EE/Ma 126b Information Theory - Homework Set #2

Ling Li, ling@cs.caltech.edu

January 15, 2001

2.1 *Differential entropy.* $h(X) = -\int f(x) \ln f(x) dx$.

(a) The exponential density, $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.

$$h(X) = -\int_0^\infty \lambda e^{-\lambda x} (\ln \lambda - \lambda x)\, dx = \Big[(\ln \lambda - 1 - \lambda x)\, e^{-\lambda x}\Big]\Big|_0^\infty = \boxed{1 - \ln \lambda}.$$

(b) The Laplace density, $f(x) = \frac{1}{2}\lambda e^{-\lambda |x|}$.

$$\begin{aligned} h(X) &= -\int_{-\infty}^{\infty} \frac{1}{2}\lambda e^{-\lambda|x|} (\ln \lambda - \ln 2 - \lambda |x|)\, dx \\ &= -\int_0^\infty \lambda e^{-\lambda x} (\ln \lambda - \ln 2 - \lambda x)\, dx \\ &= \ln 2 \int_0^\infty \lambda e^{-\lambda x} dx - \int_0^\infty \lambda e^{-\lambda x} (\ln \lambda - \lambda x)\, dx \\ &= \ln 2 + 1 - \ln \lambda = \boxed{1 - \ln \frac{\lambda}{2}}. \end{aligned}$$

(c) $X = X_1 + X_2$, where $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ are independent. Thus

$$f(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(t-\mu_1)^2}{2\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(x-t-\mu_2)^2}{2\sigma_2^2}}\, dt = \frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}} e^{-\frac{(x-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}},$$

i.e., $X \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. So

$$h(X) = \boxed{\frac{1}{2} \ln \left(2\pi e(\sigma_1^2 + \sigma_2^2)\right)}.$$

2.2 *Mutual information for correlated normals.* $X \sim N(0, \sigma^2)$ and $Y \sim N(0, \sigma^2)$, so $h(X) = h(Y) = \frac{1}{2} \log 2\pi e \sigma^2$. For $-1 < \rho < 1$, we have

$$\begin{vmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{vmatrix} = (1-\rho^2)\sigma^4 > 0.$$

So $h(X, Y) = \frac{1}{2} \log(2\pi e)^2 (1-\rho^2)\sigma^4$. Thus

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

For example, $\boxed{I(X;Y) = 0 \text{ when } \rho = 0}$. However, $I(X;Y) \to \infty$ when $\rho^2 \to 1$, implies that $\boxed{I(X;Y) = \infty \text{ when } \rho = \pm 1}$. Since $I(X;Y)$ is the amount of information that X says about Y or Y says about X, it is natural that it is 0 when X and Y are independent ($\rho = 0$); when $\rho = \pm 1$, X and Y are fully correlated and X gives exactly all the bits of Y, so $I(X;Y)$ is infinity since infinity number of bits is needed to exactly describe a continuous random variable.

2.3 *Uniformly distributed noise.*

$$f_X(x) = \begin{cases} 1, & -1/2 \le x \le 1/2; \\ 0, & \text{otherwise.} \end{cases} \quad , \quad f_Z(z) = \begin{cases} a^{-1}, & -a/2 \le z \le a/2; \\ 0, & \text{otherwise.} \end{cases}$$

From $Y = X + Z$, and X and Z are independent, we have

$$f_Y(y) = \int_{-\infty}^{\infty} f_X(x) f_Z(y-x)dx = \int_{-1/2}^{1/2} f_Z(y-x)dx = \int_{y-1/2}^{y+1/2} f_Z(z)dz.$$

When $0 < a \le 1$,

$$f_Y(y) = \begin{cases} a^{-1}(\frac{1+a}{2} + y), & \frac{-1-a}{2} \le y < \frac{-1+a}{2}; \\ 1, & \frac{-1+a}{2} \le y \le \frac{1-a}{2}; \\ a^{-1}(\frac{1+a}{2} - y), & \frac{1-a}{2} < y \le \frac{1+a}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

and with variable change,

$$\begin{aligned} h(Y) &= -\int f_Y(y) \ln f_Y(y) dy \\ &= -a \int_0^1 y \ln(y) dy - 0 - a \int_0^1 y \ln(y) dy \\ &= \frac{a}{2} \text{ nats.} \end{aligned}$$

When $a > 1$,

$$f_Y(y) = \begin{cases} a^{-1}(\frac{1+a}{2} + y), & \frac{-1-a}{2} \le y < \frac{1-a}{2}; \\ a^{-1}, & \frac{1-a}{2} \le y \le \frac{-1+a}{2}; \\ a^{-1}(\frac{1+a}{2} - y), & \frac{-1+a}{2} < y \le \frac{1+a}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

and with variable change,

$$\begin{aligned} h(Y) &= -\int f_Y(y) \ln f_Y(y) dy \\ &= -a \int_0^{a^{-1}} y \ln y dy - \int_{\frac{1-a}{2}}^{\frac{-1+a}{2}} a^{-1} \ln a^{-1} - a \int_0^{a^{-1}} y \ln y dy \\ &= (\frac{1}{2a} + \ln a) \text{ nats.} \end{aligned}$$

Since $h(Y|X) = h(Z|X) = h(Z) = \ln a$ nats, we get

$$\boxed{I(X;Y) = h(Y) - h(Y|X) = \begin{cases} \frac{a}{2} - \ln a, & 0 < a \le 1; \\ \frac{1}{2a}, & a > 1. \end{cases} \quad \text{(nats)}}.$$

2.4 *Quantized random variables.* Let random variable T be the decay time (in years). The cumulative distribution function is

$$F(t) = \Pr\{T \le t\} = 1 - 2^{-\frac{t}{80}} = 1 - e^{-\frac{t \ln 2}{80}}, \quad t \ge 0,$$

and the probability density function is

$$f(t) = F'(t) = \frac{\ln 2}{80} 2^{-\frac{t \ln 2}{80}} = \frac{1}{D} e^{-\frac{t}{D}},$$

where $D = 80 \log e$. Thus by Problem 2.1a,

$$h(T) = 1 + \ln D \text{ nats} = \log(eD) \text{ bits} \approx 8.29 \text{ bits}.$$

To describe T to 3 *digit* accuracy, we need roughly

$$h(T) + 3 \log 10 = \log(80000 e \log e) \approx \boxed{18.26 \text{ bits}}.$$

2.5 *Scaling.* Let $\mathbf{Y} = A\mathbf{X}$. Then $f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{|A|} f_{\mathbf{X}}(A^{-1}\mathbf{y})$, where $|A| = |\det(A)|$ is the absolute value of the determinant. Thus

$$\begin{aligned}
h(A\mathbf{X}) &= -\int f_{\mathbf{Y}}(\mathbf{y}) \log f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y} \\
&= -\int \frac{1}{|A|} f_{\mathbf{X}}(A^{-1}\mathbf{y}) \log \left(\frac{1}{|A|} f_{\mathbf{X}}(A^{-1}\mathbf{y}) \right) d\mathbf{y} \\
&= \log |A| - \int \frac{1}{|A|\,|A^{-1}|} f_{\mathbf{X}}(\mathbf{x}) \log f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&= \log |A| + h(\mathbf{X}),
\end{aligned}$$

after a change of variables in the integral. □