Lecture 4

Universal Probability

4.1 Mean and Pointwise Universality

Let $P$ be a collection of probability measures on $\mathcal{X}^\infty$. Each $P \in P$ specifies a random process $(X_n)_{n=1}^\infty$ with $n$-th order density (pmf/pdf) $p(x^n) = P(X^n = x^n)$. A sequence $(q(x^n))_{n=1}^\infty$ of densities is said to be mean universal with respect to $P$ if

$$\lim_{n \to \infty} \frac{1}{n} D(p(x^n) \| q(x^n)) = 0, \quad P \in P$$

(4.1)

When $P$ is the collection of stationary ergodic probability measures, then we simply say $(q(x^n))_{n=1}^\infty$ is mean universal (without any qualifier). If $(q(x^n))_{n=1}^\infty$ is consistent, namely,

$$\sum_{x \in \mathcal{X}} q(x^n) = q(x^{n-1}), \quad x^{n-1} \in \mathcal{X}^{n-1}, \quad n = 1, 2, \ldots,$$

then, by the Kolmogorov extension theorem, there exists a unique random process with probability measure $Q$ on $\mathcal{X}^\infty$ with $Q(X^n = x^n) = q(x^n)$. Hence, for a consistent $(q(x^n))$ its universality w.r.t. $P$ can be rewritten as

$$D(P \| Q) = 0, \quad P \in P.$$

Example 4.1. Let $P$ be finite, say, $P = \{P_1, P_2, \ldots, P_k\}$. Then

$$Q = \frac{1}{k} \sum_{i=1}^k P_i,$$

is mean universal w.r.t. $P$. To see this, consider

$$D(p_i(x^n) \| q(x^n)) = \sum_{x^n} p_i(x^n) \log \frac{p_i(x^n)}{q(x^n)}$$

$$= \sum_{x^n} p_i(x^n) \log \frac{p_i(x^n)}{\sum_{j=1}^k p_j(x^n)}$$

$$= \log k + \sum_{x^n} p_i(x^n) \log \frac{p_i(x^n)}{\sum_{j=1}^k p_j(x^n)}$$

$$\leq \log k.$$
Hence, for every $P_i \in \mathcal{P}$,
\[
\overline{D}(P_i \| Q) = \lim_{n \to \infty} \frac{1}{n} D(P_i(x^n) \| q(x^n)) \leq \lim_{n \to \infty} \frac{\log k}{n} = 0.
\]

Note that the convergence is uniform in $i$.

A sequence $(q(x^n))_{n=1}^\infty$ of densities is said to be pointwise universal w.r.t. $\mathcal{P}$ if
\[
\limsup_{n \to \infty} \frac{1}{n} \log \frac{p(X^n)}{q(X^n)} \leq 0 \quad P\text{-a.s.,} \quad P \in \mathcal{P}. \tag{4.2}
\]

Note that $(q(x^n))_{n=1}^\infty$ is mean universal w.r.t. $\mathcal{P}$ if
\[
\lim_{n \to \infty} E \left[ \frac{1}{n} \log \frac{p(X^n)}{q(X^n)} \right] = 0, \quad P \in \mathcal{P},
\]
where the expectation is w.r.t. $P$. Thus, mean universality denotes the $L_1$ convergence of $(1/n) \log(p(X^n)/q(X^n))$, while pointwise universality denotes its almost sure convergence. Consequently, neither universality nor pointwise universality implies the other. However, it is easy to construct a mean universal density (sequence) from a pointwise universal one (see Problem 4.1) and we typically regard pointwise universality as the stronger notion.

**Example 4.2.** Let $\mathcal{P}$ and $Q$ be defined as in Example 4.1. Then $Q$ is pointwise universal w.r.t. $\mathcal{P}$.

Suppose that the sequences $(q(x^n))_{n=1}^\infty$ consists of subprobability densities, that is, $q(x^n) \geq 0, x^n \in \mathcal{X}^n$, and $\sum q(x^n) < 1$ (for the discrete case) or $\int q(x^n) dx^n < 1$ (for the continuous case). If $(q(x^n))_{n=1}^\infty$ satisfies (4.2), then its normalized version
\[
q'(x^n) = \frac{q(x^n)}{\sum_{x^n} q(x^n)}
\]
is pointwise universal (see Problem 4.2). Hence, we extend the notion of pointwise universality from probability to subprobability and say $(q(x^n))_{n=1}^\infty$ is pointwise universal.

### 4.2 Lempel–Ziv Probability

A parsing of an infinite sequence $(x_n)_{n=1}^\infty \in \mathcal{X}^\infty$ is a sequence of finite-length phrases $y_1, y_2, \ldots \in \mathcal{X}^*$ such that For example, 010011... can be parsed into $(0, 1, 00, 11, \ldots)$, where $y_1 = 0, y_2 = 1, y_3 = 00$, and so on.
**PROBLEMS**

4.1. Let \((q(x^n))_{n=1}^{\infty}\) be pointwise universal w.r.t. \(\mathcal{P}\), and \(u(x^n) \equiv 1/|\mathcal{X}|^n\) be the uniform probability density. Show that

\[
q'(x^n) = \frac{n-1}{n} q(x^n) + \frac{1}{n} u(x^n), \quad n = 1, 2, \ldots,
\]

is both mean and pointwise universal w.r.t. \(\mathcal{P}\).

4.2. Let \((q(x^n))_{n=1}^{\infty}\) be a sequence of subprobability densities such that (4.2) holds. Show that its normalized version

\[
q'(x^n) = \frac{q(x^n)}{\sum_{x^n \in \mathcal{X}} q(x^n)}, \quad n = 1, 2, \ldots,
\]

is pointwise universal w.r.t. \(\mathcal{P}\).