

PROBABILITY THEORY

MATH 154

Unit 11: Jensen-Hölder-Minkowski

11.1. A continuous function $h : \mathbb{R} \rightarrow \mathbb{R}$ is called **convex**, if there exists for all $x_0 \in \mathbb{R}$ a linear map $l(x) = ax + b$ such that $l(x_0) = h(x_0)$ and for all $x \in \mathbb{R}$ the inequality $l(x) \leq h(x)$ holds for all x .

11.2. Examples:

a) A linear function $h(x) = ax + b$ is convex, b) $h(x) = x^2$ is convex, c) $h(x) = e^x$ is convex, d) $h(x) = |x|$ is convex, e) $h(x) = -x^2$ is not convex, f) $h(x) = -\log(x)$ is convex on $(0, \infty)$.

Theorem 1 (Jensen inequality). For $X \in \mathcal{L}^1$ and h convex, $E[h(X)] \geq h(E[X])$.

Proof. Fix $x_0 = E[X]$. By definition of convexity, there is a linear function $l(x)$, producing a lower bound for h at x_0 meaning $l(x) \leq h(x)$. By the linearity and monotonicity of expectation, we get $h(E[X]) = l(E[X]) = E[l(X)] \leq E[h(X)]$. If $h(X)$ should not be in $\mathcal{L}$, the statement still holds with the understanding $E[h(X)] = \infty$. $\square$

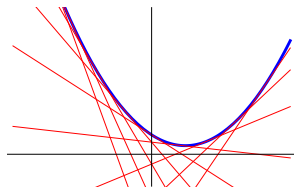


FIGURE 1. A convex function.

11.3. For $h(x) = x^2$, this gives $E[X^2] \geq E[X]^2$ which is equivalent to $\text{Var}[X] = E[X^2] - E[X]^2 \geq 0$. That the variance is non-negative was already clear from $\text{Var}[X] = E[(X - E[X])^2]$.

11.4. For $h(x) = -\log(x)$ and a two point probability space $\Omega = \{0, 1\}$ with a random variable X satisfying $P[X = \{0\}] = x$, $P[X = \{1\}] = y$ and $P[A] = |A|/2$, we get the **geometric-arithmetic inequality** $\sqrt{xy} \leq (x + y)/2$.

11.5. Given $p \leq q$. Define $h(x) = |x|^{q/p}$. It is convex. Jensen's inequality gives $E[|X|^q] = E[h(|X|^p)] \geq h(E[|X|^p]) = E[|X|^p]^{q/p}$. This implies that $\|X\|_q := E[|X|^q]^{1/q} \geq E[|X|^p]^{1/p} = \|X\|_p$ for $p \leq q$ and so

Theorem 2 (L^p stratification).

$$\mathcal{L}^\infty \subset \mathcal{L}^q \subset \mathcal{L}^p \subset \mathcal{L}^1$$

for $p \leq q$.

The smallest space is $\mathcal{L}^\infty$ which is the space of all bounded random variables.

11.6. The stratification result follows from the following inequality:

Theorem 3 (Hölder inequality). *Given $p, q \in [1, \infty]$ with $p^{-1} + q^{-1} = 1$ and $X \in \mathcal{L}^p$ and $Y \in \mathcal{L}^q$. Then $XY \in \mathcal{L}^1$ and $\|XY\|_1 \leq \|X\|_p \|Y\|_q$.*

Proof. The random variables X, Y are defined over a probability space $(\Omega, \mathcal{A}, P)$. The identity $p^{-1} + q^{-1} = 1$ is equivalent to $q + p = pq$ or $q(p - 1) = p$.

Without loss of generality we can restrict us to $X, Y \geq 0$ because replacing X with $|X|$ and Y with $|Y|$ does not change anything. We can also assume $\|X\|_p > 0$ because otherwise $X = 0$, where both sides are zero. We can write therefore X instead of $|X|$ and assume X is not zero. The key idea of the proof is to introduce a new probability measure

$$Q = \frac{X^p P}{E[X^p]}.$$

If $P[A] = \int_A 1 dP(x)$ then $Q[A] = [\int_A X^p(x) dP(x)] / E[X^p]$ so that $Q[\Omega] = E[X^p] / E[X^p] = 1$ and Q is a probability measure. Let E_Q denote the expectation with respect to this new measure. Define the new random variable $U = 1_{\{X > 0\}} Y / X^{p-1}$. Jensen's inequality applied to the convex function $h(x) = x^q$ gives

$$(1) \quad E_Q[U]^q \leq E_Q[U^q].$$

Using $E_Q[U] = E_Q[\frac{Y}{X^{p-1}}] = \frac{E[XY]}{E[X^p]}$ and $E_Q[U^q] = E_Q[\frac{Y^q}{X^{q(p-1)}}] = E_Q[\frac{Y^q}{X^p}] = \frac{E[Y^q]}{E[X^p]}$, Equation (1) can be rewritten as $\frac{E[XY]^q}{E[X^p]^q} \leq \frac{E[Y^q]}{E[X^p]}$ which implies

$$E[XY] \leq E[Y^q]^{1/q} E[X^p]^{1-1/q} = E[Y^q]^{1/q} E[X^p]^{1/p}.$$

This is equivalent to $\|XY\|_1 \leq \|X\|_p \|Y\|_q$. □

11.7. A special case of the Hölder's inequality is the **Cauchy-Schwarz** inequality

$$\|XY\|_1 \leq \|X\|_2 \cdot \|Y\|_2.$$

The semi-norm property of $\mathcal{L}^p$ follows from the following fact:

Theorem 4 (Minkowski inequality (1896)). *Given $p \in [1, \infty]$ and $X, Y \in \mathcal{L}^p$. Then*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

Proof. We use Hölder's inequality from below to get

$$E[|X + Y|^p] \leq E[|X| |X + Y|^{p-1}] + E[|Y| |X + Y|^{p-1}] \leq \|X\|_p C + \|Y\|_p C,$$

where $C = \| |X + Y|^{p-1} \|_q = E[|X + Y|^p]^{1/q}$ which leads to the claim. □