Lecture 19 Differential Entropy

Textbook 8.1-8.6

November 11th, 2022

Let X be a random variable with cumulative distribution function $F(x) = \Pr(X \leq x)$. If $F(x)$ is continuous, the random variable is said to be continuous. Let $f(x) = F'(x)$ when the derivative is defined. If $\int_{-\infty}^{\infty} f(x) = 1$, $f(x)$ is called the *probability density function* for X. The set where $f(x) > 0$ is called the *support set* of X.

The differential entropy $h(X)$ of a continuous random variable X with density $f(x)$ is defined as

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

where S is the support of the random variable.

Uniform distribution

Consider a random variable distributed uniformly from 0 to a so that its density is $1/a$ from 0 to a and 0 elsewhere. Then

$$h(X) = -\int_0^a \frac{1}{a} \log \frac{1}{a} dx = \log a.$$

Normal distribution

Let $X \sim \phi(x) = (1/\sqrt{2\pi\sigma^2})e^{-x^2/2\sigma^2}$. Then calculating the differential entropy in nats, we obtain

$$\begin{aligned} h(\phi) &= -\int \phi \ln \phi \\ &= -\int \phi(x)[-\frac{x^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma^2}] \\ &= \frac{EX^2}{2\sigma^2} + \frac{1}{2}\ln 2\pi\sigma^2 \\ &= \frac{1}{2} + \frac{1}{2}\ln 2\pi\sigma^2 \\ &= \frac{1}{2}\ln 2\pi e\sigma^2 \text{ nats.} \end{aligned}$$

Changing the base of the logarithm, we have

$$h(\phi) = \frac{1}{2}\log 2\pi e\sigma^2 \text{ bits.}$$

Definition

The differential entropy of a set $X_1, X_2, \ldots, X_n$ of random variables with density $f(x_1, x_2, \ldots, x_n)$ is defined as

$$h(X_1, X_2, \cdots, X_n) = -\int f(x^n) \log f(x^n) dx^n.$$

Definition

If X, Y have a joint density function $f(x, y)$, we can define the conditional differential entropy $h(X|Y)$ as

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

Since in general $f(x|y) = f(x, y)/f(y)$, we can also write

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

But we must be careful if any of the differential entropies are infinite.

Entropy of a multivariate normal distribution

Let $X_1, X_2, \ldots, X_n$ have a multivariate normal distribution with mean μ and covariance matrix K. Then

$$h(X_1, X_2, \cdots, X_n) = h(\mathcal{N}_n(\mu, K)) = \frac{1}{2}\log(2\pi e)^n |K| \text{ bits},$$

where $|K|$ denotes the determinant of K.

Proof.

The joint probability density function is

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

Proof.

Then

$$
\begin{aligned}
h(f) &= -\int f(x)[-\frac{1}{2}(\mathbf{x}-\mu)^T K^{-1}(\mathbf{x}-\mu) - \ln(\sqrt{2\pi})^n |K|^{\frac{1}{2}}]d\mathbf{x}. \\
&= \frac{1}{2}E[\sum_{i,j}(X_i-\mu_i)(K^{-1})_{ij}(X_j-\mu_j)] + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}E[\sum_{i,j}(X_i-\mu_i)(X_j-\mu_j)(K^{-1})_{ij}] + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\sum_{i,j}E[(X_i-\mu_i)(X_j-\mu_j)](K^{-1})_{ij} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\sum_{i,j}K_{ji}(K^{-1})_{ij} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\sum_{j}I_{jj} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{n}{2} + \frac{1}{2}\ln(2\pi)^n|K| \\
&= \frac{1}{2}\ln(2\pi e)^n|K| \text{ nats} \\
&= \frac{1}{2}\log(2\pi e)^n|K| \text{ bits.}
\end{aligned}
$$

□

Definition

The relative entropy $D(f \| g)$ between two densities f and g is defined by

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

Note that $D(f \| g)$ is finite only if the support set of f is contained in the support set of g.

Definition

The mutual information $I(X;Y)$ between two random variables with joint density $f(x,y)$ is defined as

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

From the definition it is clear that

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

and

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

Example

Let $(X, Y) \sim \mathcal{N}(0, K)$, *where*

$$K = \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}$$

Then we have that $h(X) = h(Y) = \frac{1}{2}\log(2\pi e)\sigma^2$ *and* $h(X, Y) = \frac{1}{2}\log(2\pi e)^2|K| = \frac{1}{2}\log(2\pi e)^2\sigma^4(1-\rho^2)$, *and therefore*

$$I(X;Y) = h(X) + h(Y) - h(X, Y) = -\frac{1}{2}\log(1-\rho^2).$$

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

with equality if and only if $f = g$ almost everywhere.

Proof.

Let S be the support set of f. Then

$$
\begin{aligned}
-D(f\|g) &= \int_S f \log \frac{g}{f} \\
&\leq \log \int_S f \frac{g}{f} \\
&= \log \int_S g \\
&\leq \log 1 = 0
\end{aligned}
$$

We have equality if and only if we have equality in Jensen's inequality, which occurs if and only if $f = g$ a.e. □

Corollary

$I(X;Y) \geq 0$ *with equality if and only if* X *and* Y *are independent.*

Corollary

$h(X|Y) \leq h(X)$ *with equality if and only if* X *and* Y *are independent.*

Chain rule for differential entropy

Theorem

$$h(X_1, X_2, \cdots, X_n) = \sum_{i=1}^{n} h(X_i | X_1, X_2, \cdots, X_{i-1}).$$

Corollary

$$h(X_1, X_2, \cdots, X_n) \leq \sum h(X_i),$$

with equality if and only if $X_1, X_2, \ldots, X_n$ *are independent.*

Application: Hadamard's inequality

If we let $X \sim \mathcal{N}(0, K)$ be a multivariate normal random variable, calculating the entropy in the above inequality gives us

$$|K| \leq \Pi_{i=1}^{n} K_{ii}.$$

Theorem

$$h(X + c) = h(X).$$

Theorem

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

Proof.

Let $Y = aX$. Then $f_Y(y) = \frac{1}{|a|} f_X(\frac{y}{a})$, and

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

after a change of variables in the integral. □

Corollary

$$h(A\mathbf{X}) = h(\mathbf{X}) + \log |\mathit{det}(A)|.$$

Theorem

Let the random vector $\mathbf{X} \in \mathbb{R}^n$ *have zero mean and covariance* $K = E\mathbf{X}\mathbf{X}^t$*. Then* $h(\mathbf{X}) \leq \frac{1}{2}\log(2\pi e)^n|K|$*, with equality if and only if* $\mathbf{X} \sim \mathcal{N}(0, K)$*.*

Proof.

Let $g(\mathbf{x})$ be any density satisfying $\int g(\mathbf{x})x_i x_j d\mathbf{x} = K_{ij}$ for all i, j. Let ϕ_K be the density of a $\mathcal{N}(0, K)$ vector, where we set $\mu = 0$. Note that $\log \phi_K(\mathbf{x})$ is a quadratic form and $\int x_i x_j \phi_K(\mathbf{x}) d\mathbf{x} = K_{ij}$. Then

$$\begin{aligned} 0 &\le D(g \| \phi_K) \\ &= \int g \log(g/\phi_K) \\ &= -h(g) - \int g \log \phi_K \\ &= -h(g) - \int \phi_K \log \phi_K \\ &= -h(g) + h(\phi_K) \end{aligned}$$

where the substitution $\int g \log \phi_K = \int \phi_K \log \phi_K$ follows from the fact that g and ϕ_K yields the same moments of the quadratic form $\log \phi_K(\mathbf{x})$. □