An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Sept. 5, 2016

Lecture #13: Rate Distortion Theory

Outline of the lecture

- Introduction

Outline of the lecture

- Introduction
- Calculation of rate distortion function

Outline of the lecture

- Introduction
- Calculation of rate distortion function
- Simultaneous description of independent Gaussian RVs.

Introduction

- Distortion measure is a measure of distance between the random variable and its representation.

Introduction

- Distortion measure is a measure of distance between the random variable and its representation.
- A distortion measure is a mapping

$$d : X \times \hat{X} \rightarrow R^+$$

from the set of source alphabet-reproduction alphabet pairs into the set of nonnegative real numbers.

Introduction

- Distortion measure is a measure of distance between the random variable and its representation.
- A distortion measure is a mapping

$$d : X \times \hat{X} \rightarrow R^+$$

from the set of source alphabet-reproduction alphabet pairs into the set of nonnegative real numbers.

- A distortion measure is said to be bounded if the maximum value of the distortion is finite

$$d_{\max} = \max_{x \in X, \hat{x} \in \hat{X}} d(x, \hat{x}) < \infty$$

Introduction

- Examples of common distortion functions:

Introduction

- Examples of common distortion functions:
  - Hamming distortion

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

Introduction

- Examples of common distortion functions:
  - Hamming distortion

$$d(x,\hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

  - Squared error distortion

$$d(x,\hat{x}) = (x-\hat{x})^2$$

Introduction

- Examples of common distortion functions:
  - Hamming distortion

$$d(x,\hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

  - Squared error distortion

$$d(x,\hat{x}) = (x-\hat{x})^2$$

- The distortion between sequences x^n and $\hat{x}^n$ is defined by

$$d(x^n,\hat{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i,\hat{x}_i)$$

Introduction

- Examples of common distortion functions:
  - Hamming distortion

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

  - Squared error distortion

$$d(x, \hat{x}) = (x - \hat{x})^2$$

- The distortion between sequences x^n and $\hat{x}^n$ is defined by

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^{n} d(x_i, \hat{x}_i)$$

- We are interested to know: given a source distribution and a distortion measure, what is the minimum expected distortion achievable at a particular rate?

Definition

- A $(2^{nR}, n)$ rate distortion code consists of following encoding function

$$f_n : X^n \rightarrow \{1, 2, \cdots, 2^{nR}\}$$

and following decoding function

$$g_n : \{1, 2, \cdots, 2^{nR}\} \rightarrow \hat{X}^n$$

Definition

- A $(2^{nR}, n)$ rate distortion code consists of following encoding function
$$f_n : X^n \to \{1, 2, \cdots, 2^{nR}\}$$
and following decoding function
$$g_n : \{1, 2, \cdots, 2^{nR}\} \to \hat{X}^n$$
- Distortion for the $(2^{nR}, n)$ rate distortion code is defined as
$$D = \sum_{x^n} p(x^n) d(x^n, g_n(f_n(x^n)))$$

Definition

- A rate distortion pair (R, D) is said to be achievable if there exists sequence of $(2^{nR}, n)$ rate distortion codes (f_n, g_n) with $\lim_{n\to\infty} Ed(X^n, g_n(f_n(X^n))) \leq D$.

Definition

- A rate distortion pair (R, D) is said to be achievable if there exists sequence of $(2^{nR}, n)$ rate distortion codes (f_n, g_n) with $\lim_{n\to\infty} Ed(X^n, g_n(f_n(X^n))) \leq D$.
- The rate distortion region for a source is the closure of the set of achievable rate distortion pairs (R, D).

Definition

- A rate distortion pair (R, D) is said to be achievable if there exists sequence of $(2^{nR}, n)$ rate distortion codes (f_n, g_n) with $\lim_{n\to\infty} Ed(X^n, g_n(f_n(X^n))) \leq D$.
- The rate distortion region for a source is the closure of the set of achievable rate distortion pairs (R, D).
- The rate distortion function R(D) is the infimum of rates R such that (R, D) is in the rate distortion region of the source for a given distortion D.

Definition

- A rate distortion pair (R, D) is said to be achievable if there exists sequence of $(2^{nR}, n)$ rate distortion codes (f_n, g_n) with $\lim_{n \to \infty} Ed(X^n, g_n(f_n(X^n))) \leq D$.
- The rate distortion region for a source is the closure of the set of achievable rate distortion pairs (R, D).
- The rate distortion function R(D) is the infimum of rates R such that (R, D) is in the rate distortion region of the source for a given distortion D.
- The distortion rate function D(R) is the infimum of all distortions D such that (R,D) is in the rate distortion region of the source for a given rate R.

Definition

- The information rate distortion function $R^I(D)$ for a source X with distortion measure $d(x, \hat{x})$ is defined as

$$R^I(D) = \min_{p(\hat{x}|x): \sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D} I(X; \hat{X})$$

where minimization is over all conditional distributions $p(\hat{x}|x)$ for which the joint distribution $p(x, \hat{x}) = p(x)p(\hat{x}|x)$ satisfies the expected distortion constraint.

Definition

- The information rate distortion function $R^I(D)$ for a source X with distortion measure $d(x, \hat{x})$ is defined as

$$R^I(D) = \min_{p(\hat{x}|x): \sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D} I(X; \hat{X})$$

where minimization is over all conditional distributions $p(\hat{x}|x)$ for which the joint distribution $p(x, \hat{x}) = p(x)p(\hat{x}|x)$ satisfies the expected distortion constraint.

- The rate distortion function for an i.i.d. source X with distribution p(x) and bounded distortion function $d(x, \hat{x})$ is equal to the associated information rate distortion function. Thus

$$R(D) = R^I(D) = \min_{p(\hat{x}|x): \sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x}) \leq D} I(X; \hat{X})$$

is the minimum achievable rate at distortion D.

Rate Distortion Function

- Let X be $N(0, \sigma^2)$, By rate distortion theorem, we have

$$R(D) = \min_{f(\hat{x}|x): E(\hat{X}-X)^2 \leq D} I(X, \hat{X})$$

Rate Distortion Function

- Let X be $N(0, \sigma^2)$, By rate distortion theorem, we have

$$R(D) = \min_{f(\hat{x}|x):E(\hat{X}-X)^2 \le D} I(X, \hat{X})$$

- We know that

$$\begin{aligned} I(X, \hat{X}) &= h(X) - h(X|\hat{X}) \\ &= \frac{1}{2}\log(2\pi e)\sigma^2 - h(X - \hat{X}|\hat{X}) \\ &\ge \frac{1}{2}\log(2\pi e)\sigma^2 - h(X - \hat{X}) \\ &\ge \frac{1}{2}\log(2\pi e)\sigma^2 - h(N(0, E(X - \hat{X})^2)) \\ &= \frac{1}{2}\log(2\pi e)\sigma^2 - \frac{1}{2}\log(2\pi e)E(X - \hat{X})^2 \\ &\ge \frac{1}{2}\log(2\pi e)\sigma^2 - \frac{1}{2}\log(2\pi e)D \\ &= \frac{1}{2}\log\frac{\sigma^2}{D} \end{aligned}$$

Rate Distortion Function

- If $D \le \sigma^2$, we choose $X = \hat{X} + Z$, $\hat{X} \sim N(0, \sigma^2 - D), Z \sim N(0, D)$, where $\hat{X}$ and Z are independent. For this we have

$$I(X, \hat{X}) = \frac{1}{2}\log\frac{\sigma^2}{D}$$

and $E(X - \hat{X})^2 = D$.

Rate Distortion Function

- If $D \leq \sigma^2$, we choose $X = \hat{X} + Z$, $\hat{X} \sim N(0, \sigma^2 - D), Z \sim N(0, D)$, where $\hat{X}$ and Z are independent. For this we have

$$I(X, \hat{X}) = \frac{1}{2} \log \frac{\sigma^2}{D}$$

  and $E(X - \hat{X})^2 = D$.
- If $D > \sigma^2$, we chose X=0 with probability 1, achieving R(D)=0.

Rate Distortion Function

- If $D \leq \sigma^2$, we choose $X = \hat{X} + Z$, $\hat{X} \sim N(0, \sigma^2 - D), Z \sim N(0, D)$, where $\hat{X}$ and Z are independent. For this we have

$$I(X, \hat{X}) = \frac{1}{2} \log \frac{\sigma^2}{D}$$

  and $E(X - \hat{X})^2 = D$.
- If $D > \sigma^2$, we chose X=0 with probability 1, achieving R(D)=0.
- We can rewrite the distortion in terms of the rate

$$D(R) = \sigma^2 2^{-2R}$$

Rate Distortion Function

- The rate distortion function for a Bernoulli(p) source with Hamming distortion is given by

$$R(D) = \begin{cases} H(p) - H(D) & 0 \leq D \leq \min\{p, 1-p\} \\ 0 & D > \min\{p, 1-p\} \end{cases}$$

Rate Distortion Function

- The rate distortion function for a Bernoulli(p) source with Hamming distortion is given by

$$R(D) = \begin{cases} H(p) - H(D) & 0 \leq D \leq \min\{p, 1-p\} \\ 0 & D > \min\{p, 1-p\} \end{cases}$$

- X $\sim$ Bernoulli(p). Without loss of generality, we consider $p < \frac{1}{2}$. $X \oplus \hat{X} = 1$ is equivalent to $X \neq \hat{X}$.

Rate Distortion Function

- We first will find a lower bound on $I(X,\hat{X})$ and then show that this lower bound is achievable.

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(p) - H(X \oplus \hat{X}|\hat{X}) \\ &\geq H(p) - H(X \oplus \hat{X}) \\ &\geq H(p) - H(D) \end{aligned}$$

since $Pr(X \neq \hat{X}) \leq D$ and H(D) increases with D for $D \leq \frac{1}{2}$.

Rate Distortion Function

- We first will find a lower bound on $I(X,\hat{X})$ and then show that this lower bound is achievable.

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= H(p) - H(X \oplus \hat{X}|\hat{X}) \\ &\geq H(p) - H(X \oplus \hat{X}) \\ &\geq H(p) - H(D) \end{aligned}$$

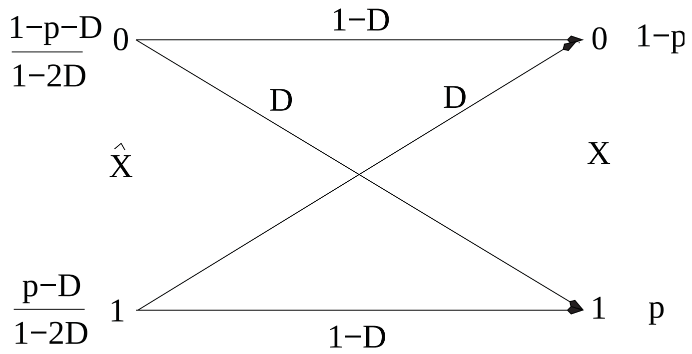
since $Pr(X \neq \hat{X}) \leq D$ and H(D) increases with D for $D \leq \frac{1}{2}$.

- Thus

$$R(D) \leq H(p) - H(D)$$

Rate Distortion Function

- For $0 \leq D \leq p$, we can achieve the value of the rate distortion function by choosing $(X, \hat{X})$ to have joint distribution given by the binary symmetric channel as shown below



Rate Distortion Function

- We choose the distribution of $\hat{X}$ at the input of the channel so that the output distribution of X is the specified distribution. Let $r = Pr(\hat{X} = 1)$. We choose r such that $r(1 - D) + (1 - r)D = p$ or $r = \frac{p-D}{1-2D}$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Rate Distortion Function

- We choose the distribution of $\hat{X}$ at the input of the channel so that the output distribution of X is the specified distribution. Let $r = Pr(\hat{X} = 1)$. We choose r such that $r(1-D) + (1-r)D = p$ or $r = \frac{p-D}{1-2D}$.
- If $D \leq p \leq \frac{1}{2}$, then $Pr(\hat{X} = 1) \geq 0$ and $Pr(\hat{X} = 0) \geq 0$. We have

$$I(X;\hat{X}) = H(X) - H(X|\hat{X}) = H(p) - H(D).$$

  and expected distortion is $P(X \neq \hat{X}) = D$.

Rate Distortion Function

- We choose the distribution of $\hat{X}$ at the input of the channel so that the output distribution of X is the specified distribution. Let $r = Pr(\hat{X} = 1)$. We choose r such that $r(1-D) + (1-r)D = p$ or $r = \frac{p-D}{1-2D}$.
- If $D \leq p \leq \frac{1}{2}$, then $Pr(\hat{X} = 1) \geq 0$ and $Pr(\hat{X} = 0) \geq 0$. We have

$$I(X;\hat{X}) = H(X) - H(X|\hat{X}) = H(p) - H(D).$$

  and expected distortion is $P(X \neq \hat{X}) = D$.
- If $D \geq p$, we achieve $R(D) = 0$ by letting $\hat{X} = 0$ with probability 1. In this case $I(X;\hat{X}) = 0$ and $D = p$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Rate Distortion Function

- We choose the distribution of $\hat{X}$ at the input of the channel so that the output distribution of X is the specified distribution. Let $r = Pr(\hat{X} = 1)$. We choose r such that $r(1-D) + (1-r)D = p$ or $r = \frac{p-D}{1-2D}$.
- If $D \leq p \leq \frac{1}{2}$, then $Pr(\hat{X} = 1) \geq 0$ and $Pr(\hat{X} = 0) \geq 0$. We have

$$I(X;\hat{X}) = H(X) - H(X|\hat{X}) = H(p) - H(D).$$

  and expected distortion is $P(X \neq \hat{X}) = D$.
- If $D \geq p$, we achieve $R(D) = 0$ by letting $\hat{X} = 0$ with probability 1. In this case $I(X;\hat{X}) = 0$ and $D = p$.
- Similarly, if $D \geq 1-p$, we can achieve $R(D) = 0$ by setting $\hat{X} = 1$ with probability 1.

Rate Distortion Function

- Consider a source X uniformly distributed on the set $\{1, 2, \cdots, m\}$, Find the rate distortion function for this source with Hamming distortion, i.e.

$$d(x,\hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

Rate Distortion Function

- Consider a source X uniformly distributed on the set $\{1, 2, \cdots, m\}$, Find the rate distortion function for this source with Hamming distortion, i.e.

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

- Consider any joint distribution that satisfies the distortion constraint D.

Rate Distortion Function

- Consider a source X uniformly distributed on the set $\{1, 2, \cdots, m\}$, Find the rate distortion function for this source with Hamming distortion, i.e.

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

- Consider any joint distribution that satisfies the distortion constraint D.
- Since $D = Pr\{X \neq \hat{X}\}$, we have by Fano's inequality

$$H(X|\hat{X}) \leq H(D) + D\log(m-1)$$

Rate Distortion Function

- Consider a source X uniformly distributed on the set $\{1, 2, \cdots, m\}$, Find the rate distortion function for this source with Hamming distortion, i.e.

$$d(x, \hat{x}) = \begin{cases} 0 & \text{if } x = \hat{x} \\ 1 & \text{if } x \neq \hat{x} \end{cases}$$

- Consider any joint distribution that satisfies the distortion constraint D.
- Since $D = Pr\{X \neq \hat{X}\}$, we have by Fano's inequality

$$H(X|\hat{X}) \leq H(D) + D \log(m-1)$$

- Thus, we have

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &\geq \log m - H(D) - D\log(m-1) \end{aligned}$$

Rate Distortion Function

- We can achieve this lower bound by choosing $p(\hat{x})$ to be uniform distribution and conditional distribution of $p(\hat{x}|x)$ to be

$$p(\hat{x}|x) = \begin{cases} 1 - D & \text{if } \hat{x} = x \\ D/(m-1) & \text{if } \hat{x} \neq x \end{cases}$$

Rate Distortion Function

- We can achieve this lower bound by choosing $p(\hat{x})$ to be uniform distribution and conditional distribution of $p(\hat{x}|x)$ to be

$$p(\hat{x}|x) = \begin{cases} 1 - D & \text{if } \hat{x} = x \\ D/(m-1) & \text{if } \hat{x} \neq x \end{cases}$$

- The bound is satisfied with equality for $D \leq 1 - \frac{1}{m}$, hence

$$R(D) = \begin{cases} \log m - H(D) - D\log(m-1) & \text{if } 0 \leq D \leq 1 - \frac{1}{m} \\ 0 & \text{if } D > 1 - \frac{1}{m} \end{cases}$$

Rate Distortion Function

- Information rate distortion function is defined as

$$R(D) = \min_{q(\hat{x}|x): \sum_{(x,\hat{x})} p(x)q(\hat{x}|x)d(x,\hat{x}) \leq D} I(X;\hat{X})$$

where the minimization is over all conditional distributions $q(\hat{x}|x)$ for which the joint distribution $p(x)q(\hat{x}|x)$ satisfies the expected distortion constraint.

Rate Distortion Function

- Information rate distortion function is defined as

$$R(D) = \min_{q(\hat{x}|x):\sum_{(x,\hat{x})} p(x)q(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X})$$

  where the minimization is over all conditional distributions $q(\hat{x}|x)$ for which the joint distribution $p(x)q(\hat{x}|x)$ satisfies the expected distortion constraint.
- This is a minimization of a convex function over the convex set of all $q(\hat{x}|x) \geq 0$ satisfying $\sum_{\hat{x}} q(\hat{x}|x) = 1$ for all x and $\sum q(\hat{x}|x)p(x)d(x,\hat{x}) \leq D$

Rate Distortion Function

- Information rate distortion function is defined as

$$R(D) = \min_{q(\hat{x}|x):\sum_{(x,\hat{x})} p(x)q(\hat{x}|x)d(x,\hat{x})\leq D} I(X;\hat{X})$$

  where the minimization is over all conditional distributions $q(\hat{x}|x)$ for which the joint distribution $p(x)q(\hat{x}|x)$ satisfies the expected distortion constraint.
- This is a minimization of a convex function over the convex set of all $q(\hat{x}|x) \geq 0$ satisfying $\sum_{\hat{x}} q(\hat{x}|x) = 1$ for all x and $\sum q(\hat{x}|x)p(x)d(x,\hat{x}) \leq D$
- Using Lagrange multiplier method, we setup the functional

$$\begin{aligned} J(q) &= \sum_{x}\sum_{\hat{x}} p(x)q(\hat{x}|x)\log\frac{q(\hat{x}|x)}{\sum_x p(x)q(\hat{x}|x)} \\ &+ \lambda\sum_{x}\sum_{\hat{x}} p(x)q(\hat{x}|x)d(x,\hat{x}) + \sum_{x}\nu(x)\sum_{\hat{x}} q(\hat{x}|x) \end{aligned}$$

Rate Distortion Function

- If we let $q(\hat{x}) = \sum_x p(x)q(\hat{x}|x)$ be the distribution on $\hat{X}$ induced by $q(\hat{x}|x)$, we can write J(q) as

$$\begin{aligned} J(q) &= \sum_x \sum_{\hat{x}} p(x)q(\hat{x}|x) \log \frac{q(\hat{x}|x)}{q(\hat{x})} \\ &+ \lambda \sum_x \sum_{\hat{x}} p(x)q(\hat{x}|x)d(x,\hat{x}) + \sum_x \nu(x) \sum_{\hat{x}} q(\hat{x}|x) \end{aligned}$$

Rate Distortion Function

- If we let $q(\hat{x}) = \sum_x p(x)q(\hat{x}|x)$ be the distribution on $\hat{X}$ induced by $q(\hat{x}|x)$, we can write J(q) as

$$\begin{aligned} J(q) &= \sum_x \sum_{\hat{x}} p(x)q(\hat{x}|x) \log \frac{q(\hat{x}|x)}{q(\hat{x})} \\ &+ \lambda \sum_x \sum_{\hat{x}} p(x)q(\hat{x}|x)d(x,\hat{x}) + \sum_x \nu(x) \sum_{\hat{x}} q(\hat{x}|x) \end{aligned}$$

- Differentiating with respect to $q(\hat{x}|x)$, we have

$$\begin{aligned} \frac{\partial J}{\partial q(\hat{x}|x)} &= p(x) \log \frac{q(\hat{x}|x)}{q(\hat{x})} + p(x) - \sum_{x'} p(x')q(\hat{x}|x') \frac{1}{q(\hat{x})} p(x) \\ &+ \lambda p(x)d(x,\hat{x}) + \nu(x) = 0 \end{aligned}$$

Rate Distortion Function

- If we let $q(\hat{x}) = \sum_x p(x)q(\hat{x}|x)$ be the distribution on $\hat{X}$ induced by $q(\hat{x}|x)$, we can write J(q) as

$$\begin{aligned} J(q) &= \sum_x \sum_{\hat{x}} p(x)q(\hat{x}|x)\log\frac{q(\hat{x}|x)}{q(\hat{x})} \\ &+ \lambda \sum_x \sum_{\hat{x}} p(x)q(\hat{x}|x)d(x,\hat{x}) + \sum_x \nu(x) \sum_{\hat{x}} q(\hat{x}|x) \end{aligned}$$

- Differentiating with respect to $q(\hat{x}|x)$, we have

$$\begin{aligned} \frac{\partial J}{\partial q(\hat{x}|x)} &= p(x)\log\frac{q(\hat{x}|x)}{q(\hat{x})} + p(x) - \sum_{x'} p(x')q(\hat{x}|x')\frac{1}{q(\hat{x})}p(x) \\ &+ \lambda p(x)d(x,\hat{x}) + \nu(x) = 0 \end{aligned}$$

- Setting $\log\mu(x) = \nu(x)/p(x)$, we obtain

$$p(x)\left[\log\frac{q(\hat{x}|x)}{q(\hat{x})} + \lambda d(x,\hat{x}) + \log\mu(x)\right] = 0$$

Rate Distortion Function

- Hence

$$q(\hat{x}|x) = \frac{q(\hat{x})e^{-\lambda d(x,\hat{x})}}{\mu(x)}$$

item Since $\sum_{\hat{x}} q(\hat{x}|x) = 1$, we have

$$\mu(x) = \sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}$$

or

$$q(\hat{x}|x) = \frac{q(\hat{x})e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}}$$

Rate Distortion Function

- Hence

$$q(\hat{x}|x) = \frac{q(\hat{x})e^{-\lambda d(x,\hat{x})}}{\mu(x)}$$

item Since $\sum_{\hat{x}} q(\hat{x}|x) = 1$, we have

$$\mu(x) = \sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}$$

or

$$q(\hat{x}|x) = \frac{q(\hat{x})e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}}$$

- Multiplying both sides by p(x) and summing over all x, we get

$$q(\hat{x}) = q(\hat{x}) \sum_{x} \frac{p(x)e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}}$$

Rate Distortion Function

- If $q(\hat{x}) > 0$, we obtain

$$\sum_{x} \frac{p(x)e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}} = 1$$

Rate Distortion Function

- If $q(\hat{x}) > 0$, we obtain

$$\sum_x \frac{p(x)e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}} = 1$$

- Applying Kuhn Tucker conditions, we have

$$\frac{\partial J}{\partial q(\hat{x}|x)} \begin{cases} = 0 & \text{if} q(\hat{x}|x) > 0 \\ \geq 0 & \text{if} q(\hat{x}|x) = 0 \end{cases}$$

Rate Distortion Function

- If $q(\hat{x}) > 0$, we obtain

$$\sum_x \frac{p(x)e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}} = 1$$

- Applying Kuhn Tucker conditions, we have

$$\frac{\partial J}{\partial q(\hat{x}|x)} \begin{cases} = 0 & \text{if} q(\hat{x}|x) > 0 \\ \geq 0 & \text{if} q(\hat{x}|x) = 0 \end{cases}$$

- Substituting the value of the derivative we get

$$\begin{aligned} \sum_x \frac{p(x)e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}} &= 1 \quad if \quad q(\hat{x}) > 0 \\ \sum_x \frac{p(x)e^{-\lambda d(x,\hat{x})}}{\sum_{\hat{x}} q(\hat{x})e^{-\lambda d(x,\hat{x})}} &\leq 1 \quad if \quad q(\hat{x}) = 0 \end{aligned}$$

Simultaneous description of independent Gaussian RVs.

- Consider m independent normal random sources $X_1, \cdots, X_m$, where X_i are $\sim N(0, \sigma_i^2)$ with squared error distortion.

Simultaneous description of independent Gaussian RVs.

- Consider m independent normal random sources $X_1, \cdots, X_m$, where X_i are $\sim N(0, \sigma_i^2)$ with squared error distortion.
- Assume that we are given R bits with which to represent this random vector. How should the bits be alloted to various components to minimize the total distortion.

Simultaneous description of independent Gaussian RVs.

- Consider m independent normal random sources $X_1, \cdots, X_m$, where X_i are $\sim N(0, \sigma_i^2)$ with squared error distortion.
- Assume that we are given R bits with which to represent this random vector. How should the bits be alloted to various components to minimize the total distortion.
- We have

$$R(D) = \min_{f(\hat{x}^m|x^m):Ed(X^m,\hat{X}^m)\leq D} I(X^m;\hat{X}^m)$$

where $d(x^m, \hat{x}^m) = \sum_{i=1}^m (x_i - \hat{x}_i)^2$

Simultaneous description of independent Gaussian RVs.

- We have

$$\begin{aligned}
I(X^m;\hat{X}^m) &= h(X^m) - h(X^m|\hat{X}^m) \\
&= \sum_{i=1}^m h(X_i) - \sum_{i=1}^m h(X_i|X^{i-1},\hat{X}^m) \\
&\geq \sum_{i=1}^m h(X_i) - \sum_{i=1}^m h(X_i|\hat{X}_i) \\
&= \sum_{i=1}^m I(X_i;\hat{X}_i) \\
&\geq \sum_{i=1}^m R(D_i) \\
&= \sum_{i=1}^m \left(\frac{1}{2}\log\frac{\sigma_i^2}{D_i}\right)^+
\end{aligned}$$

where $D_i = E(X_i - \hat{X}_i)^2$

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Simultaneous description of independent Gaussian RVs.

- Hence the problem of finding the rate distortion function can be reduced to

$$R(D) = \min_{D_i = D} \sum_{i=1}^{m} \max \left\{ \frac{1}{2} \log \frac{\sigma_i^2}{D_i}, 0 \right\}$$

Simultaneous description of independent Gaussian RVs.

- Hence the problem of finding the rate distortion function can be reduced to

$$R(D) = \min_{D_i = D} \sum_{i=1}^{m} \max \left\{ \frac{1}{2} \log \frac{\sigma_i^2}{D_i}, 0 \right\}$$

- Using Lagrange multipliers, we construct the functional as

$$J(D) = \sum_{i=1}^{m} \frac{1}{2} \log \frac{\sigma_i^2}{D_i} + \lambda \sum_{i=1}^{m} D_i$$

Simultaneous description of independent Gaussian RVs.

- Hence the problem of finding the rate distortion function can be reduced to

$$R(D) = \min_{D_i = D} \sum_{i=1}^{m} \max\left\{\frac{1}{2}\log\frac{\sigma_i^2}{D_i}, 0\right\}$$

- Using Lagrange multipliers, we construct the functional as

$$J(D) = \sum_{i=1}^{m} \frac{1}{2}\log\frac{\sigma_i^2}{D_i} + \lambda \sum_{i=1}^{m} D_i$$

- Differentiating with respect to D_i, and equating to 0, we get

$$\frac{\partial J}{\partial D_i} = -\frac{1}{2}\frac{1}{D_i} + \lambda = 0 \implies D_i = \lambda'$$

Simultaneous description of independent Gaussian RVs.

- Let $X_i \sim N(0, \sigma_i^2), i = 1, 2, \cdots, m$ be independent Gaussian random variables, and let the distortion measure be $d(x^m, \hat{x}^m) = \sum_{i=1}^{m}(x_i - \hat{x}_i)^2$.

Simultaneous description of independent Gaussian RVs.

- Let $X_i \sim N(0, \sigma_i^2), i = 1, 2, \cdots, m$ be independent Gaussian random variables, and let the distortion measure be $d(x^m, \hat{x}^m) = \sum_{i=1}^{m} (x_i - \hat{x}_i)^2$.
- Then the rate distortion function is given by

$$R(D) = \sum_{i=1}^{m} \frac{1}{2} \log \frac{\sigma_i^2}{D_i}$$

where

$$D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2 \\ \sigma_i^2 & \text{if } \lambda \geq \sigma_i^2 \end{cases}$$

where λ is chosen so that $\sum_{i=1}^{m} D_i = D$.

Simultaneous description of independent Gaussian RVs.

- Let $X_i \sim N(0, \sigma_i^2), i = 1, 2, \cdots, m$ be independent Gaussian random variables, and let the distortion measure be $d(x^m, \hat{x}^m) = \sum_{i=1}^{m} (x_i - \hat{x}_i)^2$.
- Then the rate distortion function is given by

$$R(D) = \sum_{i=1}^{m} \frac{1}{2} \log \frac{\sigma_i^2}{D_i}$$

where

$$D_i = \begin{cases} \lambda & \text{if } \lambda < \sigma_i^2 \\ \sigma_i^2 & \text{if } \lambda \geq \sigma_i^2 \end{cases}$$

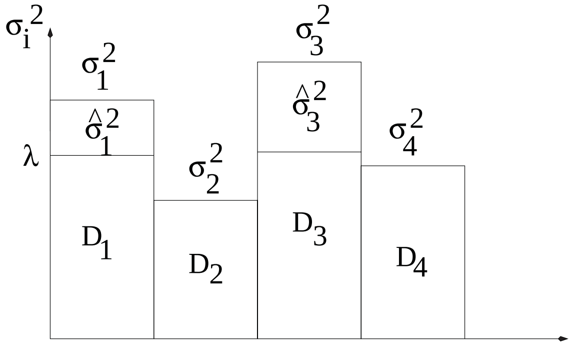
where λ is chosen so that $\sum_{i=1}^{m} D_i = D$.
- This gives rise to a kind of reverse waterfilling.

Simultaneous description of independent Gaussian RVs.

- We choose a constant λ and only describe those random variables with variances greater than λ.

Simultaneous description of independent Gaussian RVs.

- We choose a constant λ and only describe those random variables with variances greater than λ.
- No bits are used to describe random variables with variance less than λ.



Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory