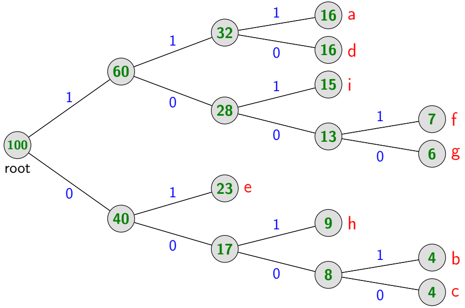
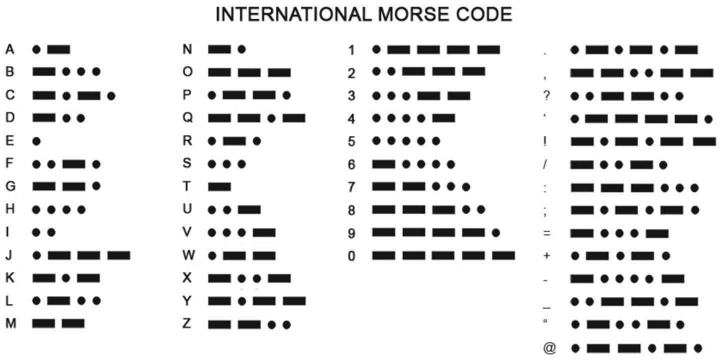
Lossless Coding

a_k	p_k	$\boldsymbol{b}_k$
a	0.16	111
b	0.04	0001
c	0.04	0000
d	0.16	110
e	0.23	01
f	0.07	1001
g	0.06	1000
h	0.09	001
i	0.15	101



Morse Code (first version around 1837)

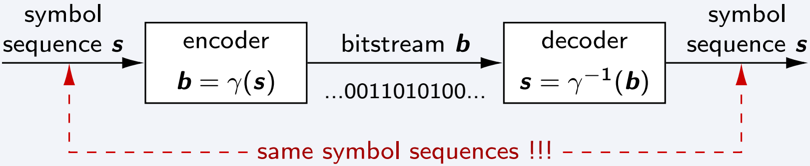


- Length of a dot is one unit
- Length of a dash is three units
- Space between parts of the same letter is one unit
- Space between letters is three units
- Space between words is seven units

Overview

Lossless Coding

- **Reversible mapping** of symbol sequences into bitstreams
- Each symbol must be taken from a countable alphabet (not necessarily finite)
- Original symbol sequence can be exactly recovered from bit sequence (in contrast to lossy coding)
- Also called **noiseless coding** or **entropy coding**



➡ **Bit rate reduction only possible if source data have statistical properties that can be exploited for data compression**

Lossless Coding Framework

Framework

- Discrete random process $\boldsymbol{S} = \{S_n\}$ (discrete-time & discrete-amplitude)
- Consider messages $\boldsymbol{s}^{(L)}$ of L symbols

$$\boldsymbol{s}^{(L)} = \{s_0, s_1, s_2, \cdots, s_{L-1}\}$$

- Messages represent finite-length realizations of random process $\boldsymbol{S}$

Lossless Coding

- Mapping of messages $\boldsymbol{s}^{(L)}$ onto bitstreams $\boldsymbol{b}^{(K)}$ with K bits

$$\boldsymbol{b}^{(K)} = \{b_0, b_1, b_2, \cdots, b_{K-1}\} \qquad \text{with} \qquad \forall k, b_k \in \{0, 1\}$$

- Mapping γ from messages onto bitstreams is an invertible deterministic function

$$\boldsymbol{b}^{(K)} = \gamma\Big(\boldsymbol{s}^{(L)}\Big) \qquad \Longleftrightarrow \qquad \boldsymbol{s}^{(L)} = \gamma^{-1}\Big(\boldsymbol{b}^{(K)}\Big) \tag{1}$$

Realization of Lossless Coding

Practical Realization

- Split messages $\boldsymbol{s}^{(L)}$ into fixed- or variable-length sub-sequences $\left\{\boldsymbol{s}_k^{(N_k)}\right\}$

$$\boldsymbol{s}_k^{(N_k)} = \{s_k, s_{k+1}, s_{k+2}, \cdots, s_{k+N_k-1}\}$$

- Assign codewords $\boldsymbol{b}_k^{\ell_k}$ to sub-sequences $\boldsymbol{s}_k^{(N_k)}$
- Form bitstream $\boldsymbol{b}^{(K)}$ by concatenating the codewords

$$\boldsymbol{b}^{(K)} = \left\{\boldsymbol{b}_1^{(\ell_1)}, \boldsymbol{b}_2^{(\ell_2)}, \boldsymbol{b}_3^{(\ell_3)}, \cdots\right\}$$

Lossless Code

- Encoder mapping:

$$\boldsymbol{b}_k^{(\ell_k)} = \gamma\left(\boldsymbol{s}_k^{(N_k)}\right) \tag{2}$$

- Decoder mapping:

$$\boldsymbol{s}_k^{(N_k)} = \gamma^{-1}\left(\boldsymbol{b}_k^{(\ell_k)}\right) \tag{3}$$

Example: Variable-Length Coding for Scalars

- Symbol alphabet: $\mathcal{A} = \{A, B, M, N\}$
- Lossless code: Codeword table

code A

letter	codeword
A	00
B	01
M	10
N	11

code B

letter	codeword
A	0
B	110
M	111
N	10

code C

letter	codeword
A	0
B	010
M	100
N	10

- Example message: $\boldsymbol{s}$ = "BANANAMAN"

➡ Bitstream (code A): $\boldsymbol{b}$ = "010011001100100011" (18 bits)

➡ Bitstream (code B): $\boldsymbol{b}$ = "1100100100111010" (16 bits)

➡ Bitstream (code C): $\boldsymbol{b}$ = "0100100100100010" (16 bits)

➡ **Goal: Minimize average codeword length**

$$\bar{\ell} = \mathrm{E}\{\, \ell(S) \,\} = \sum_k p_k \cdot \ell_k \tag{4}$$

Example: Variable-Length Coding for Scalars

- Symbol alphabet: $\mathcal{A} = \{\text{A}, \text{B}, \text{M}, \text{N}\}$
- Lossless code: Codeword table

code A

letter	codeword
A	00
B	01
M	10
N	11

code B

letter	codeword
A	0
B	110
M	111
N	10

code C

letter	codeword
A	0
B	010
M	100
N	10

Decoding

- Code A: $\boldsymbol{b}$ = "010011001100100011" ➡ $\boldsymbol{s}$ = "BANANAMAN"
- Code B: $\boldsymbol{b}$ = "110010010011101 0" ➡ $\boldsymbol{s}$ = "BANANAMAN"
- Code C: $\boldsymbol{b}$ = "010010010010001 0" ➡ $\boldsymbol{s}$ = "B or AN ..."

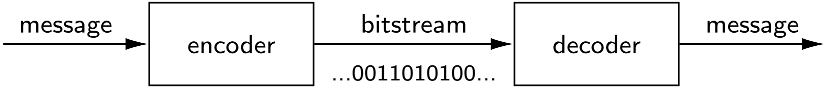
➡ **Necessary condition: Unique decodability:**

Each bitstream uniquely represents a single message!

Outline

- Review: Probability, Random Variables, Random Processes
- Scalar Variable Length Codes
  - Prefix Codes
  - Unqiue Decodability
  - Entropy: Lower Limit for Average Codeword Length
  - Optimal Prefix Codes: Huffman Codes
  - Conditional Prefix Codes
- Variable-Length Codes for Vectors
  - Block Codes
  - V2V Codes
  - Entropy Rate
- Arithmetic Codes
  - Shannon-Fano-Elias Coding
  - Finite Precision: Arithmetic Coding

Mathematical Description of Source Coding



Transmission of new information to receiver

- Message is unknown by receiver

➡ Source can be modeled as a **random process**

Modeling of information sources as random processes

➡ Description using mathematical framework of probability theory

➡ Requires reasonable assumptions with respect to source of information

➡ Characterization of performance by probabilistic averages

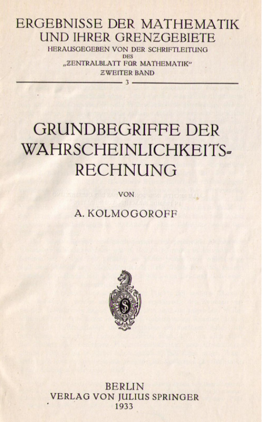
➡ Basis for **mathematical theory of communication**

Probability Theory

- Branch of mathematics for describing and modeling random events
- Modern probability theory (the axiomatic definition of probability) introduced by KOLMOGOROV

Review:

- ➡ Probability
- ➡ Random variables
- ➡ Discrete random processes



ERGEBNISSE DER MATHEMATIK
UND IHRER GRENZGEBIETE
HERAUSGEGEBEN VON DER SCHRIFTLEITUNG
DES
„ZENTRALBLATT FÜR MATHEMATIK"
ZWEITER BAND
3

GRUNDBEGRIFFE DER
WAHRSCHEINLICHKEITS-
RECHNUNG

VON

A. KOLMOGOROFF

BERLIN
VERLAG VON JULIUS SPRINGER
1933

Random Experiment

Definition

- Any experiment with uncertain outcome ζ
- **Sample space** $\mathcal{O}$: Union of all possible outcomes ζ (**Certain event** $\mathcal{O}$)
- **Event** $\mathcal{A}$: Union of zero or more possible outcomes ζ ($\mathcal{A} \subseteq \mathcal{O}$)

Examples

Roulette

- Sample space $\mathcal{O} = \{0, 32, 15, 19, 4, 21, \cdots, 26\}$ (37 possible outcomes)
- Example event $\mathcal{A} = \{32\}$ (red 32)
- Example event $\mathcal{B} = \{15, 4 \cdots, 26\}$ (any black number)

Voltage Measurement

- Sample space $\mathcal{O} = [0\,\mathrm{V};\ \infty\,\mathrm{V})$ (infinitely many possible outcomes)
- Example event $\mathcal{A} = [0\,\mathrm{V};\ 10\,\mathrm{V}]$ (not greater than 10 V)
- Example event $\mathcal{B} = (1\,\mathrm{V};\ 3\,\mathrm{V})$ (between 1 V and 3 V)

Probability

Definition

- Measure $\mathrm{P}(\mathcal{A})$ assigned to events $\mathcal{A}$ of a random experiment
- Satisfies the following three axioms

Axioms

1. Probabilities are non-negative real numbers

$$\mathrm{P}(\mathcal{A}) \geq 0, \quad \forall \mathcal{A} \subseteq \mathcal{O} \tag{5}$$

2. Certain event $\mathcal{O}$ has a probability equal to 1

$$\mathrm{P}(\mathcal{O}) = 1 \tag{6}$$

3. Probability of two disjoint events $\mathcal{A}$ and $\mathcal{B}$

$$\mathcal{A} \cap \mathcal{B} = \emptyset \quad \Longrightarrow \quad \mathrm{P}(\mathcal{A} \cup \mathcal{B}) = \mathrm{P}(\mathcal{A}) + \mathrm{P}(\mathcal{B}) \tag{7}$$

Probability Estimation

Empirical Probability

- Repeatable random experiment
- Relative frequency of an event $\mathcal{A}$ in N trials

$$\frac{N(\mathcal{A})}{N} = \frac{\text{number of trials in which } \mathcal{A} \text{ was observed}}{\text{number of total trials}} \tag{8}$$

- Empirical probability

$$\mathrm{P}(\mathcal{A}) = \lim_{N\to\infty} \frac{N(\mathcal{A})}{N} \tag{9}$$

Practical Probability Estimation

- Repeat a random experiment N times
- Use the approximation

$$\mathrm{P}(\mathcal{A}) = \frac{N(\mathcal{A})}{N}$$

→ Estimation quality depends on the size of N

Conditional Probability

Definition

- Probability of an event $\mathcal{A}$ given that another event $\mathcal{B}$ has occured
- Referred to as **conditional probability** $\mathrm{P}(\mathcal{A}\,|\,\mathcal{B})$
- Kolmogorov's definition

$$\mathrm{P}(\mathcal{A}\,|\,\mathcal{B}) = \frac{\mathrm{P}(\mathcal{A}\cap\mathcal{B})}{\mathrm{P}(\mathcal{B})}, \qquad \text{for} \quad \mathrm{P}(\mathcal{B}) > 0 \tag{10}$$

Bayes' Theorem

- For $\mathrm{P}(\mathcal{A}) > 0$ and $\mathrm{P}(\mathcal{B}) > 0$,

$$\mathrm{P}(\mathcal{A}\,|\,\mathcal{B}) = \mathrm{P}(\mathcal{B}\,|\,\mathcal{A}) \cdot \frac{\mathrm{P}(\mathcal{A})}{\mathrm{P}(\mathcal{B})} \tag{11}$$

- Can be directly derived from definition of conditional probability

$$\mathrm{P}(\mathcal{A}\cap\mathcal{B}) = \mathrm{P}(\mathcal{B}\cap\mathcal{A}) \quad\Longrightarrow\quad \mathrm{P}(\mathcal{A}\,|\,\mathcal{B})\cdot\mathrm{P}(\mathcal{B}) = \mathrm{P}(\mathcal{B}\,|\,\mathcal{A})\cdot\mathrm{P}(\mathcal{A})$$

Independence of Events

Definition

- Two events $\mathcal{A}$ and $\mathcal{B}$ are said to be **independent** if and only if

$$\mathrm{P}(\mathcal{A} \cap \mathcal{B}) = \mathrm{P}(\mathcal{A}) \cdot \mathrm{P}(\mathcal{B}) \tag{12}$$

Consequence

- Using reformulated definition of conditional probability

$$\mathrm{P}(\mathcal{A} \,|\, \mathcal{B}) = \frac{\mathrm{P}(\mathcal{A} \cap \mathcal{B})}{\mathrm{P}(\mathcal{B})}, \qquad \text{for} \;\; \mathrm{P}(\mathcal{B}) > 0$$

- Two events $\mathcal{A}$ and $\mathcal{B}$, with $\mathrm{P}(\mathcal{B}) > 0$, are independent if and only if

$$\mathrm{P}(\mathcal{A} \,|\, \mathcal{B}) = \mathrm{P}(\mathcal{A}) \tag{13}$$

Random Variable

Definition

- 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

Examples for Random Variables

Roulette

- Random variable S: $S(\zeta) = \{$"number of pocket the ball lands"$\}$
- Random variable C: $C(\zeta) = \{0 : \text{"0"},\ 1 : \text{"red"},\ 2 : \text{"black"}\}$

Digital Camera

- Random variable V: $V(\zeta) = \{$"voltage (in V) of a photocell"$\}$
- Random variable Y: $Y(\zeta) = \{$"discrete sample value at output"$\}$

Note: A random variable may take ...

- a finite, a countably infinite, or an uncountable number of values

Cumulative Distribution Function

Definition

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

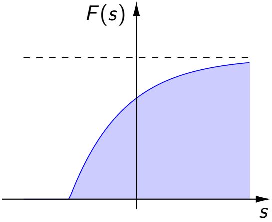
$$F_S(s) = \mathrm{P}(S \leq s) = \mathrm{P}(\{\zeta : S(\zeta) \leq s\}) \tag{14}$$

- $F_S(s)$ is also referred to as **distribution** of a random variable S

Properties

- $F_S(s)$ is non-decreasing
- $F_S(-\infty) = 0$
- $F_S(\infty) = 1$
- $\mathrm{P}(a < S \leq b) = F_S(b) - F_S(a)$

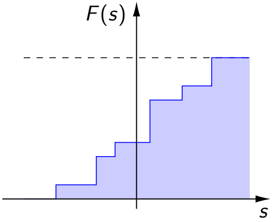
Examples: Cumulative Distribution Functions



Continuous function

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

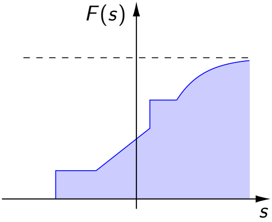
➡ **Continuous random variable**



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

Joint Cumulative Distribution Function

Multiple Random Variables

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

$$F_{XY}(x, y) = \mathrm{P}(X \leq x,\ Y \leq y) \tag{15}$$

- Straightforward extension to $N > 2$ random variables

Random Vectors

- Random vector $\boldsymbol{S}$: Vector of multiple random variables
- N-dimensional random vector: $\boldsymbol{S} = \boldsymbol{S}^{(N)} = \{S_0, S_1, \cdots, S_{N-1}\}$
- N-dimensional joint cdf / joint distribution

$$F_{\boldsymbol{S}}(\boldsymbol{s}) = \mathrm{P}(\boldsymbol{S} \leq \boldsymbol{s}) = \mathrm{P}(S_0 \leq s_0,\ S_1 \leq s_1,\ \cdots, S_{N-1} \leq s_{N-1}) \tag{16}$$

- Joint cdf / joint distribution of two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$F_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{P}(\boldsymbol{X} \leq \boldsymbol{x},\ \boldsymbol{Y} \leq \boldsymbol{x}) \tag{17}$$

Conditional Cumulative Distribution Function

Conditional Cdfs / Conditional Distribution

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

$$F_{S|\mathcal{B}}(s \,|\, \mathcal{B}) = \mathrm{P}(S \leq s \,|\, \mathcal{B}) = \frac{\mathrm{P}(\{S \leq s\} \cap \mathcal{B})}{\mathrm{P}(\mathcal{B})} \tag{18}$$

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

$$F_{X|Y}(x \,|\, y) = \mathrm{P}(X \leq x \,|\, Y \leq y) = \frac{\mathrm{P}(X \leq x,\ Y \leq y)}{\mathrm{P}(Y \leq y)} = \frac{F_{XY}(x, y)}{F_Y(y)} \tag{19}$$

Random Vectors

- Conditional cdf of a random vector $\boldsymbol{X}$ given another random vector $\boldsymbol{Y}$

$$F_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \mathrm{P}(\boldsymbol{X} \leq \boldsymbol{x} \,|\, \boldsymbol{Y} \leq \boldsymbol{y}) = \frac{\mathrm{P}(\boldsymbol{X} \leq \boldsymbol{x},\ \boldsymbol{Y} \leq \boldsymbol{y})}{\mathrm{P}(\boldsymbol{Y} \leq \boldsymbol{y})} = \frac{F_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y})}{F_{\boldsymbol{Y}}(\boldsymbol{y})} \tag{20}$$

Discrete Random Variables: Probability Mass Function

Discrete Random Variables

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

➜ S can only take values of a countable alphabet

$$\mathcal{A}_S = \{a_0, a_1, \cdots\} \tag{21}$$

Definition: Probability Mass Function of a Discrete Random Variable S

- Probability mass function (pmf) is defined on alphabet $\mathcal{A}_S$

$$p_S(a) = \mathrm{P}(\{\zeta \in \mathcal{O} : S(\zeta) = a\}) \tag{22}$$

- Property

$$\sum_{\forall a \in \mathcal{A}_S} p_S(a) = \mathrm{P}(\mathcal{O}) = 1 \tag{23}$$

- Simplified notation

$$p_k = p_S(a_k) \tag{24}$$

Cdf for Discrete Random Variables

Cumulative Distribution Function

- Cdf for discrete random variables S

$$F_S(s) = \sum_{\forall a \in \mathcal{A}_S \, : \, a \leq s} p_S(a) \tag{25}$$

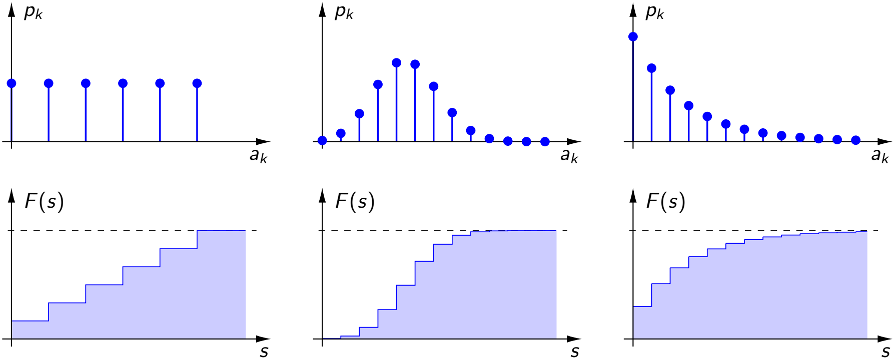
Cumulative Mass Function

- Discrete version of cdf
- Cumulative mass function (cmf) defined on alphabet $\mathcal{A}_S = \{a_0, a_1, \cdots\}$

$$c_k = c(a_k) = \sum_{\forall a \in \mathcal{A}_S \, : \, a \leq a_k} p_S(a) \tag{26}$$

Examples for Discrete Distributions

Uniform	**Binomial**	**Geometric**
$p_k = \frac{1}{M}$	$p_k = \binom{n}{k}\, p^k (1-p)^{n-k}$	$p_k = (1-p)^k\, p$
$(0 \le k < M)$	$(0 \le k \le n)$	$(k \ge 0)$



Joint Probability Mass Function

Multiple Random Variables

- Joint pmf of two discrete random variables X and Y

$$p_{XY}(x, y) = \mathrm{P}(X = x, Y = y) \tag{27}$$

- Straightforward extension to $N > 2$ random variables

Random Vectors

- Consider random vector $\boldsymbol{S} = \boldsymbol{S}^{(N)} = \{S_0, S_1, \cdots, S_{N-1}\}$
- N-dimensional joint pmf

$$p_{\boldsymbol{S}}(\boldsymbol{s}) = \mathrm{P}(\boldsymbol{S} = \boldsymbol{s}) \tag{28}$$

- Joint pmf of two random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$

$$p_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y}) = \mathrm{P}(\boldsymbol{X} = \boldsymbol{x},\ \boldsymbol{Y} = \boldsymbol{y}) \tag{29}$$

Conditional Probability Mass Function

Conditional Pmf

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

$$p_{S|\mathcal{B}}(a \,|\, \mathcal{B}) = \mathrm{P}(S = a \,|\, \mathcal{B}) = \frac{\mathrm{P}(\{S = a\} \cap \mathcal{B})}{\mathrm{P}(\mathcal{B})} \tag{30}$$

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

$$p_{X|Y}(x \,|\, y) = \mathrm{P}(X = x \,|\, Y = y) = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(Y = y)} = \frac{p_{XY}(x, y)}{p_Y(y)} \tag{31}$$

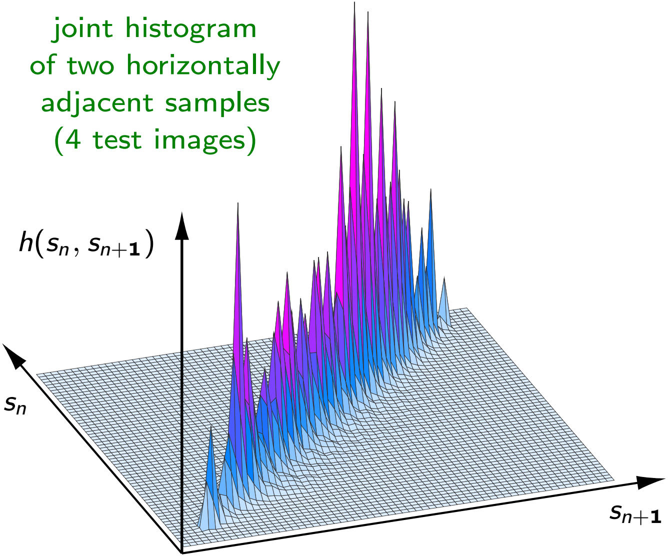
Random Vectors

- Conditional pmf of a random vector $\boldsymbol{X}$ given another random vector $\boldsymbol{Y}$

$$p_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x} \,|\, \boldsymbol{y}) = \frac{p_{\boldsymbol{XY}}(\boldsymbol{x}, \boldsymbol{y})}{p_{\boldsymbol{Y}}(\boldsymbol{y})} \tag{32}$$

Example for a Joint Pmf (Empirical Estimation)

joint histogram
of two horizontally
adjacent samples
(4 test images)



Expected Value / Expectation

Definition

- Expected value of a function $g(S)$ of a random variable S
- **Discrete random variable S** with alphabet $\mathcal{A}$

$$\mathrm{E}\{\, g(S) \,\} = \sum_{\forall a \in \mathcal{A}} g(a)\, p_S(a) \tag{33}$$

Multiple Random Variables

- Expected value of function $g(X, Y)$ of two random variable X and Y

$$\mathrm{E}\{\, g(X, Y) \,\} = \sum_{x,y} g(x, y)\, p_{XY}(x, y) \tag{34}$$

- Straightforward extension to more than two random variables

Conditional Expectations

Given an Event

- Expected value of function $g(S)$ given event $\mathcal{B}$, with $\mathrm{P}(\mathcal{B}) > 0$

$$\mathrm{E}\{\, g(S) \,|\, \mathcal{B} \,\} = \sum_{a} g(a)\, p_{S|\mathcal{B}}(a \,|\, \mathcal{B}) \tag{35}$$

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

$$\mathrm{E}\{\, g(X) \,|\, y \,\} = \mathrm{E}\{\, g(X) \,|\, Y = y \,\} = \sum_{x} g(x)\, p_{X|Y}(x, y) \tag{36}$$

➡ $\mathrm{E}\{\, g(X) \,|\, y \,\}$ is a deterministic function of y

Conditional Expectations

Given another Random Variable

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

$$\mathrm{E}\{\, g(X) \,|\, Y \,\} = \sum_x g(x)\, p_{X|Y}(x, Y) \tag{37}$$

➡ $\mathrm{E}\{\, g(X) \,|\, Y \,\}$ is another random variable

Properties of Expected Values

Important Properties

- Linearity

$$\mathrm{E}\{\, a X + b Y \,\} = a\,\mathrm{E}\{\, X \,\} + b\,\mathrm{E}\{\, Y \,\} \tag{38}$$

- For independent random variables X and Y

$$\mathrm{E}\{\, XY \,\} = \mathrm{E}\{\, X \,\}\,\mathrm{E}\{\, Y \,\} \tag{39}$$

- Iterative expectation rule

$$\mathrm{E}\{\, \mathrm{E}\{\, g(X) \,|\, Y \,\} \,\} = \mathrm{E}\{\, g(X) \,\} \tag{40}$$

Proofs: *Homework*.

Important Expected Values

Mean, Variance, Covariance

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

$$\mu_S = \mathrm{E}\{\, S \,\} \qquad \text{and} \qquad \sigma_S^2 = \mathrm{E}\{\, (S - \mu_S)^2 \,\} \tag{41}$$

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

$$\sigma_{XY}^2 = \mathrm{cov}(X, Y) = \mathrm{E}\{\, (X - \mu_X)\,(Y - \mu_Y) \,\} \tag{42}$$

$$\phi_{XY} = \frac{\sigma_{XY}^2}{\sigma_X\, \sigma_Y} \tag{43}$$

Average Codeword Length

- Average codeword length for given lossless code

$$\bar{\ell} = \mathrm{E}\{\, \ell(S) \,\} = \sum_{a \in \mathcal{A}} p(a) \cdot \ell(a) \tag{44}$$

Random Processes

Discrete-Time Random Process

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

$$F_k^{(N)}(\boldsymbol{s}) = \mathrm{P}\left(\boldsymbol{S}_k^{(N)} \leq \boldsymbol{s}\right) = \mathrm{P}(S_k \leq s_0, S_{k+1} \leq s_1, \cdots, S_{k+N-1} \leq s_{N-1}) \tag{45}$$

Discrete Random Processes

Discrete Random Processes

- Random variables $\{S_n\}$ are of discrete type
- Each random variable S_n has an alphabet $\mathcal{A}_n$
- N-th order joint pmf

$$p_k^{(N)}(\boldsymbol{a}) = \mathrm{P}(S_k = a_0, S_{k+1} = a_1, \cdots, S_{k+N-1} = a_{N-1}) \tag{46}$$

with

$$F_k^{(N)}(\boldsymbol{s}) = \sum_{\forall a_0 \in \mathcal{A}_0 : a_0 \leq s_0, \cdots} p_k^{(N)}(a_0, a_1, \cdots, a_{N-1}) \tag{47}$$

Stationary Process

Stationary Random Processes

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

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

$$\forall k, i: \qquad F_k^{(N)}(\boldsymbol{s}) = F_i^{(N)}(\boldsymbol{s}) \tag{48}$$

$$p_k^{(N)}(\boldsymbol{a}) = p_i^{(N)}(\boldsymbol{a}) \tag{49}$$

- In this course: Only consider stationary processes

➡ Simplify notations

  - N-th order cdf: $F_N(\boldsymbol{s})$
  - N-th order pmf: $p_N(\boldsymbol{a})$

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

Independent and Identically Distributed (IID) Random Process

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

$$F_N(\boldsymbol{s}) = \prod_{k=0}^{N-1} F_S(s_k), \qquad p_N(\boldsymbol{a}) = \prod_{k=0}^{N-1} p_S(a_k) \tag{50}$$

Markov Processes

Random Process with Markov Property

- Markov Property: Future outcomes do not depend on past outcomes, but only on the present outcome

$$\mathrm{P}(S_n \leq s_n \,|\, S_{n-1} \leq s_{n-1}, \cdots) = \mathrm{P}(S_n \leq s_n \,|\, S_{n-1} \leq s_{n-1}) \quad (51)$$

Stationary Markov Process

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

$$F(s_n \,|\, s_{n-1}) = \mathrm{P}(S_n \leq s_n \,|\, S_{n-1} \leq s_{n-1}) \quad (52)$$

- With marginal cdf $F_S(s)$, the N-th order joint cdf is given by

$$F_N(\boldsymbol{s}) = F_S(s_0) \prod_{k=1}^{N-1} F(s_k | s_{k-1}) \quad (53)$$

Stationary Discrete Markov Processes

Markov Property for Discrete Random Processes

- Statistical properties completely specified by 1-st order conditional pmf

$$p(a_n \mid a_{n-1}) = \mathrm{P}(S_n = a_n \mid S_{n-1} = a_{n-1}) \tag{54}$$

- With marginal pmf $p_S(a)$, the N-th order joint pmf is given by

$$p_N(\boldsymbol{a}) = p_S(a_0) \prod_{k=1}^{N-1} p(a_k|a_{k-1}) \tag{55}$$

Example: Stationary Discrete Markov Process

Specified by conditional pmf $p(a_n|a_{n-1})$

a	a_0	a_1	a_2
$p(a\|a_0)$	0.90	0.05	0.05
$p(a\|a_1)$	0.15	0.80	0.05
$p(a\|a_2)$	0.25	0.15	0.60

Homework:

Determine marginal pmf
$p_S(a) = \mathrm{P}(S = a)$

Summary

Lossless Coding

- Reversible mapping: Symbol sequence $\Longleftrightarrow$ bitstream
- Necessary condition: Unique decodability
- Optimization criterion: Minimize average codeword length

Probability

- Axiomatic definition, empirical probability
- Conditional probability and independence

Discrete Random Variables

- Cumulative distribution function (cdf): Staircase function
- Probability mass function (pmf)
- Expected values: Mean, variance, covariance, average codeword length

Discrete Random Processes

- Stationary, memoryless, iid, Markov processes