Electrical Engineering 229A Lecture 18 Notes

Daniel Raban

October 26, 2021

1 Differential Entropy and the Additive White Gaussian Noise Channel Model

1.1 Differential entropy

Let X be a real-valued random variable with density, i.e.

$$\mathbb{P}(X \in [a,b)) = \int_a^b f(x)\, dx$$

for some nonnegative function f.

Definition 1.1. The **differential entropy** of X is

$$h(f) := -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx.$$

This need not be well-defined (an example is provided in Handout 7), so when we talk about $h(f)$, we will assume it exists.

Example 1.1. Let $X \sim \mathrm{Unif}([a,b])$ with density

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a,b] \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$h(f) = \int_a^b \frac{1}{b-a} \log(b-a)\, dx = \log(b-a).$$

Note that if $b - a < 1$, this is *negative*. So $h(f)$ is very different from entropy.

Example 1.2. Let $X \sim N(\mu, \sigma^2)$ with density

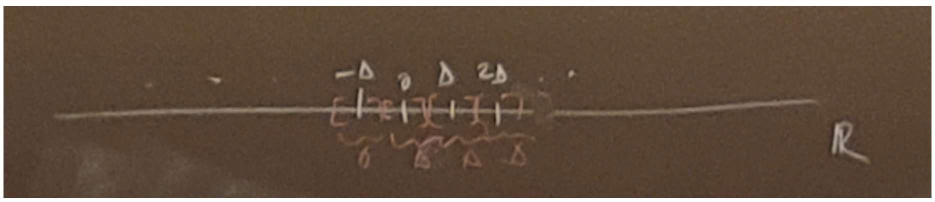
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma)}.$$

Then

$$\begin{aligned} h(f) &= (\log e) \int_{-\infty}^{\infty} f(x) \left[\frac{(x-\mu)^2}{2\sigma^2} + \frac{1}{2}\ln(2\pi\sigma^2) \right] dx \\ &= (\log e) \left[\frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2) \right] \\ &= \frac{1}{2}\log(2\pi e\sigma^2). \end{aligned}$$

1.2 Connection to entropy

Here is the connection between differential entropy and an underlying entropy. Imagine quantizing $\mathbb{R}$ at scale Δ with $\Delta \to 0$.



We get a discrete probability distribution having probability

$$\int_{k\Delta - \frac{\Delta}{2}}^{k\Delta + \frac{\Delta}{2}} f(x)\,dx \quad \text{at } k$$

as an approximation to a random variable with density f. Think of the entropy of this approximation. This is

$$-\sum_{k\in\mathbb{Z}} (\Delta f(k\Delta) + o(\Delta)) \log(\Delta f(k\Delta) + o(\Delta)) - \log\Delta - \Delta \sum_{k\in\mathbb{Z}} f(k\Delta)\log f(k\Delta) + o(\Delta).$$

So we can think of $h(f)$ as the amount of entropy of a quantized pproximation about $-\log\Delta$ as $\Delta \to 0$.

This $-\log\Delta$ is a problem because $-\log\Delta \to \infty$ as $\Delta \to 0$.

1.3 Relative entropy

However, this quantization problem does not show up for relative entropy.

Definition 1.2. Given two probability densities f and g, the **relative entropy** is

$$D(f \mid\mid g) := \int_{-\infty}^{\infty} f(x) \log\frac{f(x)}{g(x)}\,dx.$$

Note that in writing

$$\sum_{k\in\mathbb{Z}}(\Delta f(k\Delta)+o(\Delta))\log\frac{\Delta f(k\Delta)+o(\Delta)}{\Delta g f(k\Delta)+o(\Delta)},$$

the Δs in the log cancel.

The relative entropy will be nonnegative by convexity of $u \mapsto u\log u$ because it is

$$\int_{-\infty}^{\infty} g(x)\frac{f(x)}{g(x)}\log\frac{f(x)}{g(x)}\,dx.$$

1.4 Joint differential entropy

Definition 1.3. The **joint differential entropy** of $X_1,\dots,X_n$ (real-valued random variables with a joint density f) is

$$h(X_1,\dots,X_n) = -\mathbb{E}[\log f(X_1,\dots,X_n)].$$

Example 1.3. The most important example is when $X_1,\dots,X_n$ are jointly Gaussian random variables with invertible covariance matrix:

$$\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix}, K\right),$$

where K is a symmetric, positive definite matrix. The joint density is

$$f(x_1,\dots,x_n) = \frac{1}{(2\pi)^{n/2}(\det K)^{1/2}}e^{-\frac{1}{2}(x-m)^\top K^{-1}(x-m)}.$$

The joint differential entropy is

$$h(X_1,\dots,X_n) = \frac{1}{2}\log((2\pi e)^n\det K).$$

This can be understood by diagonalizing K. $K = U^\top D U$, where $U^\top U = I$ and $D = \mathrm{diag}(\sigma_1^2,\dots,\sigma_n^2)$. Then

$$h(X_1,\dots,X_n) = \sum_{\ell=1}^n \frac{1}{2}\log(2\pi e\sigma_\ell^2).$$

1.5 Mutual information

If X and Y have joint density $f(x,y)$, then they will have marginal densities $f(x)$ and $f(y)$ respectively.

Definition 1.4. The **mutual information** is defined as

$$I(X;Y) = D(f(x,y) \mid\mid f(x)f(y)).$$

This will turn out to be

$$I(X;Y) = h(X) + h(Y) - h(X,Y),$$

when this expression makes sense. This will also be

$$I(X;Y) = h(X) - h(X \mid Y)$$

, if these quantities exist, where $h(X \mid Y)$ is the conditional differential entropy.

Definition 1.5. The **conditional differential entropy** is

$$h(X \mid Y) = \int_{-\infty}^{\infty} f(y)h(X \mid Y = y)\, dy,$$

where

$$h(X \mid Y = y) = -\int_{-\infty}^{\infty} f(x \mid y) \log f(x \mid y)\, dx.$$

1.6 Chain rules for differential entropy

We can write some chain rules.

Proposition 1.1 (Chain rule for differential entropy)**.** *When all these quantities make sense,*

$$h(X_1, \dots, X_n) = h(X_1) + h(X_2 \mid X_1) + h(X_3 \mid X_1, X_2) + \cdots + h(X_n \mid X_1, \dots, X_{n-1}).$$

Proposition 1.2 (Chain rule for mutual information)**.** *When* $(X, Y_1, \dots, Y_n)$ *have a joint density,*

$$I(X;Y_1, \dots, Y_n) = I(X;Y_1) + I(X;Y_2 \mid Y_1) + I(X;Y_3 \mid Y_1, Y_2) + \cdots + I(X;Y_n \mid Y_1, \dots, Y_{n-1}).$$

1.7 Basic properties of differential entropy

Proposition 1.3. *For any constant c, $h(X + c) = h(X)$.*

Proof. Adding c just translates the density. □

Proposition 1.4. *If $c \neq 0$, then $h(cX) = h(X) - \log |c|$.*

Proof. The density of cX is $\frac{1}{|c|}f(x/c)$. So

$$\begin{aligned} h(cX) &= \int_{-\infty}^{\infty} \frac{1}{|c|} f(x/c) \log \frac{1}{|c|} f(x/c)\, dx \\ &= h(X) - \log|c|. \end{aligned}$$

□

Remark 1.1. This is consistent with $X \sim N(0, \sigma^2) \iff cX \sim 0, c^2\sigma^2)$. Here, $h(X) = \frac{1}{2}\log(2\pi e\sigma^2)$ and $h(cX) = \frac{1}{2}\log(2\pi e\sigma^2) + \log|c|$.

Proposition 1.5. *If $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = \sigma^2$, then*

$$h(X) \le \frac{1}{2}\log(2\pi e\sigma^2).$$

This upper bound is the entropy of the Gaussian.

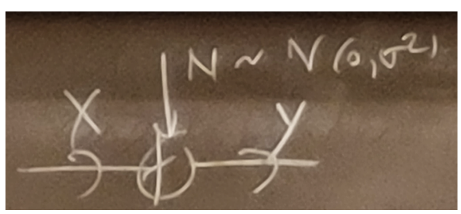
Proof. Let $\phi(x)$ denote the $N(0, \sigma^2)$ density, i.e. $\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}$. Write

$$\begin{aligned} 0 &\le D(f \mid\mid \phi) \\ &= \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{\phi(x)} \\ &= -h(f) + (\log e) \int_{-\infty}^{\infty} f(x) \left[\frac{1}{2}\ln(2\pi\sigma^2) + \frac{x^2}{2\sigma^2} \right] dx \\ &= -h(f) + \frac{1}{2}\log(2\pi e\sigma^2). \end{aligned}$$

□

1.8 The additive white Gaussian noise channel model

This is a discrete time model. At each channel use, the input is a real number, say $x \in \mathbb{R}$. The output is a real number Y. Conditioned on $X = x$, $Y \sim \mathcal{N}(x, \sigma^2)$, where σ^2 is the variance of the noise.



Consider an input power constrained scenario and block based communication: We have an encoding map

$$e_n : [M_n] \to \mathbb{R}$$

and a decoding map

$$d_n : \mathbb{R}^n \to [M_n], \qquad Y^n(m).$$

Here X_n is the output of e_n, and Y_n is the input of d_n.

Conditioned on $X^n(m) = x^n$,

$$Y^n \sim \mathcal{N}(x^n, \sigma^2 I),$$

i.e. the noise is iid over time. In other words,

$$f(y^n \mid x^n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y_i - x_i)/(2\sigma)}.$$

The **power constraint** P requires that each $X^n(m)$ satisfies

$$\sum_{i=1}^{n} (X_i(n))^2 \leq nP.$$

Intuitively, M_n can be on the scale of

$$\frac{V_n(\sqrt{n(P + \sigma^2)})}{V_n(\sqrt{n\sigma^2})},$$

where $V_n(R)$ denotes the colume of the ball in $\mathbb{R}^n$ of radius R.