LECTURE 15

Last time:

- Source-Channel Coding Theorem
- Feedback Capacity

Lecture outline

- Continuous Random Variables
- Differential Entropy
- AEP for continuous random variables

Continuous random variables

We consider continuous random variables with probability density functions (pdfs)

X has pdf $f_X(x)$

Cumulative distribution function (CDF)

$F_X(x) = P(X \leq x) = \int_{\infty}^{x} f_X(t)dt$

Function of a random variable

- In general, for $Y = g(X)$

Get CDF of Y: $F_Y(y) = \mathbf{P}(Y \leq y)$ Differentiate to get

$$f_Y(y) = \frac{dF_Y}{dy}(y)$$

- If g is strictly monotonic, assume the inverse function $x = h(y)$ is differentiable,

$$f_Y(y) = f_X(h(y)) \left| \frac{dx}{dy}(y) \right|$$

Example

X: uniform on $[0, 2]$,

$$Y = X^3$$

To compute the pdf of Y: for $y \in [0, 8]$,

- general approach

$$\begin{aligned} F_Y(y) &= P(Y < y) = P(X^3 < y) \\ &= P(X < y^{1/3}) = \frac{1}{2} y^{1/3} \end{aligned}$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{6} y^{-2/3}$$

-

$$\begin{aligned} f_Y(y) &= f_X(x) \left| \frac{dx}{dy} \right| \\ &= \frac{1}{2}\frac{1}{3} y^{-2/3} \end{aligned}$$

Key: the pdf of a cts random variable depends on how it's value is represented.

Differential Entropy

Question How much information is contained in a continuous random variable?

Consider a random variable X with a continuous density $f(x)$, divide the range of X into bins of length Δ. For each bin, there exists a value x_i,

$$f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)dx$$

Construct a discrete quantization of X as a r.v. X^Δ

$$X^\Delta = x_i, \qquad \text{for } X \in [i\Delta, (i+1)\Delta)$$

$$\begin{aligned} H(X^\Delta) &= -\sum p_i \log p_i \\ &= -\sum f(x_i)\Delta \log(f(x_i)\Delta) \\ &= -\sum f(x_i) \log f(x_i)\Delta - \log \Delta \end{aligned}$$

The more precisely X is quantized, the more information X^Δ contains.

Differential Entropy

Let $\Delta \to 0$.

$$H(X^{\Delta}) + \log \Delta \to -\int f(x) \log f(x) dx := h(X)$$

- $h(X)$ does NOT give the absolute amount of information contained in X.
- Differential entropy allows us to compare the randomness of two continuous random variables, when they are quantized to the same precision.

Corollary The entropy of a n-bit quantization of a cts r.v. X is approximately $h(X) + n$.

Examples:

- Uniform distribution, $X \sim U(a,b)$,

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

$$E[X] = \frac{b-a}{2}, \qquad \text{var}[X] = \frac{(b-a)^2}{12}$$

$$\begin{aligned} h(X) &= -\int_a^b \frac{1}{b-a} \log \frac{1}{b-a} dx \\ &= \log(b-a) \end{aligned}$$

- Gaussian random variable $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]$$

$$\begin{aligned} h(X) &= -\int f_X(x) \log f_X(x) dx \\ &= -\int f_X(x) \left(\log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(x-\mu)^2\right) dx \\ &= \log\sqrt{2\pi\sigma^2} + \frac{1}{2} \\ &= \frac{1}{2} \log 2\pi e\sigma^2 \end{aligned}$$

Example

Use differential entropy to compare the randomness of rv.'s

X is uniform on $[0,1]$. We quantize X into one of 8 bins of length 1/8 each. Let the bin number be $Y \in \{0,\ldots,7\}$.

$$\begin{aligned} h(X) &= 0 \\ h(X|Y) &= \log\frac{1}{8} = -3 \end{aligned}$$

By observing Y, we obtain 3 bits information about X.

Y can be thought as the first 3 digits of the binary expansion of X.

Joint and Conditional Differential Entropy, Mutual Information

- Joint differential entropy

$$h(X, Y) = -\int f_{X,Y}(x, y) \log f_{X,Y}(x, y) dx dy$$

- Conditional differential entropy

$$h(X|Y) = -\int f_{X,Y}(x, y) \log f_{X|Y}(x|y) dx dy$$

- Relative entropy of two densities f and g:

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

- Mutual Information

$$I(X; Y) = \int f_{X,Y}(x, y) \log \frac{f_{X,Y}(x, y)}{f_X(x) f_Y(y)} dx dy$$

Exercise as $\Delta \to 0$,

$$\begin{aligned} I(X^{\Delta}; Y^{\Delta}) &\to I(X; Y) \\ &= H(X^{\Delta}) - H(X^{\Delta}|Y^{\Delta}) \\ &\approx h(X) - \log \Delta - (h(X|Y) - \log \Delta) \end{aligned}$$

Properties of Differential Entropy

Chain Rule

$$h(X,Y) = h(X) + h(Y|X) = h(Y) + h(X|Y)$$

$$\begin{aligned} h(X,Y) &= E_{X,Y}[-\log f_{X,Y}(x,y)] \\ &= E_{X,Y}[-\log f_X(x) - \log f_{Y|X}(y|x)] \\ &= E_X[-\log f_X(x)] + E_{X,Y}[-\log f_{Y|X}(y|x)] \\ &= h(X) + h(Y|X) \end{aligned}$$

Similarly

$$I(X,Y;Z) = I(X;Z) + I(Y;Z|X)$$

Information Inequality

$$D(f||g) \geq 0$$

$$\begin{aligned} D(f||g) &= \int f(x) \log \frac{f(x)}{g(x)} dx \\ &= -\int f(x) \log \frac{g(x)}{f(x)} dx \\ &\geq \log \left[\int f(x) \frac{g(x)}{f(x)} dx \right] \\ &= 0 \end{aligned}$$

Corollary Mutual information is non-negative

$$I(X;Y) = D(f_{X,Y} \| f_X f_Y) \geq 0$$

Corollary Conditioning reduces entropy

$$h(X) - h(X|Y) = I(X;Y) \geq 0$$

Corollary Independence bound

$$h(X_1, X_2, \ldots, X_n) \leq \sum_{i=1}^{n} h(X_i)$$

Question

- $h(X) \geq 0$?
- What is the differential entropy of a discrete random variable?
- $h(X + a) = h(X)$?
- $h(2X) = h(X)$?

Scaling and Translation

- Translation does not change differential entropy

$$h(X + a) = h(X)$$

- Scaling changes differential entropy

$$h(aX) = h(X) + \log |a|$$

Example $X \sim N(0, 1)$,

$$\begin{aligned} h(X) &= \frac{1}{2} \log 2\pi e \\ h(\sigma X + \mu) &= \frac{1}{2} \log(2\pi e \sigma^2) \end{aligned}$$

Proof $Y = aX$, $f_Y(y) = \frac{1}{|a|} f_X(y/a)$,

$$\begin{aligned} h(Y) &= -\int f_Y(y) \log f_Y(y) dy \\ &= -\int \frac{1}{|a|} f_X(y/a) \log \left(\frac{1}{|a|} f_X(y/a) \right) dy \\ &= -\int f_X(x) \log f_X(x) dx + \log |a| \end{aligned}$$

Generalize

- If $Y = g(X)$,

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|$$

- $h(Y) = h(X) - E\left[\log \left|\frac{dx}{dy}\right|\right]$

Example $\underline{X}$ is i.i.d. Gaussian random vector $N(\underline{0}, I)$, $\underline{Y} = A\underline{X}$, for a fixed invertible matrix A. $Y \sim N(\underline{0}, K_Y = AA')$.

$$\begin{aligned} h(X) &= \frac{1}{2}\log(2\pi e)^n \\ h(Y) &= h(X) - \log|A^{-1}| = \frac{1}{2}\log(2\pi e)^n + \log|A| \\ &= \frac{1}{2}\log(2\pi e)^n |K_Y| \end{aligned}$$

Shannon "In the discrete case the entropy measures in an absolute way the randomness of the chance variable; in the continuous case the measurement is relative to the coordinate system"

Concavity and Convexity

Differential Entropy $h(X)$ is a concave function of the density $f_X(x)$.

Mutual information $I(X;Y)$ is concave in f_X for any fixed $f_{Y|X}$; and is convex in $f_{Y|X}$ for any fixed f_X.

Example: Maximize the Differential Entropy

Question 1 if X takes value in $[a, b]$, what distribution maximizes the differential entropy?

Uniform distribution on $[a, b]$ maximizes the differential entropy.

Question 2 if X has mean $E[X] = \mu$ and variance $\text{var}[X] = \sigma^2$, what distribution of X maximizes the differential entropy?

$$\max_f \left[-\int f(x) \log f(x) dx \right]$$

subject to the contraint

$$\int f(x)dx = 1$$
$$\int xf(x)dx = \mu$$
$$\int (x-\mu)^2 f(x)dx = \sigma^2$$

Can be solved by Lagrange method to conclude $X \sim N(\mu, \sigma^2)$.

Gaussian distribution maximizes the differential entropy for the same first and second order moment.

Assume X' has the same first and second order moment as the Gaussian random variable X. Let the density of X be f and density of X' be g.

$$
\begin{aligned}
& D(X'||X) = D(g||f) \\
= \; & \int g(x) \log \frac{g(x)}{f(x)} dx \\
= \; & \int g(x) \log g(x) dx \\
& - \int g(x) \left[\log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(x-\mu)^2 \right] dx \\
= \; & -h(X') - \int f(x) \left[\log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(x-\mu)^2 \right] dx \\
= \; & h(X) - h(X')
\end{aligned}
$$