Probabilities and entropy of some small neural networks for boolean functions†

Ida G. Sprinkhuizen-Kuyper and Egbert J. W. Boers

Department of Computer Science
Leiden University
{kuyper,boers}@wi.leidenuniv.nl

Abstract

The a priori probabilities of boolean functions with two inputs are calculated for a one-layer neural network and for the 2-2-1 network. The calculations are done both interpreting boolean values as 0 and 1 and as -1 and 1. For the 2-2-1 network the probabilities of the trivial functions (same output for all patterns) are much larger than those of the other functions, while the probabilities of the not linearly separable functions are much smaller. Calculation of the conditional entropy of the neural network based on these probabilities results in an example where the conditional entropy increases while learning examples.

1 Introduction

The a priori probabilities for a neural network representing certain functions can give information about the learnability and the generalization ability of the network when learning those functions from examples. For small networks these probabilities can be calculated explicitly. In this paper the a priori probabilities of boolean functions with two inputs are calculated for a one-layer neural network and for the 2-2-1 network. The calculations are done both for interpreting boolean values as 0 and 1 and as -1 and 1. For the one-layer network we will find that

- for boolean values equal to 0 and 1, the a priori probabilities for the trivial functions (same output for all patterns) are much larger than those for the other linearly separable functions, while
- for boolean values equal to -1 and 1, the a priori probabilities are more equal for all functions.

†. Technical Report 96-30, Dept. of Computer Science, Leiden University. Available as ftp://ftp.wi.leidenuniv.nl/pub/CS/TechnicalReports/1996/tr96-30.ps.gz

For both interpretations of the boolean values the probabilities of the trivial functions in the 2-2-1 network are much larger than those of the other functions, while the probabilities of the XOR-like functions are much smaller.

In their study of the generalization ability of layered feedforward neural networks Denker *et al.* (1987) introduce the entropy S_m and the average generalization ability G_m as function of the size m of the training set. The entropy S_m is a measure of the functional diversity of the chosen architecture restricted so that it correctly represents the m examples of the training set. In Denker *et al.* (1987) it is suggested that the entropy S_m decreases when m increases, while in Solla (1992) and Schwartz *et al.* (1990) it is said explicitly. During his work for his master's thesis Claas (1996) tried to prove that indeed the entropy as defined by Denker *et al.* (1987) has to decrease with m. He could not find a proof and thus we tried the opposite: we tried to find a counter example hoping to get more insight in the behaviour of S_m. We were indeed able to construct a counter example. So S_m will not in general decrease with m. The case is that S_m is a so-called conditional entropy (McEliece, 1977). On average S_m will decrease, but presenting a special example can lead to an increase of the entropy. This example of increasing entropy is found for the 2-2-1 network interpreting the boolean values as 0 and 1.

Section 2 contains the probabilities for the one-layer network and section 3 those for the 2-2-1 network. In both sections the boolean values are interpreted as 0 and 1. Section 4 gives comparable results for the interpretation of the boolean values as -1 and 1. In section 5 the definition of (conditional) entropy is given following Schwartz *et al.* (1990). In section 6 we give examples that this entropy can increase. Section 7 contains the conclusions. Appendices A and B contain the calculations of the probabilities for the 2-2-1 network based on those for the one-layer network.

2 Probabilities for a one-layer network

In this section the boolean values are interpreted as 0 (false) and 1 (true).

Consider the one-layer network with two inputs X_1 and X_2 and one output node Y as given in figure 1. The extra input X_0 has value 1 and serves as a threshold value of the

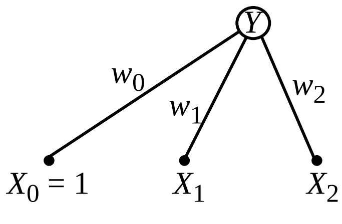

Figure 1. A one-layer network with two inputs and one output node.

output node. This network has three weights w_0, w_1, w_2. We take the transfer function of the output node equal to the sigmoid $g(x) = 1/(1 + e^{-x})$, so the output of the network is $g(w_0 + w_1X_1 + w_2X_2)$ as function of the inputs X_1 and X_2. We interpret the output as zero if $g(x) < \varepsilon$ (i.e. $x < -\alpha$) and as one if $g(x) > 1 - \varepsilon$ (i.e. $x > \alpha$).

This network is able to learn/represent all linearly separable boolean functions of 2 variables. So for example the trivial function $f(X_1, X_2) = 0$ for all $X_1, X_2 \in \{0, 1\}$ is represented if the weights satisfy the following inequalities:

$$g(w_0) < \varepsilon \quad \Leftrightarrow \quad w_0 < -\alpha$$

$$g(w_0 + w_1) < \varepsilon \quad \Leftrightarrow \quad w_0 + w_1 < -\alpha$$

$$g(w_0 + w_2) < \varepsilon \quad \Leftrightarrow \quad w_0 + w_2 < -\alpha$$

$$g(w_0 + w_1 + w_2) < \varepsilon \quad \Leftrightarrow \quad w_0 + w_1 + w_2 < -\alpha$$

Restricting the weight space to a cube $-N < w_i < N$ results in a volume of measure $N^2(N-\alpha) + N(N-\alpha)^2 + 1/6\,(N-\alpha)^3$ for the part of the weight space correctly representing the function with all outputs equal to zero (Claas, 1996).

We computed the volumes of the parts of the weight space corresponding to all representable boolean functions and their corresponding probabilities. Taking the limit $N \to \infty$ resulted in the probabilities given in table 1. Note that if $N \to \infty$, the exact value of α is no longer important.

Both the XOR function and the NOT XOR function have probability zero, since they cannot be represented by the network.

n	outputs of f_n	$P_0(f_n)$
0	(0,0,0,0)	13/48
1	(0,0,0,1)	1/48
2	(0,0,1,0)	2/48
3	(0,0,1,1)	2/48
4	(0,1,0,0)	2/48
5	(0,1,0,1)	2/48
6	(0,1,1,0)	0
7	(0,1,1,1)	2/48
8	(1,0,0,0)	2/48
9	(1,0,0,1)	0
10	(1,0,1,0)	2/48
11	(1,0,1,1)	2/48
12	(1,1,0,0)	2/48
13	(1,1,0,1)	2/48
14	(1,1,1,0)	1/48
15	(1,1,1,1)	13/48

Table 1. The probabilities of the functions of the network of figure 1. The number n corresponds to the interpretation of the output pattern as binary number. In the output pattern the first digit is the output for input (0, 0), the second digit that of $(0, 1)$, the third digit is the output of (1, 0) and the most right digit corresponds to the output for both inputs equal to 1. In the most right column the probabilities are given.

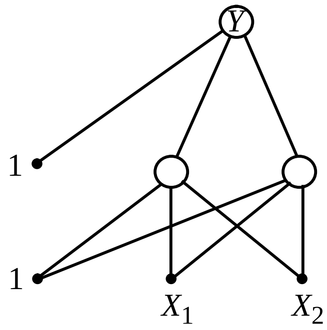

Figure 2. The 2-2-1 network.

3 Probabilities in the 2-2-1 network

Also in this section the boolean values are interpreted as 0 (false) and 1 (true).

Using the probabilities of table 1, we were able to compute the probabilities of the 2-2-1 network (see figure 2). These probabilities are given in table 2 (the computations are given in appendix A). From this table it is clear that the ratio of the probabilities of the trivial functions f_0 and f_{15} to those of the other functions is large (≥ 20), while the probability of the not linearly separable functions (f_6 and f_9) is small with respect to those of the other functions.

n	outputs of f_n	$P_0(f_n)$
0	(0,0,0,0)	11013/27648
1	(0,0,0,1)	271/27648
2	(0,0,1,0)	508/27648
3	(0,0,1,1)	484/27648
4	(0,1,0,0)	508/27648
5	(0,1,0,1)	484/27648
6	(0,1,1,0)	24/27648
7	(0,1,1,1)	532/27648
8	(1,0,0,0)	532/27648
9	(1,0,0,1)	24/27648
10	(1,0,1,0)	484/27648
11	(1,0,1,1)	508/27648
12	(1,1,0,0)	484/27648
13	(1,1,0,1)	508/27648
14	(1,1,1,0)	271/27648
15	(1,1,1,1)	11013/27648

Table 2. The probabilities of the functions of the 2-2-1 network of figure 2. These probabilities are based on the probabilities in table 1.

4 Probabilities for boolean values -1 and 1

When the inputs of the network for a boolean function are -1 (false) and 1 (true) instead of 0 and 1, it is also possible to calculate the a priori probabilities for the one-layer network of figure 1. So for example the function f_0 is represented if the weights satisfy the following inequalities:

$$g(w_0 - w_1 - w_2) < \varepsilon \qquad \Leftrightarrow \qquad w_0 - w_1 - w_2 < -\alpha$$

$$g(w_0 + w_1 - w_2) < \varepsilon \qquad \Leftrightarrow \qquad w_0 + w_1 - w_2 < -\alpha$$

$$g(w_0 - w_1 + w_2) < \varepsilon \qquad \Leftrightarrow \qquad w_0 - w_1 + w_2 < -\alpha$$

$$g(w_0 + w_1 + w_2) < \varepsilon \qquad \Leftrightarrow \qquad w_0 + w_1 + w_2 < -\alpha$$

Here we suppose that the transfer function $g(x)$ is a sigmoide between the values -1 and 1, e.g. $g(x) = \tanh(x)$.

If α is equal to zero the volume of the part of the cube $-N < w_i < N$, $i = 0, 1, 2$, for which the weights result in f_0, is equal to $\frac{2}{3}N^3$. This volume results in an a priori probability $P_0(f_0)$ of $1/12$. Table 3 contains the a priori probabilities of the boolean functions of two inputs for the network of figure 1. It is clear from this table that the probability distribution is more regular than that of table 1. Similarly to the calculations (see appendix B) for the 2-2-1 network with inputs and outputs between 0 and 1, we calculated the probabilities for this network with inputs and outputs between -1 and 1. The a priori probabilities for the 2-2-1 network with inputs and outputs -1 and 1 are given in table 4. Again the prob-

n	outputs of f_n	$P_0(f_n)$
0	(-1,-1,-1,-1)	1/12
1	(-1,-1,-1,1)	1/16
2	(-1,-1,1,-1)	1/16
3	(-1,-1,1,1)	1/12
4	(-1,1,-1,-1)	1/16
5	(-1,1,-1,1)	1/12
6	(-1,1,1,-1)	0
7	(-1,1,1,1)	1/16
8	(1,-1,-1,-1)	1/16
9	(1,-1,-1,1)	0
10	(1,-1,1,-1)	1/12
11	(1,-1,1,1)	1/16
12	(1,1,-1,-1)	1/12
13	(1,1,-1,1)	1/16
14	(1,1,1,-1)	1/16
15	(1,1,1,1)	1/12

Table 3. The probabilities of the boolean functions of the network of figure 1. The difference with table 1 is that here the inputs are -1 and 1 instead of 0 and 1.

n	outputs of f_n	$P_0(f_n)$
0	(-1,-1,-1,-1)	461/2304
1	(-1,1-,-1,1)	110/2304
2	(1-,-1,1,-1)	110/2304
3	(-1,-1,1,1)	121/2304
4	(-1,1,-1,-1)	110/2304
5	(-1,1,-1,1)	121/2304
6	(-1,1,1,-1)	9/2304
7	(-1,1,1,1)	110/2304
8	(1,-1,-1,-1)	110/2304
9	(1,-1,-1,1)	9/2304
10	(1,-1,1,-1)	121/2304
11	(1,-1,1,1)	110/2304
12	(1,1,-1,-1)	121/2304
13	(1,1,-1,1)	110/2304
14	(1,1,1,-1)	110/2304
15	(1,1,1,1)	461/2304

Table 4. The probabilities of the functions of the 2-2-1 network of figure 2. These probabilities are based on the probabilities in table 1.

abilities of the trivial functions are large with respect to those of the other functions, however the differences are less extreme than in the case of table 2 (ratio ≥ 3.8 instead of ≥ 20).

5 The entropy of a neural network

Defining the entropy of a neural network, we follow Schwartz *et al.* (1990) (see also Herz *et al.* (1991, section 6.5)).

Consider an ensemble of layered networks with fixed architecture and varying weights. Such an ensemble is described by its configuration space $\{\mathbf{W}\}$: every point $\mathbf{W}$ is a list of values for all weights needed to select a network design within the chosen architecture. The resulting network realizes a specific input-output function, $y = f_{\mathbf{W}}(x)$.

A density $\rho_0(\mathbf{W})$ on the weight space constrains the effective volume of the configuration space to

$$Z_0 = \int \rho_0(\mathbf{W})\, d\mathbf{W}$$

Regions corresponding to the implementation of the function f are identified by the masking function

$$\Theta_f(\mathbf{W}) = \begin{cases} 1 & \text{if} \quad f_{\mathbf{W}} = f \\ 0 & \text{if} \quad f_{\mathbf{W}} \neq f \end{cases}$$

and occupy a volume

$$Z(f) = \int \Theta_f(\mathbf{W})\,\rho_0(\mathbf{W})\,d\mathbf{W}$$

The specification of an architecture and its corresponding configuration space thus defines a probability on the space of functions:

$$P_0(f) = \frac{Z(f)}{Z_0}$$

which results from a full exploration of configuration space. $P_0(f)$ is the probability that a randomly chosen network in configuration space will realize the function *f*. The class of functions implementable by a given architecture is

$$\boldsymbol{F} = \{f | P_0(f) \neq 0\}$$

The entropy of the distribution

$$S_0 = -\sum_{\{f\}} P_0(f)\,{}^2\log P_0(f)$$

is a measure of the functional diversity of the chosen architecture. The maximum value of S_0 is ${}^2\log(n_{\boldsymbol{F}})$, where $n_{\boldsymbol{F}}$ is the number of functions in class $\boldsymbol{F}$, and is attained when all realizable functions are equally likely, and corresponds to the uniform distribution, $P_0(f) = 1/n_{\boldsymbol{F}}$ for all $f \in \boldsymbol{F}$.

Supervised learning results in a monotonic reduction of the effective volume of the configuration space. An example $\xi^\alpha = (x^\alpha, y^\alpha)$ of the desired function $\tilde{f}$ is learned by removing from $\boldsymbol{F}$ every function that contradicts it. A sequence of m input-output pairs $(\xi^1, \ldots, \xi^m)$, which are examples of $\tilde{f}$ thus defines a sequence of classes of functions,

$$\boldsymbol{F}_m \subseteq \boldsymbol{F}_{m-1} \subseteq \ldots \boldsymbol{F}_1 \subseteq \boldsymbol{F}$$

where every function $f \in \boldsymbol{F}_m$ correctly classifies all of the training examples ξ^α, $1 \leq \alpha \leq m$. The effective volume of configuration space is reduced to

$$Z_m = \int \sum_{f \in \boldsymbol{F}_m} \Theta_f(\mathbf{W})\,\rho_0(\mathbf{W})\,d\mathbf{W}$$

by learning a training set of size *m*.

The probability on the space of functions is modified by learning and becomes

$$P_m(f) = \frac{Z(f)}{Z_m}, \text{ for } f \in \boldsymbol{F}_m.$$

The entropy of the distribution after m training examples,

$$S_m = -\sum_{\{f\}} P_m(f)\,{}^2\log P_m(f)$$

reflects the narrowing of the probability distribution: $S_m < S_0$. The entropy decrease $\eta_m = S_{m-1} - S_m$ defines the efficiency of learning the *m*th example.

The optimal case of $S_m = 0$ corresponds to the elimination of all ambiguity about the function to be implemented.

Thus far we followed Schwartz *et al.* (1990). If all functions have equal probability it is clear that $S_m = {}^2\log|\boldsymbol{F}_m|$, $|\boldsymbol{F}_m|$ being the cardinality of $\boldsymbol{F}_m$, and thus in that case S_m will decrease when $\boldsymbol{F}_m$ becomes smaller. However, in general all functions will not have equal probability as we showed in sections 2 to 4.

6 An example of increasing entropy

We start with an artificial example:

Consider the set of functions $\boldsymbol{F}_m = \{f_1, f_2, f_3, f_4, f_5, f_6\}$, with probabilities $P_m(f_6) = 1/2$, $P_m(f_i) = 1/10$, $i = 1\ldots5$. The entropy S_m is equal to:

$$S_m = -\frac{1}{2} \cdot {}^2\log\left(\frac{1}{2}\right) - 5 \cdot \frac{1}{10} \cdot {}^2\log\left(\frac{1}{10}\right) = \frac{1}{2} \cdot {}^2\log 20 \approx 2.16$$

Suppose the (m+1)th example accepts f_1 to f_5, but rejects f_6. Then we get the new set of functions $\boldsymbol{F}_{m+1} = \{f_1, f_2, f_3, f_4, f_5\}$ with probabilities $P_{m+1}(f_i) = 1/5$, $i = 1\ldots5$. The corresponding entropy S_{m+1} is:

$$S_{m+1} = -5 \cdot \frac{1}{5} \cdot {}^2\log\left(\frac{1}{5}\right) = {}^2\log 5 \approx 2.32$$

since $5 > \sqrt{20}$ it is clear that in this case $S_{m+1} > S_m$.

In (McEliece, 1977) the conditional entropy of X, given $Y = y$ is defined as:

$$H(X|Y{=}y) = -\sum_x p(x|y) \log p(x|y)$$

and the conditional entropy $H(X|Y)$ is its expectation:

$$H(X|Y) = \sum_y p(y) H(X|Y{=}y)$$

Example 1.7 in (McEliece, 1977) also shows that the conditional entropy $H(X|Y{=}y)$ can be larger than the original entropy $H(X)$. It is also proved that the mutual information

$$I(X;Y) = H(X) - H(X|Y)$$

is positive (theorem 1.3 in (McEliece, 1977)). So on the average (learning arbitrary functions) the entropy S_m will decrease, while for learning some concrete function it is possible that sometimes S_m increases.

A concrete example of increasing entropy for a neural network follows from the probabilities of the 2-2-1 network as given in table 2. Presenting the training examples: $(0, 0) \to 0$, $(0, 1) \to 0$, and $(1, 0) \to 1$ results in values for the entropies: $S_2 \approx 0.636$ and $S_3 \approx 1.000$. So here we found a real example of a neural network for which the entropy increases by learning an example. The clue of this example is that after two examples the trivial function f_0 is allowed, resulting in a low entropy because of the high probability of the trivial function. After the third example two functions with almost equal probability remain.

7 Conclusion

We calculated the probabilities of boolean functions of 2 inputs for a one-layer network and for the 2-2-1 network both interpreting boolean values as 0 and 1 and as -1 and 1. For the one-layer network the probability distribution for the linearly separable boolean functions is much more regular when interpreting the boolean values as -1 ans 1 than when using the interpretation 0 and 1. Thus we expect that on the average for the one-layer network nontrivial functions are easier learned with the -1, 1 interpretation than with the 0, 1 interpretation of boolean values. The trivial functions will be easier learned by a one-layer network with the 0, 1 interpretation of boolean values.

For the two-layer 2-2-1 network the probabilities of the trivial functions are larger than those of the other functions for both interpretations. So to speed up learning it seems essential to find as soon as possible a weight configuration that excludes the trivial functions.

The entropy introduced by Denker *et al.* (1987) for learning by examples is a so-called conditional entropy. It is possible that presenting concrete examples will result in an increase of the entropy. Finally, after presenting enough examples so that all ambiguity about the function is eliminated, the entropy will become zero. Especially when the probabilities of the representable functions vary strongly, sometimes the entropy will increase. When all probabilities are equal, the entropy will always decrease. By computing the probabilities for a simple neural network representing boolean functions of two variables, we showed that these probabilities can vary strongly in practice. Especially the trivial functions (all patterns giving the same output) have a high probability. Thus for a neural network it is possible to be confronted with the effect that the entropy S_m increases sometimes. We showed that for the 2-2-1 network this effect indeed occurs.

From a learning perspective it is interesting to think about the meaning of the remaining entropy after learning a training set containing a fixed number of examples. As long as one of the trivial functions is allowed by the training set, the entropy can be smaller than when both trivial functions are excluded, because of the large probability of the trivial functions. So, from the perspective of the entropy (functional diversity) it can sometimes be a good strategy to choose the training set such that it allows one of the trivial functions, because the entropy will be smaller. However, when a nontrivial function has to be learned it is probably more realistic to choose a training set that excludes the trivial functions. The entropy can be larger, but the value of the remaining entropy will be more realistic with respect to the amount of work that has to be done to learn the desired function. Also the probability of the desired function will be higher for such a training set.

Acknowledgements

We like to thank W.J. Claas for his work and the discussions on this subject during the period he was working for his master's thesis. We also like to thank G. Cybenko (visiting our institute for the "Kloosterman chair") who told me about the notion of conditional entropy (McEliece, 1977).

References

Claas, W.J. (1996). *Leren met terugkoppeling, het gemiddeld generaliserend vermogen van feedforward netwerken*. Master's Thesis, Internal Report 96-01, Department of Computer Science, Leiden University.

Denker, J., Schwartz, D., Wittner, B., Solla, S., Howard, R., Jackel, L., & Hopfield, J. (1987). "Large Automatic Learning, Rule Extraction, and Generalisation". *Complex Systems* 1, 877–922.

Hertz, J., Krogh, A., & Palmer, R.G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley: Redwood, CA.

McEliece, R.J. (1977). *The Theory of Information and Coding*. Vol. 3 of: Rota, G.-C. (ed.): Encyclopedia of Mathematics and its Applications, Addison-Wesley.

Solla, S. (1992). "Supervised Learning: A Theoretical Framework". In: Casdagli, M. & Eubank, S. (eds.): *Nonlinear Modeling and Forecasting, SFI Studies in the Science of Complexity, Proc. Vol. XII*, Addison-Wesley: Redwood, CA.

Schwartz, D.B., Samalam, V.K., Solla, S.A., & Denker, J.S. (1990). "Exhaustive Learning". *Neural Computation* 2, 374–358.

Appendix A Probabilities of the 2-2-1 network (inputs 0 and 1)

Table 5 contains the probabilities for the functions f_0 to f_7 as function of the function of the hidden nodes (in table 5 the first hidden node $H_1 = f_0$ and the second hidden node runs from f_0 to f_7). The probabilities not given in this table with $H_1 = f_0$ can be derived by the properties that the probabilities for $H_2 = f_i$ are equal to those for $H_2 = f_{15-i}$ and also the probabilties for f_i are equal to those for f_{15-i}. The tables 6 to 11 contain the probabilities for the first hidden node representing f_1 to f_5, and f_7 (the probability that H_1 represents f_6

H_1	H_2	probabilities								probability of this combination of H_1 and H_2
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_0	f_0	1/2	0	0	0	0	0	0	0	13/48 · 13/48
f_0	f_1	3/8	1/8	0	0	0	0	0	0	13/48 · 1/48
f_0	f_2	3/8	0	1/8	0	0	0	0	0	13/48 · 2/48
f_0	f_3	3/8	0	0	1/8	0	0	0	0	13/48 · 2/48
f_0	f_4	3/8	0	0	0	1/8	0	0	0	13/48 · 2/48
f_0	f_5	3/8	0	0	0	0	1/8	0	0	13/48 · 2/48
f_0	f_6	-	-	-	-	-	-	-	-	0
f_0	f_7	3/8	0	0	0	0	0	0	1/8	13/48 · 2/48

Table 5. The probabilities of the functions f_0 to f_7 as function of the hidden nodes. In this table the first hidden node H_1 represents the function f_0. The probabilities of the functions f_8 to f_{15} follow from the fact that the probability of f_i is equal to the probability of f_{15-i}. The probabilities for the second hidden node representing f_{15-i} are equal to those for the second hidden node representing f_i.

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_1	f_0	3/8	1/8	0	0	0	0	0	0	$1/48 \cdot 13/48$
f_1	f_1	17/48	7/48	0	0	0	0	0	0	$1/48 \cdot 1/48$
f_1	f_2	7/24	1/12	1/12	1/24	0	0	0	0	$1/48 \cdot 2/48$
f_1	f_3	5/16	1/16	1/24	1/12	0	0	0	0	$1/48 \cdot 2/48$
f_1	f_4	7/24	1/12	0	0	1/12	1/24	0	0	$1/48 \cdot 2/48$
f_1	f_5	5/16	1/16	0	0	1/24	1/12	0	0	$1/48 \cdot 2/48$
f_1	f_6	-	-	-	-	-	-	-	-	0
f_1	f_7	5/16	1/16	0	0	0	0	1/24	1/12	$1/48 \cdot 2/48$

Table 6. The probabilities for the first hidden node representing f_1

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_2	f_0	3/8	0	1/8	0	0	0	0	0	$2/48 \cdot 13/48$
f_2	f_1	7/24	1/12	1/12	1/24	0	0	0	0	$2/48 \cdot 1/48$
f_2	f_2	17/48	0	7/48	0	0	0	0	0	$2/48 \cdot 2/48$
f_2	f_3	5/16	1/24	1/16	1/12	0	0	0	0	$2/48 \cdot 2/48$
f_2	f_4	7/24	0	1/12	0	1/12	0	1/24	0	$2/48 \cdot 2/48$
f_2	f_5	7/24	0	1/12	0	0	1/12	0	1/24	$2/48 \cdot 2/48$
f_2	f_6	-	-	-	-	-	-	-	-	0
f_2	f_7	5/16	0	1/16	0	0	1/24	0	1/12	$2/48 \cdot 2/48$

Table 7. The probabilities for the first hidden node representing f_2

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_3	f_0	3/8	0	0	1/8	0	0	0	0	$2/48 \cdot 13/48$
f_3	f_1	5/16	1/16	1/24	1/12	0	0	0	0	$2/48 \cdot 1/48$
f_3	f_2	5/16	1/24	1/16	1/12	0	0	0	0	$2/48 \cdot 2/48$
f_3	f_3	17/48	0	0	7/48	0	0	0	0	$2/48 \cdot 2/48$
f_3	f_4	7/24	0	0	1/12	1/12	0	0	1/24	$2/48 \cdot 2/48$
f_3	f_5	13/48	1/48	1/24	1/24	1/24	1/24	0	1/24	$2/48 \cdot 2/48$
f_3	f_6	-	-	-	-	-	-	-	-	0
f_3	f_7	5/16	0	0	1/16	1/24	0	0	1/12	$2/48 \cdot 2/48$

Table 8. The probabilities for the first hidden node representing f_3

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_4	f_0	3/8	0	0	0	1/8	0	0	0	$2/48 \cdot 13/48$
f_4	f_1	7/24	1/12	0	0	1/12	1/24	0	0	$2/48 \cdot 1/48$
f_4	f_2	7/24	0	1/12	0	1/12	0	1/24	0	$2/48 \cdot 2/48$
f_4	f_3	7/24	0	0	1/12	1/12	0	0	1/24	$2/48 \cdot 2/48$
f_4	f_4	17/48	0	0	0	7/48	0	0	0	$2/48 \cdot 2/48$
f_4	f_5	5/16	1/24	0	0	1/16	1/12	0	0	$2/48 \cdot 2/48$
f_4	f_6	-	-	-	-	-	-	-	-	0
f_4	f_7	5/16	0	0	1/24	1/16	0	0	1/12	$2/48 \cdot 2/48$

Table 9. The probabilities for the first hidden node representing f_4

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_5	f_0	3/8	0	0	0	0	1/8	0	0	$2/48 \cdot 13/48$
f_5	f_1	5/16	1/16	0	0	1/24	1/12	0	0	$2/48 \cdot 1/48$
f_5	f_2	7/24	0	1/12	0	0	1/12	0	1/24	$2/48 \cdot 2/48$
f_5	f_3	13/48	1/48	1/24	1/24	1/24	1/24	0	1/24	$2/48 \cdot 2/48$
f_5	f_4	5/16	1/24	0	0	1/16	1/12	0	0	$2/48 \cdot 2/48$
f_5	f_5	17/48	0	0	0	0	7/48	0	0	$2/48 \cdot 2/48$
f_5	f_6	-	-	-	-	-	-	-	-	0
f_5	f_7	5/16	0	1/24	0	0	1/16	0	1/12	$2/48 \cdot 2/48$

Table 10. The probabilities for the first hidden node representing f_5

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_7	f_0	3/8	0	0	0	0	0	0	1/8	$2/48 \cdot 13/48$
f_7	f_1	5/16	1/16	0	0	0	0	1/24	1/12	$2/48 \cdot 1/48$
f_7	f_2	5/16	0	1/16	0	0	1/24	0	1/12	$2/48 \cdot 2/48$
f_7	f_3	5/16	0	0	1/16	1/24	0	0	1/12	$2/48 \cdot 2/48$
f_7	f_4	5/16	0	0	1/24	1/16	0	0	1/12	$2/48 \cdot 2/48$
f_7	f_5	5/16	0	1/24	0	0	1/16	0	1/12	$2/48 \cdot 2/48$
f_7	f_6	-	-	-	-	-	-	-	-	0
f_7	f_7	17/48	0	0	0	0	0	0	7/48	$2/48 \cdot 2/48$

Table 11. The probabilities for the first hidden node representing f_7

is zero).

From these tables the a priori probabilities for the boolean functions in the 2-2-1 network can be derived. For example in table 5 we see that when both hidden nodes represent f_0 the probability of the network representing f_0 is equal to 1/2. The probabilty that both hidden nodes represent f_0 is equal to $(13/48)^2$. So we find here a contribution of $1/2 \cdot (13/48)^2$ to the probability of f_0 for the complete network. This same probability occurs when the first hidden node represents f_0 and the second hidden node represents f_{15} and vice versa and also when both hidden nodes represent f_{15} , so each probability found in the tables 5 to 11 has to be multiplied by a factor 4. Calculating the probabilities from the tables 5 to 11 results in:

$$
\begin{aligned}
P_0(f_0) = {} & 4\Big[\Big(1/2\cdot\Big(\frac{13}{48}\Big)^2+\frac{3}{8}\cdot\frac{13}{48}\cdot\frac{1}{48}+5\cdot\frac{3}{8}\cdot\frac{13}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{3}{8}\cdot\frac{1}{48}\cdot\frac{13}{48}+\frac{17}{48}\cdot\frac{1}{48}\cdot\frac{1}{48}+\frac{7}{24}\cdot\frac{1}{48}\cdot\frac{2}{48}+{} \\
& 3\cdot\frac{5}{16}\cdot\frac{1}{48}\cdot\frac{2}{48}+\frac{7}{24}\cdot\frac{1}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{3}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{7}{24}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{17}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+2\cdot\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+{} \\
& 2\cdot\frac{7}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{3}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{1}{48}+2\cdot\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{17}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{7}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+{} \\
& \frac{13}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{3}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{7}{24}\cdot\frac{2}{48}\cdot\frac{1}{48}+2\cdot\frac{7}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{17}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+2\cdot\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+{} \\
& \Big(\frac{3}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{7}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{13}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+2\cdot\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{17}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+{} \\
& \Big(\frac{3}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{1}{48}+4\cdot\frac{5}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{17}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)\Big] \\
= {} & \frac{11013}{12\cdot 48^2}
\end{aligned}
$$

$$
\begin{aligned}
P_0(f_1) = {} & 4\Big[\frac{1}{8}\cdot\frac{13}{48}\cdot\frac{1}{48}+\Big(\frac{1}{8}\cdot\frac{1}{48}\cdot\frac{13}{48}+\frac{7}{48}\cdot\frac{1}{48}\cdot\frac{1}{48}+2\cdot\frac{1}{12}\cdot\frac{1}{48}\cdot\frac{2}{48}+3\cdot\frac{1}{16}\cdot\frac{1}{48}\cdot\frac{2}{48}\Big)+{} \\
& \Big(\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+{} \\
& \Big(\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{1}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{1}{48}\Big] \\
= {} & \frac{271}{12\cdot 48^2}
\end{aligned}
$$

$$
\begin{aligned}
P_0(f_2) = {} & 4\Big[\frac{1}{8}\cdot\frac{13}{48}\cdot\frac{2}{48}+\Big(\frac{1}{12}\cdot\frac{1}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{1}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{7}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+{} \\
& 2\cdot\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+2\cdot\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+{} \\
& \frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}+\Big(\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}+2\cdot\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)\Big] \\
= {} & \frac{508}{12\cdot 48^2}
\end{aligned}
$$

$$
\begin{aligned}
P_0(f_3) = {} & 4\Big[\frac{1}{8}\cdot\frac{13}{48}\cdot\frac{2}{48}+\Big(\frac{1}{24}\cdot\frac{1}{48}\cdot\frac{2}{48}+\frac{1}{12}\cdot\frac{1}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{1}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+{} \\
& \frac{1}{12}\cdot\frac{2}{48}\cdot\frac{1}{48}+2\cdot\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{7}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\Big(\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}+{} \\
& \frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\Big(\frac{1}{16}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}\Big)\Big] \\
= {} & \frac{484}{12\cdot 48^2}
\end{aligned}
$$

$$P_0(f_4) = P_0(f_2) = \frac{508}{12 \cdot 48^2}$$

$$P_0(f_5) = P_0(f_3) = \frac{484}{12 \cdot 48^2}$$

$$P_0(f_6) = 4\left[\frac{1}{24}\cdot\frac{1}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{1}{48}\right] = \frac{24}{12 \cdot 48^2}$$

$$\begin{aligned}P_0(f_7) = 4\Bigg[&\frac{1}{8}\cdot\frac{13}{48}\cdot\frac{2}{48}+\frac{1}{12}\cdot\frac{1}{48}\cdot\frac{2}{48}+\left(\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}\right)+\left(2\cdot\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}\right)+\\&\left(\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}\right)+\left(2\cdot\frac{1}{24}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}\right)+\left(\frac{1}{8}\cdot\frac{2}{48}\cdot\frac{13}{48}+\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{1}{48}+\right.\\&\left.4\cdot\frac{1}{12}\cdot\frac{2}{48}\cdot\frac{2}{48}+\frac{7}{48}\cdot\frac{2}{48}\cdot\frac{2}{48}\right)\Bigg]\\=\ &\frac{532}{12 \cdot 48^2}\end{aligned}$$

Appendix B Probabilities for inputs -1 and 1

Using the probabilities from table 4 in order to compute the probabilities in the 2-2-1 network interpreting the boolean values as -1 (false) and 1 (true) results in the tables 12 to 18. From these tables the following a priori probabilities are derived:

H_1	H_2	probabilities f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	probability of this combination
f_0	f_0	1/2	0	0	0	0	0	0	0	$1/12 \cdot 1/12$
f_0	f_1	7/24	5/24	0	0	0	0	0	0	$1/12 \cdot 1/16$
f_0	f_2	7/24	0	5/24	0	0	0	0	0	$1/12 \cdot 1/16$
f_0	f_3	7/24	0	0	5/24	0	0	0	0	$1/12 \cdot 1/12$
f_0	f_4	7/24	0	0	0	5/24	0	0	0	$1/12 \cdot 1/16$
f_0	f_5	7/24	0	0	0	0	5/24	0	0	$1/12 \cdot 1/12$
f_0	f_6	-	-	-	-	-	-	-	-	0
f_0	f_7	7/24	0	0	0	0	0	0	5/24	$1/12 \cdot 1/16$

Table 12. The probabilities of the functions f_0 to f_7 as function of the hidden nodes when the inputs (and outputs) are equal to -1 and 1 and the first hidden node represents f_0.

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_1	f_0	7/24	5/24	0	0	0	0	0	0	$1/16 \cdot 1/12$
f_1	f_1	5/24	7/24	0	0	0	0	0	0	$1/16 \cdot 1/16$
f_1	f_2	7/48	7/48	7/48	1/16	0	0	0	0	$1/16 \cdot 1/16$
f_1	f_3	7/48	7/48	1/16	7/48	0	0	0	0	$1/16 \cdot 1/12$
f_1	f_4	7/48	7/48	0	0	7/48	1/16	0	0	$1/16 \cdot 1/16$
f_1	f_5	7/48	7/48	0	0	1/16	7/48	0	0	$1/16 \cdot 1/12$
f_1	f_6	-	-	-	-	-	-	-	-	0
f_1	f_7	7/48	7/48	0	0	0	0	1/16	7/48	$1/16 \cdot 1/16$

Table 13. The probabilities for the first hidden node representing f_1 (inputs equal to -1 and 1)

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_2	f_0	7/24	0	5/24	0	0	0	0	0	$1/16 \cdot 1/12$
f_2	f_1	7/48	7/48	7/48	1/16	0	0	0	0	$1/16 \cdot 1/16$
f_2	f_2	5/24	0	7/24	0	0	0	0	0	$1/16 \cdot 1/16$
f_2	f_3	7/48	1/16	7/48	7/48	0	0	0	0	$1/16 \cdot 1/12$
f_2	f_4	7/48	0	7/48	0	7/48	0	1/16	0	$1/16 \cdot 1/16$
f_2	f_5	7/48	0	7/48	0	0	7/48	0	1/16	$1/16 \cdot 1/12$
f_2	f_6	-	-	-	-	-	-	-	-	0
f_2	f_7	7/48	0	7/48	0	0	1/16	0	7/48	$1/16 \cdot 1/16$

Table 14. The probabilities for the first hidden node representing f_2 (inputs equal to -1 and 1)

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_3	f_0	7/24	0	0	5/24	0	0	0	0	$1/12 \cdot 1/12$
f_3	f_1	7/48	7/48	1/16	7/48	0	0	0	0	$1/12 \cdot 1/16$
f_3	f_2	7/48	1/16	7/48	7/48	0	0	0	0	$1/12 \cdot 1/16$
f_3	f_3	5/24	0	0	7/24	0	0	0	0	$1/12 \cdot 1/12$
f_3	f_4	7/48	0	0	7/48	7/48	0	0	1/16	$1/12 \cdot 1/16$
f_3	f_5	1/12	1/16	1/16	1/12	1/16	1/12	0	1/16	$1/12 \cdot 1/12$
f_3	f_6	-	-	-	-	-	-	-	-	0
f_3	f_7	7/48	0	0	7/48	1/16	0	0	7/48	$1/12 \cdot 1/16$

Table 15. The probabilities for the first hidden node representing f_3 (inputs equal to -1 and 1)

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_4	f_0	7/24	0	0	0	5/24	0	0	0	$1/16 \cdot 1/12$
f_4	f_1	7/48	7/48	0	0	7/48	1/16	0	0	$1/16 \cdot 1/16$
f_4	f_2	7/48	0	7/48	0	7/48	0	1/16	0	$1/16 \cdot 1/16$
f_4	f_3	7/48	0	0	7/48	7/48	0	0	1/16	$1/16 \cdot 1/12$
f_4	f_4	5/24	0	0	0	7/24	0	0	0	$1/16 \cdot 1/16$
f_4	f_5	7/48	1/16	0	0	7/48	7/48	0	0	$1/16 \cdot 1/12$
f_4	f_6	-	-	-	-	-	-	-	-	0
f_4	f_7	7/48	0	0	1/16	7/48	0	0	7/48	$1/16 \cdot 1/16$

Table 16. The probabilities for the first hidden node representing f_4 (inputs equal to -1 and 1)

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_5	f_0	7/24	0	0	0	0	5/24	0	0	$1/12 \cdot 1/12$
f_5	f_1	7/48	7/48	0	0	1/16	7/48	0	0	$1/12 \cdot 1/16$
f_5	f_2	7/48	0	7/48	0	0	7/48	0	1/16	$1/12 \cdot 1/16$
f_5	f_3	1/12	1/16	1/16	1/12	1/16	1/12	0	1/16	$1/12 \cdot 1/12$
f_5	f_4	7/48	1/16	0	0	7/48	7/48	0	0	$1/12 \cdot 1/16$
f_5	f_5	5/24	0	0	0	0	7/24	0	0	$1/12 \cdot 1/12$
f_5	f_6	-	-	-	-	-	-	-	-	0
f_5	f_7	7/48	0	1/16	0	0	7/48	0	7/48	$1/12 \cdot 1/16$

Table 17. The probabilities for the first hidden node representing f_5 (inputs equal to -1 and 1)

H_1	H_2	probabilities								probability of this combination
		f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	
f_7	f_0	7/24	0	0	0	0	0	0	5/24	$1/16 \cdot 1/12$
f_7	f_1	7/48	7/48	0	0	0	0	1/16	7/48	$1/16 \cdot 1/16$
f_7	f_2	7/48	0	7/48	0	0	1/16	0	7/48	$1/16 \cdot 1/16$
f_7	f_3	7/48	0	0	7/48	1/16	0	0	7/48	$1/16 \cdot 1/12$
f_7	f_4	7/48	0	0	1/16	7/48	0	0	7/48	$1/16 \cdot 1/16$
f_7	f_5	7/48	0	1/16	0	0	7/48	0	7/48	$1/16 \cdot 1/12$
f_7	f_6	-	-	-	-	-	-	-	-	0
f_7	f_7	5/48	0	0	0	0	0	0	7/24	$1/16 \cdot 1/16$

Table 18. The probabilities for the first hidden node representing f_7 (inputs equal to -1 and 1)

$$
\begin{aligned}
P_0(f_0) \;=\; & 4\Big[\Big(\frac{1}{2}\cdot\frac{1}{12}\cdot\frac{1}{12}+4\cdot\frac{7}{24}\cdot\frac{1}{12}\cdot\frac{1}{16}+2\cdot\frac{7}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{24}\cdot\frac{1}{16}\cdot\frac{1}{12}+\frac{5}{24}\cdot\frac{1}{16}\cdot\frac{1}{16}+3\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+\\
& 2\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{24}\cdot\frac{1}{16}\cdot\frac{1}{12}+3\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+\frac{5}{24}\cdot\frac{1}{16}\cdot\frac{1}{16}+2\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\\
& \Big(\frac{7}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}+4\cdot\frac{7}{48}\cdot\frac{1}{12}\cdot\frac{1}{16}+\frac{5}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}+\frac{1}{12}\cdot\frac{1}{12}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{24}\cdot\frac{1}{16}\cdot\frac{1}{12}+\\
& 3\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+2\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}+\frac{5}{24}\cdot\frac{1}{16}\cdot\frac{1}{16}\Big)+\Big(\frac{7}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}+4\cdot\frac{7}{48}\cdot\frac{1}{12}\cdot\frac{1}{16}+\\
& \frac{1}{12}\cdot\frac{1}{12}\cdot\frac{1}{12}+\frac{5}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{24}\cdot\frac{1}{16}\cdot\frac{1}{12}+3\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+2\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}+\frac{5}{24}\cdot\frac{1}{16}\cdot\frac{1}{16}\Big)\Big]\\
=\; & \frac{5532}{12\cdot 48^2} \;=\; \frac{461}{48^2}
\end{aligned}
$$

$$
\begin{aligned}
P_0(f_1) \;=\; & 4\Big[\frac{5}{24}\cdot\frac{1}{12}\cdot\frac{1}{16}+\Big(\frac{5}{24}\cdot\frac{1}{16}\cdot\frac{1}{12}+\frac{7}{24}\cdot\frac{1}{16}\cdot\frac{1}{16}+3\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+2\cdot\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\\
& \Big(\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{48}\cdot\frac{1}{12}\cdot\frac{1}{16}+\frac{1}{16}\cdot\frac{1}{12}\cdot\frac{1}{16}+\frac{1}{16}\cdot\frac{1}{12}\cdot\frac{1}{12}\Big)+\\
& \Big(\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}+\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{48}\cdot\frac{1}{12}\cdot\frac{1}{16}+\frac{1}{16}\cdot\frac{1}{12}\cdot\frac{1}{12}+\frac{1}{16}\cdot\frac{1}{12}\cdot\frac{1}{16}\Big)+\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{16}\Big]\\
=\; & \frac{1320}{12\cdot 48^2} \;=\; \frac{110}{48^2}
\end{aligned}
$$

$$
P_0(f_2) \;=\; P_0(f_1) \;=\; \frac{110}{48^2}
$$

$$
\begin{aligned}
P_0(f_3) \;=\; & 4\Big[\frac{5}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}+\Big(\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{16}+\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\Big(\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{16}+\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}\Big)+\Big(\frac{5}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}+\\
& 4\cdot\frac{7}{48}\cdot\frac{1}{12}\cdot\frac{1}{16}+\frac{7}{24}\cdot\frac{1}{12}\cdot\frac{1}{12}+\frac{1}{12}\cdot\frac{1}{12}\cdot\frac{1}{12}\Big)+\Big(\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}+\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{16}\Big)+\frac{1}{12}\cdot\frac{1}{12}\cdot\frac{1}{12}+\\
& \Big(\frac{7}{48}\cdot\frac{1}{16}\cdot\frac{1}{12}+\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{16}\Big)\Big]\\
=\; & \frac{1452}{12\cdot 48^2} \;=\; \frac{121}{48^2}
\end{aligned}
$$

$$
P_0(f_4) \;=\; P_0(f_1) \;=\; \frac{110}{48^2}
$$

$$
P_0(f_5) \;=\; P_0(f_3) \;=\; \frac{121}{48^2}
$$

$$
P_0(f_6) \;=\; 4\Big[4\cdot\frac{1}{16}\cdot\frac{1}{16}\cdot\frac{1}{16}\Big] \;=\; \frac{1}{256} \;=\; \frac{9}{48^2}
$$

$$
P_0(f_7) \;=\; P_0(f_1) \;=\; \frac{110}{48^2}
$$