ulm university universität uulm

Faculty of Engineering, Computer Science and Psychology
Institute of Communications Engineering
Prof. Dr.-Ing. Martin Bossert/ M.Sc. Cornelia Ott

Summer Term 2018

Solution: Test Examination

Applied Information Theory

July 30, 2018

- The exam duration is 90 minutes.
- No aids are permitted.
- All four problems will be evaluated.
- In total, 90 points can be reached.
- The solutions of different problems must be written on separate sheets.
- If not stated otherwise, all solutions must be justified.

Problem 1: (Information Theory Basics)

a) Let X, Y, X_1, X_2 be discrete random variables, such that X_1, X_2 are statistically independent. Decide whether the following statements are true or false. Justify for your answer.

1. $H(XY) = H(X) + H(Y)$
2. $I(X_1X_2; Y) = I(X_1; Y) + I(X_2; Y) - I(X_1; X_2)$
3. $I(X, Y) \geq \max\{H(X|Y), H(Y|X)\}$

Solution:

1. False, for example: choose $X = Y$ with $H(X) \neq 0$
   $\Rightarrow H(XX) = H(X) \neq 2 \cdot H(X)$
2. True:

$$\begin{aligned}
&I(X_1; Y) + I(X_2; Y) - I(X_1; X_2) \\
&= \Big(H(X_1) - H(X_1|Y)\Big) + \Big(H(X_2) - H(X_2|Y)\Big) - \Big(H(X_1) + H(X_2) - H(X_1X_2)\Big) \\
&= H(X_1X_2) - \underbrace{\Big(H(X_1|Y) + H(X_2|Y)\Big)}_{=H(X_1X_2|Y)\ (X_1,\ X_2 \text{ ind.})} \\
&= I(X_1X_2; Y)
\end{aligned}$$

3. False, for example: choose X, Y statistically independent, $H(X) \neq 0$
   $\Rightarrow I(X, Y) = 0 < H(X) = H(X|Y) \leq \max\{H(X|Y), H(Y|X)\}$

b) Let X, Y be discrete random variables. Complete the following proof. Make clear which formulas and properties you are using.

$$\begin{aligned}
H(X|Y) - H(X) &= \sum_{i=1}^{k} \sum_{j=1}^{l} f_{XY}(x_i, y_j) \log_2 \left(\frac{f_X(x_i) f_Y(y_j)}{f_{XY}(x_i, y_j)} \right) \\
&\vdots \\
&\leq 0
\end{aligned}$$

Solution:

$$
\begin{aligned}
H(X|Y) - H(X) &= -\sum_{i=1}^{K}\sum_{j=1}^{L} f_{XY}(x_i, y_j)\log_2 f_{X|Y}(x_i \mid y_j) + \sum_{i=1}^{K} f_X(x_i)\log_2 f_X(x_i) \\
&= -\sum_{i=1}^{K}\sum_{j=1}^{L} f_{XY}(x_i, y_j)\log_2 \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)} + \sum_{i=1}^{K}\log_2 f_X(x_i)\sum_{j=1}^{L} f_{XY}(x_i, y_j) \\
&= \sum_{i=1}^{K}\sum_{j=1}^{L} f_{XY}(x_i, y_j)\left(-\log_2 \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)} + \log_2 f_X(x_i)\right) \\
&= \sum_{i=1}^{K}\sum_{j=1}^{L} f_{XY}(x_i, y_j)\log_2 \frac{f_X(x_i) f_Y(y_j)}{f_{XY}(x_i, y_j)} \\
&\leq \sum_{i=1}^{K}\sum_{j=1}^{L} f_{XY}(x_i, y_j)\left(\frac{f_X(x_i) f_Y(y_j)}{f_{XY}(x_i, y_j)} - 1\right)\log_2 e \\
&= \left(\sum_{i=1}^{K}\sum_{j=1}^{L} f_X(x_i) f_Y(y_j) - \sum_{i=1}^{K}\sum_{j=1}^{L} f_{XY}(x_i, y_j)\right)\log_2 e \\
&= (1-1)\log_2 e = 0.
\end{aligned}
$$

□

c) Let $X, Y : \Omega \to \{1, \dots, 6\}$ be random variables describing two independent die rolls. Sort the following random variables according to their uncertainty.

$$(X, Y); X + Y; X + 10 \cdot Y; (X + Y)\text{mod } 2$$

Solution:

- (X, Y) contains the full information about X and Y.
- $X + 10 \cdot Y$ contains the information about Y in the first digit and the information about X in the second digit.
- $X + Y$ contains only partial information about X and Y, e.g. $X + Y = 3$ could be $(1, 2)$ or $(2, 1)$.
- $(X + Y)\text{mod } 2$ obviously contains the least information about X and Y.

Thus we get

$$H(X, Y) = H(X + 10 \cdot Y) > H(X + Y) > H\big((X + Y)\text{mod } 2\big).$$

d) Consider an urn filled with four balls with label 0, two balls with label 1 and two balls with label 2. Calculate $I(X; Y)$ for the following random variables.

$$
\begin{aligned}
X &: \Omega \to \{0, 1, 2\} \\
Y &: \Omega \to \{\neq 0, 0\}
\end{aligned}
$$

Solution:

$$I(X; Y) = H(X) + H(Y) - \underbrace{H(X, Y)}_{=H(X)} = H(Y) = h(0.5) = 1$$

e) Let $\mathcal{M}$ be the set of all possible messages and $\mathcal{C}$ the set of all ciphers.
Consider a symmetric cryptosystem consisting of an encryption E and a decryption D using the same key $k \in \mathcal{K}$, satisfying

$$D(E(m,k),k) = m \qquad \forall m \in \mathcal{M}, k \in \mathcal{K}.$$

Which of the following statements is only true for this system if it is perfectly secure?

1. $H(\mathcal{M},\mathcal{K}) = H(\mathcal{K},\mathcal{C})$.
2. The encryption $E : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ is injective in its first argument.
3. $p(m) = p(m|c) \qquad \forall m \in \mathcal{M}, c \in \mathcal{C}$.

Solution:

1. $H(\mathcal{M},\mathcal{K}) = H(\mathcal{M},\mathcal{K},\mathcal{C}) = H(\mathcal{K},\mathcal{C})$ is always true, because the encryption E calculates $c \in \mathcal{C}$ from $m \in \mathcal{M}$ and $k \in \mathcal{K}$ and the decryption D calculates $m \in \mathcal{M}$ from $c \in \mathcal{C}$ and $k \in \mathcal{K}$.
2. The encryption E is always injective in its first argument, otherwise no decryption would be possible as E would not be invertible for a fixed k.
3. This equals $H(\mathcal{M}) = H(\mathcal{M}|\mathcal{C})$ and implies $I(\mathcal{M};\mathcal{C}) = H(\mathcal{M}) - H(\mathcal{M}|\mathcal{C}) = 0$ and is thus only true if the system is perfectly secure.

Problem 2: (Source Coding)

a) Given is the code $\{0, 010, 0101\}$. Can this code be decoded uniquely? If yes, give a decoding algorithm. If no, give a sequence which cannot be decoded uniquely.

Solution: Yes, it can be decoded uniquely. Decoding could work as follows. Read from left ro right. If the current and the next symbol are both 0 (**case (i)**), decode the current symbol to 0. Otherwise, look if the third-next symbol is a 0 (**case (ii)**) or 1 (**case (iii)**) and decode the current and its subsequent 2 or 3 symbols to 010 or 0101 respectively. The latter decision works since the codeword 010 must be followed by a 0 (no sequence starts with a 1). E.g.,

	00100010100101
case (i):	0|0100010100101
case (ii):	0|010|0010100101
case (i):	0|010|0|010100101
case (iii):	0|010|0|0101|00101
case (i):	0|010|0|0101|0|0101
case (iii):	0|010|0|0101|0|0101|

b) The code {10,11,00,101,1010,1011} cannot be decoded uniquely. Does a code with codewords of exactly these lengths exist, which can be uniquely decoded?

Solution: We look at the Kraft inequality

$$\sum_i 2^{-w_i} = 3 \cdot 2^{-2} + 1 \cdot 2^{-3} + 2 \cdot 2^{-4} = 1 \leq 1.$$

Thus, a prefix-free—and therefore uniquely decodable—code with these lengths exists.

c) A suffix-free code is a code in which no codeword is suffix of any other codeword.

1. Explain why suffix-free codes are always uniquely decodable.
2. Explain a disadvantage of suffix-free codes in comparison to prefix-free codes.
3. Is every uniquely decodable code either prefix- or suffix-free? If yes, give a reason for this statement. If no, give a counterexample.

Solution:

1. If we read a sequence and the codewords from right to left, the code is prefix-free and we can therefore decode uniquely.
2. We have to receive the entire sequence first before we can decode (not instantaneously decodable).
3. No, see the example in Exercise **a)**. 0 is both a prefix and suffix of 010, but the code is uniquely decodable.

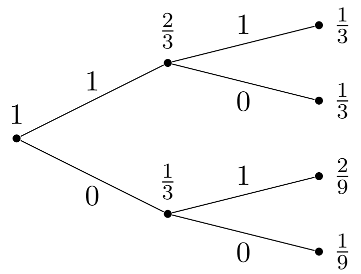
d) Given is the following source with alphabet $\{a, b, c, d\}$ and probabilities:

a	b	c	d
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{9}$	$\frac{1}{9}$

1. Calculate the entropy of the source and explain the meaning of your result.
2. Construct a prefix-free code for the given source, using the Huffman algorithm.
3. What is the expected codeword length of this code?
4. What is the main advantage of Huffman in comparison to the Shannon-Fano algorithm?

Solution:

1. The entropy can be calculated as follows.
$$H(X) = -\sum_i p_i \log_2(p_i) = 2 \cdot \tfrac{1}{3}\log_2(3) + \tfrac{2}{9}\log_2(\tfrac{9}{2}) + \tfrac{1}{9}\log_2(9) \approx 1.8911.$$
This means that we need at least 1.8911 bits on average to encode a source symbol.
2. One possibility is ($\exists$ more):



3. From to the pathlength lemma, we get:
$$\sum_i P_i = 1 + \tfrac{1}{3} + \tfrac{2}{3} = 2.$$
4. The Huffman algorithm always returns an optimal code tree.

e) Prove that the q-ary Shannon–Fano algorithm fulfills
$$E[W] < \frac{H(X)}{\log_2(q)} + 1,$$
where $E[W]$ is the expected codeword length and $H(X)$ is the uncertainty of the source.

Solution:

Proof: In the q-ary Shannon–Fano algorithm, we choose the codeword lengths as
$$w_i = \lceil -\log_q(p_i) \rceil,$$
where p_i is the probability of the ith source symbol. Using $\lceil x \rceil < x + 1 \ \forall x \in \mathbb{R}$, we obtain
$$\begin{aligned} E[W] &= \sum_i p_i w_i = \sum_i p_i \lceil -\log_q(p_i) \rceil < \sum_i p_i(-\log_q(p_i) + 1) \\ &= -\sum_i p_i \log_q(p_i) + \sum_i p_i \\ &= -\sum_i p_i \frac{\log_2(p_i)}{\log_2(q)} + 1 = \frac{H(X)}{\log_2(q)} + 1. \end{aligned}$$

□

Problem 3: (Channel Coding)

a) Consider 4 independent, parallel, time-discrete and additive Gaussian channels. The noise variances are given by $N_1 = 1$, $N_2 = 2$, $N_3 = 4$, $N_4 = 7$.

1. Find the values of the signal powers $S_1, \dots, S_4$ which maximize the sum capacity of the channel
$$C = \sum_{i=1}^{4} \frac{1}{2} \log_2 \left(1 + \frac{S_i}{N_i}\right)$$
under the sum power constraint
$$\sum_{i=1}^{4} S_i \leq 8.$$
2. For which distribution of the input signals is this capacity achieved? Briefly justify whether or not this is a distribution which can be used in a realistic transmission system.

Solution:

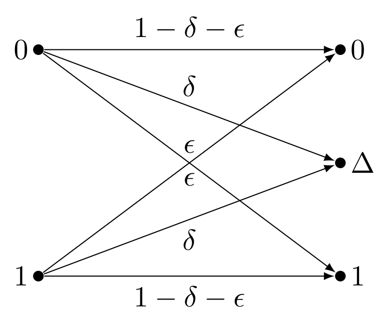
1. We use waterfilling to find the solution. Using the algorithm and notation of Exercise 7.1, we obtain

i	$E_{\mathsf{free}}(i)$
0	8
1	$8 - 1 \cdot (2-1) = 7$
2	$7 - 2 \cdot (4-2) = 3$
3	$3 - 3 \cdot (7-4) = -6 < 0$ (break!)

Thus, $B = N_3 + \frac{E_{\mathsf{free}}(2)}{3} = 4 + 1 = 5$, and using $S_i = \max\{0, B - N_i\}$, we obtain
$$S_1 = 4,\ S_2 = 3,\ S_3 = 1,\ S_4 = 0.$$
Hint: Illustrate the result as in the exercises in order to validate your solution.

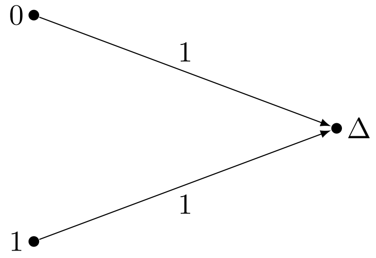
2. This capacity is achieved if the input signal i is Gaussian distributed with zero mean and variance S_i (i.e., $\sim \mathcal{N}(0, S_i)$). In realistic transmission systems we often have a discrete rather than an absolutely continuous input distribution, so the capacity would not be achieved exactly.

b) Derive the capacity of the following channel. Justify each step.



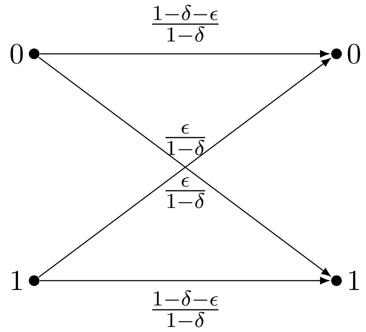
Solution: The channel can be decomposed into two strongly symmetric channels:

- With probability $q_1 := \delta$, the channel outputs an erasure, which can be modelled as the channel



  with capacity $C_1 = 0$.

- With probability $q_2 = 1 - \delta$, the channel is a binary symmetric channel of the form



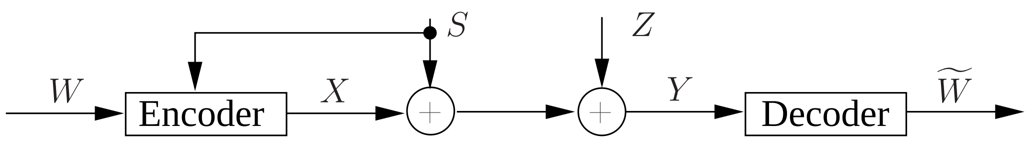
  with capacity $C_2 = 1 - h(\frac{\epsilon}{1-\delta})$

Hence, the overall capacity becomes

$$C = q_1 C_1 + q_2 C_2 = (1 - \delta) \cdot (1 - h(\tfrac{\epsilon}{1-\delta})).$$

c) Briefly explain the Tomlinson–Harashima precoding scheme with the help of a sketch. Specify which signals are known at the transmitter/receiver.

Solution: The block diagram of the Tomlinson–Harashima precoding (THP) scheme looks as follows:



Here,

- W is the source signal (known at the transmitter, unknown at the receiver).
- S is an interference signal (known at the transmitter, unknown at the receiver).
- X is the transmit signal (known at the transmitter, unknown at the receiver).
- Z is the noise (unknown at the transmitter, unknown at the receiver).
- Y is the received signal (unknown at the transmitter, known at the receiver).
- $\widetilde{W}$ is the decoded signal (unknown at the transmitter, known at the receiver).

In THP, the encoder simply subtracts S from W and applies a modulo operation to the result (i.e., $X = \mathrm{mod}(W - S)$) in order to match the power constraint. The decoder applies a modulo operation to the received signal Y.

Thus, the interference S is cancelled and the decoded signal $\widetilde{W}$ does not depend on it. The disadvantage is that due to the modulo operations, the noise is transformed by a non-linear function, which results in a capacity loss.

d) Prove the Shannon limit for the AWGN channel, i.e. show that error-free transmission is possible if and only if

$$\frac{E_b}{N_0} := \frac{S}{N_0 R} > -1.6\text{dB},$$

where R is the transmission rate, S is the signal power and N_0 is the noise power spectral density.

Solution: The capacity of a bandlimited AWGN channel with bandwidth W is given by

$$C_{\mathsf{AWGN}}(W) = W \log_2 \left(1 + \frac{S}{N_0 W}\right).$$

This is a monotonically increasing function in W, so we can upper bound it by

$$C_{\mathsf{AWGN},W} \leq C_\infty := \lim_{W \to \infty} W \log_2 \left(1 + \frac{S}{N_0 W}\right) \overset{\text{l'Hôpital}}{=} \frac{S}{N_0 \ln 2}.$$

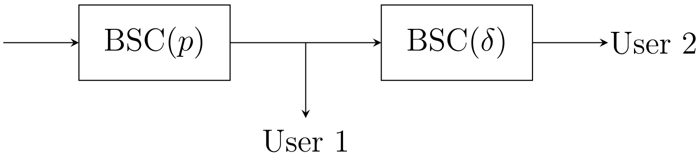
Due to the channel coding theorem, transmission is possible if and only if $R < C_\infty$ (i.e. if $R < C_{\mathsf{AWGN},W}$ for some W). We know that

$$R < C_\infty = \frac{S}{N_0 \ln 2} \quad \Leftrightarrow \quad \frac{E_b}{N_0} = \frac{S}{N_0 R} > \ln 2 \approx -1.6\text{dB}.$$

□

Problem 4: (Multi-User Information Theory)

a) Given the following channel model:

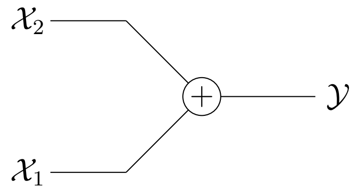


What are the maximum achievable rates R_1 and R_2 to User 1 and 2 respectively?

Solution:

- $R_1 \leq 1 - h(p)$
- $R_2 \leq 1 - h(\epsilon)$, where $\epsilon = p(1-\delta) + \delta(1-p)$

b) Given the following additive channel: such that $\mathcal{X}_1 \in \{0,1\}$ and $\mathcal{X}_2 \in \{0,1\}$,



1. What is the maximum rates R_1 and R_2 if only one user is allowed to transmit?
2. Which possible values can be received when both User 1 and User 2 are transmitting at the same time?
3. What are the maximum rates $\tilde{R}_1$ and $\tilde{R}_2$ in this case (assume transmitted symbols are all equally probable)?
4. Explain how full cooperation between User1 and User2 can be done? what is the maximum achievable sum of rates R_Σ?
5. Sketch the region of achievable rates for a fully cooperative system with TDMA!

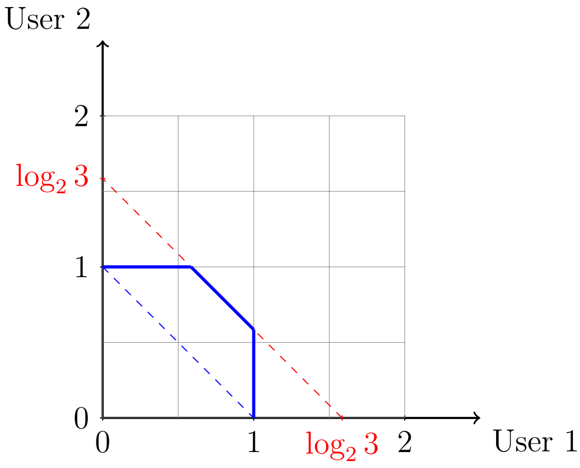
Solution:

1. $R_1 = R_2 = 1$.
2. $\mathcal{Y} \in \{0,1,2\}$
3. For $(\mathcal{X}_1, \mathcal{X}_2) \in \{(1,0),(0,1)\}$, the receiver is not able to recover the transmitted symbols. This can be modeled as an erasure channel for each user with an erasure probability $p = \frac{1}{2} \cdot \frac{1}{2} \cdot 2 = \frac{1}{2}$. Therefore:

$$\tilde{R}_1 = \tilde{R}_2 = 1 - p = \frac{1}{2}.$$

4. With full cooperation, the confusion resulting from the symbols $\{(1,0),(0,1)\}$ is avoided, leaving only three possible symbols to be transmitted. Thus, $R_\Sigma = \log_2 3 = 1.585$.

5.



c) Let a M/M/3/5 system be given with an arrival rate $\lambda = 4s^{-1}$.

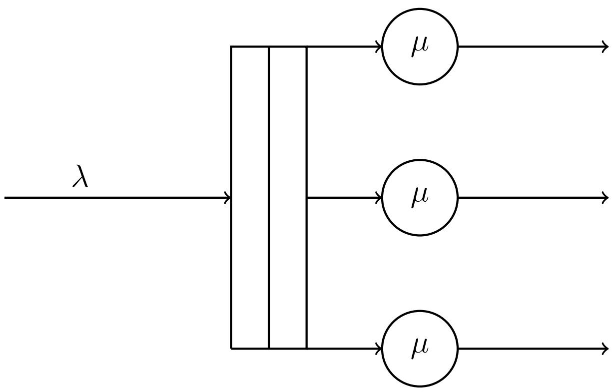
1. Sketch the system.
2. What is highest service rate μ at which the system is not stable?

Let the average service time be $X = 0.5s$.

1. Give the Markov chain, where the states indicate the number of users in the system.
2. Calculate the loss probability P_V.

Solution:

1. The System has 3 processing units and 2 waiting slots.

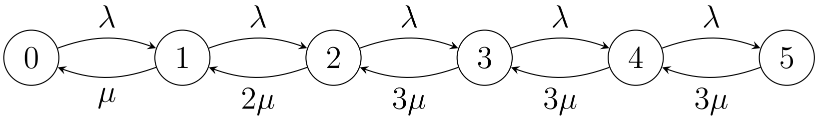


2. The utilization ρ must be smaller than 1 for the system to be stable, that means

$$\frac{\lambda}{3 \cdot \mu} < 1$$
$$\Leftrightarrow \quad \frac{4}{3}s^{-1} = \frac{\lambda}{3} < \mu.$$

Thus $\mu = \frac{4}{3}s^{-1}$ is the highest service rate at which the system is not stable.

3. The maximal processing rate is 3μ, which can be achieved if at least 3 users are in the system.



4. The loss probability is given through the steady state probability of state 5

$$P_V = p_5.$$

The steady state probabilities can be calculated using the following system of equations (with $\rho = \frac{\lambda}{\mu}$).

$$\begin{aligned}
p_0 \cdot \lambda &= p_1 \cdot \mu \quad &\Rightarrow p_1 = \frac{\lambda}{\mu} p_0 = \rho \cdot p_0 \\
p_1 \cdot \lambda &= 2 \cdot p_2 \cdot \mu \quad &\Rightarrow p_2 = \frac{\lambda}{2 \cdot \mu} p_1 = \frac{\rho^2}{2} p_0 \\
p_2 \cdot \lambda &= 3 \cdot p_3 \cdot \mu \quad &\Rightarrow p_3 = \frac{\lambda}{3 \cdot \mu} p_2 = \frac{\rho^3}{2 \cdot 3} p_0 \\
p_3 \cdot \lambda &= 3 \cdot p_4 \cdot \mu \quad &\Rightarrow p_4 = \frac{\lambda}{3 \cdot \mu} p_3 = \frac{\rho^4}{2 \cdot 3^2} p_0 \\
p_4 \cdot \lambda &= 3 \cdot p_5 \cdot \mu \quad &\Rightarrow p_5 = \frac{\lambda}{3 \cdot \mu} p_4 = \frac{\rho^5}{2 \cdot 3^3} p_0 \\
1 &= p_0 + p_1 + p_2 + p_3 + p_4 + p_5
\end{aligned}$$

$$\Rightarrow p_0 = \left(1 + \rho + \frac{\rho^2}{2} + \frac{\rho^3}{2 \cdot 3} + \frac{\rho^4}{2 \cdot 3^2} + \frac{\rho^5}{2 \cdot 3^3}\right)^{-1}$$

$$\begin{aligned}
\Rightarrow P_V = p_5 &= \frac{\rho^5}{2 \cdot 3^3} p_0 \\
&= \frac{\rho^5}{2 \cdot 3^3} \cdot \left(1 + \rho + \frac{\rho^2}{2} + \frac{\rho^3}{2 \cdot 3} + \frac{\rho^4}{2 \cdot 3^2} + \frac{\rho^5}{2 \cdot 3^3}\right)^{-1}
\end{aligned}$$

With $\lambda = 4s^{-1}$ and $\mu = X^{-1} = 2s^{-1}$ we have

$$\rho = \frac{4}{2} = 2.$$

Inserting this value yields

$$\begin{aligned}
p_0 &= \left(1 + 2 + \frac{2^2}{2} + \frac{2^3}{2 \cdot 3} + \frac{2^4}{2 \cdot 3^2} + \frac{2^5}{2 \cdot 3^3}\right)^{-1} \\
&= \left(1 + 2 + 2 + \frac{4}{3} + \frac{8}{9} + \frac{16}{27}\right)^{-1} \\
&= \left(\frac{5 \cdot 27}{27} + \frac{4 \cdot 9}{27} + \frac{8 \cdot 3}{27} + \frac{16}{27}\right)^{-1} \\
&= \frac{27}{211} \\
P_V &= \frac{2^5}{2 \cdot 3^3} \cdot \frac{27}{211} = \frac{16}{211}
\end{aligned}$$

Useful Formulas:

Binary entropy

$$h(p) = -p\log_2(p) - (1-p)\log_2(1-p).$$

IT Inequality

$$\log_b(r) \leq (r-1)\log_b(e) \quad \forall r > 0, b \in \mathbb{N}.$$

Capacity of a BSC with crossover probability p

$$C_{\mathsf{BSC}} = 1 - h(p).$$

Capacity of a BEC with erasure probability δ

$$C_{\mathsf{BEC}} = 1 - \delta.$$

Capacity of a time discrete Gaussian channel with noise power N and signal power S

$$C_{\mathsf{Gauss}} = \frac{1}{2}\log_2\left(1 + \frac{S}{N}\right).$$

Capacity of a bandlimited Gaussian channel with bandwidth W, noise power N_0W and signal power S

$$C_{\mathsf{Gauss},W} = W\log_2\left(1 + \frac{S}{N_0W}\right).$$