Differential Entropy

Basic Definitions

- The differential entropy is the entropy of a continuous random variable, say X. It is defined by

$$h(X) = -\int_S f_X(x) \log f_X(x) dx,$$

where $f(x)$ is the probability density function and S is the support set, $S = \{x | f(x) > 0\}$.

- The unit depends on the base of $\log$.
- We assume that $f(x)$ and the integral exist unless explicitly stated.

Examples

- Uniform distribution $U(0, a)$

$$f_U(x) = \begin{cases} \frac{1}{a}, & 0 \le x \le a, \\ 0 & \text{otherwise} \end{cases} \quad \to h(U) = \log a.$$

  - Note that $0 \log 0 = 0$.
  - Differential entropy can be negative.
  - $2^{h(X)} (= a)$ is the volume of support set.

- Normal distribution $N(\mu, \sigma^2)$

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad \to h(N) = \frac{1}{2} \log(2\pi e \sigma^2).$$

AEP for Continuous Random Variables

- (Theorem) Let $X_1, X_2, \ldots, X_n$ be a sequence of samples drawn i.i.d. from $f(x)$. Then

$$-\frac{1}{n}\log f(X_1, X_2, \ldots, X_n) \to E[-\log f(X)] = h(X).$$

- One can define the typical set of given ϵ and n by

$$A_\epsilon^{(n)} = \left\{(x_1, x_2, \ldots, x_n) : \left| -\frac{\log f(x_1, x_2, \ldots, x_n)}{n} - h(X) \right| \leq \epsilon \right\},$$

where $f(x_1, \ldots, x_n) = \prod_i f(x_i)$.

Properties of Typical Set

- Given ϵ, for n sufficiently large,

$$Pr(A_\epsilon^{(n)}) > 1 - \epsilon.$$

- Define the volume of a set to be $\text{Vol}(A) = \int_A dx_1 \dots dx_n$. Then the volume of a typical set is bounded by,

$$\text{Vol}(A_\epsilon^{(n)}) \leq 2^{n(h(X)+\epsilon)}.$$

- Given ϵ, for n sufficiently large,

$$\text{Vol}(A_\epsilon^{(n)}) \geq (1-\epsilon)2^{n(h(X)-\epsilon)}.$$

- Thus, the volume of a typical set is decided by the differential entropy.

Relation of Differential Entropy to Discrete Entropy

- Consider a random variable X with density $f(x)$.
- We can create a discrete (or quantized) version of X by the scheme described in the text.
- The differential entropy and the discrete entropy is related to the quantization level.

$$H(X^\Delta) + \log \Delta \to h(X).$$

- $h(X) + n$ is the number of bits on average to describe X to n-bit accuracy.

Joint and Conditional Differential Entropy

- The joint differential entropy is defined by

$$h(X_1, \ldots, X_n) = -\int f(x_{1:n}) \log f(x_{1:n}) dx_1 \ldots dx_n.$$

- The conditional differential entropy is defined by

$$h(X|Y) = -\int f(x, y) \log f(x|y) dx dy = -E \log f(X|Y).$$

- We have

$$h(X|Y) = h(X, Y) - h(Y).$$

Entropy of A Gaussian Random Vector

- Let $\mathbf{X} = (X_1, \ldots, X_n)$ have a multi-variate normal distribution, denoted by $N(\mu, \Sigma)$

$$f(\mathbf{x}) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)}.$$

- Then

$$h(\mathbf{X}) = \frac{1}{2} \log[(2\pi e)^n |\Sigma|].$$

- See the text for a straightforward proof.

Relative Entropy and Mutual Information

- The relative entropy between two densities f and g is defined by

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} dx.$$

- The mutual information between two continuous random variables X and Y is defined by

$$I(X;Y) = \int f(x,y) \log \frac{f(x,y)}{f(x)f(y)} dxdy.$$

- $I(X;Y)$ is the limit of the mutual information between quantized version $I(X^{\Delta};Y^{\Delta})$.

Properties

- $h(X, Y) = h(X) + h(Y|X)$.
- $I(X; Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$.
- $D(f||g) \geq 0$.
- $I(X; Y) = D(f(x, y)||f(x)f(y)) \geq 0$.
- $I(X; Y) = 0$ iff X and Y are independent.
- $h(X|Y) \leq h(X)$.
- $h(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} h(X_i|X_1, \ldots, X_{i-1}) \leq \sum_{i=1}^{n} h(X_i)$.

Hadamard's Inequality

- Let $\mathbf{X} \sim N(0, \Sigma)$. Then

$$|\Sigma| \leq \prod_{i=1}^{n} \Sigma_{ii},$$

by the last inequality of the previous slide.

Maximum Entropy Distribution

- Let $\mathbf{X}$ be a random vector with zero mean and covariance K. Then
$$h(\mathbf{X}) \leq \frac{1}{2} \log[(2\pi e)^n |K|].$$
The rhs is the entropy of a normal distribution with the same covariance.

- **Proof:** See the text.