Shannon Entropy, Diversity Measures, and Primitive Roots

Shannon Entropy and Diversity Measures

Shannon Entropy, introduced by Claude Shannon in the context of information theory, measures the level of uncertainty or randomness in a probability distribution. For a discrete distribution

$$P=\{p_1,p_2,\ldots,p_n\},$$

the Shannon entropy is defined as:

$$H(P) = -\sum_{i=1}^{n} p_i \log_b p_i,$$

where the base b of the logarithm determines the units of entropy (e.g., b=2 gives entropy in bits, while b= e gives entropy in nats). This formula expresses the *expected amount of information* or uncertainty associated with the outcome of a random variable.

Entropy and related measures are fundamental in **cryptography**, **data compression**, and **statistics**, because they quantify how unpredictable or diverse a dataset is.

Other Diversity Measures

Several generalizations and alternatives to Shannon entropy are widely used:

- **Rényi Entropy**:
  A generalization of Shannon entropy, defined as

$$H_\alpha(P) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} p_i^\alpha\right),$$

where $\alpha>0$, $\alpha\neq 1$.

As $\alpha\to 1$, Rényi entropy converges to Shannon entropy.

- **Tsallis Entropy**:
  Another generalization, expressed as

which reduces to Shannon entropy when $q \to 1$.
This measure is often applied in physics and complex systems.

- **Simpson's Index**:
  Commonly used in ecology, defined as
-

$$D = \sum_{i=1}^{n} p_i^2,$$

which represents the probability that two randomly chosen individuals belong to the same category. Its complement, $1 - D$, is also used as a diversity measure.

Primitive Roots

In number theory, a **primitive root modulo p** (where p is prime) is an integer g with the following property:

For every integer a that is coprime to p ($\gcd(a,p) = 1$), there exists an integer k such that

$$g^k \equiv a \pmod{p}.$$

In other words, g is a generator of the multiplicative group of integers modulo p, denoted $(\mathbb{Z}/p\mathbb{Z})^\times$.
This means that every nonzero residue modulo p can be expressed as some power of g.

Primitive roots always exist for prime moduli and play an essential role in **cryptography** (e.g., Diffie–Hellman key exchange, ElGamal encryption), since they allow the construction of secure cyclic groups used in modular arithmetic–based algorithms.

$$S_q(P) = \frac{1}{q-1}\left(1 - \sum_{i=1}^{n} p_i^q\right),$$