Continuous Random Variables

- Differential entropy
- Maximum entropy distributions
- Normal distribution
- Multivariate normal distribution
- Gaussian process
- Differential entropy rate

Differential Entropy

- Let X be a continuous random variable with PDF $f_X(x)$. The differential entropy of X is defined as

$$h(X) = E\{-\log_2 f_X\} = -\int_{\mathcal{R}} f_X(x) \log_2 f_X(x) dx$$

- Example: Differential Entropy for uniform distribution

$$f_X(x) = \frac{1}{a} \mathbf{1}_{[0,a]}(x)$$

$$h(X) = \log_2(a)$$

- Note, differential entropy can be negative!

Properties of Differential Entropy

- Translation does not change the differential entropy

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

- Scaling by *a* increases the differential entropy by $\log_2|a|$

$$h(aX) = h(X) + \log_2 |a|$$

Joint and Conditional Differential Entropy

- Joint differential entropy of X and Y

$$h(X,Y) = E\left\{-\log_2 f_{X,Y}\right\}$$

- Conditional differential entropy of X given Y

$$h(X|Y) = E\left\{-\log_2 f_{X|Y}\right\}$$

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

- Conditioning reduces differential entropy

$$h(X|Y) \leq h(X)$$

with equality iff X and Y are independent.

Mutual Information

- Mutual information between X and Y

$$I(X;Y) = E\left\{\log_2 \frac{f_{X,Y}}{f_X f_Y}\right\}$$

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

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

- Mutual information is non-negative

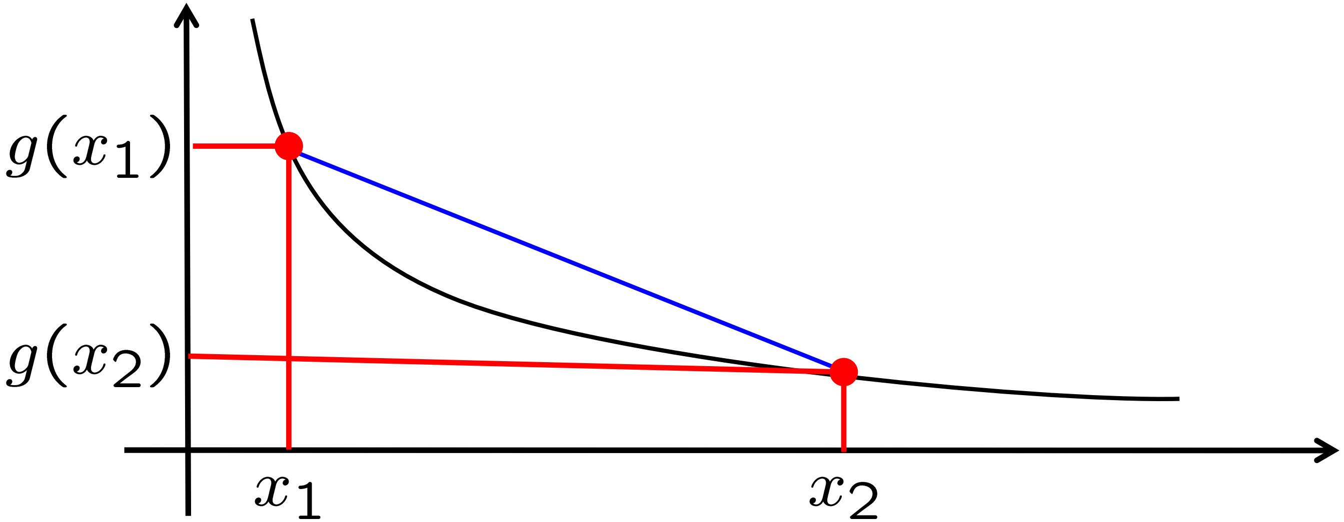
$$I(X;Y) \geq 0$$

with equality iff X and Y are independent.

Remark: Convex Functions

- A function g(x) is said to be **convex** over an interval [a,b] if for every x_1, $x_2 \in$ [a,b] and $0 \le \lambda \le 1$

$$\lambda g(x_1) + [1-\lambda] g(x_2) \geq g(\lambda x_1 + [1-\lambda] x_2)$$



- A function is said to be **strictly convex** if equality holds only for λ = 0 or λ = 1.

Remark: Jensen's Inequality

- If g is a **convex** function and X is a random variable, then

$$E\left\{g(X)\right\} \geq g\left(E\{X\}\right)$$

- Moreover, if g is **strictly convex**, then equality implies that E{X}=X with probability 1, i.e., X is a constant.

- Example: Two mass point distribution

$$p_1 g(x_1) + p_2 g(x_2) \geq g(p_1 x_1 + p_2 x_2)$$

- If g is strictly convex, equality holds only for the point $x_1 = x_2 := x$, that is, X is a constant.

Maximum Entropy Distribution for given σ^2

- For given variance, find PDF that maximizes entropy

$$\inf_{f_X} \{-h(X)\} \quad \text{s.t.} \quad E\{X^2\} = \sigma^2$$

- Solve unconstrained problem $\inf_{f_X} J(X)$ with $\lambda > 0$ and

$$\begin{aligned}
J(X) &= E\left\{\ln f_X(X) + \lambda(X^2 - \sigma^2)\right\} \\
&= E\left\{-\ln \frac{e^{-\lambda(X^2-\sigma^2)}}{f_X(X)}\right\} \\
&\geq -\ln E\left\{\frac{e^{-\lambda(X^2-\sigma^2)}}{f_X(X)}\right\} \\
&= -\ln E\{Y\}
\end{aligned}$$

Maximum Entropy Distribution for given σ^2

- Jensen: Lower bound is tight iff Y is a constant.

$$\frac{e^{-\lambda(x^2-\sigma^2)}}{f_X(x)} = \text{const.} \quad \forall x \in \mathcal{R}$$

- We obtain the PDF

$$f_X(x) = ce^{-\lambda x^2}$$

- **Normal distribution** with variance σ^2 that integrates to 1

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

Differential Entropy of Normal Distribution

- For given variance, maximum differential entropy is

$$
\begin{aligned}
h(X) &= E\{-\log_2 f_X\} \\
&= \frac{1}{2}\log_2(2\pi\sigma^2) + \frac{E\left\{X^2\right\}}{2\sigma^2}\log_2 e \\
&= \frac{1}{2}\log_2(2\pi\sigma^2) + \frac{1}{2}\log_2 e \\
&= \frac{1}{2}\log_2(2\pi e\sigma^2)
\end{aligned}
$$

Multivariate Normal Distribution

- Let $\mathbf{X}=(X_1, X_2, \ldots, X_n)^T$ have a multivariate normal distribution with zero mean and covariance matrix C

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|C|^{\frac{1}{2}}} e^{-\frac{1}{2}\mathbf{x}^T C^{-1}\mathbf{x}}$$

- Change of coordinate system: $\mathbf{Y} = T^T \mathbf{X}$, where $T^T C\, T=\Lambda$ being diagonal with Eigenvalues λ_i; note that $|T| = 1$

$$\mathbf{x}^T C^{-1}\mathbf{x} = \mathbf{y}^T T^T C^{-1} T \mathbf{y} = \mathbf{y}^T \Lambda^{-1}\mathbf{y}$$

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{\lambda_i}} e^{-\frac{1}{2}\frac{\mathbf{y}_i^2}{\lambda_i}}$$

Multivariate Normal Distribution

- Differential entropy of a multivariate normal random variable

$$
\begin{aligned}
h(\mathbf{Y}) &= h(T^T \mathbf{X}) \\
&= h(\mathbf{X}) + \log_2 |T^T| \\
&= h(\mathbf{X})
\end{aligned}
$$

$$
\begin{aligned}
h(\mathbf{X}) &= \frac{n}{2} \log_2(2\pi e) + \sum_{i=1}^{n} \frac{1}{2} \log_2 \lambda_i \\
&= \frac{n}{2} \log_2(2\pi e) + \frac{1}{2} \log_2 |C| \\
&= \frac{1}{2} \log_2 [(2\pi e)^n |C|]
\end{aligned}
$$

Bivariate Normal Random Variable

- Covariance matrix C

$$C = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 \end{bmatrix} \qquad |\rho_{12}| \le 1$$

- Mutual information between components

$$\begin{aligned} I(X_1;X_2) &= h(X_1) + h(X_2) - h(X_1,X_2) \\ &= \frac{1}{2}\log_2 \frac{\left|\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\right|}{|C|} \\ &= -\frac{1}{2}\log_2(1-\rho_{12}^2) \end{aligned}$$

- $I(X_1;X_2) = 0$ for $\rho_{12} = 0$

Continuous-Valued Stochastic Process

- Continuous-valued discrete-time stochastic process $\{X_i\}$, X_i 2 $\mathcal{R}^n$, is an indexed sequence of continuous random variables X_1, X_2, …, X_n with joint PDF

$$f_{X_1,X_2,\ldots,X_n} = \mathrm{Pr}\left\{(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n)\right\}$$

$$\forall \quad (x_1, x_2, \ldots, x_n) \in \mathcal{R}^n$$

- A stochastic process is said to be stationary if the joint PDF is invariant with respect to shifts

$$f_{X_1,X_2,\ldots,X_n} = f_{X_{1+l},X_{2+l},\ldots,X_{n+l}} \quad \forall \quad l \in \mathcal{Z}$$

Differential Entropy Rate

- How does the differential entropy of X^n grow with n?
- Differential entropy rate of a continuous-valued stochastic process $\{X_i\}$ is defined by

$$h(\{X_i\}) = \lim_{n\to\infty} \frac{1}{n} h(X_1, X_2, \ldots, X_n)$$

- Using the chain rule

$$h(\{X_i\}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} h(X_i | X_{i-1}, \ldots, X_1)$$

Gaussian Stochastic Process

- Consider a zero-mean cyclostationary Gaussian process $\{X_i\}$ whose statistical properties repeat with period n

$$
\begin{aligned}
h(\{X_i\}) &= \frac{1}{n} h(X_1, X_2, \ldots, X_n) \\
&= \frac{1}{2} \log_2(2\pi e) + \frac{1}{n}\frac{1}{2} \log_2 |C^{(n)}|
\end{aligned}
$$

- Let the covariance matrix be **circulant** with $C_{kl} = E\{X_k X_l\}$

$$
C^{(n)} = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & c_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{bmatrix} \qquad C_{kl} = c_{(l-k) \bmod n}
$$

Remark: Eigenvalues of Circulant Matrix

- Let C have Eigenvalues λ_ν and Eigenvectors $t^{(\nu)}$

$$Ct^{(\nu)} = \lambda_\nu t^{(\nu)}$$

$$\sum_{l=1}^{n} C_{kl} t_l \quad = \quad \lambda t_k \quad \forall \quad k = 1, \ldots, n$$

$$\sum_{l=0}^{n-k} c_l t_{l+k} + \sum_{l=n-k+1}^{n-1} c_l t_{l+k-n} \quad = \quad \lambda t_k \quad \forall \quad k = 1, \ldots, n$$

- Solve difference equation by using complex roots of unity

$$\lambda_\nu = \sum_{l=0}^{n-1} c_l e^{-j\frac{2\pi}{n} l\nu} := \mathrm{DFT}\{c_l\}$$

Gaussian Stochastic Process

- Consider a zero-mean cyclostationary Gaussian process with period n

$$
\begin{aligned}
h(\{X_i\}) &= \frac{1}{2}\log_2(2\pi e) + \frac{1}{n}\frac{1}{2}\log_2|C^{(n)}| \\
&= \frac{1}{2}\log_2(2\pi e) + \frac{1}{n}\sum_{\nu=0}^{n-1}\frac{1}{2}\log_2\lambda_\nu \\
&= \frac{1}{2}\log_2(2\pi e) + \frac{1}{n}\sum_{\nu=0}^{n-1}\frac{1}{2}\log_2 \mathrm{DFT}\{c_l\}_\nu
\end{aligned}
$$

Remark: Szegö's Theorem

- Let the period n grow and consider a stationary Gaussian stochastic process with autocorrelation sequence and PSD

$$\phi_{XX}[l] = E\{X_i X_{i+l}\} \qquad \Phi_{XX}(\omega) = \sum_{l \in \mathcal{Z}} \phi_{XX}[l] e^{-j\omega l}$$

- Then, the covariance matrix is Toeplitz with Eigenvalues λ_ν

$$C^{(n)} = \begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_{n-1} \\ \phi_{-1} & \phi_0 & \phi_1 & \cdots & \phi_{n-2} \\ \phi_{-2} & \phi_{-1} & \phi_0 & \cdots & \phi_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{1-n} & \phi_{2-n} & \phi_{3-n} & \cdots & \phi_0 \end{bmatrix}$$

- The distribution of discrete Eigenvalues converges

$$\lim_{n \to \infty} \frac{1}{n} \sum_{\nu=0}^{n-1} g(\lambda_\nu) = \frac{1}{2\pi} \int_{-\pi}^{\pi} g\left(\Phi(\omega)\right) d\omega$$

Gaussian Stochastic Process

- Given a zero-mean stationary Gaussian process $\{X_i\}$ with

$$\phi_{XX}[l] = E\{X_i X_{i+l}\} \qquad \Phi_{XX}(\omega) = \sum_{l\in\mathcal{Z}} \phi_{XX}[l] e^{-j\omega l}$$

- The differential entropy rate is

$$\begin{aligned}
h(\{X_i\}) &= \frac{1}{2}\log_2(2\pi e) + \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=0}^{n-1} \frac{1}{2} \log_2 \lambda_\nu \\
&= \frac{1}{2}\log_2(2\pi e) + \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=0}^{n-1} \frac{1}{2} \log_2 \mathrm{DFT}\{c_l\}_\nu \\
&= \frac{1}{2}\log_2(2\pi e) + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log_2 \Phi_{XX}(\omega) d\omega
\end{aligned}$$