variable, $\mathsf{E}|X| < +\infty$, X taking values in an interval $I \subset \mathbb{R}$, $\phi : I \to \mathbb{R}$ convex. Then $\phi \circ X$ is a random variable, $\mathsf{E}X \in I$, $\mathsf{E}[\phi \circ X] \in (-\infty, +\infty]$ is well-defined, and*

$$\mathsf{E}[\phi \circ X] \geq \phi(\mathsf{E}X).$$

Moreover, if $\mathsf{E}X \in I \backslash \overset{\circ}{I}$, or else if ϕ is strictly convex on some open interval containing $\mathsf{E}X$ and $\mathsf{E}[\phi \circ X] = \phi(\mathsf{E}X)$, then necessarily $X = \mathsf{E}X$, P-a.s.[1]

Proof. $\phi|_{\overset{\circ}{I}}$ is continuous and $I \backslash \overset{\circ}{I}$ and $\overset{\circ}{I}$ are Borel subsets of $\mathbb{R}$. Thus ϕ is Borel measurable, and so $\phi \circ X$ a random variable. Next, let $a_1 := \inf I \in [-\infty, +\infty)$ and $a_2 := \sup I \in (-\infty, +\infty]$ (note that, necessarily, $I \neq \emptyset$). If $a_1 = -\infty$ (resp. $a_2 = +\infty$), clearly $\mathsf{E}X > a_1$ (resp. $\mathsf{E}X < a_2$), since, by assumption, $\mathsf{E}|X| < +\infty$. Otherwise, note that $\mathsf{E}X = a_1$ (resp. $\mathsf{E}X = a_2$) implies $X = a_1$ (resp. $X = a_2$), P-a.s., so that $a_1 \in I$ (resp. $a_2 \in I$). It follows that $\mathsf{E}X \in I$. Further to this, note that for some $\{a, b\} \subset \mathbb{R}$, $\phi \circ X \geq aX + b$, hence $(\phi \circ X)^- \leq (aX + b)^- \leq |aX + b| \leq |a||X| + |b|$, so that $\mathsf{E}(\phi \circ X)^- < +\infty$.

We next wish to prove $\mathsf{E}[\phi \circ X] \geq \phi(\mathsf{E}X)$. If $\mathsf{E}X \in I \backslash \overset{\circ}{I}$, this is clear, since then $X = \mathsf{E}X$, P-a.s. Otherwise, $\mathsf{E}X \in \overset{\circ}{I}$, and $\phi|_{\overset{\circ}{I}}$ is the pointwise supremum of the restrictions to $\overset{\circ}{I}$ of the affine minorants of ϕ.[2] Let the set of the latter be denoted $\mathcal{A}$. Then $\phi \circ X \geq a \circ X$, and hence $\mathsf{E}[\phi \circ X] \geq a(\mathsf{E}X)$ for all $a \in \mathcal{A}$. Now take the supremum over $a \in \mathcal{A}$ to obtain $\mathsf{E}[\phi \circ X] \geq \phi(\mathsf{E}X)$.

Finally, suppose that, in fact, $\mathsf{E}[\phi \circ X] = \phi(\mathsf{E}X)$ and $\mathsf{E}X \in \overset{\circ}{I}$, with ϕ strictly convex on an open interval containing $\mathsf{E}X$. Then there exists some $a \in \mathcal{A}$, such that the only zero of $\phi - a$ is $\mathsf{E}X$. We obtain $\mathsf{E}[\phi \circ X - a \circ X] = \phi(\mathsf{E}X) - a(\mathsf{E}X) = 0$, hence $\phi \circ X = a \circ X$, and so $X = \mathsf{E}X$, P-a.s. $\square$

[1] $\overset{\circ}{I}$ is the interior of I, i.e. it is the interval I without its endpoints.

[2] This is a consequence of the fact that ϕ has "nondecreasing difference quotients", in the precise sense that $\frac{\phi(t)-\phi(s)}{t-s} \leq \frac{\phi(u)-\phi(s)}{u-s} \leq \frac{\phi(u)-\phi(t)}{u-t}$, whenever $\{s, t, u\} \subset I$ and $s < t < u$. Moreover, $\phi|_{\overset{\circ}{I}}$ then admits a finite left and right derivative function, the latter pointwise no smaller than the former.