ENTROPY

**Definition 0.1.** Fix a Polish space $X$. For any $\mu$ and $\nu$ in $\mathcal{P}(X)$, define

$$H(\nu|\mu) \overset{\text{def}}{=} \begin{cases} \int_{z \in X} \ln \frac{d\nu}{d\mu}(z) \nu(dz) & \text{if } \nu \ll \mu \\ \infty & \text{else,} \end{cases}$$

where the integral is interpreted in the sense of Lebesgue. Some introductory comments are in order.

**Lemma 0.2.** For any Polish space $X$ and any $\nu$ and $\mu$ in $\mathcal{P}(X)$, $H(\nu|\mu)$ is well-defined and nonnegative. We have that $H(\nu|\mu) = 0$ if and only if $\nu = \mu$.

**Proof.** Define $f(x) \overset{\text{def}}{=} x \ln x$ for all $x > 0$, and define $f(0) = 0$; it is easy to see that $f$ is convex and continuous on $[0, \infty)$.

We first claim that when $\nu \ll \mu$, $H(\nu|\mu)$ is well-defined. To do this, we will show that

$$\int_{z \in A} \left\{ - \ln \frac{d\nu}{d\mu}(z) \right\} \nu(dz) < \infty,$$

where

$$A \overset{\text{def}}{=} \left\{ z \in X : 0 \leq \frac{d\nu}{d\mu}(z) \leq 1 \right\}.$$

If $\nu(A) = 0$, then this integral is zero (by standard construction of Lebesgue integrals as limits of integrals of increasing simple functions). If $\nu(A) > 0$, then $\mu(A) > 0$. The convexity of $f$ implies that $-f$ is concave, so

$$\int_{z \in A} \left\{ - \ln \frac{d\nu}{d\mu}(z) \right\} \nu(dz) = \int_{z \in A} \left\{ - f \left( \frac{d\nu}{d\mu}(z) \right) \right\} \mu(dz) \leq -f \left( \int_{z \in A} \frac{\frac{d\nu}{d\mu}(z) \mu(dz)}{\mu(A)} \right) \mu(A) \leq -f \left( \frac{\nu(A)}{\mu(A)} \right) \mu(A) < \infty,$$

proving that $H(\nu|\mu)$ is well-defined. To show that it is positive, we again use the convexity of $f$ to see that

$$\int_{z \in X} \ln \frac{d\nu}{d\mu}(z) \nu(dz) = \int_{z \in X} f \left( \frac{d\mu}{d\nu}(z) \right) \mu(dz) \geq f \left( \int_{z \in X} \frac{d\mu}{d\nu}(z) \mu(dz) \right) = f(1) = 0.$$

Since $f$ is strictly convex, equality holds if and only if $\frac{d\nu}{d\mu}$ is constant, in which case $\frac{d\nu}{d\mu} \equiv 1$, and thus $H(\nu|\mu) = 0$. \hfill \square

**Theorem 0.3** (Entropy Duality). Fix a Polish space $X$. For any $\nu$ and $\mu$ in $\mathcal{P}(X)$, we have that

(1) $$H(\nu|\mu) = \sup_{\phi \in \mathcal{B}(X)} \left\{ \int_{z \in X} \phi(z) \nu(dz) - \log \int_{z \in X} e^{\phi(z)} \mu(dz) \right\}$$

and for any $\phi \in \mathcal{B}(X)$, we have that

(2) $$\log \int_{z \in X} e^{\phi(z)} \mu(dz) = \sup_{\nu \in \mathcal{P}(X)} \left\{ \int_{z \in X} \phi(z) \nu(dz) - H(\nu|\mu) \right\}.$$

**Proof.** First note that if $\nu \in \mathcal{P}(X)$ is such that $H(\nu|\mu) < \infty$ and $\phi \in \mathcal{B}(X)$, then by Jensen’s inequality

$$\int_{z \in X} \phi(z) \nu(dz) - H(\nu|\mu) = \int_{z \in X} \left\{ \phi(z) - \ln \frac{d\nu}{d\mu}(z) \right\} \nu(dz)$$

$$= \int_{z \in X} \left\{ \frac{d\nu}{d\mu}(z) \phi(z) \nu(dz) - \ln \frac{d\nu}{d\mu}(z) \nu(dz) \right\} \leq \int_{z \in X} e^{\phi(z)} \frac{d\nu}{d\mu}(z) \nu(dz) \leq \ln \int_{z \in X} e^{\phi(z)} \mu(dz)$$


(note that $\nu \left\{ z \in X : \frac{d\nu}{d\mu}(z) > 0 \right\} = 1$). Thus the left-hand sides of (1) and (2) are greater than or equal to the right-hand sides.

Let's next show that the left-hand side of (1) is less than the right-hand side. Define $\nu \in \mathcal{B}(X)$ via

$$
\nu(A) \overset{\text{def}}{=} \int_{x \in A} e^{f(x)} \mu(dx) \quad A \in \mathcal{B}(X)
$$

Then

$$
\int_{x \in X} \phi(x) \nu(dx) = \sum_{x \in X} \phi(x) \nu(x) = \int_{x \in X} \phi(x) \mu(dx) = H(\nu|\mu).
$$

Thus the left-hand side of (2) is less than or equal to the right-hand side.

The final step is to show that the left-hand side of (1) is less than or equal to the right-hand side. To do so, first assume that $\nu \ll \mu$. We will need to approximate. Define

$$
\phi_n(x) \overset{\text{def}}{=} \begin{cases} 
\ln \frac{d\nu}{d\mu}(x) & \text{if } |\ln \frac{d\nu}{d\mu}(x)| < N \\
N & \text{if } \ln \frac{d\nu}{d\mu}(x) > N \\
-N & \text{if } \ln \frac{d\nu}{d\mu}(x) < -N.
\end{cases}
$$

Then we have that

$$
\lim_n \int_{x \in X} \phi_n(x) \nu(dx) = H(\nu|\mu).
$$

We also note that

$$
0 \leq e^{\phi_n(x)} \leq 1 + \frac{d\nu}{d\mu}(x);
$$

thus by dominated convergence,

$$
\lim \ln \int_{x \in X} e^{\phi_n(x)} \mu(dx) = 0.
$$

Thus

$$
H(\nu|\mu) = \lim_n \left\{ \int_{x \in X} \phi_n(x) \nu(dx) - \ln \int_{x \in X} e^{\phi_n(x)} \mu(dx) \right\},
$$

which proves that the left-hand side of (2) is less than or equal to the right-hand side in this case (when $\nu \ll \mu$). Finally, assume that $\nu \not\ll \mu$. Then $H(\nu|\mu) = \infty$ and there is an $A \in \mathcal{B}(X)$ such that $\nu(A) > 0$. Define now $\phi_n \overset{\text{def}}{=} n\chi_A$ for all $n$. Then

$$
\lim_n \left\{ \int_{x \in X} \phi_n(x) \nu(dx) - \ln \int_{x \in X} e^{\phi_n(x)} \mu(dx) \right\} = \lim_n n\nu(A) = \infty.
$$

This finishes the proof. \qed