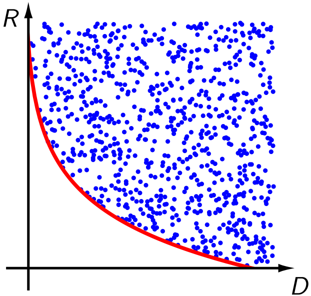
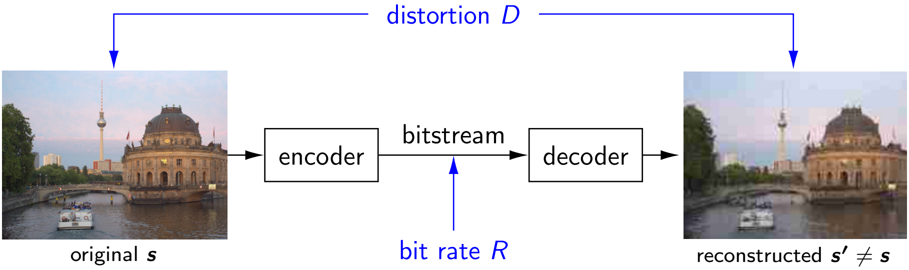
Rate-Distortion Theory



Lossy Coding



Lossy coding is characterized by two aspects:

- Bit rate R: Average number of bits per sample (or per time unit)
- Distortion D: Measure for deviation between original signal $\boldsymbol{s}$ and reconstructed signal $\boldsymbol{s}'$

Question: What is the maximum achievable coding efficiency?
(trade-off between bit rate R and distortion D)

Rate-Distortion Theory

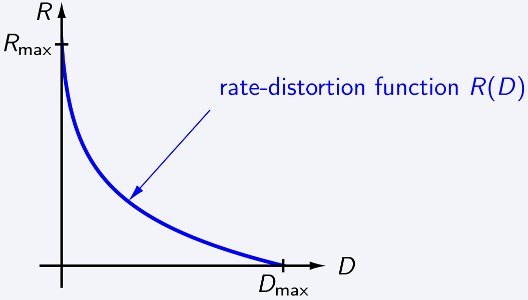
Rate-Distortion Function

- Theoretical bound for lossy compression
  - ➡ Minimum bit rate R for given distortion D
  - ➡ Minimum distortion D for given bit rate R

➡ Property of a source (no consideration of a specific coding method)

Example:

Discrete source



➡ **Require: Probabilistic model of the source**

Probabilistic Modeling of Sources

Source Coding in Practice

- Encoder and decoder are computer programs

➡ Actual input signals are **discrete-time** and **discrete-amplitude** signals

Real-world signals

- In most cases: Continuous-time and continuous-amplitude signals
- Discrete signals are obtained by sampling and quantization
- Typical scenarios: Initial quantization has negligible effect on source coding

Theoretical Analysis of Lossy Source Coding

- Will mostly use models for **discrete-time** and **continuous-amplitude** signals
- Main reason: Mathematical tractability

➡ Interpretation: Consider signal before initial quantization

➡ **Models for discrete-time and continuous-amplitude random processes**

Review: Random Variables and CDFs

Random Variables

- A **random variable** S is a function of the sample space $\mathcal{O}$ that assigns a real value $S(\zeta)$ to each outcome $\zeta \in \mathcal{O}$ of a random experiment

Cumulative Distribution Function

- **Cumulative distribution function** (cdf) of a random variable S

$$F_S(s) = \mathrm{P}(S \leq s) = \mathrm{P}(\{\zeta : S(\zeta) \leq s\}) \tag{1}$$

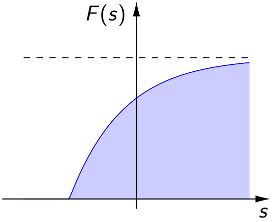
- Joint cdf / joint distribution of two random variables X and Y

$$F_{XY}(x, y) = \mathrm{P}(X \leq x,\ Y \leq y) \tag{2}$$

- Conditional cdf of a random variable X given another random variable Y

$$F_{X|Y}(x\,|\,y) = \mathrm{P}(X \leq x\,|\,Y \leq y) = \frac{\mathrm{P}(X \leq x,\ Y \leq y)}{\mathrm{P}(Y \leq y)} = \frac{F_{XY}(x, y)}{F_Y(y)} \tag{3}$$

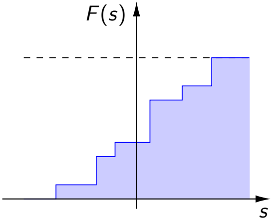
Examples: Cumulative Distribution Functions



Continuous function

- Random variable S can take all values inside one or more non-zero intervals

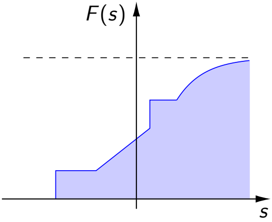
➡ **Continuous random variable**



Staircase function

- Random variable S can only take a countable number of values

➡ **Discrete random variable**



Mixed type

- Random variable S can take all values inside one or more intervals and a countable number of additional values

Probability Density Function (PDF)

Continuous Random Variables

- A random variable S is called a **continuous random variable**, if and only if its cdf $F_S(s)$ is a continuous function

Definition: Probability Density Function

- Probability density function (pdf) of a continuous random variable S

$$f_S(s) = \frac{\partial}{\partial s} F_S(s) \quad \Longleftrightarrow \quad F_S(s) = \int_{-\infty}^{s} f_S(t)\,\mathrm{d}t \tag{4}$$

Properties

- $f_S(s) \geq 0, \; \forall s$
- $\int_{-\infty}^{\infty} f_S(t)\,\mathrm{d}t = 1$
- $\mathrm{P}(a < S \leq b) = \int_a^b f_S(t)\,\mathrm{d}t$

Examples for Continuous Distributions (Zero Mean)

Uniform

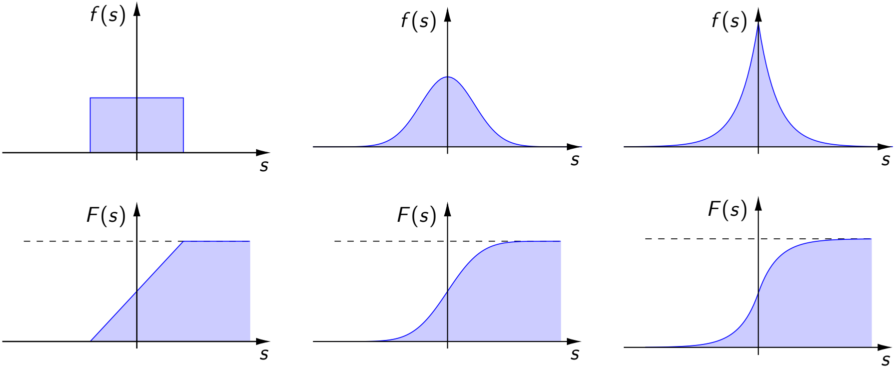
$$f(s) = \begin{cases} \frac{1}{2a} & : \ |s| \le a \\ 0 & : \ \text{otherwise} \end{cases}$$

Gaussian

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

Laplacian

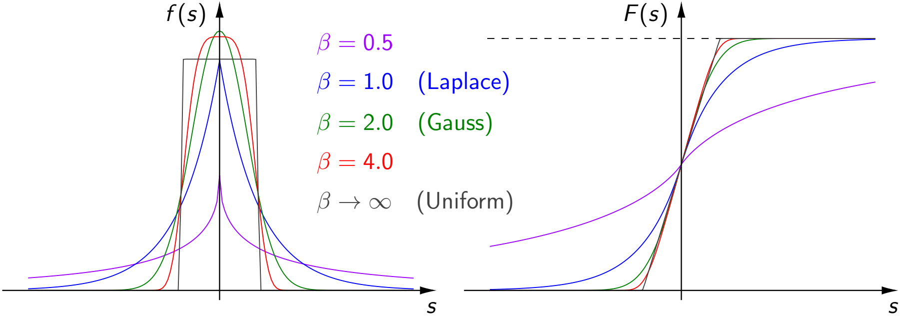
$$f(s) = \frac{1}{\sqrt{2\sigma^2}}\, e^{-\sqrt{\frac{2}{\sigma^2}}\,|s|}$$



Generalized Gaussian Distribution

Shape parameter $\beta \in (0, \infty)$:

$$f(s) = \frac{\beta}{2\,\alpha\,\Gamma(1/\beta)}\, e^{-\left(\frac{|s-\mu|}{\alpha}\right)^{\beta}} \qquad \text{with} \qquad \Gamma(x) = \int_0^{\infty} e^{-t}\, t^{x-1}\, \mathrm{d}t$$



➡ **Suitable approximation for many distributions**

Joint and Conditional Probability Density Function

Joint Probability Density Function

- Joint pdf of two random variables X and Y

$$f_{XY}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{XY}(x,y) \tag{5}$$

Conditional Probability Density Function

- Conditional pdf of a random variable S given event $\mathcal{B}$, with $\mathrm{P}(\mathcal{B}) > 0$

$$f_{S|\mathcal{B}}(s\,|\,\mathcal{B}) = \frac{\partial}{\partial x} F_{S|\mathcal{B}}(s\,|\,\mathcal{B}) \tag{6}$$

- Conditional pdf of a random variable X given another random variable Y

$$f_{X|Y}(x\,|\,y) = \frac{\partial}{\partial x} F_{X|Y}(x\,|\,y) = \frac{\frac{\partial^2}{\partial x\,\partial y} F_{XY}(x,y)}{\frac{\partial}{\partial y} F_Y(y)} = \frac{f_{XY}(x,y)}{f_Y(y)} \tag{7}$$

Pdf for Discrete Random Variables

Dirac Delta Function $\delta(x)$

- Generalized function (or distribution) with following properties

$$\delta(x) = \begin{cases} +\infty & : \quad x = 0 \\ 0 & : \quad x \neq 0 \end{cases} \qquad \text{with} \qquad \int_{-\infty}^{\infty} \delta(x)\, \mathrm{d}x = 1 \tag{8}$$

➡ Sifting or sampling property

$$\int_{-\infty}^{\infty} f(x)\, \delta(x-a)\, \mathrm{d}x = f(a) \tag{9}$$

Pdf for discrete random variables

- In strict sense: Pdf is not defined for discrete random variables
- Pdf using Dirac delta function δ

$$f_S(s) = \sum_{\forall a \in \mathcal{A}_S} \delta(s-a)\, p_S(a) \tag{10}$$

Expectations for Continuous Random Variables

Expected Value for Continuous Random Variables

- Expected value of a function $g(S)$ of a continuous random variable S

$$\mathrm{E}\{\, g(S) \,\} = \int_{-\infty}^{\infty} g(s)\, f_S(s)\, \mathrm{d}s \tag{11}$$

Joint and Conditional Expectations

- Analog to discrete case (integral over pdf instead of sum over pmf)
- Expected value of function $g(X, Y)$ of two random variable X and Y

$$\mathrm{E}\{\, g(X, Y) \,\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f_{XY}(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{12}$$

- Expected value of $g(X)$ given another random variable Y

$$\mathrm{E}\{\, g(X) \,|\, Y \,\} = \int_{-\infty}^{\infty} g(x)\, f_{X|Y}(x, Y)\, \mathrm{d}x \tag{13}$$

Properties of Expected Values

Same Properties as for Discrete Case

- Linearity

$$\mathrm{E}\{\, aX + bY \,\} = a\,\mathrm{E}\{\, X \,\} + b\,\mathrm{E}\{\, Y \,\} \tag{14}$$

- For independent random variables X and Y

$$\mathrm{E}\{\, XY \,\} = \mathrm{E}\{\, X \,\}\,\mathrm{E}\{\, Y \,\} \tag{15}$$

- Iterative expectation rule

$$\mathrm{E}\{\, \mathrm{E}\{\, g(X) \,|\, Y \,\} \,\} = \mathrm{E}\{\, g(X) \,\} \tag{16}$$

Proofs are the same as for discrete case.

Important Expected Values

- **Mean** μ_S and **variance** σ_S^2 of a random variable S

$$\mu_S = \mathrm{E}\{\, S \,\} \qquad \text{and} \qquad \sigma_S^2 = \mathrm{E}\{\, (S - \mu_S)^2 \,\} \tag{17}$$

- **Covariance** σ_{XY}^2 of two random variables X and Y, and **correlation coefficient** ϕ_{XY} between two random variables X and Y

$$\sigma_{XY}^2 = \mathrm{cov}(X, Y) = \mathrm{E}\{\, (X - \mu_X)\,(Y - \mu_Y) \,\} \tag{18}$$

$$\phi_{XY} = \frac{\sigma_{XY}^2}{\sigma_X\, \sigma_Y} \tag{19}$$

- **Cross-covariance matrix** $\boldsymbol{C}_{\boldsymbol{XY}}$ of two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, and **auto-covariance matrix** $\boldsymbol{C}_{\boldsymbol{XX}}$ of a random vector $\boldsymbol{X}$

$$\boldsymbol{C}_{\boldsymbol{XY}} = \mathrm{cov}(\boldsymbol{X}, \boldsymbol{Y}) = \mathrm{E}\{\, (\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})\,(\boldsymbol{Y} - \boldsymbol{\mu}_{\boldsymbol{Y}})^{\mathrm{T}} \,\} \tag{20}$$

$$\boldsymbol{C}_{\boldsymbol{XX}} = \mathrm{cov}(\boldsymbol{X}, \boldsymbol{X}) = \mathrm{E}\{\, (\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})\,(\boldsymbol{X} - \boldsymbol{\mu}_{\boldsymbol{X}})^{\mathrm{T}} \,\} \tag{21}$$

Continuous Random Processes

Discrete-Time Continuous-Amplitude Random Process

- Series of random experiments at time instants t_n, with $n = 0, 1, 2, \cdots$
- For each experiment: Continuous random variable $S_n = S(t_n)$
- **Continuous random process**: Series of continuous random variables

$$\boldsymbol{S} = \{S_n\}$$

Characterization of Statistical Properties

- Consider N-dimensional random vector $\boldsymbol{S}_k^{(N)} = \{S_k, S_{k+1}, \cdots, S_{k+N-1}\}$
- N-th order joint cdf / N-th order joint pdf

$$F_k^{(N)}(\boldsymbol{s}) \;=\; \mathrm{P}\Big(S_k \leq s_0, S_{k+1} \leq s_1, \cdots, S_{k+N-1} \leq s_{N-1}\Big) \tag{22}$$

$$f_k^{(N)}(\boldsymbol{s}) \;=\; \frac{\partial^N}{\partial s_0 \cdots \partial s_{N-1}} F_k^{(N)}(\boldsymbol{s}) \tag{23}$$

Stationary Process

Stationary Random Processes

- Statistical properties are **invariant to a shift in time**

➡ Joint cdf, pdf, pmf do not depend on start index k

$$\forall k, i: \qquad F_k^{(N)}(\boldsymbol{s}) = F_i^{(N)}(\boldsymbol{s}) \tag{24}$$

$$f_k^{(N)}(\boldsymbol{s}) = f_i^{(N)}(\boldsymbol{s}) \tag{25}$$

- In this course: Only consider stationary processes

➡ Simplify notations

- N-th order cdf: $F_N(\boldsymbol{s})$
- N-th order pdf: $f_N(\boldsymbol{s})$

Characteristics of Stationary Random Processes

Auto-Covariance Matrix

N-th auto-covariance matrix $\boldsymbol{C}_N$

$$\boldsymbol{C}_N = \mathrm{E}\left\{ \left(\boldsymbol{S}^{(N)} - \boldsymbol{\mu}_N\right) \left(\boldsymbol{S}^{(N)} - \boldsymbol{\mu}_N\right)^{\mathrm{T}} \right\} \quad \text{with} \quad \boldsymbol{\mu}_N = (\mu_S, \cdots, \mu_S)^{\mathrm{T}} \tag{26}$$

is a symmetric Toeplitz matrix

$$\boldsymbol{C}_N = \sigma_S^2 \begin{pmatrix} 1 & \phi_1 & \phi_2 & \cdots & \phi_{N-1} \\ \phi_1 & 1 & \phi_1 & \cdots & \phi_{N-2} \\ \phi_2 & \phi_1 & 1 & \cdots & \phi_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{N-1} & \phi_{N-2} & \phi_{N-3} & \cdots & 1 \end{pmatrix} \tag{27}$$

with **correlation coefficients**

$$\phi_{k,\ell} = \frac{\mathrm{cov}(S_k, S_\ell)}{\sigma_S^2} = \frac{1}{\sigma_S^2}\,\mathrm{E}\{\,(S_k - \mu_S)(S_\ell - \mu_S)\,\} = \phi_{|k-\ell|} \tag{28}$$

Memoryless and IID Processes

Memoryless Random Process

- Random process $\boldsymbol{S} = \{S_n\}$ for which all random variables S_n are independent of each other

➡ Diagonal auto-covariance matrix

$$\boldsymbol{C}_N = \sigma_S^2 \cdot \boldsymbol{I} \qquad (\boldsymbol{I}: \text{ identity matrix}) \tag{29}$$

Independent and Identically Distributed (IID) Random Process

- **Stationary** and **memoryless**
- With $F_S(s)$, and $f_S(s)$ being the cdf, and pdf for a single random variable $S = S_n$ (called marginal cdf and pdf), we have

$$F_N(\boldsymbol{s}) = \prod_{k=0}^{N-1} F_S(s_k), \qquad f_N(\boldsymbol{s}) = \prod_{k=0}^{N-1} f_S(s_k), \tag{30}$$

Markov Processes

Random Process with Markov Property

- Markov Property

$$\mathrm{P}(S_n \leq s_n \,|\, S_{n-1} \leq s_{n-1}, \cdots) = \mathrm{P}(S_n \leq s_n \,|\, S_{n-1} \leq s_{n-1}) \tag{31}$$

Stationary Markov Process

- Stationary random process with Markov Property
- Statistical properties completely specified by 1-st order conditional cdf/pdf

$$F(s_n \,|\, s_{n-1}) \;=\; \mathrm{P}(S_n \leq s_n \,|\, S_{n-1} \leq s_{n-1}) \tag{32}$$

$$f(s_n \,|\, s_{n-1}) \;=\; \frac{\partial}{\partial s_n} F(s_n \,|\, s_{n-1}) \tag{33}$$

- N-th order joint cdf and pdf are given by

$$F_N(\boldsymbol{s}) = F_S(s_0) \prod_{k=1}^{N-1} F(s_k|s_{k-1}), \qquad f_N(\boldsymbol{s}) = f_S(s_0) \prod_{k=1}^{N-1} f(s_k|s_{k-1}) \tag{34}$$

Autoregressive (AR) processes

AR(p) model

- Autoregressive model of order p for random variables S_n with mean μ

$$S_n = Z_n + \mu + \sum_{k=1}^{p} \varrho_k \cdot (S_{n-k} - \mu) \tag{35}$$

where $\boldsymbol{Z} = \{Z_n\}$ is a zero-mean iid process (innovation process) and $\varrho_1, \cdots, \varrho_p$ are the model parameters

Special case: AR(1) model

- AR(1) process

$$S_n = Z_n + \mu + \varrho \cdot (S_{n-1} - \mu) \tag{36}$$

➡ Important type of stat. Markov model for continous random processes

➡ Completely specified by

- mean μ, correlation coefficient ϱ and
- pdf $f_Z(z)$ of iid innovation process $\{Z_n\}$

Auto-Covariance for AR(1) Processes

Correlation Coefficients

- **Correlation coefficient** ϱ between successive random variables S_{n-1} and S_n

$$\varrho = \phi_{n,n-1} = \frac{1}{\sigma_S^2} \mathrm{E}\{ (S_n - \mu_S)(S_{n-1} - \mu_S) \} \tag{37}$$

- Correlation coefficient between any two random variables S_k and S_ℓ

$$\phi_{k,\ell} = \phi_{|k-\ell|} = \varrho^{|k-\ell|} \tag{38}$$

N-th order Auto-Covariance Matrix

- Depends only on correlation coefficient ϱ and variance σ_S^2

$$\boldsymbol{C}_N = \sigma_S^2 \begin{pmatrix} 1 & \varrho & \varrho^2 & \cdots & \varrho^{N-1} \\ \varrho & 1 & \varrho & \cdots & \varrho^{N-2} \\ \varrho^2 & \varrho & 1 & \cdots & \varrho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varrho^{N-1} & \varrho^{N-2} & \varrho^{N-3} & \cdots & 1 \end{pmatrix} \tag{39}$$

Gaussian Processes

Continuous Gaussian Random Process

- All finite collections of random variables S_n are Gaussian random vectors
- N-th order pdf is given by auto-covariance matrix $\boldsymbol{C}_N$ and mean μ_S

$$f_{\boldsymbol{S}}(\boldsymbol{s}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{C}_N|}} \, e^{-\frac{1}{2}(\boldsymbol{s}-\boldsymbol{\mu}_S)^{\mathrm{T}} \boldsymbol{C}_N^{-1} (\boldsymbol{s}-\boldsymbol{\mu}_S)} \qquad \text{with} \qquad \boldsymbol{\mu}_S = \begin{pmatrix} \mu_S \\ \vdots \\ \mu_S \end{pmatrix} \tag{40}$$

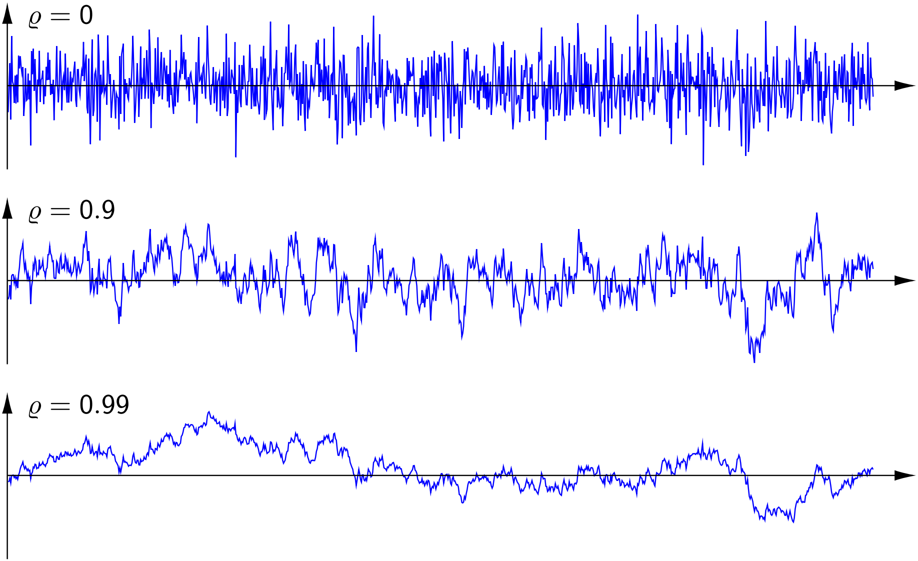
Stationary Gauss-Markov Process

- Stationary Markov process that is also a Gaussian random process
- Can be constructed with Gaussian iid process $\boldsymbol{Z} = \{Z_n\}$ according to

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

- Statistics properties are completely specified by **mean** μ_S, **variance** σ_S^2, and **correlation coefficent** ϱ

Examples of Gauss-Markov Processes (1000 Samples)



Summary

Rate-Distortion Theory

- Theoretical bounds for lossy source coding
- Rate-distortion function $R(D)$

Continuous Random Variables

- Can take an uncountable number of values
- Cumulative distribution function (cdf) is a continuous function
- Probability density function (pdf): Derivative of cdf
- Expected values: Mean, variance, covariance

Discrete-Time Continuous-Amplitude Random Processes

- Stationary, memoryless, iid, Markov processes (as in discrete case)
- Covariance matrix for stationary processes: Toeplitz matrix
- Autoregressive AR(p) processes: Model for dependencies
- AR(1) process: First-order autoregressive process, Markov process
- Gaussian processes
- Stationary Gauss-Markov process: AR(1) with Gaussian pdf