Information and entropy

The information function

Let ξ be a countable partition of a probability space $(X, \mathcal{F})$. The information function $I_\xi : X \to \mathbb{R}$ is

$$I_\xi(x) = -\log \mu(\xi(x))$$

Given a measurable partition η, the conditional information of ξ given η is the function $I_{\xi|\eta} : X \to \mathbb{R}$ given by

$$I_{\xi|\eta}(x) = -\log \mu_x^\eta(\xi(x))$$

Note that when $\eta(x) > 0$, we have $\mu_x^\eta(A) = \mu(A \cap \eta(x))/\mu(\eta(x))$, hence

$$I_{\xi|\eta}(x) = -\log \frac{\mu(\xi(x) \cap \eta(x))}{\mu(\eta(x))}$$

Also observe that $\mu_x^{\xi \vee \eta}(A) = \mu_x^\eta(A \cap \xi(x))/\mu_x^\eta(\xi(x))$

Lemma 1. .

1. *$I_{\xi|\eta} = 0$ a.e. if and only if $\eta \succeq \xi \bmod \mu$.*
2. *$I_{\xi|\eta} = I_\xi$ a.e. if and only if $\xi \perp \eta$.*
3. *$I_{\xi \vee \xi' | \eta} = I_{\xi|\eta} + I_{\xi'|\eta \vee \xi}$ a.e.*

Theorem 2. *Let ξ be a countable partition. If $\eta_1 \preceq {}_2\eta \preceq \ldots$ are measurable partitions and $\eta_\infty = \bigvee \eta_n$, or if $\eta_1 \succeq \eta_2 \succeq \eta_3 \ldots$ and $\eta_\infty = \bigwedge \eta_n$, then*

$$I(\xi|\eta_n) \to I(\varepsilon|\eta_\infty) \qquad \mu\text{-a.e.}$$

Proof. Immediate from $\mu_x^{\eta_n} \to \mu_x^{\eta_\infty}$ (w.r.t. the algebra of test functions 1_A, $A \in \xi$). □

Entropy

The *entropy* of a countable partition ξ is

$$H_\mu(\xi) = \int I_\xi d\mu = -\sum_{A \in \xi} \mu(A) \log \mu(A)$$

The *conditional entropy* with respect to a measurable partition η is

$$\begin{aligned} H_\mu(\xi|\eta) &= \int I_{\xi|\eta} d\mu \\ &= \int \left(\int I_{\xi|\eta} d\mu_x^\eta \right) d\mu(x) \\ &= -\int \left(\sum_{A \in \xi} \mu_x^\eta(A) \log \mu_x^\eta(A) \right) d\mu(x) \\ &= \int H_{\mu_x^\eta}(\xi) d\mu(x) \end{aligned}$$

Lemma 3. .

1. $0 \leq H(\xi), H(\xi|\eta) \leq \infty$.
2. $H(\xi) = 0$ *if and only if* $\xi = \{X\} \bmod \mu$, *and* $H(\xi|\eta) = 0$ *if and only if* $\eta \succeq \xi$
3. $H(\xi \vee \xi'|\eta) = H(\xi'|\eta) + H(\xi|\eta \vee \xi')$

Proof. Exercise ((3) is proved by integrating the corresponding formula for information). □

Lemma 4. .

1. *If* $\eta_1 \preceq \eta_2$ *are measurable partitions and* ξ *is countable then* $H(\xi|\eta_1) \geq H(\xi|\eta_2)$. *Equality if and only if* $\mu_y^{\eta_2}(A_i)$ *is constant* $\mu_x^{\eta_1}$*-a.e.* y*, hence equal to* $\mu_x^{\xi_1}(A_i)$ *(*ξ *is conditionally independent of* η_2 *given* η_1*).*
2. $H(\xi|\eta) \leq H(\xi)$ *with equality if and only if* $\xi, \mathcal{B}_\eta$ *are independent.*
3. *If* $\xi = \{A_1, \ldots, A_k\}$ *then* $H(\xi) \leq \log k$ *with equality if and only if* $\mu(A_i) = 1/k$.

Proof. These are conseqsuences of (strict) convexity of $u(t) = t \log t$. Since $\eta_1 \preceq \eta_2$ we have $\mu_x^{\eta_1} = \int \mu_y^{\eta_2} d\mu_x^{\eta_1}(y)$ a.s. (this can be verified by integrating functions against both measures and getting the same answer). Therefore by convexity, for every $A \in \xi$ we have $u(\mu_x^{\eta_1}(A)) \leq \int u(\mu_y^{\eta_2}(A)) d\mu_x^{\eta_1}(y)$. Summing over $A \in \xi$ this gives

$$H_{\mu_x^{\eta_1}}(\xi) \geq \int H_{\mu_y^{\eta_2}}(\xi) d\mu_x^{\eta_1}(y)$$

Integrating over x gives (1). The last part of (1) and also (2) follow by strict convexity. For (3) note that

$$-\frac{1}{k}\log k = u(\frac{1}{k}) = u(\sum \frac{1}{k}\mu(A_i)) \leq \sum \frac{1}{k} u(\mu(A_i)) = -\frac{1}{k} H(\xi)$$

Equality holds if and only if all $\mu(A_i)$ are equal. □

Lemma 5. *If* $\eta_1 \preceq \eta_2 \preceq \ldots$ *are measurable partitions,* ξ *a countable partition, and* $H(\xi|\eta_1) = \int I_{\xi|\eta_1} d\mu < \infty$*, then*

$$\int \sup_n I_{\xi|\eta_n} d\mu < \infty$$

Proof. See pp. 16-17 of Ledrappier's lecture slides. □

Proposition 6. .

1. *If* $\eta_1 \preceq \eta_2 \preceq \ldots$ *and* $\eta_\infty = \bigvee \eta_n$ *are measurable partitions,* ξ *a countable partition, and* $H(\xi|\eta_1) < \infty$*, then* $H(\xi|\eta_n) \searrow H(\xi, \eta_\infty)$.

2. *If* $\eta_1 \succeq \eta_2 \succeq \ldots$ *and* $\eta_\infty = \bigwedge \eta_n$ *are measurable partitions,* ξ *a countable partition, then* $H(\xi|\eta_n) \nearrow H(\xi, \eta_\infty)$.

Proof. (1) We know that $I_{\xi|\eta_n} \to I_{\xi|\eta_\infty}$a.e. and the previous lemma allows us to integrate (dominated convergence). Montonicity by previous prop.

(2) We saw that $H_{m_x^{\eta_n}}(\xi) \geq \int H_{\mu_y^{\eta_{n+1}}}(\xi) d\mu_x^{\eta_n}(y)$. This means that $x \mapsto H_{\mu_x^{\eta_n}}(\xi)$ is a sub-martingale with respect to the decreasing sequnece of σ-algebras $\mathcal{B}_{\eta_n}$. By a version of the martingale theorem the sequence converges a.e. and in L^1 if it makes sense. □