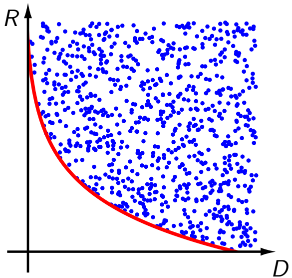
Rate-Distortion Theory



Last Lecture: Scalar Quantization & Lloyd Quantizer

Scalar Quantization

- Individual quantization of each input sample: $s' = Q(s)$
- Quantizer is characterized by K reconstruction levels s'_k and $K-1$ decision thresholds u_k

Lloyd Quantizer

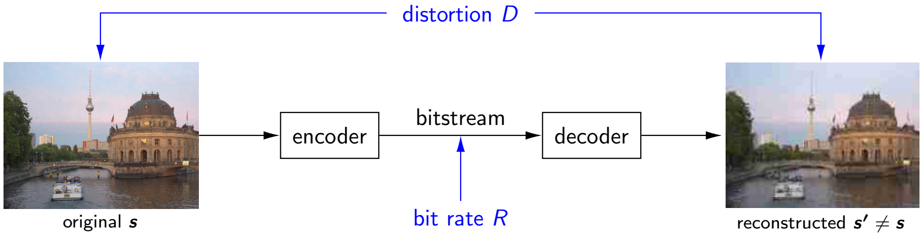
- Minimizes distortion D for given number K of quantization intervals
- Two optimization criterions
  - Centroid condition (MSE): $s'_k = \mathrm{E}\{\, S \,|\, S \in \mathcal{I}_k \,\}$
  - Nearest neighbor condition (MSE): $u_k = (s'_k + s'_{k-1})/2$
- Lloyd quantizer design: Iterate between the two optimization criterions

High-Rate Approximation: Lloyd Quantizer and Fixed-Length Coding

- Panter and Dite approximation for operational distortion-rate function

$$D_F(R) = \varepsilon_F^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_F^2 = \frac{1}{12\sigma^2} \left(\int_{-\infty}^{\infty} \sqrt[3]{f(s)} \,\mathrm{d}s \right)^3$$

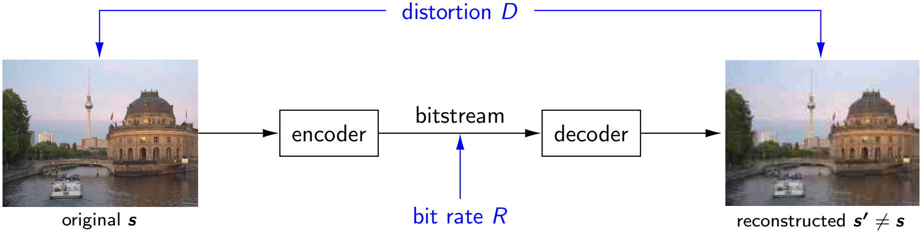
Rate-Distortion Theory: Theoretical Bounds for Lossy Coding



Maximum Coding Efficiency for given Source ?

- What is the minimum bit rate $R(D)$ required for achieving a maximum distortion D ?
- What is the minimum possible distortion $D(R)$ when coding at a maximum rate R ?

Rate-Distortion Theory: Theoretical Bounds for Lossy Coding

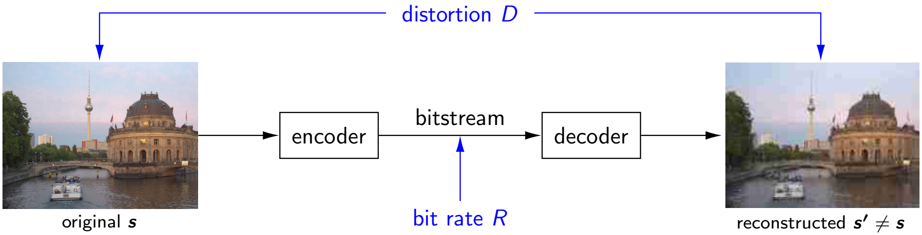


Maximum Coding Efficiency for given Source ?

- What is the minimum bit rate $R(D)$ required for achieving a maximum distortion D ?
- What is the minimum possible distortion $D(R)$ when coding at a maximum rate R ?

➜ **Rate-distortion theory** gives answers without considering any specific coding method

Rate-Distortion Theory: Theoretical Bounds for Lossy Coding



Maximum Coding Efficiency for given Source ?

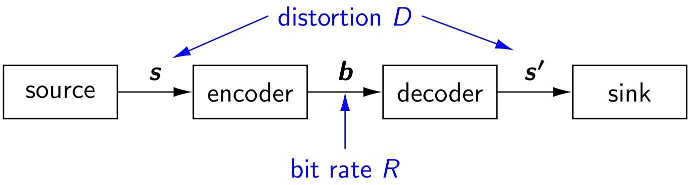
- What is the minimum bit rate $R(D)$ required for achieving a maximum distortion D ?
- What is the minimum possible distortion $D(R)$ when coding at a maximum rate R ?

➡ **Rate-distortion theory** gives answers without considering any specific coding method

➡ Require a probabilistic model of the source (multi-variate pdf)

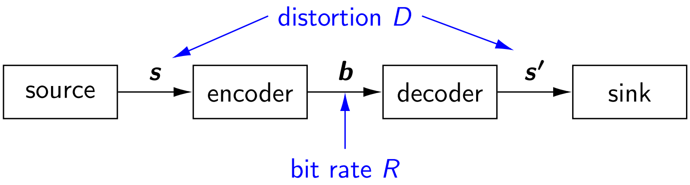
Probabilistic Modeling for Lossy Source Coding

- Source: Statistical properties of source signal $\boldsymbol{s}$ are characterized by random process $\boldsymbol{S}$
- Encoder: Deterministic process for generating bitstream $\boldsymbol{b} = f_{\text{enc}}(\boldsymbol{s})$
- Decoder: Deterministic process for generating reconstructed signal $\boldsymbol{s}' = f_{\text{dec}}(\boldsymbol{b})$
- Sink: Receiver of reconstructed signal $\boldsymbol{s}'$



Probabilistic Modeling for Lossy Source Coding

- Source: Statistical properties of source signal $\boldsymbol{s}$ are characterized by random process $\boldsymbol{S}$
- Encoder: Deterministic process for generating bitstream $\boldsymbol{b} = f_{\text{enc}}(\boldsymbol{s})$
- Decoder: Deterministic process for generating reconstructed signal $\boldsymbol{s}' = f_{\text{dec}}(\boldsymbol{b})$
- Sink: Receiver of reconstructed signal $\boldsymbol{s}'$

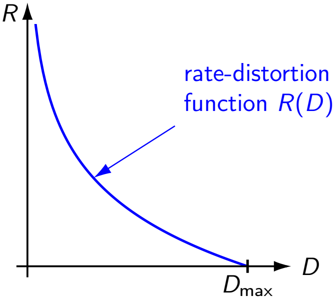


Probabilistic Modeling for Source

- Stationary random process $\boldsymbol{S} = \{\cdots, S_k, S_{k+1}, S_{k+2}, \cdots\}$ with given N-dimensional pdf

$$f_N(\boldsymbol{s}) = f(s_k, s_{k+1}, \cdots, s_{k+N-1})$$

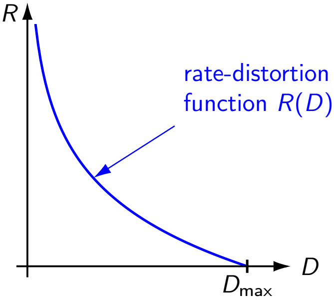
Rate-Distortion Function $R(D)$ and Distortion-Rate Function $D(R)$

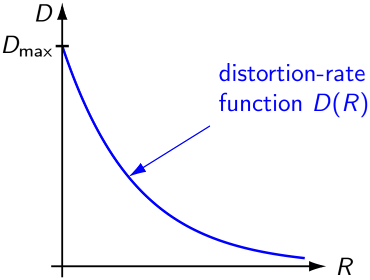


Rate-Distortion Function $R(D)$

- Specifies minimum required bit rate R for given maximum distortion D
- Property of the source

Rate-Distortion Function $R(D)$ and Distortion-Rate Function $D(R)$





Rate-Distortion Function $R(D)$

- Specifies minimum required bit rate R for given maximum distortion D
- Property of the source

Distortion-Rate Function $D(R)$

- Specifies minimum achievable distortion D for given maximum bit rate R
- Inverse of rate-distortion function $R(D)$

First Definition: Operational Rate-Distortion Function

Source given by random process S

- Each source code Q is associated with a rate-distortion point

$$(R, D) = \big(\, r(Q),\, \delta(Q) \,\big)$$

Note: For a given source $\boldsymbol{S}$ and known encoder and decoder, $r(Q)$ and $\delta(Q)$ represent probabilistic averages for the rate and distortion

First Definition: Operational Rate-Distortion Function

Source given by random process S

- Each source code Q is associated with a rate-distortion point

$$(R, D) = \big(r(Q), \delta(Q) \big)$$

Note: For a given source $\boldsymbol{S}$ and known encoder and decoder, $r(Q)$ and $\delta(Q)$ represent probabilistic averages for the rate and distortion

➡ **Achievable rate distortion point** (R, D)

$$\exists Q : \quad r(Q) \leq R \quad \wedge \quad \delta(Q) \leq D$$

First Definition: Operational Rate-Distortion Function

Source given by random process S

- Each source code Q is associated with a rate-distortion point

$$(R, D) = \big(\, r(Q),\, \delta(Q) \,\big)$$

Note: For a given source $\boldsymbol{S}$ and known encoder and decoder, $r(Q)$ and $\delta(Q)$ represent probabilistic averages for the rate and distortion

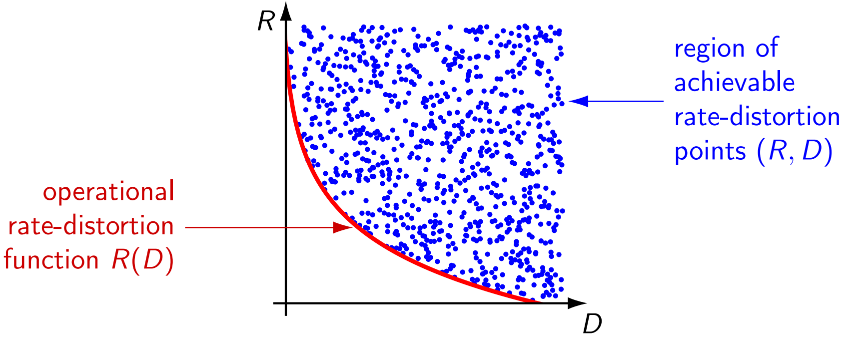
➡ Achievable rate distortion point (R, D)

$$\exists Q : \quad r(Q) \le R \quad \wedge \quad \delta(Q) \le D$$

➡ Operational rate-distortion function $R(D)$ and distortion-rate function $D(R)$

$$\boxed{R(D) = \inf_{Q:\, \delta(Q) \le D} r(Q)} \qquad \text{and} \qquad \boxed{D(R) = \inf_{Q:\, r(Q) \le R} \delta(Q)}$$

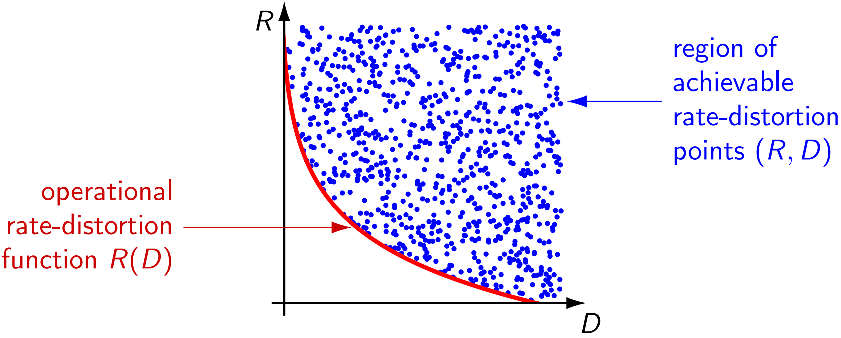
Illustration: Operational Rate-Distortion Function



Operational rate-distortion function $R(D)$ / **distortion-rate function** $D(R)$

- Divides R-D space into region of achievable and region of non-achievable rate-distortion points
- Specifies the greatest lower bound for lossy coding

Illustration: Operational Rate-Distortion Function



Operational rate-distortion function $R(D)$ / **distortion-rate function** $D(R)$

- Divides R-D space into region of achievable and region of non-achievable rate-distortion points
- Specifies the greatest lower bound for lossy coding

➡ **Impossible to evaluate** (minimization over all possible codes)

Mutual Information for Discrete Random Variables

Definition: Mutual Information (discrete case)

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B)$$

Mutual Information for Discrete Random Variables

Definition: Mutual Information (discrete case)

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B)$$

Interpretation

- $H(A)$: Uncertainty about random variable A

Mutual Information for Discrete Random Variables

Definition: Mutual Information (discrete case)

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B)$$

Interpretation

- $H(A)$: Uncertainty about random variable A
- $H(A\,|\,B)$: Uncertainty about random variable A after observing B

Mutual Information for Discrete Random Variables

Definition: Mutual Information (discrete case)

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B)$$

Interpretation

- $H(A)$: Uncertainty about random variable A
- $H(A\,|\,B)$: Uncertainty about random variable A after observing B

➞ $I(A;B)$: Reduction of uncertainty about A due to the observation of B

Mutual Information for Discrete Random Variables

Definition: Mutual Information (discrete case)

- Mutual information between two discrete random variables A and B

$$I(A;B) = H(A) - H(A\,|\,B)$$

Interpretation

- $H(A)$: Uncertainty about random variable A
- $H(A\,|\,B)$: Uncertainty about random variable A after observing B

→ $I(A;B)$: Reduction of uncertainty about A due to the observation of B

Mutual information $I(A;B)$:

Average amount of information that a random variable B carries about another random variable A

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B)$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$\begin{aligned} I(A;B) &= H(A) - H(A\,|\,B) \\ &= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \end{aligned}$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\}
\end{aligned}
$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} = \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\}
\end{aligned}
$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$\begin{aligned} I(A;B) &= H(A) - H(A\,|\,B) \\ &= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\ &= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} = \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\} \\ &= \sum_a \sum_b p_{AB}(a,b) \log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \end{aligned}$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} = \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\} \\
&= \sum_a \sum_b p_{AB}(a,b)\,\log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \\
&= D_{KL}\Big(p_{AB} \,\Big|\Big|\, p_A\,p_B \Big)
\end{aligned}
$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} = \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\} \\
&= \sum_a \sum_b p_{AB}(a,b) \log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \\
&= D_{KL}\Big(p_{AB} \,\Big|\Big|\, p_A\,p_B \Big) \geq 0
\end{aligned}
$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$\begin{aligned} I(A;B) &= H(A) - H(A\,|\,B) \\ &= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\ &= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} = \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\} \\ &= \sum_a \sum_b p_{AB}(a,b)\, \log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \\ &= D_{KL}\Big(p_{AB} \,\Big|\Big|\, p_A\, p_B \Big) \geq 0 \end{aligned}$$

➡ Mutual information is symmetric

$$\begin{aligned} I(A;B) &= H(A) - H(A\,|\,B) \\ &= H(B) - H(B\,|\,A) \end{aligned}$$

Symmetry of Mutual Information

- Mutual information between two discrete random variables A and B

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|\,B) \Big\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} = \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\} \\
&= \sum_a \sum_b p_{AB}(a,b)\, \log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \\
&= D_{KL}\Big(p_{AB} \,\Big|\Big|\, p_A\,p_B \Big) \geq 0
\end{aligned}
$$

➡ Mutual information is symmetric

$$
\begin{aligned}
I(A;B) &= H(A) - H(A\,|\,B) \\
&= H(B) - H(B\,|\,A)
\end{aligned}
$$

➡ Random variable A carries the same amount of information about B as B carries about A

Properties of Mutual Information

- General formulation: Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

Properties of Mutual Information

- General formulation: Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

- Relationship to Kullback-Leibler Divergence

$$I(A;B) = D\Big(\, p_{AB} \,\Big|\Big|\, p_A\, p_B \,\Big) \geq 0$$

Properties of Mutual Information

- General formulation: Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

- Relationship to Kullback-Leibler Divergence

$$I(A;B) = D\Big(\, p_{AB} \,\Big|\Big|\, p_A\, p_B \,\Big) \geq 0$$

- Relationship to marginal entropy

$$I(A;B) \leq H(A) \qquad \text{and} \qquad I(A;B) \leq H(B)$$

Properties of Mutual Information

- General formulation: Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

- Relationship to Kullback-Leibler Divergence

$$I(A;B) = D\Big(\, p_{AB} \,\Big|\Big|\, p_A\, p_B \,\Big) \geq 0$$

- Relationship to marginal entropy

$$I(A;B) \leq H(A) \qquad \text{and} \qquad I(A;B) \leq H(B)$$

- Independent random variables A and B

$$I(A;B) = 0$$

Properties of Mutual Information

- General formulation: Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

- Relationship to Kullback-Leibler Divergence

$$I(A;B) = D\Big(\, p_{AB} \,\Big|\Big|\, p_A\, p_B \,\Big) \geq 0$$

- Relationship to marginal entropy

$$I(A;B) \leq H(A) \qquad \text{and} \qquad I(A;B) \leq H(B)$$

- Independent random variables A and B

$$I(A;B) = 0$$

- Deterministic functional relationship $B = f(A)$

$$B = f(A) \quad \Longrightarrow \quad I(A;B) = H(B)$$

Mutual Information for Continuous Random Variables ?

Definition of Mutual information

- Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

Mutual Information for Continuous Random Variables ?

Definition of Mutual information

- Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

➡ Continuous random variables

- Discrete entropy $H(\cdot)$ and discrete conditional entropy $H(\cdot|\cdot)$ are not defined
- $H(A)$ and $H(A\,|\,\cdot)$ both approach infinity if A is continuous

Mutual Information for Continuous Random Variables ?

Definition of Mutual information

- Discrete random variables A and B

$$I(A;B) \;=\; H(A) - H(A\,|\,B) \;=\; H(B) - H(B\,|\,A)$$

➡ Continuous random variables

- Discrete entropy $H(\cdot)$ and discrete conditional entropy $H(\cdot|\cdot)$ are not defined
- $H(A)$ and $H(A\,|\,\cdot)$ both approach infinity if A is continuous

Question: Can we define mutual information for continuous random variables

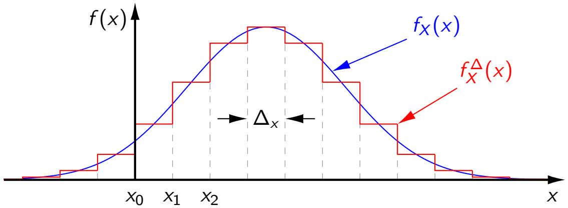
- Via approximation by discrete random variables

➡ Quantize pdfs with a quantizaton step size Δ

➡ Calculate mutual information for resulting discrete random variables

➡ Consider limit for $\Delta \to 0$

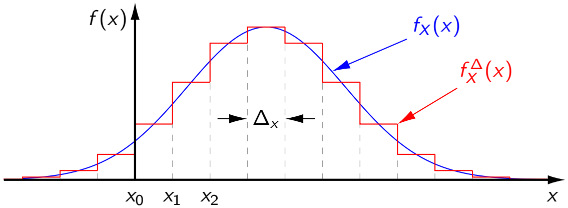
Discrete Approximation for Continuous Random Variables



- Discrete approximation $f_X^{\Delta}(x)$ of probability density function $f_X(x)$

$$\forall x: \; x_k \leq x < x_{k+1}, \qquad f_X^{\Delta}(x) = \frac{1}{\Delta_x} \cdot \mathrm{P}(x_k \leq X < x_{k+1})$$

Discrete Approximation for Continuous Random Variables



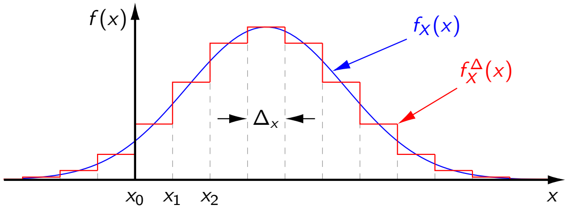
- Discrete approximation $f_X^\Delta(x)$ of probability density function $f_X(x)$

$$\forall x: \; x_k \leq x < x_{k+1}, \qquad f_X^\Delta(x) = \frac{1}{\Delta_x} \cdot \mathrm{P}(x_k \leq X < x_{k+1})$$

➡ Pmf $p_{X_\Delta}(x_k)$ for discrete random variable X_Δ with alphabet $\{x_k\}$

$$p_{X_\Delta}(x_k) = f_X^\Delta(x_k) \cdot \Delta_x$$

Discrete Approximation for Continuous Random Variables



- Discrete approximation $f_X^{\Delta}(x)$ of probability density function $f_X(x)$

$$\forall x: \; x_k \leq x < x_{k+1}, \qquad f_X^{\Delta}(x) = \frac{1}{\Delta_x} \cdot \mathrm{P}(x_k \leq X < x_{k+1})$$

➡ Pmf $p_{X_\Delta}(x_k)$ for discrete random variable X_Δ with alphabet $\{x_k\}$

$$p_{X_\Delta}(x_k) = f_X^{\Delta}(x_k) \cdot \Delta_x$$

➡ Similarly: Joint pmf $p_{X_\Delta Y_\Delta}(x_k, y_i)$ for two discrete approximations X_Δ and Y_Δ

$$p_{X_\Delta Y_\Delta}(x_k, y_i) = f_{XY}^{\Delta}(x_k, y_i) \cdot \Delta_x \cdot \Delta_y$$

Mutual Information for Continuous Random Variables

- Mutual information $I(X_\Delta; Y_\Delta)$ for discrete approximations X_Δ and Y_Δ

$$I(X_\Delta; Y_\Delta) = \sum_{\forall x_k} \sum_{\forall y_i} p_{X_\Delta Y_\Delta}(x_k, y_i) \cdot \log_2 \frac{p_{X_\Delta Y_\Delta}(x_k, y_i)}{p_{X_\Delta}(x_k)\, p_{Y_\Delta}(y_i)}$$

Mutual Information for Continuous Random Variables

- Mutual information $I(X_\Delta; Y_\Delta)$ for discrete approximations X_Δ and Y_Δ

$$\begin{aligned} I(X_\Delta; Y_\Delta) &= \sum_{\forall x_k} \sum_{\forall y_i} p_{X_\Delta Y_\Delta}(x_k, y_i) \cdot \log_2 \frac{p_{X_\Delta Y_\Delta}(x_k, y_i)}{p_{X_\Delta}(x_k)\, p_{Y_\Delta}(y_i)} \\ &= \sum_{\forall x_k} \sum_{\forall y_i} f_{XY}^{\Delta}(x_k, y_i) \cdot \log_2 \frac{f_{XY}^{\Delta}(x_k, y_i)}{f_X^{\Delta}(x_k)\, f_Y^{\Delta}(y_i)} \cdot \Delta_x \cdot \Delta_y \end{aligned}$$

Mutual Information for Continuous Random Variables

- Mutual information $I(X_\Delta; Y_\Delta)$ for discrete approximations X_Δ and Y_Δ

$$I(X_\Delta; Y_\Delta) = \sum_{\forall x_k} \sum_{\forall y_i} p_{X_\Delta Y_\Delta}(x_k, y_i) \cdot \log_2 \frac{p_{X_\Delta Y_\Delta}(x_k, y_i)}{p_{X_\Delta}(x_k)\, p_{Y_\Delta}(y_i)}$$

$$= \sum_{\forall x_k} \sum_{\forall y_i} f_{XY}^{\Delta}(x_k, y_i) \cdot \log_2 \frac{f_{XY}^{\Delta}(x_k, y_i)}{f_X^{\Delta}(x_k)\, f_Y^{\Delta}(y_i)} \cdot \Delta_x \cdot \Delta_y$$

- Mutual information $I(X; Y)$ for continuous random variables X and Y

$$I(X; Y) = \lim_{\substack{\Delta_x \to 0 \\ \Delta_y \to 0}} I(X_\Delta; Y_\Delta)$$

Mutual Information for Continuous Random Variables

- Mutual information $I(X_\Delta; Y_\Delta)$ for discrete approximations X_Δ and Y_Δ

$$\begin{aligned} I(X_\Delta; Y_\Delta) &= \sum_{\forall x_k} \sum_{\forall y_i} p_{X_\Delta Y_\Delta}(x_k, y_i) \cdot \log_2 \frac{p_{X_\Delta Y_\Delta}(x_k, y_i)}{p_{X_\Delta}(x_k)\, p_{Y_\Delta}(y_i)} \\ &= \sum_{\forall x_k} \sum_{\forall y_i} f_{XY}^{\Delta}(x_k, y_i) \cdot \log_2 \frac{f_{XY}^{\Delta}(x_k, y_i)}{f_X^{\Delta}(x_k)\, f_Y^{\Delta}(y_i)} \cdot \Delta_x \cdot \Delta_y \end{aligned}$$

- Mutual information $I(X; Y)$ for continuous random variables X and Y

$$I(X; Y) = \lim_{\substack{\Delta_x \to 0 \\ \Delta_y \to 0}} I(X_\Delta; Y_\Delta)$$

➡ **Mutual information for continuous random variables**

$$\boxed{I(X; Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) \, \log_2 \frac{f_{XY}(x, y)}{f_X(x)\, f_Y(y)} \, \mathrm{d}x \, \mathrm{d}y}$$

Mutual Information: Comparison of Discrete and Continuous Case

Discrete Random Variables A and B

$$I(A;B) = \sum_a \sum_b p_{AB}(a,b)\,\log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)}$$

Continuous Random Variables X and Y

$$I(A;B) = \int\!\!\int f_{XY}(x,y)\,\log_2 \frac{f_{XY}(x,y)}{f_X(x)\,f_Y(y)}\,\mathrm{d}x\,\mathrm{d}y$$

Mutual Information: Comparison of Discrete and Continuous Case

Discrete Random Variables A and B

$$I(A;B) = \sum_a \sum_b p_{AB}(a,b)\,\log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)}$$

$$= \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\}$$

Continuous Random Variables X and Y

$$I(A;B) = \int\!\!\int f_{XY}(x,y)\,\log_2 \frac{f_{XY}(x,y)}{f_X(x)\,f_Y(y)}\,\mathrm{d}x\,\mathrm{d}y$$

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

Mutual Information: Comparison of Discrete and Continuous Case

Discrete Random Variables A and B

$$I(A;B) = \sum_a \sum_b p_{AB}(a,b)\,\log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)}$$

$$= \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\}$$

$$= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\}$$

Continuous Random Variables X and Y

$$I(A;B) = \int\!\!\int f_{XY}(x,y)\,\log_2 \frac{f_{XY}(x,y)}{f_X(x)\,f_Y(y)}\,\mathrm{d}x\,\mathrm{d}y$$

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

$$= \mathrm{E}\left\{ \log_2 \frac{f_{X|Y}(X\,|\,Y)}{f_X(X)} \right\}$$

Mutual Information: Comparison of Discrete and Continuous Case

Discrete Random Variables A and B

$$
\begin{aligned}
I(A;B) &= \sum_a \sum_b p_{AB}(a,b)\,\log_2 \frac{p_{AB}(a,b)}{p_A(a)\,p_B(b)} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\,p_B(B)} \right\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} \\
&= \mathrm{E}\big\{ -\log_2 p_A(A) \big\} - \mathrm{E}\big\{ -\log_2 p_{A|B}(A\,|B) \big\}
\end{aligned}
$$

Continuous Random Variables X and Y

$$
\begin{aligned}
I(A;B) &= \iint f_{XY}(x,y)\,\log_2 \frac{f_{XY}(x,y)}{f_X(x)\,f_Y(y)}\,\mathrm{d}x\,\mathrm{d}y \\
&= \mathrm{E}\left\{ \log_2 \frac{f_{XY}(X,Y)}{f_X(X)\,f_Y(Y)} \right\} \\
&= \mathrm{E}\left\{ \log_2 \frac{f_{X|Y}(X\,|\,Y)}{f_X(X)} \right\} \\
&= \mathrm{E}\big\{ -\log_2 f_X(X) \big\} - \mathrm{E}\big\{ -\log_2 f_{X|Y}(X\,|Y) \big\}
\end{aligned}
$$

Mutual Information: Comparison of Discrete and Continuous Case

Discrete Random Variables A and B

$$
\begin{aligned}
I(A;B) &= \sum_a \sum_b p_{AB}(a,b)\, \log_2 \frac{p_{AB}(a,b)}{p_A(a)\, p_B(b)} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\, p_B(B)} \right\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|B) \Big\} \\
&= H(A) - H(A\,|\,B)
\end{aligned}
$$

Discrete Entropy

$$
\begin{aligned}
H(A) &= \mathrm{E}\big\{ -\log_2 p_A(A) \big\} \\
H(A\,|\,B) &= \mathrm{E}\big\{ -\log_2 p_{A|B}(A\,|\,B) \big\}
\end{aligned}
$$

Continuous Random Variables X and Y

$$
\begin{aligned}
I(A;B) &= \iint f_{XY}(x,y)\, \log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)}\, \mathrm{d}x\, \mathrm{d}y \\
&= \mathrm{E}\left\{ \log_2 \frac{f_{XY}(X,Y)}{f_X(X)\, f_Y(Y)} \right\} \\
&= \mathrm{E}\left\{ \log_2 \frac{f_{X|Y}(X\,|\,Y)}{f_X(X)} \right\} \\
&= \mathrm{E}\Big\{ -\log_2 f_X(X) \Big\} - \mathrm{E}\Big\{ -\log_2 f_{X|Y}(X\,|Y) \Big\}
\end{aligned}
$$

Mutual Information: Comparison of Discrete and Continuous Case

Discrete Random Variables A and B

$$
\begin{aligned}
I(A;B) &= \sum_a \sum_b p_{AB}(a,b)\, \log_2 \frac{p_{AB}(a,b)}{p_A(a)\, p_B(b)} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{AB}(A,B)}{p_A(A)\, p_B(B)} \right\} \\
&= \mathrm{E}\left\{ \log_2 \frac{p_{A|B}(A\,|\,B)}{p_A(A)} \right\} \\
&= \mathrm{E}\Big\{ -\log_2 p_A(A) \Big\} - \mathrm{E}\Big\{ -\log_2 p_{A|B}(A\,|B) \Big\} \\
&= H(A) - H(A\,|\,B)
\end{aligned}
$$

Discrete Entropy

$$
\begin{aligned}
H(A) &= \mathrm{E}\big\{ -\log_2 p_A(A) \big\} \\
H(A\,|\,B) &= \mathrm{E}\big\{ -\log_2 p_{A|B}(A\,|\,B) \big\}
\end{aligned}
$$

Continuous Random Variables X and Y

$$
\begin{aligned}
I(A;B) &= \iint f_{XY}(x,y)\, \log_2 \frac{f_{XY}(x,y)}{f_X(x)\, f_Y(y)}\, \mathrm{d}x\, \mathrm{d}y \\
&= \mathrm{E}\left\{ \log_2 \frac{f_{XY}(X,Y)}{f_X(X)\, f_Y(Y)} \right\} \\
&= \mathrm{E}\left\{ \log_2 \frac{f_{X|Y}(X\,|\,Y)}{f_X(X)} \right\} \\
&= \mathrm{E}\Big\{ -\log_2 f_X(X) \Big\} - \mathrm{E}\Big\{ -\log_2 f_{X|Y}(X\,|Y) \Big\} \\
&= h(X) - h(X\,|\,Y)
\end{aligned}
$$

Differential Entropy

$$
\begin{aligned}
h(X) &= \mathrm{E}\big\{ -\log_2 f_X(X) \big\} \\
h(X\,|\,Y) &= \mathrm{E}\big\{ -\log_2 f_{X|Y}(X\,|\,Y) \big\}
\end{aligned}
$$

Mutual Information for Random Vectors

- Consider vectors (or blocks) of random variables

$$\boldsymbol{A} = \{A_1, A_2, A_3, \cdots, A_N\} \qquad \text{and} \qquad \boldsymbol{B} = \{B_1, B_2, B_3, \cdots, B_M\}$$

Mutual Information for Random Vectors

- Consider vectors (or blocks) of random variables

$$\boldsymbol{A} = \{A_1, A_2, A_3, \cdots, A_N\} \qquad \text{and} \qquad \boldsymbol{B} = \{B_1, B_2, B_3, \cdots, B_M\}$$

➡ Mutual Information for Discrete Random Vectors

$$\begin{aligned} I(\boldsymbol{A};\boldsymbol{B}) &= H(\boldsymbol{A}) - H(\boldsymbol{A}\,|\,\boldsymbol{B}) \\ &= H(\boldsymbol{B}) - H(\boldsymbol{B}\,|\,\boldsymbol{A}) = \mathrm{E}\left\{ \log_2 \frac{p_{\boldsymbol{AB}}(\boldsymbol{A},\boldsymbol{B})}{p_{\boldsymbol{A}}(\boldsymbol{A})\, p_{\boldsymbol{B}}(\boldsymbol{B})} \right\} \end{aligned}$$

Mutual Information for Random Vectors

- Consider vectors (or blocks) of random variables

$$\boldsymbol{A} = \{A_1, A_2, A_3, \cdots, A_N\} \qquad \text{and} \qquad \boldsymbol{B} = \{B_1, B_2, B_3, \cdots, B_M\}$$

➡ Mutual Information for Discrete Random Vectors

$$\begin{aligned} I(\boldsymbol{A};\boldsymbol{B}) &= H(\boldsymbol{A}) - H(\boldsymbol{A}\,|\,\boldsymbol{B}) \\ &= H(\boldsymbol{B}) - H(\boldsymbol{B}\,|\,\boldsymbol{A}) = \mathrm{E}\left\{ \log_2 \frac{p_{\boldsymbol{AB}}(\boldsymbol{A},\boldsymbol{B})}{p_{\boldsymbol{A}}(\boldsymbol{A})\,p_{\boldsymbol{B}}(\boldsymbol{B})} \right\} \end{aligned}$$

➡ Mutual Information for Continuous Random Vectors

$$\begin{aligned} I(\boldsymbol{A};\boldsymbol{B}) &= h(\boldsymbol{A}) - h(\boldsymbol{A}\,|\,\boldsymbol{B}) \\ &= h(\boldsymbol{B}) - h(\boldsymbol{B}\,|\,\boldsymbol{A}) = \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{AB}}(\boldsymbol{A},\boldsymbol{B})}{f_{\boldsymbol{A}}(\boldsymbol{A})\,f_{\boldsymbol{B}}(\boldsymbol{B})} \right\} \end{aligned}$$

Mutual Information for Mixed Case

Consider mixed case:

- Discrete random variable A (or discrete random vector)
- Continuous random variable X (or continuous random vector)

Mutual Information for Mixed Case

Consider mixed case:

- Discrete random variable A (or discrete random vector)
- Continuous random variable X (or continuous random vector)

Mutual Information of A and X

- Can be expressed using discrete entropies or differential entropies

$$I(A; X) = H(A) - H(A \mid X) = h(X) - h(X \mid A)$$

Mutual Information for Mixed Case

Consider mixed case:

- Discrete random variable A (or discrete random vector)
- Continuous random variable X (or continuous random vector)

Mutual Information of A and X

- Can be expressed using discrete entropies or differential entropies

$$I(A;X) \;=\; H(A) - H(A\,|\,X) \;=\; h(X) - h(X\,|\,A)$$

- Conditional entropies $H(A\,|\,X)$ and $h(X\,|\,A)$ are given by

$$H(A\,|\,X) \;=\; \mathrm{E}\big\{ -\log_2 p_{A|X}(A\,|\,X) \big\} = \int_{x=-\infty}^{\infty} f_X(x) \left(-\sum_a p_{A|X}(a\,|\,x) \log_2 p_{A|X}(a\,|\,x) \right) \mathrm{d}x$$

$$h(X\,|\,A) \;=\; \mathrm{E}\big\{ -\log_2 f_{X|A}(X\,|\,A) \big\} = \sum_a p_A(a) \left(-\int_{x=-\infty}^{\infty} f_{X|A}(x\,|\,a) \log_2 f_{X|A}(x\,|\,a)\, \mathrm{d}x \right)$$

Distortion Measures

General Distortion Measures

- Measure for deviation between N input samples $\boldsymbol{s} = \{s_1, s_2, \cdots, s_N\}$ and the corresponding N reconstructed samples $\boldsymbol{s'} = \{s'_1, s'_2, \cdots, s'_N\}$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) \geq 0 \qquad \text{(equality if and only if } \boldsymbol{s} = \boldsymbol{s'}\text{)}$$

Distortion Measures

General Distortion Measures

- Measure for deviation between N input samples $\boldsymbol{s} = \{s_1, s_2, \cdots, s_N\}$ and the corresponding N reconstructed samples $\boldsymbol{s'} = \{s'_1, s'_2, \cdots, s'_N\}$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) \geq 0 \qquad \text{(equality if and only if } \boldsymbol{s} = \boldsymbol{s'}\text{)}$$

Additive Distortion Measures

- Define single-symbol distortion measure

$$d_1(s, s') \geq 0 \qquad \text{(equality if and only if } s = s'\text{)}$$

Distortion Measures

General Distortion Measures

- Measure for deviation between N input samples $\boldsymbol{s} = \{s_1, s_2, \cdots, s_N\}$ and the corresponding N reconstructed samples $\boldsymbol{s'} = \{s'_1, s'_2, \cdots, s'_N\}$

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) \geq 0 \qquad \text{(equality if and only if } \boldsymbol{s} = \boldsymbol{s'}\text{)}$$

Additive Distortion Measures

- Define single-symbol distortion measure

$$d_1(s, s') \geq 0 \qquad \text{(equality if and only if } s = s'\text{)}$$

➞ N-th order distortion $d_N(\boldsymbol{s}, \boldsymbol{s'})$ is the average of the single-symbol distortions

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) = \frac{1}{N} \sum_{k=1}^{N} d_1(s_k, s'_k)$$

Common Additive Distortion Measures

Difference Distortion Measures

- Single symbol distortion $d_1(s, s')$ is calculated based on absolute differences

$$d_1(s, s') = \left| s - s' \right|^p \qquad \text{with} \qquad p > 0$$

→ Average distortion for vectors/blocks of N samples

$$d_N(\boldsymbol{s}, \boldsymbol{s'}) = \frac{1}{N} \sum_{k=1}^{N} \left| s_k - s'_k \right|^p = \frac{1}{N} \left\| \boldsymbol{s} - \boldsymbol{s'} \right\|_p^p$$

Common Additive Distortion Measures

Difference Distortion Measures

- Single symbol distortion $d_1(s, s')$ is calculated based on absolute differences

$$d_1(s, s') = \left|s - s'\right|^p \qquad \text{with} \qquad p > 0$$

➡ Average distortion for vectors/blocks of N samples

$$d_N\left(\boldsymbol{s}, \boldsymbol{s}'\right) = \frac{1}{N} \sum_{k=1}^{N} \left|s_k - s'_k\right|^p = \frac{1}{N} \left\|\boldsymbol{s} - \boldsymbol{s}'\right\|_p^p$$

Most Common: **Mean Squared Error** (i.e., $p = 2$)

$$d_1(s, s') = \left(s - s'\right)^2$$

$$d_N\left(\boldsymbol{s}, \boldsymbol{s}'\right) = \frac{1}{N} \sum_{k=0}^{N-1} \left(s_k - s'_k\right)^2 = \frac{1}{N} \left\|\boldsymbol{s} - \boldsymbol{s}'\right\|_2^2 = \frac{1}{N} \left(\boldsymbol{s} - \boldsymbol{s}'\right)^{\mathrm{T}} \left(\boldsymbol{s} - \boldsymbol{s}'\right)$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S'}$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}') \,\}$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}')\,\} \;\; = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\; d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}')\,\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = \mathrm{E}\left\{ \log_s \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')} \right\}$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}')\,\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\; d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = \mathrm{E}\left\{ \log_s \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')} \right\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, \log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}')\,\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = \mathrm{E}\left\{ \log_s \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')} \right\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, \log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$\text{with} \quad f_{\boldsymbol{S}'}(\boldsymbol{s}') = \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}')\,\} \;=\; \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\; d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = \mathrm{E}\left\{ \log_s \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')} \right\} \;=\; \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, \log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s})\, f_{\boldsymbol{S}'}(\boldsymbol{s}')}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$\text{with} \quad f_{\boldsymbol{S}'}(\boldsymbol{s}') = \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}$$

- Joint pdf $f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')$ can be written as

$$f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$$

Distortion and Mutual Information as Expected Values

Distortion and Mutual Information for Source Coding

- Consider given source $\boldsymbol{S}$ and source code Q (reconstructed signal is denoted by random process $\boldsymbol{S}'$)

➡ N-th order distortion $\delta_N(Q)$ and N-th order mutual information $I_N(Q)$

$$\delta_N(Q) = \mathrm{E}\{\, d_N(\boldsymbol{s}, \boldsymbol{s}') \,\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') \, d_N(\boldsymbol{s}, \boldsymbol{s}') \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = \mathrm{E}\left\{ \log_s \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s}) \, f_{\boldsymbol{S}'}(\boldsymbol{s}')} \right\} \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') \, \log_2 \frac{f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')}{f_{\boldsymbol{S}}(\boldsymbol{s}) \, f_{\boldsymbol{S}'}(\boldsymbol{s}')} \, \mathrm{d}\boldsymbol{s} \, \mathrm{d}\boldsymbol{s}'$$

$$\text{with} \quad f_{\boldsymbol{S}'}(\boldsymbol{s}') = \int_{\mathbb{R}^N} f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') \, \mathrm{d}\boldsymbol{s}$$

- Joint pdf $f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}')$ can be written as

$$f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$$

➡ Statistical properties of any source code Q can be completely specified by N-th order conditional pdf $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$

Distortion and Mutual Information for Source Codes

Consider coding of given source $\boldsymbol{S}$ with any source code Q

- Statistical properties of source $\boldsymbol{S}$ are given by N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$

Distortion and Mutual Information for Source Codes

Consider coding of given source $\boldsymbol{S}$ with any source code Q

- Statistical properties of source $\boldsymbol{S}$ are given by N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$
- Statistical properties of Q can be described by N-th order conditional pdf $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$

$$\text{source code } Q: \qquad f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$$

Distortion and Mutual Information for Source Codes

Consider coding of given source $\boldsymbol{S}$ with any source code Q

- Statistical properties of source $\boldsymbol{S}$ are given by N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$
- Statistical properties of Q can be described by N-th order conditional pdf $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$

$$\text{source code } Q: \qquad f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$$

➡ N-th order distortion $\delta_N(Q)$ and mutual information $I_N(Q)$

$$\delta_N(Q) = \delta_N(g_N^Q) \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\; d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

Distortion and Mutual Information for Source Codes

Consider coding of given source $\boldsymbol{S}$ with any source code Q

- Statistical properties of source $\boldsymbol{S}$ are given by N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$
- Statistical properties of Q can be described by N-th order conditional pdf $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$

$$\text{source code } Q: \qquad f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$$

➡ N-th order distortion $\delta_N(Q)$ and mutual information $I_N(Q)$

$$\delta_N(Q) = \delta_N(g_N^Q) \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\; d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = I_N(g_N^Q) \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\; \log_2 \frac{g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})}{\left(\int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\;\; \mathrm{d}\boldsymbol{s} \right)}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

Distortion and Mutual Information for Source Codes

Consider coding of given source $\boldsymbol{S}$ with any source code Q

- Statistical properties of source $\boldsymbol{S}$ are given by N-th order pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$
- Statistical properties of Q can be described by N-th order conditional pdf $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$

$$\text{source code } Q: \qquad f_{\boldsymbol{S}\boldsymbol{S}'}(\boldsymbol{s}, \boldsymbol{s}') = f_{\boldsymbol{S}}(\boldsymbol{s}) \cdot g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$$

➡ N-th order distortion $\delta_N(Q)$ and mutual information $I_N(Q)$

$$\delta_N(Q) = \delta_N(g_N^Q) \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\; d_N(\boldsymbol{s}, \boldsymbol{s}')\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

$$I_N(Q) = I_N(g_N^Q) \;=\; \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\; \log_2 \frac{g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})}{\left(\int_{\mathbb{R}^N} f_{\boldsymbol{S}}(\boldsymbol{s})\, g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})\; \mathrm{d}\boldsymbol{s} \right)}\, \mathrm{d}\boldsymbol{s}\, \mathrm{d}\boldsymbol{s}'$$

➡ Distortion and mutual information are completely determined by source pdf $f_{\boldsymbol{S}}(\boldsymbol{s})$ and conditional pdf $g_N^Q(\boldsymbol{s}' \,|\, \boldsymbol{s})$ of source code Q

Information Rate-Distortion Function

$$R(D) = \inf_{Q:\,\delta(Q)\leq D} r(Q) \qquad \text{(operational rate-distortion function)}$$

Information Rate-Distortion Function

$$R(D) = \inf_{Q:\, \delta(Q) \leq D} r(Q) \qquad \text{(operational rate-distortion function)}$$

$$\geq \inf_{Q:\, \delta(Q) \leq D} \bar{H}(\boldsymbol{S}') = \inf_{Q:\, \delta(Q) \leq D} \lim_{N \to \infty} \frac{H_N(\boldsymbol{S}')}{N} \qquad \text{(lossless coding theorem)}$$

Information Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \inf_{Q:\,\delta(Q)\le D} r(Q) && \text{(operational rate-distortion function)}\\
&\ge \inf_{Q:\,\delta(Q)\le D} \bar{H}(\boldsymbol{S}') = \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{H_N(\boldsymbol{S}')}{N} && \text{(lossless coding theorem)}\\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}') + H_N(\boldsymbol{S}'\,|\,\boldsymbol{S})}{N} && \text{(definition of mutual information)}
\end{aligned}
$$

Information Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \inf_{Q:\, \delta(Q) \le D} r(Q) && \text{(operational rate-distortion function)} \\
&\ge \inf_{Q:\, \delta(Q) \le D} \bar{H}(\boldsymbol{S}') = \inf_{Q:\, \delta(Q) \le D} \lim_{N \to \infty} \frac{H_N(\boldsymbol{S}')}{N} && \text{(lossless coding theorem)} \\
&= \inf_{Q:\, \delta(Q) \le D} \lim_{N \to \infty} \frac{I_N(\boldsymbol{S}, \boldsymbol{S}') + H_N(\boldsymbol{S}' \,|\, \boldsymbol{S})}{N} && \text{(definition of mutual information)} \\
&= \inf_{Q:\, \delta(Q) \le D} \lim_{N \to \infty} \frac{I_N(\boldsymbol{S}, \boldsymbol{S}')}{N} && \text{(code } Q\text{: deterministic function)}
\end{aligned}
$$

Information Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \inf_{Q:\,\delta(Q)\le D} r(Q) && \text{(operational rate-distortion function)} \\
&\ge \inf_{Q:\,\delta(Q)\le D} \bar{H}(\boldsymbol{S}') = \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{H_N(\boldsymbol{S}')}{N} && \text{(lossless coding theorem)} \\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}') + H_N(\boldsymbol{S}' \,|\, \boldsymbol{S})}{N} && \text{(definition of mutual information)} \\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}')}{N} && \text{(code } Q\text{: deterministic function)} \\
&= \lim_{N\to\infty} \inf_{g_N^Q:\,\delta_N(g_N^Q)\le D} \frac{I_N(g_N^Q)}{N} && \text{(use } N\text{-th order conditional pdf } g_N^Q\text{)}
\end{aligned}
$$

Information Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \inf_{Q:\,\delta(Q)\le D} r(Q) && \text{(operational rate-distortion function)}\\
&\ge \inf_{Q:\,\delta(Q)\le D} \bar{H}(\boldsymbol{S}') = \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{H_N(\boldsymbol{S}')}{N} && \text{(lossless coding theorem)}\\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}') + H_N(\boldsymbol{S}' \mid \boldsymbol{S})}{N} && \text{(definition of mutual information)}\\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}')}{N} && \text{(code } Q\text{: deterministic function)}\\
&= \lim_{N\to\infty} \inf_{g_N^Q:\,\delta_N(g_N^Q)\le D} \frac{I_N(g_N^Q)}{N} && \text{(use } N\text{-th order conditional pdf } g_N^Q\text{)}\\
&\ge \boxed{R^I(D) = \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(g_N)}{N}} && (g_N^Q \text{ for codes} = \text{subset of all } g_N)
\end{aligned}
$$

Information Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \inf_{Q:\,\delta(Q)\le D} r(Q) && \text{(operational rate-distortion function)}\\
&\ge \inf_{Q:\,\delta(Q)\le D} \bar{H}(\boldsymbol{S}') = \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{H_N(\boldsymbol{S}')}{N} && \text{(lossless coding theorem)}\\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}') + H_N(\boldsymbol{S}' \,|\, \boldsymbol{S})}{N} && \text{(definition of mutual information)}\\
&= \inf_{Q:\,\delta(Q)\le D} \lim_{N\to\infty} \frac{I_N(\boldsymbol{S},\boldsymbol{S}')}{N} && \text{(code } Q\text{: deterministic function)}\\
&= \lim_{N\to\infty} \inf_{g_N^Q:\,\delta_N(g_N^Q)\le D} \frac{I_N(g_N^Q)}{N} && \text{(use } N\text{-th order conditional pdf } g_N^Q\text{)}\\
&\ge \boxed{R^I(D) = \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(g_N)}{N}} && (g_N^Q \text{ for codes} = \text{subset of all } g_N)
\end{aligned}
$$

➞ The lower bound $R^I(D) \le R(D)$ is referred to as **information rate-distortion function**

Rate-Distortion Function

Information vs Operational Rate-Distortion Function

1. We showed: $R^{\mathrm{I}}(D)$ is a lower bound for $R(D)$

$$R(D) \geq R^{\mathrm{I}}(D) \qquad \text{with} \qquad R^{I}(D) = \lim_{N \to \infty} \inf_{g_N :\, \delta(g_N) \leq D} \frac{I_N(g_N)}{N}$$

Rate-Distortion Function

Information vs Operational Rate-Distortion Function

1. We showed: $R^{\mathrm{I}}(D)$ is a lower bound for $R(D)$

$$R(D) \geq R^{\mathrm{I}}(D) \qquad \text{with} \qquad R^{I}(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta(g_N)\leq D} \frac{I_N(g_N)}{N}$$

2. Can be shown: $R^{\mathrm{I}}(D)$ is asymptotically achievable
   - For any $D > 0$ and $\varepsilon > 0$, there exists a source code Q with

$$\delta(Q) \leq D \qquad \text{and} \qquad r(Q) \leq R^{\mathrm{I}}(D) + \varepsilon$$

   - Proof: See [Cover, Thomas, "Elements of Information Theory"]

Rate-Distortion Function

Information vs Operational Rate-Distortion Function

1 We showed: $R^{\mathrm{I}}(D)$ is a lower bound for $R(D)$

$$R(D) \geq R^{\mathrm{I}}(D) \qquad \text{with} \qquad R^{I}(D) = \lim_{N\to\infty} \inf_{g_N:\,\delta(g_N)\leq D} \frac{I_N(g_N)}{N}$$

2 Can be shown: $R^{\mathrm{I}}(D)$ is asymptotically achievable

- For any $D > 0$ and $\varepsilon > 0$, there exists a source code Q with

$$\delta(Q) \leq D \qquad \text{and} \qquad r(Q) \leq R^{\mathrm{I}}(D) + \varepsilon$$

- Proof: See [Cover, Thomas, "Elements of Information Theory"]

➡ **Information rate-distortion function $R^{\mathrm{I}}(D)$ coincides with operational rate-distortion function $R(D)$**

Rate-Distortion Function

Information vs Operational Rate-Distortion Function

1. We showed: $R^{\mathrm{I}}(D)$ is a lower bound for $R(D)$

$$R(D) \geq R^{\mathrm{I}}(D) \qquad \text{with} \qquad R^{I}(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta(g_N)\leq D} \frac{I_N(g_N)}{N}$$

2. Can be shown: $R^{\mathrm{I}}(D)$ is asymptotically achievable
   - For any $D > 0$ and $\varepsilon > 0$, there exists a source code Q with

$$\delta(Q) \leq D \qquad \text{and} \qquad r(Q) \leq R^{\mathrm{I}}(D) + \varepsilon$$

   - Proof: See [Cover, Thomas, "Elements of Information Theory"]

➡ **Information rate-distortion function $R^{\mathrm{I}}(D)$ coincides with operational rate-distortion function $R(D)$**

➡ Use term **rate-distortion function $R(D)$** for both

Fundamental Source Coding Theorem

Bounds for Lossy Source Coding

- (Information) Rate-Distortion Function $R(D)$

$$\boxed{R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\leq D} \frac{I_N(\, g_N\,)}{N}}$$

Fundamental Source Coding Theorem

Bounds for Lossy Source Coding

- (Information) Rate-Distortion Function $R(D)$
- (Information) Distortion-Rate Function $D(R)$ (inverse of $R(D)$)

$$\boxed{R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\leq D} \frac{I_N(\, g_N\,)}{N}} \quad \text{and} \quad \boxed{D(R) = \lim_{N\to\infty} \inf_{g_N:\, I_N(g_N)/N\leq R} \delta_N(\, g_N\,)}$$

Fundamental Source Coding Theorem

Bounds for Lossy Source Coding

- (Information) Rate-Distortion Function $R(D)$
- (Information) Distortion-Rate Function $D(R)$ (inverse of $R(D)$)

$$\boxed{R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\le D} \frac{I_N(\,g_N\,)}{N}} \quad \text{and} \quad \boxed{D(R) = \lim_{N\to\infty} \inf_{g_N:\, I_N(g_N)/N\le R} \delta_N(\,g_N\,)}$$

Fundamental Source Coding Theorem

- Rate-distortion function $R(D)$ and distortion-rate function $D(R)$ specify the greatest lower bound for any source code

$$\forall Q: \quad \delta(Q) \le D \quad \Longrightarrow \quad r(Q) \ge R(D)$$

$$\forall Q: \quad r(Q) \le R \quad \Longrightarrow \quad \delta(Q) \ge D(R)$$

Special Case: Lossless Source Coding Theorem

Lossless Coding of Discrete Source

- Rate-distortion function $R(D)$

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\leq D} \frac{I_N(\, g_N\,)}{N}$$

Special Case: Lossless Source Coding Theorem

Lossless Coding of Discrete Source

- Rate-distortion function $R(D)$

$$R(D) = \lim_{N \to \infty} \inf_{g_N \colon \delta_N(g_N) \leq D} \frac{I_N(\, g_N \,)}{N}$$

- Lossless coding: $\boldsymbol{S}' = \boldsymbol{S}$

$$I_N(\, g_N \,) = I_N(\boldsymbol{S}; \boldsymbol{S}) = H_N(\boldsymbol{S}) - H_N(\boldsymbol{S} \,|\, \boldsymbol{S}) = H_N(\boldsymbol{S})$$

Special Case: Lossless Source Coding Theorem

Lossless Coding of Discrete Source

- Rate-distortion function $R(D)$

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(\,g_N\,)}{N}$$

- Lossless coding: $\boldsymbol{S}' = \boldsymbol{S}$

$$I_N(\,g_N\,) = I_N(\boldsymbol{S};\boldsymbol{S}) = H_N(\boldsymbol{S}) - H_N(\boldsymbol{S}\,|\,\boldsymbol{S}) = H_N(\boldsymbol{S})$$

➡ Rate-distortion function for lossless coding $(D = 0)$

$$R(0) = \lim_{N\to\infty} \frac{H_N(\boldsymbol{S})}{N} = \bar{H}(\boldsymbol{S})$$

Special Case: Lossless Source Coding Theorem

Lossless Coding of Discrete Source

- Rate-distortion function $R(D)$

$$R(D) = \lim_{N\to\infty} \inf_{g_N:\, \delta_N(g_N)\le D} \frac{I_N(\, g_N\,)}{N}$$

- Lossless coding: $\boldsymbol{S}' = \boldsymbol{S}$

$$I_N(\, g_N\,) = I_N(\boldsymbol{S}; \boldsymbol{S}) = H_N(\boldsymbol{S}) - H_N(\boldsymbol{S} \,|\, \boldsymbol{S}) = H_N(\boldsymbol{S})$$

➡ Rate-distortion function for lossless coding $(D = 0)$

$$R(0) = \lim_{N\to\infty} \frac{H_N(\boldsymbol{S})}{N} = \bar{H}(\boldsymbol{S})$$

➡ Lossless coding theorem: Special case of fundamental source coding theorem

$$\forall Q:\ \delta(Q) = 0 \quad \Longrightarrow \quad r(Q) \ge R(0) = \bar{H}(\boldsymbol{S})$$

Special Case: Additive Distortion Measures & IID Sources

- N-th order distortion $\delta_N(\, g_N \,)$ for additive distortion measures

$$\delta_N(\, g_N \,) = \mathrm{E}\{\, d_N(\boldsymbol{S}, \boldsymbol{S'}) \,\} = \mathrm{E}\left\{ \frac{1}{N} \sum_{k=0}^{N-1} d_1(S_k, S'_k) \right\} = \mathrm{E}\{\, d_1(S, S') \,\} = \delta_1(\, g_1 \,)$$

Special Case: Additive Distortion Measures & IID Sources

- N-th order distortion $\delta_N(\, g_N \,)$ for additive distortion measures

$$\delta_N(\, g_N \,) = \mathrm{E}\{\, d_N(\boldsymbol{S}, \boldsymbol{S'}) \,\} = \mathrm{E}\left\{ \frac{1}{N} \sum_{k=0}^{N-1} d_1(S_k, S'_k) \right\} = \mathrm{E}\{\, d_1(S, S') \,\} = \delta_1(\, g_1 \,)$$

- N-th order mutual information for iid source (if $\boldsymbol{S}$ is iid, then $\boldsymbol{S'}$ is also iid)

$$I_N(\, g_N \,) = \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{SS'}}(\boldsymbol{S}, \boldsymbol{S'})}{f_{\boldsymbol{S}}(\boldsymbol{S})\, f_{\boldsymbol{S'}}(\boldsymbol{S'})} \right\} = \mathrm{E}\left\{ \log_2 \left(\frac{f_{SS'}(S, S')}{f_S(S)\, f_{S'}(S')} \right)^N \right\} = N \cdot I_1(\, g_1 \,)$$

Special Case: Additive Distortion Measures & IID Sources

- N-th order distortion $\delta_N(\, g_N \,)$ for additive distortion measures

$$\delta_N(\, g_N \,) = \mathrm{E}\{\, d_N(\boldsymbol{S}, \boldsymbol{S'}) \,\} = \mathrm{E}\left\{ \frac{1}{N} \sum_{k=0}^{N-1} d_1(S_k, S'_k) \right\} = \mathrm{E}\{\, d_1(S, S') \,\} = \delta_1(\, g_1 \,)$$

- N-th order mutual information for iid source (if $\boldsymbol{S}$ is iid, then $\boldsymbol{S'}$ is also iid)

$$I_N(\, g_N \,) = \mathrm{E}\left\{ \log_2 \frac{f_{\boldsymbol{SS'}}(\boldsymbol{S}, \boldsymbol{S'})}{f_{\boldsymbol{S}}(\boldsymbol{S})\, f_{\boldsymbol{S'}}(\boldsymbol{S'})} \right\} = \mathrm{E}\left\{ \log_2 \left(\frac{f_{SS'}(S, S')}{f_S(S)\, f_{S'}(S')} \right)^N \right\} = N \cdot I_1(\, g_1 \,)$$

➡ Rate-Distortion Function for IID Sources and Additive Distortion Measures

$$R(D) = \lim_{N\to\infty} \inf_{g_N :\, \delta_N(g_N) \le D} \frac{I_N(\, g_N \,)}{N} \quad \Longrightarrow \quad \boxed{R(D) = \inf_{g_1 :\, \delta_1(g_1) \le D} I_1(\, g_1 \,)}$$

Rate-Distortion Function for Discrete Sources

Properties of $R(D)$ for discrete sources

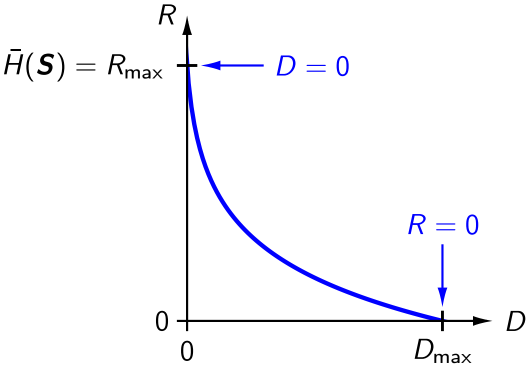
- Domain: $[\,0,\,+\infty\,)$
- Non-increasing function
- Convex function
- Maximum rate (lossless coding)

$$R(0) = R_{\max} = \bar{H}(\boldsymbol{S})$$

- There exists is maximum distortion $D_{\max}$

$$\exists\, D_{\max}: \quad R(D) = \begin{cases} > 0 & : \quad D < D_{\max} \\ 0 & : \quad D \geq D_{\max} \end{cases}$$

➜ MSE distortion measure: $D_{\max} = \sigma^2$



Rate-Distortion Function for Continuous Sources

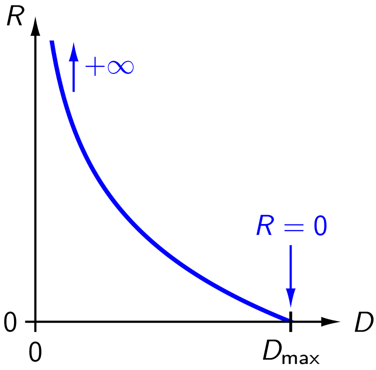
Properties of $R(D)$ for continuous sources

- Domain: $[\,0,\, +\infty\,)$
- Non-increasing function
- Convex function
- Unlimited rate

$$\lim_{D \to 0} R(D) = +\infty$$

- There exists is maximum distortion $D_{\max}$

$$\exists\, D_{\max} : \quad R(D) = \begin{cases} > 0 & : \quad D < D_{\max} \\ \;\; 0 & : \quad D \geq D_{\max} \end{cases}$$



➡ MSE distortion measure: $D_{\max} = \sigma^2$

Discussion of Rate-Distortion Functions

Greatest Lower Bound for Lossless Coding

- Operational / information rate-distortion function

$$R(D) = \inf_{Q:\,\delta(Q)\le D} r(Q) = \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(\,g_N\,)}{N}$$

- Operational RDF: ➡ Obvious definition, but impossible to evaluate
- Information RDF: ➡ Property of source (no need to consider codes)
  ➡ Still impossible to evaluate directly
  ➡ Numerical minimization for discrete sources: Blahut-Arimoto algorithm

How can we proceed?

- Derive lower bound for rate-distortion function
- For some sources and distortion measures:
  ➡ Show that lower bound is achievable

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{\, -\log_2 f_X(X) \,\} = -\int_{-\infty}^{\infty} f_X(x) \, \log_2 f_X(x) \, \mathrm{d}x$$

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{-\log_2 f_X(X)\} = -\int_{-\infty}^{\infty} f_X(x)\,\log_2 f_X(x)\,\mathrm{d}x$$

Example: Differential Entropy of Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases}$$

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{ -\log_2 f_X(X) \} = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x) \, \mathrm{d}x$$

Example: Differential Entropy of Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases}$$

➡ Differential entropy:

$$h(S) \;=\; -\int_{-\infty}^{\infty} f(s) \log_2 f(s) \, \mathrm{d}s$$

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{ -\log_2 f_X(X) \} = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x) \,\mathrm{d}x$$

Example: Differential Entropy of Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases}$$

→ Differential entropy:

$$h(S) \;=\; -\int_{-\infty}^{\infty} f(s) \log_2 f(s) \,\mathrm{d}s \;=\; -\int_{-A/2}^{A/2} \frac{1}{A} \log_2 \frac{1}{A} \,\mathrm{d}s$$

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{-\log_2 f_X(X)\} = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x)\,\mathrm{d}x$$

Example: Differential Entropy of Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases}$$

→ Differential entropy:

$$h(S) = -\int_{-\infty}^{\infty} f(s) \log_2 f(s)\,\mathrm{d}s = -\int_{-A/2}^{A/2} \frac{1}{A} \log_2 \frac{1}{A}\,\mathrm{d}s = \frac{1}{A} \log_2 A \int_{-A/2}^{A/2} \mathrm{d}s$$

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{ -\log_2 f_X(X) \} = -\int_{-\infty}^{\infty} f_X(x) \log_2 f_X(x) \, \mathrm{d}x$$

Example: Differential Entropy of Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases}$$

➜ Differential entropy:

$$h(S) = -\int_{-\infty}^{\infty} f(s) \log_2 f(s) \, \mathrm{d}s = -\int_{-A/2}^{A/2} \frac{1}{A} \log_2 \frac{1}{A} \, \mathrm{d}s = \frac{1}{A} \log_2 A \int_{-A/2}^{A/2} \mathrm{d}s$$

$$h(S) = \log_2 A$$

Differential Entropy

- Remember: Differential entropy of a continuous random variable X

$$h(X) = \mathrm{E}\{ -\log_2 f_X(X) \} = -\int_{-\infty}^{\infty} f_X(x) \, \log_2 f_X(x) \, \mathrm{d}x$$

Example: Differential Entropy of Uniform IID Source

- Continuous uniform iid source (with zero mean)

$$f(s) = \begin{cases} \frac{1}{A} & : \quad |s| \leq \frac{A}{2} \\ 0 & : \quad |s| > \frac{A}{2} \end{cases}$$

→ Differential entropy:

$$h(S) \;=\; -\int_{-\infty}^{\infty} f(s) \, \log_2 f(s) \, \mathrm{d}s \;=\; -\int_{-A/2}^{A/2} \frac{1}{A} \, \log_2 \frac{1}{A} \, \mathrm{d}s \;=\; \frac{1}{A} \log_2 A \int_{-A/2}^{A/2} \mathrm{d}s$$

$$h(S) \;=\; \log_2 A$$

→ Note: Differential entropy $h(S)$ can become negative

Summary: Mutual Information and Entropy

- Mutual information

$$
\begin{aligned}
\text{discrete } X: \quad & I(X;Y) = H(X) - H(X \mid Y) \\
\text{continuous } X: \quad & I(X;Y) = h(X) - h(X \mid Y)
\end{aligned}
$$

Summary: Mutual Information and Entropy

- Mutual information

$$\text{discrete } X: \quad I(X;Y) = H(X) - H(X\,|\,Y)$$

$$\text{continuous } X: \quad I(X;Y) = h(X) - h(X\,|\,Y)$$

- Discrete and differential entropies

$$H(X) = \mathrm{E}\{-\log_2 p_X(X)\} \qquad H(X\,|\,Y) = \mathrm{E}\{-\log_2 p_{X|Y}(X\,|\,Y)\}$$

$$h(X) = \mathrm{E}\{-\log_2 f_X(X)\} \qquad h(X\,|\,Y) = \mathrm{E}\{-\log_2 f_{X|Y}(X\,|\,Y)\}$$

Summary: Mutual Information and Entropy

- Mutual information

$$\begin{aligned} \text{discrete } X: \quad & I(X;Y) = H(X) - H(X \,|\, Y) \\ \text{continuous } X: \quad & I(X;Y) = h(X) - h(X \,|\, Y) \end{aligned}$$

- Discrete and differential entropies

$$H(X) = \mathrm{E}\{-\log_2 p_X(X)\} \qquad H(X \,|\, Y) = \mathrm{E}\{-\log_2 p_{X|Y}(X \,|\, Y)\}$$

$$h(X) = \mathrm{E}\{-\log_2 f_X(X)\} \qquad h(X \,|\, Y) = \mathrm{E}\{-\log_2 f_{X|Y}(X \,|\, Y)\}$$

- Relationship between discrete and differential entropy:

  Quantization X_Δ of continuous random variable X (with step size Δ)

$$h(X) = \lim_{\Delta \to 0} \Big(H(X_\Delta) + \log_2 \Delta \Big)$$

Differential Entropy Rate

N-th Order Differential Entropy

- Differential entropy for N consecutive random variables $S_k, \cdots, S_{k+N-1}$

$$h_N(\boldsymbol{S}) = h(S_k, S_{k+1}, \cdots, S_{k+N-1}) = \mathrm{E}\Big\{ -\log_2 f_{\boldsymbol{S}}(S_k, S_{k+1}, \cdots, S_{k+N-1}) \Big\}$$

Differential Entropy Rate

N-th Order Differential Entropy

- Differential entropy for N consecutive random variables $S_k, \cdots, S_{k+N-1}$

$$h_N(\boldsymbol{S}) = h(S_k, S_{k+1}, \cdots, S_{k+N-1}) = \mathrm{E}\Big\{ -\log_2 f_{\boldsymbol{S}}(S_k, S_{k+1}, \cdots, S_{k+N-1}) \Big\}$$

Differential Entropy Rate

- Definition (analog to discrete entropy rate)

$$\bar{h}(\boldsymbol{S}) = \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N}$$

Differential Entropy Rate

N-th Order Differential Entropy

- Differential entropy for N consecutive random variables $S_k, \cdots, S_{k+N-1}$

$$h_N(\boldsymbol{S}) = h(S_k, S_{k+1}, \cdots, S_{k+N-1}) = \mathrm{E}\Big\{ -\log_2 f_{\boldsymbol{S}}(S_k, S_{k+1}, \cdots, S_{k+N-1}) \Big\}$$

Differential Entropy Rate

- Definition (analog to discrete entropy rate)

$$\bar{h}(\boldsymbol{S}) = \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N}$$

➡ Special sources (proof: same as for discrete entropy)

$$\begin{aligned} \text{IID:} \qquad & \bar{h}(\boldsymbol{S}) = h(S) \\ \text{Markov:} \qquad & \bar{h}(\boldsymbol{S}) = h(S_n \,|\, S_{n-1}) \end{aligned}$$

Differential Entropy for Gaussian Source

- Marginal pdf of Gaussian random processes

$$f_G(s) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

➡ Differential entropy for Gaussian source

$$h^G(S) \;=\; \mathrm{E}\{ -\log_2 f_G(S) \}$$

Differential Entropy for Gaussian Source

- Marginal pdf of Gaussian random processes

$$f_G(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

➡ Differential entropy for Gaussian source

$$\begin{aligned} h^G(S) &= \mathrm{E}\{-\log_2 f_G(S)\} \\ &= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\log_2 e \right\} \end{aligned}$$

Differential Entropy for Gaussian Source

- Marginal pdf of Gaussian random processes

$$f_G(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

➜ Differential entropy for Gaussian source

$$\begin{aligned} h^G(S) &= \mathrm{E}\{-\log_2 f_G(S)\} \\ &= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\log_2 e \right\} \\ &= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\mathrm{E}\{(S-\mu)^2\}\,\log_2 e \end{aligned}$$

Differential Entropy for Gaussian Source

- Marginal pdf of Gaussian random processes

$$f_G(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

➜ Differential entropy for Gaussian source

$$\begin{aligned} h^G(S) &= \mathrm{E}\{ -\log_2 f_G(S) \} \\ &= \mathrm{E}\left\{ \frac{1}{2} \log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2} \log_2 e \right\} \\ &= \frac{1}{2} \log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\, \mathrm{E}\{ (S-\mu)^2 \} \log_2 e \\ &= \frac{1}{2} \log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\, \sigma^2 \log_2 e \end{aligned}$$

Differential Entropy for Gaussian Source

- Marginal pdf of Gaussian random processes

$$f_G(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

→ Differential entropy for Gaussian source

$$\begin{aligned} h^G(S) &= \mathrm{E}\{-\log_2 f_G(S)\} \\ &= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\log_2 e \right\} \\ &= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\mathrm{E}\{(S-\mu)^2\}\log_2 e \\ &= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\sigma^2\log_2 e \end{aligned}$$

$$\boxed{h^G(S) = \frac{1}{2}\log_2\left(2\pi e\,\sigma^2\right)}$$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{ -\log_2 f_G(S) \}$ taken over the pdf $f_S(s)$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{-\log_2 f_G(S)\}$ taken over the pdf $f_S(s)$

$$\mathrm{E}\{-\log_2 f_G(S)\} = -\int_{-\infty}^{\infty} f_S(s) \log_2 f_G(s) \, \mathrm{d}s$$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{-\log_2 f_G(S)\}$ taken over the pdf $f_S(s)$

$$
\begin{aligned}
\mathrm{E}\{-\log_2 f_G(S)\} &= -\int_{-\infty}^{\infty} f_S(s) \log_2 f_G(s)\,\mathrm{d}s \\
&= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\log_2 e \right\}
\end{aligned}
$$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{-\log_2 f_G(S)\}$ taken over the pdf $f_S(s)$

$$
\begin{aligned}
\mathrm{E}\{-\log_2 f_G(S)\} &= -\int_{-\infty}^{\infty} f_S(s) \log_2 f_G(s)\,\mathrm{d}s \\
&= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\log_2 e \right\} \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\mathrm{E}\{(S-\mu)^2\}\log_2 e \qquad (S \text{ has variance } \sigma^2)
\end{aligned}
$$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{\,-\log_2 f_G(S)\,\}$ taken over the pdf $f_S(s)$

$$
\begin{aligned}
\mathrm{E}\{\,-\log_2 f_G(S)\,\} &= -\int_{-\infty}^{\infty} f_S(s)\,\log_2 f_G(s)\,\mathrm{d}s \\
&= \mathrm{E}\left\{\,\frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\,\log_2 e\,\right\} \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\mathrm{E}\{\,(S-\mu)^2\,\}\,\log_2 e \qquad (S \text{ has variance } \sigma^2) \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\sigma^2\,\log_2 e
\end{aligned}
$$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{-\log_2 f_G(S)\}$ taken over the pdf $f_S(s)$

$$
\begin{aligned}
\mathrm{E}\{-\log_2 f_G(S)\} &= -\int_{-\infty}^{\infty} f_S(s)\,\log_2 f_G(s)\,\mathrm{d}s \\
&= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\,\log_2 e \right\} \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\mathrm{E}\{(S-\mu)^2\}\,\log_2 e \qquad (S \text{ has variance } \sigma^2) \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\sigma^2\,\log_2 e \\
&= \frac{1}{2}\log_2\left(2\pi e\,\sigma^2\right)
\end{aligned}
$$

Expectation Value of Negative Logarithm for Gaussian Pdf

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2
- Consider the expected value $\mathrm{E}\{-\log_2 f_G(S)\}$ taken over the pdf $f_S(s)$

$$
\begin{aligned}
\mathrm{E}\{-\log_2 f_G(S)\} &= -\int_{-\infty}^{\infty} f_S(s)\,\log_2 f_G(s)\,\mathrm{d}s \\
&= \mathrm{E}\left\{ \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{(S-\mu)^2}{2\sigma^2}\,\log_2 e \right\} \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\mathrm{E}\{(S-\mu)^2\}\,\log_2 e \qquad (S \text{ has variance } \sigma^2) \\
&= \frac{1}{2}\log_2\left(2\pi\sigma^2\right) + \frac{1}{2\sigma^2}\,\sigma^2\,\log_2 e \\
&= \frac{1}{2}\log_2\left(2\pi e\,\sigma^2\right)
\end{aligned}
$$

Note: Same expression as for $h^G(S)$

Kullback-Leibler Divergence

Recall: Discrete Random Variables

- Divergence inequality for pmfs p and q

$$D(\, p \,||\, q \,) = \mathrm{E}\left\{ \log_2 \frac{p(A)}{q(A)} \right\} = \sum_{\forall a_k} p(a_k) \, \log_2 \frac{p(a_k)}{q(a_k)} \geq 0$$

with equality if and only if $p = q$ (both pmfs are the same)

Kullback-Leibler Divergence

Recall: Discrete Random Variables

- Divergence inequality for pmfs p and q

$$D(\, p \,||\, q \,) = \mathrm{E}\left\{ \log_2 \frac{p(A)}{q(A)} \right\} = \sum_{\forall a_k} p(a_k) \, \log_2 \frac{p(a_k)}{q(a_k)} \;\geq\; 0$$

with equality if and only if $p = q$ (both pmfs are the same)

Continuous Random Variables

- Divergence inequality for pdfs f and g

$$\boxed{D(\, f \,||\, g \,) = \mathrm{E}\left\{ \log_2 \frac{f(S)}{g(S)} \right\} = \int_{-\infty}^{\infty} f(s) \, \log_2 \frac{f(s)}{g(s)} \, \mathrm{d}s \;\geq\; 0}$$

with equality if and only if $f = g$ (both pdfs are the same)

- Proof is analog to discrete case

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

➡ Differential entropy

$$h(S) \;=\; \mathrm{E}\Big\{ -\log_2 f_S(S) \Big\}$$

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

➡ Differential entropy

$$\begin{aligned} h(S) &= \mathrm{E}\Big\{ -\log_2 f_S(S) \Big\} \\ &= \mathrm{E}\Big\{ -\log_2 f_S(S) + \log_2 f_G(S) - \log_2 f_G(S) \Big\} \end{aligned}$$

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

➡ Differential entropy

$$\begin{aligned} h(S) &= \mathrm{E}\Big\{ -\log_2 f_S(S) \Big\} \\ &= \mathrm{E}\Big\{ -\log_2 f_S(S) + \log_2 f_G(S) - \log_2 f_G(S) \Big\} \\ &= \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} - D_{KL}\Big(f_S \,\Big|\Big|\, f_G \Big) \end{aligned}$$

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

➡ Differential entropy

$$\begin{aligned} h(S) &= \mathrm{E}\Big\{ -\log_2 f_S(S) \Big\} \\ &= \mathrm{E}\Big\{ -\log_2 f_S(S) + \log_2 f_G(S) - \log_2 f_G(S) \Big\} \\ &= \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} - D_{KL}\Big(f_S \,\Big|\Big|\, f_G \Big) \\ &\leq \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} \end{aligned}$$

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

➔ Differential entropy

$$
\begin{aligned}
h(S) &= \mathrm{E}\Big\{ -\log_2 f_S(S) \Big\} \\
&= \mathrm{E}\Big\{ -\log_2 f_S(S) + \log_2 f_G(S) - \log_2 f_G(S) \Big\} \\
&= \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} - D_{KL}\Big(f_S \,\Big|\Big|\, f_G \Big) \\
&\leq \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} \\
&= h^G(S) \;=\; \frac{1}{2} \log_2 \left(2\pi e\, \sigma^2 \right)
\end{aligned}
$$

Maximization of Differential Entropy

- Let S be any random variable with pdf $f_S(s)$, mean μ and variance σ^2
- Let $f_G(s)$ be the Gaussian pdf with the same mean μ and same variance σ^2

➡ Differential entropy

$$
\begin{aligned}
h(S) &= \mathrm{E}\Big\{ -\log_2 f_S(S) \Big\} \\
&= \mathrm{E}\Big\{ -\log_2 f_S(S) + \log_2 f_G(S) - \log_2 f_G(S) \Big\} \\
&= \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} - D_{KL}\Big(f_S \,\Big|\Big|\, f_G \Big) \\
&\leq \mathrm{E}\Big\{ -\log_2 f_G(S) \Big\} \\
&= h^G(S) \;=\; \frac{1}{2} \log_2 \left(2\pi e\, \sigma^2 \right)
\end{aligned}
$$

➡ **Gaussian random variable has the largest differential entropy among all random variables with the same variance σ^2**

Maximization of N-th Order Differential Entropy

- N-th order differential entropy for any source $\boldsymbol{S}$ with variance σ^2

$$h_N(\boldsymbol{S}) \;=\; h(S_k) + h(S_{k+1} \,|\, S_k) + \ldots + h(S_{k+N-1} \,|\, S_k, \cdots, S_{k+N-2})$$

Maximization of N-th Order Differential Entropy

- N-th order differential entropy for any source $\boldsymbol{S}$ with variance σ^2

$$\begin{aligned} h_N(\boldsymbol{S}) &= h(S_k) + h(S_{k+1} \,|\, S_k) + \ldots + h(S_{k+N-1} \,|\, S_k, \cdots, S_{k+N-2}) \\ &\leq N \cdot h(S) \qquad\qquad \text{(conditioning never increases entropy)} \end{aligned}$$

Maximization of N-th Order Differential Entropy

- N-th order differential entropy for any source $\boldsymbol{S}$ with variance σ^2

$$\begin{aligned} h_N(\boldsymbol{S}) &= h(S_k) + h(S_{k+1} \,|\, S_k) + \ldots + h(S_{k+N-1} \,|\, S_k, \cdots, S_{k+N-2}) \\ &\leq N \cdot h(S) \qquad\qquad \text{(conditioning never increases entropy)} \\ &\leq \frac{N}{2} \log_2 \left(2\pi e \, \sigma^2 \right) \;=\; h_N^{Gidd}(\boldsymbol{S}) \end{aligned}$$

Maximization of N-th Order Differential Entropy

- N-th order differential entropy for any source $\boldsymbol{S}$ with variance σ^2

$$
\begin{aligned}
h_N(\boldsymbol{S}) &= h(S_k) + h(S_{k+1} \,|\, S_k) + \ldots + h(S_{k+N-1} \,|\, S_k, \cdots, S_{k+N-2}) \\
&\leq N \cdot h(S) \qquad \text{(conditioning never increases entropy)} \\
&\leq \frac{N}{2} \log_2 \left(2\pi e\, \sigma^2 \right) = h_N^{Gidd}(\boldsymbol{S})
\end{aligned}
$$

Maximization of N-th order Differential Entropy

➜ **Gaussian iid process has the largest N-th order differential entropy among all stationary processes with the same variance**

$$
h_N(\boldsymbol{S}) \leq h_N^{Giid}(\boldsymbol{S}) = \frac{N}{2} \log_2 \left(2\pi e\, \sigma^2 \right)
$$

Lower Bound of Rate-Distortion Function

$$R(D) = \lim_{N \to \infty} \inf_{g_N :\, \delta_N(g_N) \le D} \frac{I_N(\boldsymbol{S}; \boldsymbol{S}')}{N}$$

Lower Bound of Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \lim_{N\to\infty} \inf_{g_N:\, \delta_N(\, g_N\,)\le D} \frac{I_N(\boldsymbol{S}; \boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \inf_{g_N:\, \delta_N(\, g_N\,)\le D} \frac{h_N(\boldsymbol{S}) - h_N(\boldsymbol{S} \mid \boldsymbol{S}')}{N}
\end{aligned}
$$

Lower Bound of Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \lim_{N\to\infty} \inf_{g_N:\, \delta_N(\, g_N\,)\le D} \frac{I_N(\boldsymbol{S}; \boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \inf_{g_N:\, \delta_N(\, g_N\,)\le D} \frac{h_N(\boldsymbol{S}) - h_N(\boldsymbol{S} \,|\, \boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} - \lim_{N\to\infty} \sup_{g_N:\, \delta_N(\, g_N\,)\le D} \frac{h_N(\boldsymbol{S} \,|\, \boldsymbol{S}')}{N}
\end{aligned}
$$

Lower Bound of Rate-Distortion Function

$$\begin{aligned}
R(D) &= \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{I_N(\boldsymbol{S};\boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \inf_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}) - h_N(\boldsymbol{S}\,|\,\boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}\,|\,\boldsymbol{S}')}{N} \\
\text{(a)} \quad &= \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}'\,|\,\boldsymbol{S}')}{N}
\end{aligned}$$

(a) Modification of the mean does not change differential entropy

Lower Bound of Rate-Distortion Function

$$
\begin{aligned}
R(D) &= \lim_{N\to\infty} \inf_{g_N:\,\delta_N(\,g_N\,)\le D} \frac{I_N(\boldsymbol{S};\boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \inf_{g_N:\,\delta_N(\,g_N\,)\le D} \frac{h_N(\boldsymbol{S}) - h_N(\boldsymbol{S}\,|\,\boldsymbol{S}')}{N} \\
&= \lim_{N\to\infty} \frac{h_N(\boldsymbol{S})}{N} - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(\,g_N\,)\le D} \frac{h_N(\boldsymbol{S}\,|\,\boldsymbol{S}')}{N} \\
\text{(a)} \quad &= \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(\,g_N\,)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}'\,|\,\boldsymbol{S}')}{N} \\
\text{(b)} \quad &\ge \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\,\delta_N(\,g_N\,)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}
\end{aligned}
$$

(a) Modification of the mean does not change differential entropy

(b) Conditioning does not increase differential entropy

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?
  - In the derivation, we used the inequality: $h_N(\boldsymbol{S}-\boldsymbol{S}' \,|\, \boldsymbol{S}') \leq h_N(\boldsymbol{S}-\boldsymbol{S}')$

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N : \delta_N(g_N) \leq D} \frac{h_N(\boldsymbol{S} - \boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?
  - In the derivation, we used the inequality: $h_N(\boldsymbol{S} - \boldsymbol{S}' \,|\, \boldsymbol{S}') \leq h_N(\boldsymbol{S} - \boldsymbol{S}')$

  ➡ Lower bound $R_L(D)$ is equal to rate-distortion function $R(D)$ if and only if the approximation error $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$ is independent of the reconstruction $\boldsymbol{S}'$

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?
  - In the derivation, we used the inequality: $h_N(\boldsymbol{S}-\boldsymbol{S}' \,|\, \boldsymbol{S}') \leq h_N(\boldsymbol{S}-\boldsymbol{S}')$

  ➡ Lower bound $R_L(D)$ is equal to rate-distortion function $R(D)$ if and only if the approximation error $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$ is independent of the reconstruction $\boldsymbol{S}'$

  ➡ Asymptotically achieved at high rates

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?
  - In the derivation, we used the inequality: $h_N(\boldsymbol{S}-\boldsymbol{S}' \,|\, \boldsymbol{S}') \leq h_N(\boldsymbol{S}-\boldsymbol{S}')$
  - ➡ Lower bound $R_L(D)$ is equal to rate-distortion function $R(D)$ if and only if the approximation error $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$ is independent of the reconstruction $\boldsymbol{S}'$
  - ➡ Asymptotically achieved at high rates

- Can we calculate it?

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?
  - In the derivation, we used the inequality: $h_N(\boldsymbol{S}-\boldsymbol{S}' \,|\, \boldsymbol{S}') \leq h_N(\boldsymbol{S}-\boldsymbol{S}')$

  ➡ Lower bound $R_L(D)$ is equal to rate-distortion function $R(D)$ if and only if the approximation error $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$ is independent of the reconstruction $\boldsymbol{S}'$

  ➡ Asymptotically achieved at high rates

- Can we calculate it?
  - For many distortion measures $d_N(\boldsymbol{s}, \boldsymbol{s}')$, it can be calculated

Shannon Lower Bound

- Lower bound $R_L(D)$ of rate-distortion function $R(D)$

$$R(D) \geq R_L(D), \qquad \boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\delta_N(g_N)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}}$$

- When does it coincide with rate-distortion function?
  - In the derivation, we used the inequality: $h_N(\boldsymbol{S}-\boldsymbol{S}' \,|\, \boldsymbol{S}') \leq h_N(\boldsymbol{S}-\boldsymbol{S}')$

  ➡ Lower bound $R_L(D)$ is equal to rate-distortion function $R(D)$ if and only if the approximation error $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$ is independent of the reconstruction $\boldsymbol{S}'$

  ➡ Asymptotically achieved at high rates

- Can we calculate it?
  - For many distortion measures $d_N(\boldsymbol{s}, \boldsymbol{s}')$, it can be calculated

  ➡ It can be calculated for all p-norm distortion measures

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

➡ MSE distortion D is given by

$$D = \mathrm{E}\{\, (S - S')^2 \,\} = \mathrm{E}\{\, Z^2 \,\} = \sigma_Z^2 + \mu_Z^2$$

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

➡ MSE distortion D is given by

$$D = \mathrm{E}\{\, (S - S')^2 \,\} = \mathrm{E}\{\, Z^2 \,\} = \sigma_Z^2 + \mu_Z^2$$

➡ Consider supremum of N-th order differential entropy

$$\sup_{g_N:\, \delta_N(g_N) \le D} h_N(\boldsymbol{S} - \boldsymbol{S}')$$

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

➡ MSE distortion D is given by

$$D = \mathrm{E}\{\, (S - S')^2 \,\} = \mathrm{E}\{\, Z^2 \,\} = \sigma_Z^2 + \mu_Z^2$$

➡ Consider supremum of N-th order differential entropy

$$\sup_{g_N:\, \delta_N(g_N) \le D} h_N(\boldsymbol{S} - \boldsymbol{S}') \;=\; \sup_{f_Z:\, \sigma_Z^2 + \mu_Z^2 \le D} h_N(\boldsymbol{Z})$$

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

➡ MSE distortion D is given by

$$D = \mathrm{E}\{\, (S - S')^2 \,\} = \mathrm{E}\{\, Z^2 \,\} = \sigma_Z^2 + \mu_Z^2$$

➡ Consider supremum of N-th order differential entropy

$$\begin{aligned} \sup_{g_N:\, \delta_N(g_N) \le D} h_N(\boldsymbol{S} - \boldsymbol{S}') \;&=\; \sup_{f_{\boldsymbol{Z}}:\, \sigma_Z^2 + \mu_Z^2 \le D} h_N(\boldsymbol{Z}) \\ \text{(a)} \quad &=\; \sup_{f_{\boldsymbol{Z}}:\, \sigma_Z^2 = D} h_N(\boldsymbol{Z}) \end{aligned}$$

(a) Modification of the mean does not change differential entropy

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

➡ MSE distortion D is given by

$$D = \mathrm{E}\{\, (S - S')^2 \,\} = \mathrm{E}\{\, Z^2 \,\} = \sigma_Z^2 + \mu_Z^2$$

➡ Consider supremum of N-th order differential entropy

$$\begin{aligned}
\sup_{g_N:\, \delta_N(g_N) \le D} h_N(\boldsymbol{S} - \boldsymbol{S}') &= \sup_{f_Z:\, \sigma_Z^2 + \mu_Z^2 \le D} h_N(\boldsymbol{Z}) \\
\text{(a)} \quad &= \sup_{f_Z:\, \sigma_Z^2 = D} h_N(\boldsymbol{Z}) \\
\text{(b)} \quad &= h_N^{Giid}(\boldsymbol{Z} \,|\, \sigma_Z^2 = D)
\end{aligned}$$

(a) Modification of the mean does not change differential entropy

(b) For given variance, Gaussian iid has maximum differential entropy

Shannon Lower Bound for MSE Distortion

- Consider random process for approximation error: $\boldsymbol{Z} = \boldsymbol{S} - \boldsymbol{S}'$

➡ MSE distortion D is given by

$$D = \mathrm{E}\{\, (S - S')^2 \,\} = \mathrm{E}\{\, Z^2 \,\} = \sigma_Z^2 + \mu_Z^2$$

➡ Consider supremum of N-th order differential entropy

$$\begin{aligned} \sup_{g_N:\, \delta_N(g_N) \le D} h_N(\boldsymbol{S} - \boldsymbol{S}') &= \sup_{f_Z:\, \sigma_Z^2 + \mu_Z^2 \le D} h_N(\boldsymbol{Z}) \\ \text{(a)} \quad &= \sup_{f_Z:\, \sigma_Z^2 = D} h_N(\boldsymbol{Z}) \\ \text{(b)} \quad &= h_N^{Giid}(\boldsymbol{Z} \,|\, \sigma_Z^2 = D) = \frac{N}{2} \log_2(2\pi e\, D) \end{aligned}$$

(a) Modification of the mean does not change differential entropy

(b) For given variance, Gaussian iid has maximum differential entropy

Shannon Lower Bound for MSE Distortion

- Using derived supremum of N-th order differential entropy

$$R_{\mathrm{L}}(D) \;=\; \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\, \delta_N(\,g_N\,)\leq D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}$$

Shannon Lower Bound for MSE Distortion

- Using derived supremum of N-th order differential entropy

$$R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\, \delta_N(\, g_N\,)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}$$

$$= \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \frac{1}{N} \cdot \frac{N}{2} \log_2(2\pi e\, D)$$

Shannon Lower Bound for MSE Distortion

- Using derived supremum of N-th order differential entropy

$$R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \sup_{g_N:\, \delta_N(g_N)\le D} \frac{h_N(\boldsymbol{S}-\boldsymbol{S}')}{N}$$

$$= \bar{h}(\boldsymbol{S}) - \lim_{N\to\infty} \frac{1}{N}\cdot\frac{N}{2}\log_2(2\pi e\, D)$$

Shannon Lower Bound for MSE Distortion

- Shannon lower bound as rate-distortion and distortion-rate function

$$\boxed{R_{\mathrm{L}}(D) = \bar{h}(\boldsymbol{S}) - \frac{1}{2}\log_2(2\pi e\, D)} \qquad \text{and} \qquad \boxed{D_{\mathrm{L}}(R) = \frac{1}{2\pi e}\cdot 2^{2\bar{h}(\boldsymbol{S})}\cdot 2^{-2R}}$$

General Form of Shannon Lower Bound (MSE)

- Shannon Lower Bound as distortion-rate function

$$\begin{aligned} D_L(R) &= \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S})} \cdot 2^{-2R} \\ &= \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S}/\sigma)} \end{aligned}$$

where $\boldsymbol{S}/\sigma$ represent the unit-variance random process

General Form of Shannon Lower Bound (MSE)

- Shannon Lower Bound as distortion-rate function

$$
\begin{aligned}
D_L(R) &= \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S})} \cdot 2^{-2R} \\
&= \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S}/\sigma)}
\end{aligned}
$$

  where $\boldsymbol{S}/\sigma$ represent the unit-variance random process

➡ The factor ε_L^2 does only depend on the shape of the pdf, not on its variance

General Form of Shannon Lower Bound (MSE)

- Shannon Lower Bound as distortion-rate function

$$\begin{aligned} D_L(R) &= \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S})} \cdot 2^{-2R} \\ &= \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S}/\sigma)} \end{aligned}$$

where $\boldsymbol{S}/\sigma$ represent the unit-variance random process

➡ The factor ε_L^2 does only depend on the shape of the pdf, not on its variance

➡ General form of Shannon lower bound for MSE distortion

$$\begin{aligned} D_L(R) &= \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \\ R_L(D) &= \frac{1}{2} \log_2\left(\frac{\varepsilon_L^2 \cdot \sigma^2}{D}\right) \end{aligned} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S}/\sigma)} = \frac{1}{2\pi e\,\sigma^2} \cdot 2^{2\bar{h}(\boldsymbol{S})}$$

Shannon Lower Bound (MSE) for IID Sources

- For iid sources, we have

$$\bar{h}(\boldsymbol{S}) = h(S)$$

Shannon Lower Bound for MSE distortion and IID sources

- Distortion-rate and rate-distortion function

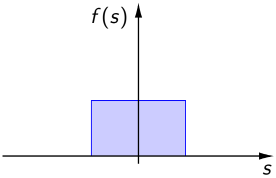
$$D_L(R) = \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R}$$
$$R_L(D) = \frac{1}{2} \log_2 \left(\frac{\varepsilon_L^2 \cdot \sigma^2}{D} \right)$$

with

$$\varepsilon_L^2 = \frac{1}{2\pi e} \cdot 2^{2h(S/\sigma)} = \frac{1}{2\pi e\, \sigma^2} \cdot 2^{2h(S)}$$

Shannon Lower Bound (MSE) for Selected IID Sources

Uniform

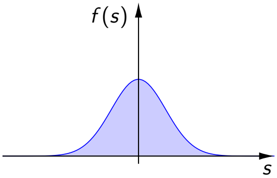


$$f(s) = \begin{cases} \frac{1}{2a} & : \; |s| \le a \\ 0 & : \; \text{otherwise} \end{cases}$$

$$h(S) = \frac{1}{2}\log_2(12\,\sigma_S^2)$$

$$D_L(R) = \underbrace{\frac{6}{\pi e}}_{\approx 0.7}\,\sigma_S^2 \cdot 2^{-2R}$$

Gaussian

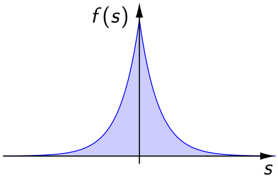


$$f(s) = \frac{1}{\sigma_S\sqrt{2\pi}}\, e^{-\frac{s^2}{2\sigma_S^2}}$$

$$h(S) = \frac{1}{2}\log_2(2\pi e\,\sigma_S^2)$$

$$D_L(R) = \sigma_S^2 \cdot 2^{-2R}$$

Laplacian



$$f(s) = \frac{1}{\sigma_S\sqrt{2}}\, e^{-\frac{\sqrt{2}}{\sigma_S}|s|}$$

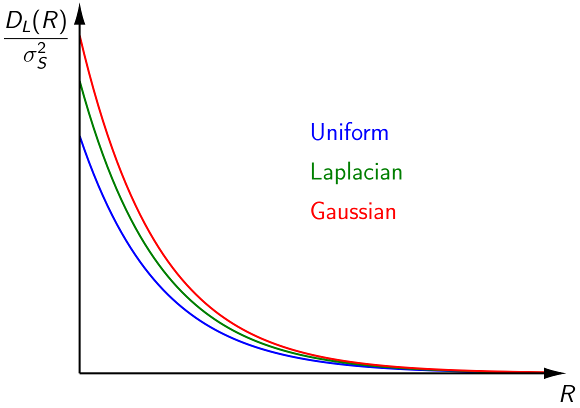
$$h(S) = \frac{1}{2}\log_2(2\,e^2\,\sigma_S^2)$$

$$D_L(R) = \underbrace{\frac{e}{\pi}}_{\approx 0.865}\,\sigma_S^2 \cdot 2^{-2R}$$

Shannon Lower Bound (MSE) for Selected IID Sources

- Shannon lower bound for MSE and IID sources

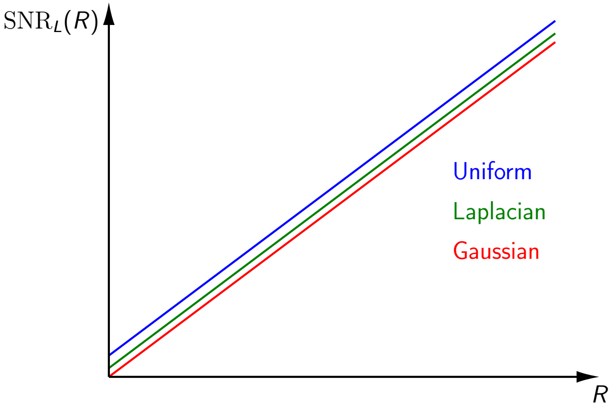
$$D_L(R) = \varepsilon_L^2 \cdot \sigma_S^2 \cdot 2^{-2R}$$



Shannon Lower Bound (MSE) for Selected IID Sources

- Shannon lower bound as signal-to-noise ratio (SNR)

$$\mathrm{SNR} = 10\,\log_{10}\frac{\sigma_S^2}{D} \qquad \Longrightarrow \qquad \mathrm{SNR}_L(R) = \underbrace{(20\log_{10}2)}_{\approx 6.02}\cdot R - (10\log_{10}\varepsilon_L^2)$$



Asymptotic Tightness of Shannon Lower Bound

- Shannon lower bound $R_L(D)$ approaches rate-distortion function $R(D)$ for high rates R

$$\lim_{R \to \infty} D(R) - D_{\mathrm{L}}(R) = 0$$

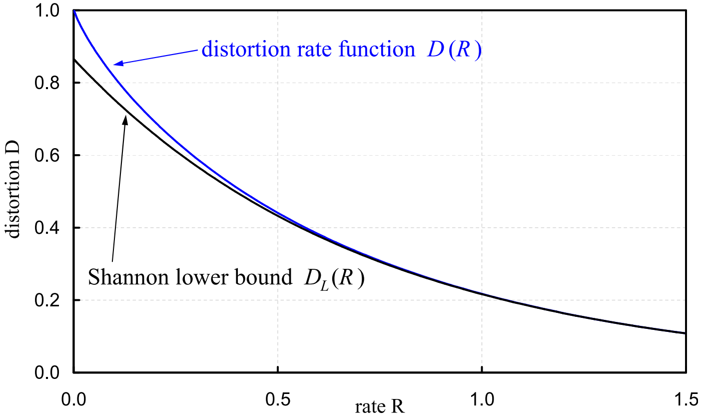
Asymptotic Tightness of Shannon Lower Bound

- Shannon lower bound $R_L(D)$ approaches rate-distortion function $R(D)$ for high rates R

$$\lim_{R\to\infty} D(R) - D_{\mathrm{L}}(R) = 0$$

Example:

Laplacian IID



High-Rate Approximations of Rate-Distortion Function

- Shannon Lower Bound for IID sources

$$\begin{aligned}
\text{Gaussian:} \quad D_L(R) &= \sigma^2 \cdot 2^{-2R} \\
\text{Laplacian:} \quad D_L(R) &= \frac{e}{\pi} \cdot \sigma^2 \cdot 2^{-2R} \\
\text{Uniform:} \quad D_L(R) &= \frac{6}{\pi e} \cdot \sigma^2 \cdot 2^{-2R}
\end{aligned}$$

High-Rate Approximations of Rate-Distortion Function

- Shannon Lower Bound for IID sources

$$\begin{aligned} \text{Gaussian:} \quad D_L(R) &= \sigma^2 \cdot 2^{-2R} \\ \text{Laplacian:} \quad D_L(R) &= \frac{e}{\pi} \cdot \sigma^2 \cdot 2^{-2R} \\ \text{Uniform:} \quad D_L(R) &= \frac{6}{\pi e} \cdot \sigma^2 \cdot 2^{-2R} \end{aligned}$$

- Shannon Lower Bound for Gaussian-Markov Sources (without derivation)

$$D_L(R) = (1 - \varrho^2) \cdot \sigma^2 \cdot 2^{-2R}$$

High-Rate Approximations of Rate-Distortion Function

- Shannon Lower Bound for IID sources

$$\begin{aligned} \text{Gaussian:} \quad D_L(R) &= \sigma^2 \cdot 2^{-2R} \\ \text{Laplacian:} \quad D_L(R) &= \frac{e}{\pi} \cdot \sigma^2 \cdot 2^{-2R} \\ \text{Uniform:} \quad D_L(R) &= \frac{6}{\pi e} \cdot \sigma^2 \cdot 2^{-2R} \end{aligned}$$

- Shannon Lower Bound for Gaussian-Markov Sources (without derivation)

$$D_L(R) = (1 - \varrho^2) \cdot \sigma^2 \cdot 2^{-2R}$$

➡ **Can use these approximations of the rate-distortion function for evaluating the coding efficiency of lossy coding techniques for high rates**

High-Rate Approximation for Lloyd Quantizer vs Shannon Lower Bound

- Shannon Lower Bound (high-rate approximation of rate-distortion function)

$$D_L(R) = \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e\,\sigma^2} \cdot 2^{2h(S)}$$

High-Rate Approximation for Lloyd Quantizer vs Shannon Lower Bound

- Shannon Lower Bound (high-rate approximation of rate-distortion function)

$$D_L(R) = \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e\, \sigma^2} \cdot 2^{2h(S)}$$

- High-rate approximation for Lloyd quantizer with fixed-length coding

$$D_F(R) = \varepsilon_F^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_F^2 = \frac{1}{12\, \sigma^2} \left(\int_{-\infty}^{\infty} \sqrt[3]{f(s)}\, \mathrm{d}s \right)^3$$

High-Rate Approximation for Lloyd Quantizer vs Shannon Lower Bound

- Shannon Lower Bound (high-rate approximation of rate-distortion function)

$$D_L(R) = \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e\, \sigma^2} \cdot 2^{2h(S)}$$

- High-rate approximation for Lloyd quantizer with fixed-length coding

$$D_F(R) = \varepsilon_F^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_F^2 = \frac{1}{12\, \sigma^2} \left(\int_{-\infty}^{\infty} \sqrt[3]{f(s)}\, \mathrm{d}s \right)^3$$

➡ Distortion increase for Lloyd quantizer (with fixed-length coding) relative to Shannon lower bound

$$\frac{D_F(R)}{D_L(R)} = \frac{\varepsilon_F^2}{\varepsilon_L^2}$$

High-Rate Approximation for Lloyd Quantizer vs Shannon Lower Bound

- Shannon Lower Bound (high-rate approximation of rate-distortion function)

$$D_L(R) = \varepsilon_L^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_L^2 = \frac{1}{2\pi e\, \sigma^2} \cdot 2^{2h(S)}$$

- High-rate approximation for Lloyd quantizer with fixed-length coding

$$D_F(R) = \varepsilon_F^2 \cdot \sigma^2 \cdot 2^{-2R} \qquad \text{with} \qquad \varepsilon_F^2 = \frac{1}{12\, \sigma^2} \left(\int_{-\infty}^{\infty} \sqrt[3]{f(s)}\, \mathrm{d}s \right)^3$$

➡ Distortion increase for Lloyd quantizer (with fixed-length coding) relative to Shannon lower bound

$$\frac{D_F(R)}{D_L(R)} = \frac{\varepsilon_F^2}{\varepsilon_L^2} \qquad \Longrightarrow \qquad \begin{array}{lll} \text{Uniform iid:} & \dfrac{\pi e}{6} \approx 1.42 & (1.53\,\text{dB}) \\ \text{Gaussian iid:} & \dfrac{\sqrt{3}\,\pi}{2} \approx 2.72 & (4.34\,\text{dB}) \\ \text{Laplacian iid:} & \dfrac{9\pi}{2e} \approx 5.20 & (7.16\,\text{dB}) \end{array}$$

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \; D \le \sigma^2 \\ 0 & : \; D > \sigma^2 \end{cases}$$

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \quad D \le \sigma^2 \\ 0 & : \quad D > \sigma^2 \end{cases}$$

- Distortion-rate function as signal-to-noise ratio (SNR)

$$\mathrm{SNR}(R) \;=\; 10\log_{10} \frac{\sigma^2}{D(R)}$$

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \quad D \le \sigma^2 \\ 0 & : \quad D > \sigma^2 \end{cases}$$

- Distortion-rate function as signal-to-noise ratio (SNR)

$$\mathrm{SNR}(R) \;=\; 10\log_{10}\frac{\sigma^2}{D(R)} \;=\; (20\log_{10} 2)\, R$$

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \quad D \le \sigma^2 \\ 0 & : \quad D > \sigma^2 \end{cases}$$

- Distortion-rate function as signal-to-noise ratio (SNR)

$$\mathrm{SNR}(R) \;=\; 10\log_{10}\frac{\sigma^2}{D(R)} \;=\; (20\log_{10} 2)\,R \;\approx\; 6.02\,R \qquad [\text{dB}]$$

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \; D \le \sigma^2 \\ 0 & : \; D > \sigma^2 \end{cases}$$

- Distortion-rate function as signal-to-noise ratio (SNR)

$$\mathrm{SNR}(R) \;=\; 10\log_{10}\frac{\sigma^2}{D(R)} \;=\; (20\log_{10} 2)\,R \;\approx\; 6.02\,R \qquad [\mathrm{dB}]$$

MSE distortion and given signal variance σ^2

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \quad D \le \sigma^2 \\ 0 & : \quad D > \sigma^2 \end{cases}$$

- Distortion-rate function as signal-to-noise ratio (SNR)

$$\mathrm{SNR}(R) \; = \; 10\log_{10}\frac{\sigma^2}{D(R)} \; = \; (20\log_{10}2)\,R \; \approx \; 6.02\,R \qquad \text{[dB]}$$

MSE distortion and given signal variance σ^2

➡ Rate-distortion function $R(D)$ is maximized for Gaussian IID sources

Rate-Distortion Function (MSE) for Gaussian IID Sources

Gaussian IID sources and MSE distortion

- Rate-distortion function coincides with Shannon lower bound

$$D(R) = \sigma^2 \cdot 2^{-2R}, \qquad R(D) = \begin{cases} \frac{1}{2}\log_2 \frac{\sigma^2}{D} & : \quad D \leq \sigma^2 \\ 0 & : \quad D > \sigma^2 \end{cases}$$

- Distortion-rate function as signal-to-noise ratio (SNR)

$$\mathrm{SNR}(R) \; = \; 10\log_{10}\frac{\sigma^2}{D(R)} \; = \; (20\log_{10}2)\,R \; \approx \; 6.02\,R \qquad [\mathrm{dB}]$$

MSE distortion and given signal variance σ^2

➡ Rate-distortion function $R(D)$ is maximized for Gaussian IID sources

➡ Gaussian IID sources are the hardest to code

Rate-Distortion Function for Stationary Gaussian and MSE

Rate-Distortion Function for Stationary Gaussian Sources

- Parametric formulation with $\theta > 0$

$$D(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\left(\Phi_{SS}(\omega), \theta\right) \mathrm{d}\omega$$

$$R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left(0, \frac{1}{2} \log_2 \frac{\Phi_{SS}(\omega)}{\theta}\right) \mathrm{d}\omega$$

with $\Phi_{SS}(\omega)$ being the Fourier series of the autocovariance function ϕ_k

$$\Phi_{SS}(\omega) = \sum_{k=-\infty}^{\infty} \phi_k \cdot e^{-\mathrm{i}\,\omega\,k} \quad \text{with} \quad \phi_k = \mathrm{E}\{\,(S_n - \mu)(S_{n+k} - \mu)\,\}$$

Rate-Distortion Function for Stationary Gaussian and MSE

Rate-Distortion Function for Stationary Gaussian Sources

- Parametric formulation with $\theta > 0$

$$D(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\left(\Phi_{SS}(\omega), \theta\right) \, \mathrm{d}\omega$$

$$R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left(0, \frac{1}{2} \log_2 \frac{\Phi_{SS}(\omega)}{\theta}\right) \, \mathrm{d}\omega$$

with $\Phi_{SS}(\omega)$ being the Fourier series of the autocovariance function ϕ_k

$$\Phi_{SS}(\omega) = \sum_{k=-\infty}^{\infty} \phi_k \cdot e^{-\mathrm{i}\,\omega\,k} \quad \text{with} \quad \phi_k = \mathrm{E}\{\, (S_n - \mu)(S_{n+k} - \mu) \,\}$$

MSE distortion and given autocovariance function ϕ_k

Rate-Distortion Function for Stationary Gaussian and MSE

Rate-Distortion Function for Stationary Gaussian Sources

- Parametric formulation with $\theta > 0$

$$D(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\left(\Phi_{SS}(\omega), \theta\right) \mathrm{d}\omega$$

$$R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left(0, \frac{1}{2} \log_2 \frac{\Phi_{SS}(\omega)}{\theta}\right) \mathrm{d}\omega$$

with $\Phi_{SS}(\omega)$ being the Fourier series of the autocovariance function ϕ_k

$$\Phi_{SS}(\omega) = \sum_{k=-\infty}^{\infty} \phi_k \cdot e^{-\mathrm{i}\,\omega\,k} \quad \text{with} \quad \phi_k = \mathrm{E}\{\,(S_n - \mu)(S_{n+k} - \mu)\,\}$$

MSE distortion and given autocovariance function ϕ_k

➡ Rate-distortion function $R(D)$ is maximized for Gaussian sources

Rate-Distortion Function for Stationary Gaussian and MSE

Rate-Distortion Function for Stationary Gaussian Sources

- Parametric formulation with $\theta > 0$

$$D(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min\left(\Phi_{SS}(\omega), \theta \right) \, \mathrm{d}\omega$$

$$R(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max\left(0, \frac{1}{2} \log_2 \frac{\Phi_{SS}(\omega)}{\theta} \right) \, \mathrm{d}\omega$$

with $\Phi_{SS}(\omega)$ being the Fourier series of the autocovariance function ϕ_k

$$\Phi_{SS}(\omega) = \sum_{k=-\infty}^{\infty} \phi_k \cdot e^{-\mathrm{i}\,\omega\,k} \quad \text{with} \quad \phi_k = \mathrm{E}\{\,(S_n - \mu)(S_{n+k} - \mu)\,\}$$

MSE distortion and given autocovariance function ϕ_k

➡ Rate-distortion function $R(D)$ is maximized for Gaussian sources

➡ Gaussian sources are the hardest to code

Special Case: Stationary Gauss-Markov Source and MSE Distortion

- Autocovariance function and its Fourier series

$$\phi_k = \sigma^2 \, \rho^{|k|} \qquad \Longleftrightarrow \qquad \Phi_{SS}(\omega) = \frac{\sigma^2 \, (1 - \varrho^2)}{1 - 2\varrho \, \cos\omega + \varrho^2}$$

Special Case: Stationary Gauss-Markov Source and MSE Distortion

- Autocovariance function and its Fourier series

$$\phi_k = \sigma^2 \, \rho^{|k|} \qquad \Longleftrightarrow \qquad \Phi_{SS}(\omega) = \frac{\sigma^2 \, (1-\varrho^2)}{1 - 2\varrho \, \cos\omega + \varrho^2}$$

- For $\varrho \geq 0$: All frequency components are coded if we choose

$$\theta \leq \min_{\forall \omega} \Phi_{SS}(\omega) = \Phi_{SS}(\pi) = \sigma^2 \, \frac{1-\varrho^2}{1+2\varrho+\varrho^2} = \sigma^2 \, \frac{1-\varrho}{1+\varrho}$$

Special Case: Stationary Gauss-Markov Source and MSE Distortion

- Autocovariance function and its Fourier series

$$\phi_k = \sigma^2 \rho^{|k|} \qquad \Longleftrightarrow \qquad \Phi_{SS}(\omega) = \frac{\sigma^2 \left(1-\varrho^2\right)}{1 - 2\varrho \cos\omega + \varrho^2}$$

- For $\varrho \geq 0$: All frequency components are coded if we choose

$$\theta \leq \min_{\forall\omega} \Phi_{SS}(\omega) = \Phi_{SS}(\pi) = \sigma^2 \frac{1-\varrho^2}{1+2\varrho+\varrho^2} = \sigma^2 \frac{1-\varrho}{1+\varrho}$$

- For this range, rate-distortion function is equal to Shannon lower bound

$$R(D) \;=\; \frac{1}{2}\log_2 \frac{\sigma^2 \left(1-\varrho^2\right)}{D} \qquad \text{for } D \leq \sigma^2 \frac{1-\varrho}{1+\varrho}$$

Special Case: Stationary Gauss-Markov Source and MSE Distortion

- Autocovariance function and its Fourier series

$$\phi_k = \sigma^2\,\rho^{|k|} \qquad \Longleftrightarrow \qquad \Phi_{SS}(\omega) = \frac{\sigma^2\,(1-\varrho^2)}{1 - 2\varrho\,\cos\omega + \varrho^2}$$

- For $\varrho \geq 0$: All frequency components are coded if we choose

$$\theta \leq \min_{\forall\omega} \Phi_{SS}(\omega) = \Phi_{SS}(\pi) = \sigma^2\,\frac{1-\varrho^2}{1+2\varrho+\varrho^2} = \sigma^2\,\frac{1-\varrho}{1+\varrho}$$

- For this range, rate-distortion function is equal to Shannon lower bound

$$R(D) \;=\; \frac{1}{2}\log_2 \frac{\sigma^2\,(1-\varrho^2)}{D} \qquad \text{for } D \leq \sigma^2\,\frac{1-\varrho}{1+\varrho}$$

$$D(R) \;=\; (1-\varrho^2)\,\sigma^2\,2^{-2R} \qquad \text{for } R \geq \log_2(1+\varrho)$$

Special Case: Stationary Gauss-Markov Source and MSE Distortion

- Autocovariance function and its Fourier series

$$\phi_k = \sigma^2 \, \rho^{|k|} \quad \Longleftrightarrow \quad \Phi_{SS}(\omega) = \frac{\sigma^2 \, (1-\varrho^2)}{1 - 2\varrho \cos\omega + \varrho^2}$$

- For $\varrho \geq 0$: All frequency components are coded if we choose

$$\theta \leq \min_{\forall \omega} \Phi_{SS}(\omega) = \Phi_{SS}(\pi) = \sigma^2 \, \frac{1-\varrho^2}{1+2\varrho+\varrho^2} = \sigma^2 \, \frac{1-\varrho}{1+\varrho}$$

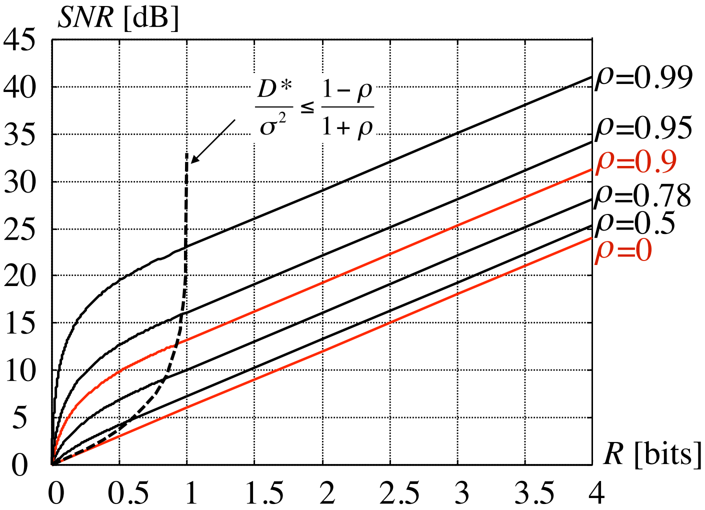
- For this range, rate-distortion function is equal to Shannon lower bound

$$R(D) \; = \; \frac{1}{2} \log_2 \frac{\sigma^2 \, (1-\varrho^2)}{D} \qquad \text{for } D \leq \sigma^2 \, \frac{1-\varrho}{1+\varrho}$$

$$D(R) \; = \; (1-\varrho^2) \, \sigma^2 \, 2^{-2R} \qquad \text{for } R \geq \log_2(1+\varrho)$$

➡ Guaranteed to be equal to Shannon lower bound for $R \geq 1$ bit/sample

Rate-Distortion Function for Stationary Gauss-Markov and MSE Distortion



Summary of Lecture

Fundamental Source Coding Theorem

- Greatest lower bound for source coding: Rate-Distortion Function $R(D)$

$$R(D) = \inf_{Q:\, \delta(Q) \le D} r(Q) = \lim_{N \to \infty} \inf_{g_N:\, \delta_N(g_N) \le D} \frac{I_N(\, g_N\,)}{N}$$

Shannon Lower Bound

- Shannon lower bound for MSE distortion

$$R_L(D) = \bar{h}(\boldsymbol{S}) - \frac{1}{2}\log_2(2\pi e\, D) \qquad \text{and} \qquad D_L(R) = \frac{1}{2\pi e} \cdot 2^{2\bar{h}(\boldsymbol{S})} \cdot 2^{-2R}$$

- Asymptotically tight: Suitable reference for performance evaluation at high rates

Rate-Distortion Function for Gaussian Sources and MSE Distortion

- Gaussian IID sources: Coincides with Shannon lower bound
- Stationary Gaussian: Parametric formulation
- Stationary Gauss-Markov: Coincides with Shannon lower bound for $R \ge \log_2(1+\varrho)$

Exercise 1: Covariance Function for AR(1) Sources

Given is a zero-mean iid process $\boldsymbol{Z} = \{Z_n\}$ with variance σ_Z^2. For a given correlation coefficient ϱ and mean μ_S, a stationary continuous AR(1) process is constructed according to

$$S_n = \mu_S + \varrho\,(S_{n-1} - \mu_S) + Z_n$$

(a) What is the variance σ_S^2 of the resulting process $\{S_n\}$?

(b) How do we have to modify the construction rule in order to get an AR(1) process with a pre-defined variance σ_S^2?

(c) Proof that

$$\mathrm{cov}(S_k, S_\ell) = \sigma_S^2 \cdot \varrho^{|k-\ell|}$$

Exercise 2: Mutual Information for Discrete Case

Given is a stationary Markov process $\boldsymbol{S} = \{S_n\}$ with the binary symbol alphabet $\mathcal{A} = \{x, y\}$. The conditional symbol probabilities $p(s_n \,|\, s_{n-1})$ are given in the table below.

s_n	$p(s_n \,\vert\, s_{n-1} = x)$	$p(s_n \,\vert\, s_{n-1} = y)$
x	$3/4$	$1/4$
y	$1/4$	$3/4$

Calculate:

- the marginal entropy $H(S_n)$,
- the joint entropy $H(S_n, S_{n+1})$ for two successive random variables,
- the conditional entropy $H(S_n \,|\, S_{n-1})$ for a random variable given the preceding random variable,
- the mutual information $I(S_n; S_{n+1})$ between two successive random variables.

Exercise 3: Mutual Information for Stationary Gauss-Markov (Optional)

Consider a stationary Gauss-Markov process $\boldsymbol{X} = \{X_n\}$ with mean μ, variance σ^2, and the correlation coefficient ϱ (correlation coefficient between two successive random variables).

Determine the mutual information $I(X_k; X_{k+N})$ between two random variables X_k and X_{k+N}, where the distance between the random variables is N times the sampling interval.

Interpret the results for the special cases $\varrho = -1$, $\varrho = 0$, and $\varrho = 1$.

Hint: It can be shown that

$$E\left\{(\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}} \cdot \boldsymbol{C}_N^{-1} \cdot (\boldsymbol{X} - \boldsymbol{\mu})\right\} = N,$$

which can be useful for the problem.

Exercise 4: Shannon Lower Bound (MSE Distortion)

Determine the Shannon lower bound for MSE distortion, as distortion-rate function, for iid processes with the following pdfs:

- The exponential pdf $\quad f_E(x) = \lambda \cdot e^{-\lambda \cdot x}$, with $x \geq 0$
- The half-normal pdf $\quad f_H(x) = \sqrt{\frac{4a}{\pi}} \cdot e^{-a \cdot x^2}$, with $x \geq 0$

Express the distortion-rate functions for the Shannon lower bound as a function of the variance σ^2.

Which of the given pdfs is easier to code (if the variance is the same)?

Verify that both pdfs are easier to code than the Gaussian iid with the same variance.

Exercise 5: Shannon Lower Bound for MAE Distortion (Optional)

Consider rate-distortion bounds for MAE (mean absolute error) distortion.

- Calculate the differential entropy for the Laplace pdf

$$f(s) = \frac{\lambda}{2} \cdot e^{-\lambda \cdot |x-\mu|}$$

  as a function of $m = E\{|X-\mu|\}$.

- Show that the Laplace pdf is the pdf with the maximum differential entropy of all pdfs with the same value of $m = E\{|X-\mu|\}$.

- Derive the Shannon lower bound for iid sources and the MAE distortion measure $d(x, x') = |x - x'|$. Formulate the Shannon lower bound as rate-distortion and as distortion-rate function.

- Calculate the Shannon lower bound for the MAE distortion measure for the Laplace iid source.