Maximum Entropy and Spectral Estimation

Introduction

- What is the distribution of velocities in the gas at a given temperature? It is the Maxwell-Boltzmann distribution.
- The maximum entropy distribution corresponds to the macrostate with the maximum number of microstates.
- Implicitly assumed is that all microstates are equally probable, which is an AEP property.

Maximum Entropy Problem and Solution

- **Problem:** Find a distribution $f^*(x)$ with maximal entropy $h(f)$ over the set of functions satisfying the following constraints.
  - $f(x) \geq 0$.
  - $\int f(x)dx = 1$.
  - $E(r_i(X)) = \int f(x)r_i(x)dx = \alpha_i$, for $i = 1, \ldots, m$.
- **Solution:**

$$f^*(x) = e^{\lambda_0 + \sum_{i=1}^m \lambda_i r_i(x)}.$$

- **Proof:** For any $g(x)$ satisfying the constraints,

$$
\begin{aligned}
h(g) &= -\int g(x)\log g(x)dx = -\int g\log\frac{g}{f^*}f^*dx \\
&= -D(g||f^*) - \int g(x)\log f^*(x)dx \\
&\leq -\int g(x)\log f^*(x)dx \\
&= -\int g(x)(\lambda_0 + \sum_{i=1}^{m}\lambda_i r_i(x))dx \\
&= -\int f^*(x)(\lambda_0 + \sum_{i=1}^{m}\lambda_i r_i(x))dx \\
&= h(f^*)
\end{aligned}
$$

Examples

- $r_i(x)$ is the exponent in the exponential. λ_i is determined by the constraints.
- Examples
  - Dice, no constraints: $p(i) = \text{const}$
  - Dice, $EX = \alpha$: $p(i) \propto e^{\lambda i}$
  - $S = [0, \infty), EX = \mu$: $f(x) \propto e^{\lambda x}$
  - $S = (-\infty, \infty), EX = \mu, EX^2 = \beta$: Gaussian $\mathcal{N}(\mu, \beta - \mu^2) \propto e^{\lambda_1 x + \lambda_2 x^2}$
  - Multivariate $EX_iX_j = K_{ij}, 1 \leq i, j \leq n$: $f(\mathbf{x}) \propto e^{\sum_{i,j} \lambda_{ij} x_i x_j}$ (Theorem 9.6.5)

Spectral Estimation

- Let $\{X_i\}$ be a zero-mean stochastic process.
- The autocorrelation function is defined by

$$R[k] = EX_iX_{i+k}$$

- The power spectral density is the Fourier transform of $R[k]$

$$S(\lambda) = \sum_k R[k]e^{-i\lambda k}, -\pi \leq \lambda \leq \pi$$

- So we can estimate the spectral density from a sample of the process.

Differential Entropy Rates

- The **differential entropy rate** of a stochastic process is defined by

$$h(\mathcal{X}) = \lim_{n\to\infty} \frac{h(X_1, \ldots, X_n)}{n},$$

if the limit exists.

- For a stationary process, the limit exists. Furthermore,

$$h(\mathcal{X}) = \lim_{n\to\infty} h(X_n | X_{n-1}, \ldots, X_1)$$

Gaussian Processes

- A Gaussian process is characterized by the property that any collection of random variables in the process is jointly Gaussian.
- For a stationary Gaussian process, we have

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

where $K^{(n)}$ is the covariance matrix for $(X_1, \ldots, X_n)$.

$$K_{ij}^{(n)} = E(X_i - EX_i)(X_j - EX_j)$$

- As $n \to \infty$, the density of eigenvalues of $K^{(n)}$ tends to a limit, which is the spectrum of the stochastic process.

Entropy Rate and Variance

- Kolmogorov showed that the entropy rate of a stationary Gaussian process is related to the spectral density by (11.39).
- So the entropy rate can be computed from the spectral density.
- Furthermore, the best estimate of X_n given the past samples (which is Gaussian) has a variance that is related to the entropy rate by (11.40).

Burg's Maximum Entropy Theorem

- The maximum entropy rate stochastic process $\{X_i\}$ satisfying the constraints

$$EX_iX_{i+k} = \alpha_k, \ k = 0, \ldots, p$$

is the p-th order Gauss-Markov process

$$X_i = -\sum_{k=1}^{p} a_k X_{i-k} + Z_i,$$

where the $Z_i's$ are i.i.d. zero-mean Gaussians $N(0, \sigma^2)$. The a_i's and σ are chosen to satisfy the constraints.

- See the text for the proof.

Yule-Walker Equations

- To choose a_k and σ, one solves the Yule-Walker equations as given in (11.51) and (11.52), which are obtained by multiplying (11.42) by $X_{i-l}, l = 0, 1, \ldots, p$ and taking expectation values.
- The Yule-Walker equations can be solved efficiently by the **Levinson-Durbin** recursions.
- The spectrum of the maximum entropy process is

$$S(l) = \frac{\sigma^2}{|1 + \sum a_k e^{-ikl}|^2},$$

which can be obtained from (11.51).