Integral Jensen inequality

Let us consider a convex set $C \subset \mathbb{R}^d$, and a convex function $f\colon C \to (-\infty, +\infty]$. For any $x_1, \dots, x_n \in C$ and $\lambda_1, \dots, \lambda_n \geq 0$ with $\sum_1^n \lambda_i = 1$, we have

$$f(\textstyle\sum_1^n \lambda_i x_i) \leq \sum_1^n \lambda_i f(x_i)\,. \tag{1}$$

For $a \in \mathbb{R}^d$, let δ_a be the Dirac measure concentrated at a, that is

$$\delta_a(E) = \begin{cases} 1 & \text{if } a \in E \\ 0 & \text{if } a \notin E \end{cases}.$$

Then $\mu := \sum_1^n \lambda_i \delta_{x_i}$ is a probability measure on C, defined for all subsets of C. Moreover, we can write $\sum_1^n \lambda_i f(x_i) = \int_C f\,d\mu$ and $\sum_1^n \lambda_i x_i = \int_C x\,d\mu(x)$, where the last integral is the integral of the vector-valued function $\psi(x) = x$. (Vector-valued integration will be recalled below.) Thus the finite Jensen inequality (1) can be written in the integral form

$$f\big(\textstyle\int_C x\,d\mu(x)\big) \leq \int_C f\,d\mu\,. \tag{2}$$

Our aim is to show that (2) holds, under suitable assumptions, also for more general probability measures on C.

Vector integration. Let (Ω, Σ, μ) be a positive measure space, and $F = (F_1, \dots, F_d)\colon \Omega \to \mathbb{R}^d$ a measurable function (that is, $F^{-1}(A) \in \Sigma$ for each open set $A \subset \mathbb{R}^d$). It is easy to see that F is measurable if and only if each F_k is a measurable function. We say that F is *integrable* on a set $E \in \Sigma$ if $F_k \in L_1(E, \mu)$ for each $k = 1, \dots, d$. It is easy to see that F is integrable on E if and only if $\int_E \|F\|\,d\mu < +\infty$. In this case, we define

$$\int_E F\,d\mu = \left(\int_E F_1\,d\mu\,,\ \dots\ ,\ \int_E F_d\,d\mu\right).$$

Observe that, for each linear functional $\ell \in (\mathbb{R}^d)^*$, we have

$$\ell\left(\int_E F\,d\mu\right) = \int_E \ell \circ F\,d\mu \tag{3}$$

(this follows immediately by representing ℓ with a vector of $\mathbb{R}^d$). As an easy consequence, we obtain that $\int_E F\,d\mu$ exists if and only if $\ell \circ F \in L_1(E, \mu)$ for each $\ell \in (\mathbb{R}^d)^*$.

Returning to our aim, our measure μ should be defined on a σ-algebra of subsets of C, for which the restriction of the identity function $x \mapsto x$ to C is measurable. The smallest such σ-algebra is the family

$$\mathcal{B}(C) = \{B \cap C : B \in \mathit{Borel}(\mathbb{R}^d)\}$$

where $\mathit{Borel}(\mathbb{R}^d)$ is the Borel σ-algebra of $\mathbb{R}^d$ (that is, the σ-algebra generated by the family of all open sets). Thus we shall consider the following family of measures:

$$\mathcal{M}_1(C) = \big\{\mu\colon \mathcal{B}(C) \to [0, +\infty) : \ \mu \text{ measure}, \ \mu(C) = 1\big\}\,.$$

The *barycenter* x_μ of a measure $\mu \in \mathcal{M}_1(C)$ is defined by

$$x_\mu = \int_C x\,d\mu(x)$$

(if the vector integral exists). Notice that by (3) we have $\ell(x_\mu) = \int_C \ell\, d\mu$ for each $\ell \in (\mathbb{R}^d)^*$.

Observation 0.1. *Let $C \subset \mathbb{R}^d$ be a nonempty convex set, $\mu \in \mathcal{M}_1(C)$.*

(a) *x_μ exists $\iff \|\cdot\| \in L_1(\mu) \iff \ell \in L_1(\mu)$ for each $\ell \in (\mathbb{R}^d)^*$.* (This follows easily from the fact that x_μ exists if and only if $\int_C |x(i)|\, d\mu(x) < +\infty$ for each $i = 1, \ldots, d$.)

(b) *If μ is concentrated on a bounded subset of C (in particular, if C is bounded), then x_μ exists.*

Proposition 0.2. *Let $C \subset \mathbb{R}^d$ be a nonempty convex set, $\mu \in \mathcal{M}_1(C)$. If the barycenter x_μ exists, then $x_\mu \in C$.*

Proof. Let us proceed by induction with respect to $n := \dim(C)$. For $n = 0$, we have $C = \{x_0\}$ and hence $\mu = \delta_{x_0}$, $x_\mu = x_0 \in C$.

Now, fix $0 \leq m < d$ and suppose that the statement holds whenever $n \leq m$. Now, let $\dim(C) = m+1$ and $x_\mu \notin C$. By the Relative Interior Theorem, $\mathrm{ri}(C) \neq \emptyset$, and hence we can suppose that $0 \in \mathrm{ri}(C)$. In this case, $L := \mathrm{span}(C) = \mathrm{aff}(C)$ and $0 \in \mathrm{int}_L(C)$. Let us consider two cases.

(a) If $x_\mu \notin L$, there exists $\ell \in (\mathbb{R}^d)^*$ such that $\ell|_L \equiv 0$ and $\ell(x_\mu) > 0$.

(b) If $x_\mu \in L$, there exists $\ell \in L^* \setminus \{0\}$ such that $\ell(x_\mu) \geq \sup \ell(C)$ (by the H-B Separation Theorem), and this ℓ can be extended to an element (denoted again by ℓ) of $(\mathbb{R}^d)^*$.

In both cases, we have

$$\int_C \big[\ell(x_\mu) - \ell(x)\big]\, d\mu(x) = \ell(x_\mu) - \int_C \ell\, d\mu = \ell(x_\mu) - \ell\left(\int_C x\, d\mu(x)\right) = 0\,.$$

Since the expression in square brackets is nonnegative, it must be μ-a.e. null. This implies that necessarily $x_\mu \in L$. But in this case, ℓ is not identically zero on L. Consequently, μ is concentrated on the set $C_1 := C \cap H$ where $H = \{x \in L : \ell(x) = \ell(x_\mu)\}$ is a hyperplane in L. Thus $\dim(C_1) \leq m$, $\mu_1 := \mu|_{C_1} \in \mathcal{M}_1(C_1)$ and, by the induction assumption, $x_\mu = x_{\mu_1} \in C_1 \subset C$. This contradiction completes the proof. ☐

Let us state an interesting corollary to the above proposition. By an *infinite convex combination* of elements of C we mean any point x_0 of the form

$$x_0 = \sum_{i=1}^{+\infty} \lambda_i c_i$$

where $c_i \in C$, $\lambda_i \geq 0$, $\sum_{j=1}^{+\infty} \lambda_j = 1$.

Corollary 0.3. *Let C be a finite-dimensional convex set in a Hausdorff t.v.s. X. Then each infinite convex combination of elements of C belongs to C.*

Proof. Let $x_0 = \sum_{i=1}^{+\infty} \lambda_i c_i$ be an infinite convex combination of elements of C. Of course, we can restrict ourselves to the subspace $Y = \mathrm{span}(C \cup \{x_0\})$. Since Y, being a finite-dimensional t.v.s., is isomorphic to $\mathbb{R}^d$, $d = \dim(Y)$, we can suppose that $X = \mathbb{R}^d$. We can also suppose that $c_i \neq c_j$ whenever $i \neq j$. Consider the measure μ concentrated on $\{c_i : i \in \mathbb{N}\}$, given by $\mu(c_i) = \lambda_i$. Since $x_\mu = x_0$, we can apply Proposition 0.2. $\square$

Before proving the Jensen inequality, we shall need the following separation lemma.

Lemma 0.4. *Let X be a locally convex t.v.s., $C \subset X$ a convex set, $f\colon C \to (-\infty, +\infty]$ a l.s.c. convex function, $x_0 \in \mathrm{dom}(f)$, and $t_0 < f(x_0)$. Then there exists a continuous affine function $a\colon X \to \mathbb{R}$ such that $a < f$ on C, and $a(x_0) > t_0$.*

Proof. Fix $\alpha \in \mathbb{R}$ such that $t_0 < \alpha < f(x_0)$. By lower semicontinuity, there exists a convex neighborhood V of x_0 such that $\alpha < f(x)$ for each $x \in V \cap C$. Extend f to the whole X defining $f(x) = +\infty$ for $x \notin C$, and consider the concave function

$$g(x) = \begin{cases} \alpha & \text{if } x \in V \\ -\infty & \text{otherwise} \end{cases}.$$

Since $g \leq f$ and g is continuous at x_0, we can use a H-B Theorem on Separation of Functions to get a continuous affine function a_1 on X such that $g \leq a_1 \leq f$. Now, $t_0 < \alpha \leq a_1(x_0) \leq f(x_0)$. Thus, for a sufficiently small $\varepsilon > 0$, the function $a := a_1 - \varepsilon$ has the required properties. $\square$

Theorem 0.5 (Jensen inequality)**.** *Let $C \subset \mathbb{R}^d$ be a nonempty convex set, $f\colon C \to (-\infty, +\infty]$ a convex l.s.c. function, and $\mu \in \mathcal{M}_1(C)$. Assume that the baricenter x_μ exists (which is automatically true for a bounded C). Then $x_\mu \in C$ and*

$$f(x_\mu) \leq \int_C f\, d\mu \tag{4}$$

(in particular, the integral on the right-hand side exists).

Proof. We already know that $x_\mu \in C$ (Proposition 0.2). If $f \equiv +\infty$, the assertion is obvious. Now, suppose that f is proper. Notice that f is $\mathcal{B}(C)$-measurable since the sublevel sets $\{x \in C : f(x) \leq \alpha\}$ are relatively closed in C. We claim that the right-hand-side integral in (4) exists. Indeed, by Lemma 0.4, there exists a continuous affine function $a\colon \mathbb{R}^d \to \mathbb{R}$ such that $a < f$ on C. Write a in the form $a(x) = \ell(x) + \beta$ where $\ell \in (\mathbb{R}^d)^*$, $\beta \in \mathbb{R}$. By Observation 0.1, $a \in L_1(\mu)$, which implies that $f^- \in L_1(\mu)$.

The effective domain $\mathrm{dom}(f) = f^{-1}(\mathbb{R})$ is convex and belongs to $\mathcal{B}(C)$. If $\mu(C \setminus \mathrm{dom}(f)) > 0$, then obviously $\int_C f\, d\mu = +\infty$ and (4) trivially holds.

Let $\mu(C \setminus \mathrm{dom}(f)) = 0$. In this case, μ is concentrated on $\mathrm{dom}(f)$, and hence, by Proposition 0.2, $x_\mu \in \mathrm{dom}(f)$. Assume that (4) is false, that is $t_0 := \int_C f\, d\mu < f(x_\mu)$. By Lemma 0.4, there exist $\ell \in (\mathbb{R}^d)^*$ and $\beta \in \mathbb{R}$ such that $\ell + \beta < f$ on C, and $\ell(x_\mu) + \beta > t_0$. But these two properties are in contradiction:

$$t_0 = \int_C f\, d\mu > \int_C (\ell + \beta)\, d\mu = \int_C \ell\, d\mu + \beta = \ell(x_\mu) + \beta > t_0 \,.$$

$\square$

Corollary 0.6 (Hermite-Hadamard inequalities)**.** *Let $f\colon [a,b]\to\mathbb{R}$ be a continuous convex function. Then*

$$f\left(\frac{a+b}{2}\right)\le \frac{1}{b-a}\int_a^b f(x)\,dx \le \frac{f(a)+f(b)}{2}\,.$$

Proof. The measure μ, defined by $d\mu=\frac{dx}{b-a}$, belongs to $\mathcal{M}_1([a,b])$. Moreover, its barycenter is $x_\mu=\frac{1}{b-a}\int_a^b x\,dx=\frac{a+b}{2}$. Thus the first inequality follows from the integral Jensen inequality.

Let us show the second inequality. Substituting $x=a+t(b-a)$, we get
$\frac{1}{b-a}\int_a^b f(x)\,dx=\int_0^1 f((1-t)a+tb)\,dt\le\int_0^1\big[(1-t)f(a)+tf(b)\big]\,dt=\frac{f(a)+f(b)}{2}\,.$ □

Image of a probability measure. Let (Ω,Σ,μ) be a probability space, and (T,τ) a topological space. Let $g\colon\Omega\to T$ be a measurable mapping, that is $g^{-1}(A)\in\Sigma$ whenever $A\subset T$ is an τ-open set. Let *Borel*(τ) denote the σ-algebra of all Borel sets in T. Since $g^{-1}(B)\in\Sigma$ for each $B\in$ *Borel*(τ), we can consider the function

$$\nu\colon \mathit{Borel}(\tau)\to[0,1]\,,\quad \nu(B)=\mu(g^{-1}(B))\,.$$

It is easy to verify that ν is a (borelian) probability measure on T, called the *image* of μ by the map g. Moreover, ν is *concentrated on the image* $g(\Omega)$, in the sense that $\nu(B)=0$ whenever $B\in$ *Borel*(τ) is disjoint from $g(\Omega)$.

Let us see how to integrate with respect to ν. Let $s=\sum_1^n\alpha_i\chi_{B_i}$ be a nonnegative simple function on T such that the sets B_i are borelian and pairwise disjoint. Then the composition $s\circ g$ is a measurable simple function on Ω and it can be represented as $s\circ g=\sum_1^n\alpha_i\chi_{g^{-1}(B_i)}$. Thus we have

$$\int_T s\,d\nu=\sum_1^n\alpha_i\nu(B_i)=\sum_1^n\alpha_i\mu(g^{-1}(B_i))=\int_\Omega s\circ g\,d\mu\,.$$

Now, if f is a nonnegative Borel-measurable function on T, we can approximate it by a pointwise converging nondecresing sequence of simple Borel-measurable functions. Passing to limits in the above formula, we get

$$\int_T f\,d\nu=\int_\Omega f\circ g\,d\mu\,. \tag{5}$$

It follows that, for an arbitrary Borel-measurable function f on T, we have:

- $\int_T f\,d\nu$ exists if and only if $\int_\Omega f\circ g\,d\mu$ exists; in this case, the two integrals are equal;
- $f\in L_1(\nu)$ if and only if $f\circ g\in L_1(\mu)$.

Let us return to the Jensen inequality. We can apply it to an image measure to obtain the following

Theorem 0.7 (Second Jensen inequality)**.** *Let (Ω,Σ,μ) be a probability measure space, and $g\colon\Omega\to\mathbb{R}^d$ a measurable mapping that is μ-integrable. Let $C\subset\mathbb{R}^d$ be a convex set such that $g(\omega)\in C$ for μ-a.e. $\omega\in\Omega$, and $f\colon C\to(-\infty,+\infty]$ a l.s.c. convex function. Then:*

- $\int_\Omega g\,d\mu\in C$;

- *$\int_\Omega f\circ g\,d\mu$ exists;*
- *we have the inequality*

$$f\left(\int_\Omega g\,d\mu\right) \le \int_\Omega f\circ g\,d\mu\,.$$

Proof. Let ν be the image of the measure μ by the mapping $g\colon \Omega\to C$. Then ν is a probability measure on C, defined on the relative Borel σ-algebra of $\mathcal{B}(C)$ C. Its barycenter is $x_\nu = \int_C x\,d\nu = \int_\Omega g\,d\mu$. The rest follows directly from Theorem 0.5 since $\int_C f\,d\nu = \int_\Omega f\circ g\,d\mu$. □

Corollary 0.8 (Examples of applications)**.** *Let (Ω,Σ,μ) be a probability measure space, and $g\in L_1(\Omega)$.*

(a) *Second Jensen inequality for $C=\mathbb{R}$ and $f(x)=|x|^p$ $(p\ge 1)$ gives*

$$\left|\int_\Omega g\,d\mu\right| \le \left(\int_\Omega |g|^p\,d\mu\right)^{1/p}.$$

(b) *Second Jensen inequality for $C=\mathbb{R}$ and $f(x)=e^x=\exp(x)$ gives*

$$\exp\left(\int_\Omega g\,d\mu\right) \le \int_\Omega e^g\,d\mu\,.$$

(c) *Second Jensen inequality for $C=(0,+\infty)$ and $f(x)=-\log x$ gives*

$$\int_\Omega \log g\,d\mu \le \log\left(\int_\Omega g\,d\mu\right).$$

Hölder via Jensen. Let us show how the well-known Hölder inequality can be derived from the Second Jensen inequality.

Let (Ω,Σ,μ) be a (not necessarily probability) nonnegative measure space. Let $p,p'\in(1,+\infty)$ be two conjugate Hölder exponents (that is, $\frac{1}{p}+\frac{1}{p'}=1$). Let f,g be nonnegative measurable functions on Ω, such that their norms in $L_p(\mu)$ and $L_{p'}(\mu)$ satisfy $0<\|f\|_p<+\infty$ and $0<\|g\|_{p'}<+\infty$. Consider the set $\Omega_1=\{g>0\}$ $(\in\Sigma)$ and the measure ν on Σ, defined by

$$d\nu = \frac{1}{\int_\Omega g^{p'}\,d\mu}\,g^{p'}\,d\mu\,.$$

Then ν is a probability measure which is concentrated on Ω_1. Applying Corollary 0.8(a) to the function $fg^{1-p'}$ (defined on the set Ω_1), we get

$$\begin{aligned}
\frac{1}{\int_\Omega g^{p'}\,d\mu}\cdot\int_\Omega fg\,d\mu &= \int_{\Omega_1} fg^{1-p'}\,d\nu\\
&\le \left(\int_{\Omega_1} f^p g^{p(1-p')}\,d\nu\right)^{1/p}\\
&= \left(\frac{1}{\int_\Omega g^{p'}\,d\mu}\cdot\int_\Omega f^p g^{p+p'-pp'}\,d\mu\right)^{1/p}\\
&= \frac{1}{\left(\int_\Omega g^{p'}\,d\mu\right)^{1/p}}\cdot\left(\int_\Omega f^p\,d\mu\right)^{1/p}
\end{aligned}$$

since $p + p' - pp' = pp'(\frac{1}{p'} + \frac{1}{p} - 1) = 0$. Multiplying by $\int_\Omega g^{p'}\, d\mu$, we obtain the Hölder inequality

$$\int_\Omega fg\, d\mu \le \|f\|_p \, \|g\|_{p'} \,.$$

Finally, notice that if our assumption on the L_p-norm of f and the $L_{p'}$-norm of g is not satisfied, the Hölder inequality is trivially satisfied (with the usual convention "$0 \cdot \infty = 0$").

A generalization of the Hermite-Hadamard inequalities. Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^d$, $B \subset \mathbb{R}^d$ a closed $\|\cdot\|$-ball centered at a point $x_0 \in \mathbb{R}^d$, and $S = \partial B$. Let m denote the Lebesgue measure in $\mathbb{R}^d$ and σ the surface measure. Then, for each continuous convex function $f\colon B \to \mathbb{R}$, we have the inequalities

$$f(x_0) \le \frac{1}{m(B)} \int_B f\, dm \le \frac{1}{\sigma(S)} \int_S f\, d\sigma \,. \tag{6}$$

Proof. By translation, we can suppose that $x_0 = 0$. For the probability measure $d\mu = \frac{dm}{m(B)}$, we have $x_\mu = 0$ by symmetry, and hence the Jensen inequality implies the first inequality in (6). Let us show the second inequality. Notice that

$$m(B) = \int_0^1 \sigma(rS)\, dr = \sigma(S) \int_0^1 r^{d-1}\, dr = \frac{1}{d}\, \sigma(S) \,.$$

Now, a similar "onion-skin integration" and convexity of f imply

$$\begin{aligned}
\int_B f\, dm &= \int_0^1 \left(\int_{rS} f\, d\sigma \right) dr \\
&= \int_0^1 \left(\int_{rS} f\big(\tfrac{1+r}{2}(\tfrac{x}{r}) + \tfrac{1-r}{2}(-\tfrac{x}{r})\big)\, d\sigma(x) \right) dr \\
&\le \int_0^1 \left(\int_{rS} \big[\tfrac{1+r}{2} f(\tfrac{x}{r}) + \tfrac{1-r}{2} f(-\tfrac{x}{r})\big]\, d\sigma(x) \right) dr \\
&= \left(\int_0^1 r^{d-1}\, dr \right) \int_S f\, d\sigma = \frac{m(B)}{\sigma(S)} \int_S f\, d\sigma \,.
\end{aligned}$$

The second inequality in (6) follows by dividing by $m(B)$. □

Possible generalizations. Some of the above results, except Proposition 0.2, can be easily generalized to the infinite-dimensional setting. The main problem is that we have to introduce appropriately the barycenter of a probability measure. The most general is the following *Pettis-integral* approach which defines the integral of a vector-valued function by requiring the equality (3) for every continuous linear functional ℓ.

Let X be a t.v.s., and $E \subset X$ a nonempty set. We shall say that a point $x_\mu \in X$ is a *barycenter* for a probability measure $\mu \in \mathcal{M}_1(E)$ (defined on the relative Borel σ-algebra $\mathcal{B}(E)$) if

$$y^*(x_\mu) = \textstyle\int_E y^*\, d\mu \qquad \text{for each } y^* \in X^*.$$

Then we have the following results. *Let X be a locally convex t.v.s., $C \subset X$ a nonempty convex set, and $\mu \in \mathcal{M}_1(C)$.*

(i) *μ has at most one barycenter x_μ.* (This follows from the fact that X^* separates the points of X.)

(ii) *If x_μ exists and C is either closed or open, then $x_\mu \in C$.* (The proof by contradiction uses the H-B Separation Theorem, in the same spirit as in Proposition 0.2.)

(iii) *If C is compact, then x_μ exists.* (This is a bit more difficult.)

(iv) *Theorem 0.5 (Jensen inequality) holds with X in place of $\mathbb{R}^d$, under the additional assumption that C is either closed or open, and f is finite.* (The proof is the same.)