Interpreting the Ising Model: The Input Matters

Haslbeck, J.M.B.; Epskamp, S.; Marsman, M.; Waldorp, L.J.

Link to publication

Routledge
Taylor & Francis Group

Interpreting the Ising Model: The Input Matters

Jonas M. B. Haslbeck, Sacha Epskamp, Maarten Marsman ⃝, and Lourens J. Waldorp

Psychological Methods Group, University of Amsterdam

ABSTRACT

The Ising model is a model for pairwise interactions between binary variables that has become popular in the psychological sciences. It has been first introduced as a theoretical model for the alignment between positive (1) and negative (−1) atom spins. In many psychological applications, however, the Ising model is defined on the domain {0, 1} instead of the classical domain $\{-1, 1\}$. While it is possible to transform the parameters of the Ising model in one domain to obtain a statistically equivalent model in the other domain, the parameters in the two versions of the Ising model lend themselves to different interpretations and imply different dynamics, when studying the Ising model as a dynamical system. In this tutorial paper, we provide an accessible discussion of the interpretation of threshold and interaction parameters in the two domains and show how the dynamics of the Ising model depends on the choice of domain. Finally, we provide a transformation that allows one to transform the parameters in an Ising model in one domain into a statistically equivalent Ising model in the other domain.

Introduction

The Ising model is a model for pairwise interactions between binary variables that originated in statistical mechanics (Glauber, 1963; Ising, 1925) but is now used in a large array of applications in the psychological sciences (e.g., Afzali et al., 2017; Borsboom & Cramer, 2013; Boschloo et al., 2015; Boschloo, Schoevers, van Borkulo, Borsboom, & Oldehinkel, 2016; Cramer et al., 2016; Dalege et al., 2016; Deserno, Borsboom, Begeer, & Geurts, 2017; Fried et al., 2015; Haslbeck & Fried, 2017; Marsman, Maris, Bechger, & Glas, 2015; Marsman, Tanis, Bechger, & Waldorp, 2019; Rhemtulla et al., 2016; Savi, van der Maas, & Maris, 2018; van der Maas, Kan, Marsman, & Stevenson, 2017).

The original Ising model has been introduced as a model for the interactions between atom spins, which can be positive (1) and negative (−1) (Brush, 1967). In this setting, with variables taking values in the domain $\{-1, 1\}$, the interaction parameters in the Ising model determine the *alignment* between variables: If an interaction parameter between two variables is positive, the two variables tend to take on the same value; on the other hand, if the interaction parameter is negative, the two variables tend to take on different values.

In most psychological applications, however, the Ising model is defined with variables taking values in the domain {0, 1}. While it is possible to transform the parameters of a given Ising model in one domain to obtain a statistically equivalent model in the other domain, the parameters in the two versions of the Ising model lend themselves to different interpretations and imply different dynamics, when studying the Ising model as a dynamical system. If unaware of those subtle differences, one might erroneously apply theoretical results from the $\{-1, 1\}$ domain to a model estimated in the {0,1} domain, or simply interpret parameters incorrectly. To prevent such confusion in the emerging psychological networks literature which makes heavy use of the Ising model, we provide a detailed discussion of both versions of the Ising model in the present tutorial paper.

We begin by discussing the different interpretations of the Ising model in the $\{-1, 1\}$ and {0, 1} domain in the Section "Different domain, different interpretation" using a simple example with two variables which allows the reader to follow all calculations while reading. We explain the differences in the interpretation of the threshold and interaction parameters in the two versions of the Ising model, and discuss in which situation

which version might be more appropriate. While most psychological applications of the Ising model use it as a statistical model, it has also been studied as a dynamical system in psychological research (e.g., Cramer et al., 2016; Dalege et al., 2016; Lunansky, van Borkulo, & Borsboom, 2019). In the Section "Different domain, different dynamics" we discuss how the dynamics of the Ising model depends on the choice of domain, and show that the domain changes the *qualitative* behavior of the model. Finally, in the Section "Transforming from $\{-1, 1\}$ to $\{0, 1\}$ and vice versa" we provide a transformation that allows one to transform the parameters in an Ising model in one domain into a statistically equivalent Ising model in the other domain.

Different domain, different interpretation

In this section we estimate an Ising model with $p = 2$ variables in both domains, $\{-1, 1\}$ and $\{0, 1\}$, and show that the resulting threshold and interaction parameters have different values and lend themselves to different interpretations. We choose the $p = 2$ variable case to make the explanation as accessible as possible. However, all results immediately extend to the general situation with p variables. The Ising model for two variables is given by

$$P(y_1, y_2) = \frac{1}{Z} \exp\{\alpha_1 y_1 + \alpha_2 y_2 + \beta_{12} y_1 y_2\}, \qquad (1)$$

where y_1, y_2 are either elements of $\{-1, 1\}$ or $\{0, 1\}$, $P(y_1, y_2)$ is the probability of the event (y_1, y_2), $\alpha_1, \alpha_2, \beta_{12}$ are parameters in $\mathbb{R}$, and Z is a normalization constant which denotes the sum of the exponent across all possible states. There are $2^p = 4$ states in this example.

To illustrate the differences across models, we generate $n = 1000$ samples of the labels A, B with the relative frequencies shown in Table 1. In applications, the labels A, B can stand for any pair of categories such as being for or against something, some event having happened or not, or a symptom being present or not. The two domains are two different ways to numerically represent these labels.

We obtain the Maximum Likelihood Estimates (MLE) of the parameters in two different ways: Once, by filling in $\{-1, 1\}$ for $\{A, B\}$; and once by filling in $\{0, 1\}$ for $\{A, B\}$. Figure 1 summarizes the two resulting models. The first column in Figure 1 shows the parameter estimates α_1, α_2,
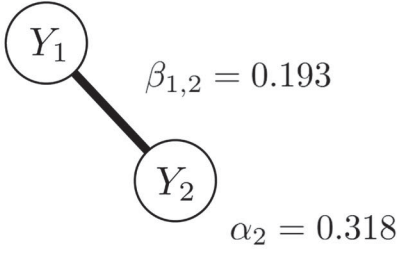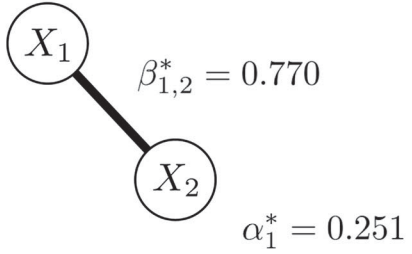
Table 1. Relative frequency of states in the example data set.

$$
\begin{array}{cc}
 & A \quad\ B \\
\begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 0.14 & 0.18 \\ 0.18 & 0.50 \end{pmatrix}
\end{array}
$$

and β_{12}, and log potentials in domain $\{-1, 1\}$. We first focus on the interpretation of the interaction parameter β_{12}. To understand the interpretation of this parameter we take a look at the log potentials for all of the four possible states $\{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$, which we obtain by plugging the four states into the expression within the exponential in Equation (1). The resulting log potentials are displayed in the second row in Figure 1 and show us the following: If β_{12} becomes larger, the probability of the states $(-1, -1), (1, 1)$ increases relative to the probability of the states $(-1, 1), (1, -1)$. This means that the interaction parameter determines the degree of *alignment* of two variables. That is, if $\beta_{12} > 0$ the *same* labels align with each other, and if $\beta_{12} < 0$ *opposite* labels align with each other. In other words, β_{12} models the probability of the states $(-1, -1), (1, 1)$ relative to the probability of the states $(-1, 1), (1, -1)$.

This is not the case in the $\{0, 1\}$ domain. The second column in Figure 1 shows that the parameter estimates $\alpha_1^*, \alpha_1^*, \beta_{12}^*$ in domain $\{0, 1\}$, and we see that they have different values than in the $\{-1, 1\}$ domain. To understand why this is the case, we again look at the interpretation of the interaction parameter β_{12}^* by inspecting the four log potentials. The key observation is that β_{12}^* only appears in the log potential of the state $(1, 1)$. What happens if β_{12}^* increases? Then the probability of the state $(1, 1)$ increases relative to the probability of all other states $(0, 1), (1, 0), (0, 0)$. In other words, β_{12}^* models the probability of state $(1, 1)$ relative to the probability of the states $(0, 1), (1, 0), (0, 0)$.

Next, we turn to the interpretation of the threshold parameters. If all interaction parameters are equal to zero, the threshold parameters in both domains indicate the tendency of a variable to be in one state or the other. That is, $\alpha, \alpha^* > 0$ implies a larger probability for the states $(1) \in \{0, 1\}, (1) \in \{-1, 1\}$ than for states $(0) \in \{0, 1\}, (-1) \in \{-1, 1\}$. If $\alpha, \alpha^* < 0$ the reverse is true, and if $\alpha, \alpha^* = 0$, the corresponding states have both probability 0.5. However, in the general case in which interaction parameters are allowed to be nonzero, the interpretation depends on the domain: In the $\{-1, 1\}$ domain the threshold parameter indicates the tendency of a variable averaged over all possible states of all other variables. In more formal terms, the threshold parameter of a given variable indicates the marginal mean of that variable. In contrast, the threshold in the $\{0, 1\}$ domain indicates the tendency of a variable when all other variables are equal to zero. We return to the different interpretations of thresholds in the Section "Different domain, different dynamics", in which we discuss the dynamics of the Ising model.

Domain	$y \in \{-1, 1\}$	$x \in \{0, 1\}$
Model Parameters	$\alpha_1 = 0.318$ Y_1 $\beta_{1,2} = 0.193$ Y_2 $\alpha_2 = 0.318$	$\alpha_2^* = 0.251$ X_1 $\beta_{1,2}^* = 0.770$ X_2 $\alpha_1^* = 0.251$
Log Potentials	$(Y_1 = -1, Y_2 = -1) : -\alpha_1 - \alpha_2 + \beta_{1,2}$ $(Y_1 = 1, Y_2 = -1) : \alpha_1 - \alpha_2 - \beta_{1,2}$ $(Y_1 = -1, Y_2 = 1) : -\alpha_1 + \alpha_2 - \beta_{1,2}$ $(Y_1 = 1, Y_2 = 1) : \alpha_1 + \alpha_2 + \beta_{1,2}$	$(X_1 = 0, X_2 = 0) : 0$ $(X_1 = 1, X_2 = 0) : \alpha_1^*$ $(X_1 = 0, X_2 = 1) : \alpha_2^*$ $(X_1 = 1, X_2 = 1) : \alpha_1^* + \alpha_2^* + \beta_{1,2}^*$

Figure 1. The threshold and interaction parameters estimated from the data generated from Table 1, and the log potentials for each combination of states, separately for the two domains $\{-1, 1\}$ and $\{0, 1\}$. The log potentials (also called energy function or Hamiltonian) are obtained by filling each state (e.g., $y_1 = -1, y_2 = -1$) in the expression within the exponential in Equation (1).

In this section, we showed that depending on its domain, the parameters of the Ising model have different interpretations. What are the consequences for applied researchers? In terms of reporting, it is important to state which domain has been used such that the reported model can be re-used in the correct way: If someone reports a set of parameters estimated from in the $\{0, 1\}$ domain, and a reader applies it to the $\{-1, 1\}$ domain they will obtain the incorrect probabilities. Note that in order to use the model one also has to report the threshold parameters. Not reporting the threshold parameters is a common problem and irrespective of the issue discussed in this paper. The only situation in which the domain does not matter is if the only goal is to compare the relative size of interaction parameters since the relative size is the same in both domains (see Section "Transforming from $\{-1, 1\}$ to $\{0, 1\}$ and vice versa").

The second consequence is that researchers have to choose which version of the Ising model is more appropriate for the phenomenon at hand. The above clarified interpretations of the Ising model in its two different domains allow to take this decision. For example, the $\{-1, 1\}$ parameterization may be more plausible for labels that are not qualitatively different, but rather opposing each other in some way such as supporting or opposing a certain viewpoint, for example agreeing or

disagreeing with a statement like "Elections should be held every two years instead of every four years." This also reflects the origin of the Ising model as a model for atom spins, which are either positive or negative. The parameterization implied by $\{0, 1\}$ could be more appropriate if the two labels are qualitatively different, such as the presence or absence of an event or a characteristic. Take psychiatric symptoms as an example: While it seems plausible that *fatigue* leads to *lack of concentration*, it is less clear whether the absence of *fatigue* also leads to the increase of *concentration*. In such a case, we can encode the possible belief that the absence of something cannot have an influence on something else by choosing the $\{0, 1\}$ domain. Importantly, the decision of which version to pick has to be based on information beyond the data, because the models are statistically equivalent and therefore cannot be distinguished by observational data. In Appendix A we prove this equivalence for the example shown in Figure 1.

While Ising models in psychological research are usually fit to cross-sectional data, one is typically interested in within-subjects dynamics. In this context, one is often interested in inferring from an estimated Ising model how to best intervene on the system. In the next section we will show how the dynamics of

the Ising model depends on its domain, and that the different versions of the Ising model make different predictions for optimal interventions.

Different domain, different dynamics

In this section we show that the domain also determines the dynamics of the Ising model, when studying as a dynamical system describing within-person dynamics. The dynamical version of the Ising model is initialized by p initial values at $t = 1$, and then each variable at time t is a function of all variables it is connected to via a nonzero interaction parameter at $t - 1$.[1] An often studied characteristic in this model is how its behavior changes when the size of the interaction parameters increases. A typical behavior of interest is the number of variables in state (1) (e.g., Cramer et al., 2016; Dalege et al., 2016).

Which behavior would we expect in the two domains $\{-1, 1\}$ and $\{0, 1\}$? From the previous section we know that in domain $\{-1, 1\}$, the interaction parameter β_{ij} models the probability of states $\{(-1, -1), (1, 1)\}$ relative to the states $\{(-1, 1), (1, -1)\}$. Now, when increasing all β_{ij}, connected variables become more synchronized, which means that all (connected) variables tend to be either *all* in state (-1) or (1). In terms of number of variables in state (1), we would therefore predict that the expected number of variables in state (1) remains unchanged, because the states (-1) and (1) occur equally often in the aligned $((-1, -1)$ and $(1, 1))$ and not aligned $((-1, 1)$ and $(1, -1))$ states. And second, we predict that the probability that at a given time point either all variables are in state (-1), or all variables are in state (1), increases. The reason is that, in the $(-1, 1)$ domain, the larger the interaction parameter, the stronger the alignment between variables. This second prediction implies that the variance of the number of states in (1) increases.

In the domain $\{0, 1\}$, the interaction parameter β_{ij}^* models the probability of the state $(1, 1)$ relative to the remaining three states $\{(0, 1), (1, 0), (0, 0)\}$. Now, when increasing β_{ij}^*, connected variables will have a higher probability to be all in state 1. Importantly, the frequency of 1s in the high probability state $(1, 1)$ is higher than in the other three states. We therefore expect that the number of variables in state (1) increases and that the probability that all variables are in state (1) increases. The second prediction implies that the variance of the number of states in (1) decreases.

We prove that the expected number of variables in state (1) remains unchanged for $\{-1, 1\}$ and increases for $\{0, 1\}$, if $\beta_{ij} > 0$ for the case $p = 2$ variables in Appendix B. Here, we show via simulation that our predictions are correct. We sample $n = 10^6$ observations from a fully connected (i.e., all interaction parameters are nonzero) Ising model with $p = 10$ variables in which all edge weights (interaction parameters) have the same size and all thresholds are set to zero. We vary both the size of the interaction parameters $\beta_{ij} \in \{0, .1, .2\}$ and the domain.[2] Figure 2 shows the distribution (over time steps) of the number of variables that are in state $\{1\}$.

The first row of Figure 2 shows the distribution of the number of variables in state (1) across time when all interaction parameters are equal to zero. We see a symmetric, unimodal distribution with mean 5 for both domains. This is what we would expect since the probability of each variable being in state (1) can be seen as an unbiased (because the thresholds are zero) coin flip that is independent of all other variables. Thus, since we have 10 variables, the means are equal to $10 \times 0.5 = 5$.

However, when increasing the interaction parameter from 0 to 0.1 (second row) the distributions become different: In domain $\{-1, 1\}$ the mean remains unchanged and the probability mass shifts from around 5 to more extreme values, resulting in increased variance. In contrast, in domain $\{0, 1\}$ the distribution shifts to the right, which implies that the mean increases and the variance slightly decreases. When further increasing the interaction parameters to 0.2 (third row), in domain $\{-1, 1\}$ most of the probability mass is concentrated on 0 and 10, while leaving the mean unchanged; in domain $\{0, 1\}$ the mean further increases and the variance further decreases. From a dynamical perspective, this means that for strongly connected Ising models (with thresholds equal to zero) the domain $\{-1, 1\}$ implies two stable states (all variables in state (-1) or (1)), while the domain $\{0, 1\}$ implies only a single stable (all variables in state (1)), whose position depends on whether interaction parameters are positive or negative. This means that the dynamic Ising model in the $\{-1, 1\}$ can switch between stable states, while $\{0, 1\}$ it always stays in the same stable state.[3]

For the general case of Ising models that are not fully connected and also have negative interaction

---

[1] Glauber dynamics (Glauber, 1963) describe a different way to sample from a dynamic Ising model. The qualitative results presented in this section also hold for Glauber dynamics.

[2] The code to reproduce the simulation and Figure 2 is available at http://github.com/jmbh/IsingVersions.

[3] The result about bistability is true for the considered fully connected Ising model with zero thresholds. It is also possible to construct a bistable Ising model in the $\{0, 1\}$ domain by choosing large negative thresholds and large positive interaction parameters. The relationship between mean/variance and changing the interaction parameter in the two domains, however, is always true.

parameters, the results described above extend to local clusters of two or more variables: In the domain $\{-1, 1\}$, increasing the interaction parameter will leave the means of all variables in the cluster unchanged, however, the variables become increasingly aligned (if interaction parameters are positive) or disaligned (if interaction parameters are negative). Alignment will lead to an increase in variance, while disalignment will lead to a decrease in variance. In contrast, in the $\{0, 1\}$ domain the mean of all variables in the cluster will increase in the case of positive interaction parameters, and decrease in the case of negative interaction parameters.

This shows that, depending on which domain is used one can come to entirely different conclusions about the dynamics of the Ising model. For example, Cramer et al. (2016) model the interactions between psychiatric symptoms with an Ising model in domain $\{0, 1\}$ and conclude that densely connected Ising models imply a larger number of active (in state (1)) symptoms and therefore represent "pathological" models. The above argument and simulation show
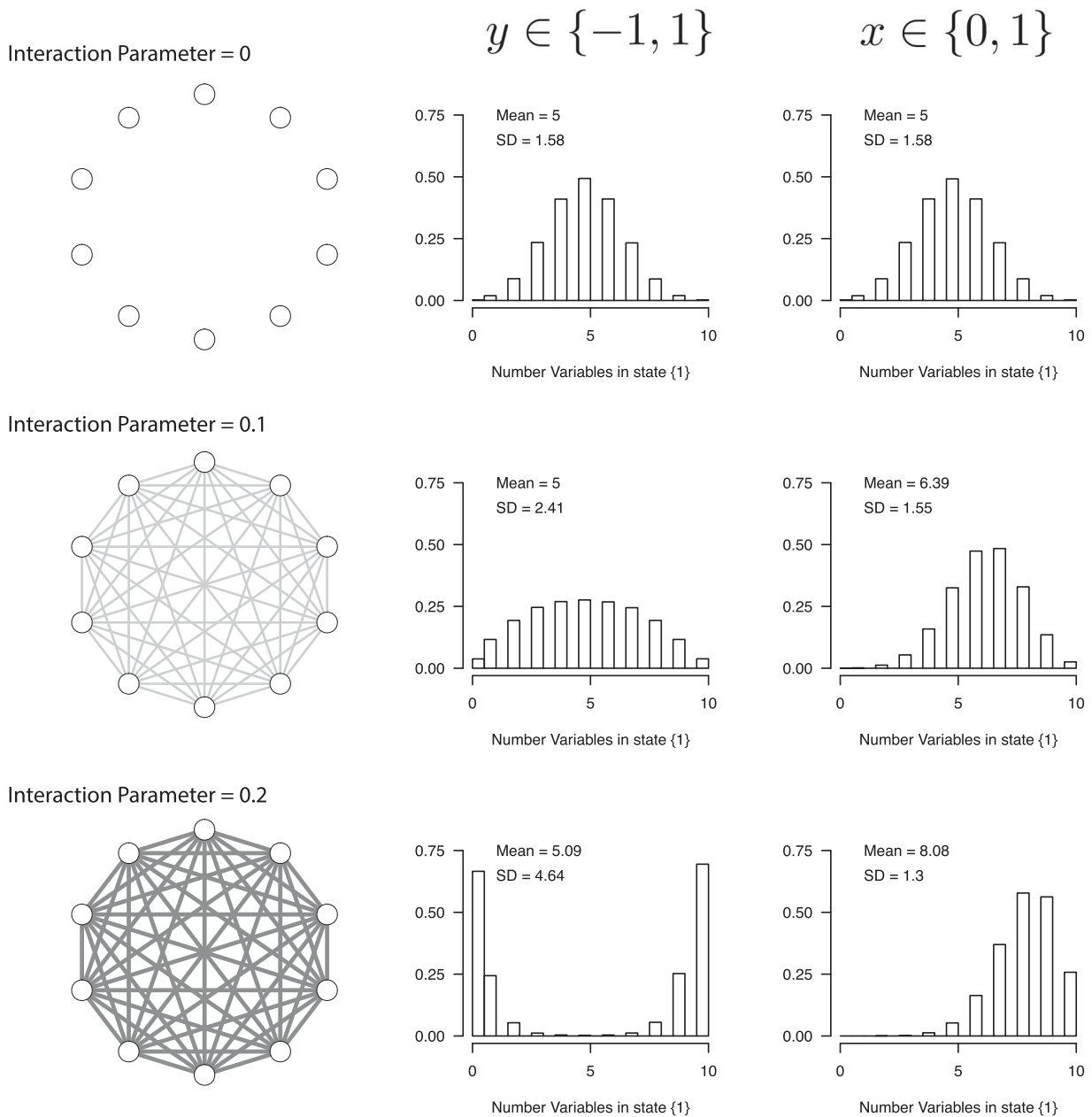
Figure 2. The distribution of the number of variables being in state one as a function of the size of the interaction parameter in a fully connected Ising model $\{0, .1, .2\}$ and used domain ($\{-1, 1\}$ and $\{0, 1\}$) of the random variable.

that this is only true when using the {0, 1} domain, which encodes the belief that the absence of a symptom cannot influence the absence of another symptom. If one decides that an alignment between variables is a more plausible interaction (as implied by the $\{-1, 1\}$ domain), then densely connected Ising models do not imply a large number of active symptoms. Instead, high density implies high variance and two stable states. Thus, the characterization of dense networks as pathological networks as in Cramer et al. (2016) hinges on choosing the {0, 1} domain.

This has important consequences: When choosing the {0, 1} domain, we would conclude that highly connected symptom networks are necessarily "bad", and interventions on the interactions between symptoms as suggested by Borsboom (2017) should always reduce symptom activation. On the other hand, in the $\{-1, 1\}$ domain highly connected symptom networks are not necessarily bad, but in fact can lead to high resilience, if the threshold parameters are large negative values. In such a situation strong interactions would keep the system in a state in which all symptoms are deactivated.

Transforming from $\{-1, 1\}$ to {0, 1} and vice versa

The Ising model is typically estimated by a sequence of p logistic regressions, which require the domain {0, 1}. However, the previous sections showed that in some situations the domain $\{-1, 1\}$ may be more appropriate. In Table 2, we present a transformation that allows one to obtain the parameterization based on domain $\{-1, 1\}$ from the parameterization based on domain {0, 1} and vice versa (see Appendix C for the derivation of the transformations). We define $\beta_{i+}^* = \sum_{j=1}^{p} \beta_{ij}^*$ as the sum over the interaction parameters associated with a given variable y_i.

Table 2 shows that the interaction parameters β_{ij} in the $\{-1, 1\}$ domain are four times smaller than the interaction parameters β_{ij}^* in the {0, 1} domain. We also see that the threshold parameter α_i is a function of both the threshold and the interaction parameters α_i^*, β_{ij}^* in the other parameterization.

Table 2. Transformation functions to obtain the threshold and interaction parameters of one parameterization from the threshold and interaction parameters of the other parameterization.

Transformation	Thresholds	Interactions
$\{0, 1\} \Rightarrow \{-1, 1\}$	$\alpha_i = \frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*$	$\beta_{ij} = \frac{1}{4}\beta_{ij}^*$
$\{-1, 1\} \Rightarrow \{0, 1\}$	$\alpha_i^* = 2\alpha_i - 2\beta_{i+}$	$\beta_{ij}^* = 4\beta_{ij}$

Parameters with asterisk indicate parameters in the {0, 1} domain.

We now apply the transformations in Table 2 to the $p = 2$ variable example in Figure 1. We choose to transform from {0, 1} to $\{-1, 1\}$:

$$a_1 = \frac{1}{2}a_1^* + \frac{1}{4}\beta_{i+}^* = \frac{0.251}{2} + \frac{0.77}{4} = 0.318$$
$$\beta_{12} = \frac{1}{4}\beta_{12}^* = \frac{.77}{4} = 0.1925 \approx 0.193$$

And indeed, we obtain the parameters obtained when estimating the Ising model in the $\{-1, 1\}$ domain (see first column in Figure 1).

From the transformation in Table 2 follows that the two models are statistically equivalent. This implies that one could also estimate the model in the $\{-1, 1\}$ domain, transform the parameters, and would obtain the parameters one would have obtained from estimating in the {0, 1} domain. Also, note that the standard errors of estimates are subject to the same transformation, and therefore one always reaches the same conclusion regarding statistical significance in both domains.

However, note that one does not necessarily arrive at statistical equivalent models when estimating in the two different domains using biased estimators. An example of a biased estimation method is the popular ℓ_1-regularized estimator (van Borkulo et al., 2015). We discuss why statistical equivalence is not guaranteed in this specific example in Appendix D. The possibility that different domains lead to models that are not statistically equivalent highlights the importance of choosing the most plausible Ising model on substantive grounds.

Conclusions

In this article, we have investigated the subtleties in choosing the domain of the Ising model. We showed that estimating the Ising model in the domains {0, 1} and $\{-1, 1\}$ results in parameters with different values and different interpretations. We also showed that the qualitative behavior of the dynamical Ising model depends on the chosen domain. Finally, we provided a transformation that explains the relation between the two parameterizations and allows one to obtain one from the other. This is useful in practice, because typically used software packages require the {0, 1} domain. This transformation also implies that the two parameterizations are statistically equivalent, which means that one cannot choose one over the other on empirical grounds. Thus, researchers should carefully reflect on which interactions between variables are plausible and choose the domain accordingly.

Article information

ORCID

Maarten Marsman (iD) http://orcid.org/0000-0001-5309-7502

References

Afzali, M. H., Sunderland, M., Batterham, P. J., Carragher, N., Calear, A., & Slade, T. (2017). Network approach to the symptom-level association between alcohol use disorder and posttraumatic stress disorder. *Social Psychiatry and Psychiatric Epidemiology*, *52*(3), 329–339. doi:10.1007/s00127-016-1331-3

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. doi:10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. doi:10.1146/annurev-clinpsy-050212-185608

Boschloo, L., Schoevers, R. A., van Borkulo, C. D., Borsboom, D., & Oldehinkel, A. J. (2016). The network structure of psychopathology in a community sample of preadolescents. *Journal of Abnormal Psychology*, *125*(4), 599–606. doi:10.1037/abn0000150

Boschloo, L., van Borkulo, C. D., Rhemtulla, M., Keyes, K. M., Borsboom, D., & Schoevers, R. A. (2015). The network structure of symptoms of the diagnostic and statistical manual of mental disorders. *PLoS One*, *10*(9), e0137621. doi:10.1371/journal.pone.0137621

Brush, S. G. (1967). History of the Lenz-Ising model. *Reviews of Modern Physics*, *39*(4), 883–893. doi:10.1103/RevModPhys.39.883

Cramer, A. O., van Borkulo, C. D., Giltay, E. J., van der Maas, H. L., Kendler, K. S., Scheffer, M., & Borsboom, D. (2016). Major depression as a complex dynamic system. *PLoS One*, *11*(12), e0167490. doi:10.1371/journal.pone.0167490

Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The causal attitude network (can) model. *Psychological Review*, *123*(1), 2–22. doi:10.1037/a0039802

Deserno, M. K., Borsboom, D., Begeer, S., & Geurts, H. M. (2017). Multicausal systems ask for multicausal approaches: A network perspective on subjective well-being in individuals with autism spectrum disorder. *Autism*, *21*(8), 960–971. doi:10.1177/1362361316660309

Foygel, R., & Drton, M. (2010). Extended Bayesian information criteria for Gaussian graphical models. *Advances in Neural Information Processing Systems*, *23*, 604–612.

Fried, E. I., Bockting, C., Arjadi, R., Borsboom, D., Amshoff, M., Cramer, A. O. J., … Stroebe, M. (2015). From loss to loneliness: The relationship between bereavement and depressive symptoms. *Journal of Abnormal Psychology*, *124*(2), 256–265. doi:10.1037/abn0000028

Glauber, R. J. (1963). Time-dependent statistics of the Ising model. *Journal of Mathematical Physics*, *4*(2), 294–307. doi:10.1063/1.1703954

Haslbeck, J., & Fried, E. I. (2017). How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. *Psychological Medicine*, *47*(16), 2767–2776. doi:10.1017/S0033291717001258

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, *31*(1), 253–258. doi:10.1007/BF02980577

Lunansky, G., van Borkulo, C. D., & Borsboom, D. (2019). Personality, resilience, and psychopathology: A model for the interaction between slow and fast network processes in the context of mental health. https://www.researchgate.net/publication/333292744_Personality_resilience_and_psychopathology_A_model_for_the_interaction_between_slow_and_fast_network_processes_in_the_context_of_mental_health

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(1), 9050. doi:10.1038/srep09050

Marsman, M., Tanis, C., Bechger, T., & Waldorp, L. (2019). Network psychometrics in educational practice. In B. Veldkamp, & C. Sluijter (Eds.), *Theoretical and practical advances in computer-based educational measurement* (pp. 93–120). Cham: Springer.

Ravikumar, P., Wainwright, M. J., & Lafferty, J. D. (2010). High-dimensional Ising model selection using l1-regularized logistic regression. *The Annals of Statistics*, *38*(3), 1287–1319. doi:10.1214/09-AOS691

Rhemtulla, M., Fried, E. I., Aggen, S. H., Tuerlinckx, F., Kendler, K. S., & Borsboom, D. (2016). Network analysis of substance abuse and dependence symptoms. *Drug and Alcohol Dependence*, *161*, 230–237. doi:10.1016/j.drugalcdep.2016.02.005

Savi, A., van der Maas, H., & Maris, G. (2018). The wiring of intelligence. *Perspectives on Psychological Science*, *14*, 1034–1061.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. doi:10.1214/aos/1176344136

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2015). A new method for constructing networks from binary data. *Scientific Reports*, *4*(1), 5918. doi:10.1038/srep05918

Van Der Maas, H., Kan, K.-J., Marsman, M., & Stevenson, C. E. (2017). Network models for cognitive development and intelligence. *Journal of Intelligence*, *5*(2), 16. doi:10.3390/jintelligence5020016v

Appendix A. Statistical Equivalence worked out for two variable example

Here we show that the two models shown in Figure 1 are statistically equivalent. Two models statistically equivalent if they output the same probability for any of states on which the models are defined.

We begin with the model estimated on the domain $\{-1, -1\}$. We first compute the potentials for the four states $\{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$:

$$\exp \{0.318(-1) + 0.318(-1) + 0.193(-1)(-1)\} = 0.6415304$$

$$\exp \{0.318(-1) + 0.318(1) + 0.193(-1)(1)\} = 0.8248249$$

$$\exp \{0.318(1) + 0.318(-1) + 0.193(1)(-1)\} = 0.8248249$$

$$\exp \{0.318(1) + 0.318(1) + 0.193(1)(1)\} = 2.29118$$

and then the normalization constant

$$Z = 0.6415304 + 0.8248249 + 0.8248249 + 2.29118$$

$$= 4.58236$$

We divide the potentials by Z and obtain the probabilities

$$P(Y_1 = -1, Y_2 = -1) = \frac{0.6415304}{Z} = 0.14$$

$$P(Y_1 = -1, Y_2 = 1) = \frac{0.8248249}{Z} = 0.18$$

$$P(Y_1 = 1, Y_2 = -1) = \frac{0.8248249}{Z} = 0.18$$

$$P(Y_1 = 1, Y_2 = 1) = \frac{2.29118}{Z} = 0.5$$

We now repeat the same with domain $\{0, 1\}$ and first compute the potentials for the states $\{(0, 0), (0, 1), (1, 0), (1, 1)\}$:

$$\exp \{0.251(0) + 0.251(0) + 0.77(0)(0)\} = 1$$

$$\exp \{0.251(0) + 0.251(1) + 0.77(0)(1)\} = 1.285714$$

$$\exp \{0.251(1) + 0.251(0) + 0.77(1)(0)\} = 1.285714$$

$$\exp \{0.251(1) + 0.251(0) + 0.77(1)(0)\} = 3.571429$$

and then the normalization constant

$$Z = 1 + 1.285714 + 1.285714 + 3.571429 = 7.142857$$

We divide the potentials by Z and obtain the probabilities

$$P(X_1 = 0, X_2 = 0) = \frac{1}{Z} = 0.14$$

$$P(X_1 = 0, X_2 = 1) = \frac{1.285714}{Z} = 0.18$$

$$P(X_1 = 1, X_2 = 0) = \frac{1.285714}{Z} = 0.18$$

$$P(X_1 = 1, X_2 = 1) = \frac{3.571429}{Z} = 0.5$$

We see that both models predict the same probabilities and are therefore statistically equivalent.

Appendix B. Increasing interaction parameters only changes the marginal probabilities domain in {0, 1}

Here we show that for an Ising model with $p = 2$ variables with $\alpha_1, \alpha_2 = 0$ and $\beta_{12} > 0$ it holds that

$$P(X_1 = -1) = P(X_2 = -1) = P(X_2 = 1) = P(X_2 = 1) \quad (2)$$

for the domain $\{-1, 1\}$, and that

$$P(X_1 = 0) = P(X_2 = 0) < P(X_1 = 1) = P(X_2 = 1) \quad (3)$$

for the domain $\{0, 1\}$.

We first show (2). We assume $\alpha_1, \alpha_2 = 0$ and $\beta_{12} > 0$. Then the Ising model is given by

$$P(X_1, X_2) = \frac{1}{Z} \exp \{\alpha_1 X_1 + \alpha_2 X_2 + \beta_{12} X_2 X_1\}$$

$$= \frac{1}{Z} \exp \{\beta_{12} X_2 X_1\},$$

where Z is the normalizing constant summing over all $2^p = 4$ states. We calculate the probability of the four possible states:

$$P(X_1 = 1, X_2 = -1) = \frac{1}{Z} \exp \{-\beta_{12}\},$$

$$P(X_1 = 1, X_2 = 1) = \frac{1}{Z} \exp \{\beta_{12}\},$$

$$P(X_1 = -1, X_2 = -1) = \frac{1}{Z} \exp \{\beta_{12}\},$$

$$P(X_1 = -1, X_2 = 1) = \frac{1}{Z} \exp \{-\beta_{12}\}.$$

And average over the state of X_2 to obtain the marginals probabilities $P(X_1)$:

$$P(X_1 = 1) = P(X_1 = 1, X_2 = -1) + P(X_1 = 1, X_2 = 1)$$

$$= \frac{1}{Z} \exp \{-\beta_{12}\} + \frac{1}{Z} \exp \{\beta_{12}\} P(X_1 = -1) = P(X_1 = -1, X_2$$

$$= -1) + P(X_1 = -1, X_2 = 1) = \frac{1}{Z} \exp \{\beta_{12}\} + \frac{1}{Z} \exp \{-\beta_{12}\}$$

We see that $P(X_1 = 1) = P(X_1 = -1)$. By symmetry the same is true for X_2, which proves our claim.

We next prove (3). We again assume $\alpha_1, \alpha_2 = 0$ and $\beta_{12} > 0$ and calculate the probabilities of the four possible states:

$$P(X_1 = 1, X_2 = 0) = \frac{1}{Z} \exp \{0\},$$

$$P(X_1 = 1, X_2 = 1) = \frac{1}{Z} \exp \{\beta_{12}\},$$

$$P(X_1 = 0, X_2 = 0) = \frac{1}{Z} \exp \{0\},$$

$$P(X_1 = 0, X_2 = 1) = \frac{1}{Z} \exp \{0\}.$$

The marginal probabilities $P(X_1)$ are:

$$P(X_1 = 1) = P(X_1 = 1, X_2 = 0) + P(X_1 = 1, X_2 = 1)$$

$$= \frac{1}{Z} \exp\{0\} + \frac{1}{Z} \exp\{\beta_{1,2}\} P(X_1 = 0) = P(X_1 = 0, X_2 = 0)$$

$$+ P(X_1 = 0, X_2 = 1) = 2\frac{1}{Z} \exp\{0\}$$

Since $\exp\{\beta_{12}\} > \exp\{0\}$, we have $P(X_1 = 1) > P(X_2 = 1)$, if $\beta_{12} > 0$. By symmetry the same is true for X_2, which proves our claim.

Note that if we assume $\beta_{12} < 0$, (2) holds again for $\{-1, 1\}$, while for $\{0, 1\}$ we have

$$P(X_1 = 0) = P(X_2 = 0) > P(X_2 = 1) = P(X_2 = 1)$$

instead.

Appendix C. Derivation of transformation from {0, 1} to $\{-1, 1\}$ and vice versa

In this section, we first introduce the Ising model for p variables with domain $\{-1, 1\}$, which is the domain used in physics applications. Next, we introduce the Ising model for p variables with domain $\{0, 1\}$, which is mostly used in the statistics literature. We connect both models by deriving a formula of the parameters of one parameterization as a function of the parameters of the other parameterization. This allows us to transform the parameterization based on domain $\{-1, 1\}$ into the parameterization of domain $\{0, 1\}$ and vice versa.

In the physics domain, variables can take on values in $\{-1, 1\}$. The probability distribution of the Ising model for p such random variables is specified by

$$p(y) = \frac{\exp\left(\sum_{i=1}^{p} \alpha_i y_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij} y_i y_j\right)}{\sum_y \exp\left(\sum_{i=1}^{p} \alpha_i y_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij} y_i y_j\right)}, \quad (4)$$

where $y, y \in \{-1, 1\}^p$, denotes a configuration of the p random variables, and the sum $\sum_y$ in the denominator denotes a sum that ranges over all 2^p possible configurations or realizations of y.

From a statistical perspective, the Ising model is a model that is completely determined by the spin variables' main effects and their pairwise interactions. A spin variable in the network tends to have a positive value ($y_i = 1$) when its main effect is positively valued ($\alpha_i > 0$), and tends to have a negative value ($y_i = -1$) when its main effect is negatively valued ($\alpha_i < 0$). Furthermore, any two variables y_i and y_j in the network tend to align their values when their interaction effect is positive ($\beta_{ij} > 0$), and tend to be in different states when their interaction effect is negative ($\beta_{ij} < 0$).

In statistical applications, the Ising model is typically used to describe the probability distribution of p binary random variables,

$$p(x) = \frac{\exp\left(\sum_{i=1}^{p} \alpha_i^* x_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* x_i x_j\right)}{\sum_x \exp\left(\sum_{i=1}^{p} \alpha_i^* x_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* x_i x_j\right)}, \quad (5)$$

where $x, x \in \{0, 1\}^p$, denotes a configuration of the p binary random variables, and again we use $\sum_x$ to denote the sum that ranges over all 2^p possible configurations or realizations of x.

Even though the model is again completely determined by main effects and pairwise interactions, its interaction parameters β^* carry a different interpretation than the interaction parameters of the Ising model for variables Y in the $\{-1, 1\}$ domain. Here, two binary variables x_i and x_j in the network tend to both equal one ($x_i x_j = 1$) when their interaction effect is positive ($\beta_{ij}^* > 0$), but their product tends to equal zero ($x_i x_j = 0$) when their interaction effect is negative ($\beta_{ij}^* < 0$). That is, whenever the interaction between two binary variables x_i and x_j in the network is negative ($\beta_{ij} < 0$), they tend to be in one of the states $\{0, 0\}, \{0, 1\}$ or $\{1, 0\}$.

Despite the different interpretations of the two Ising model formulations, one can traverse the two specifications by a simple change of variables. To wit, assume that we have obtained an Ising model for p binary variables $p(x)$ and wish to express its solution in terms of the variables in the $\{-1, 1\}$ domain, then we require the change of variables

$$x_i = \frac{1}{2}(y_i + 1) \text{ with inverse relation } y_i = 2x_i - 1. \quad (6)$$

We use this transformation in the distribution of the binary random variables,

$$p(x) = \frac{\exp\left(\sum_{i=1}^{p} \alpha_i^* x_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* x_i x_j\right)}{\sum_x \exp\left(\sum_{i=1}^{p} \alpha_i^* x_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* x_i x_j\right)}$$

$$= \frac{\exp\left(\sum_{i=1}^{p} \alpha_i^* \frac{1}{2}(y_i + 1) + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* \frac{1}{2}(y_i + 1)\frac{1}{2}(y_j + 1)\right)}{\sum_y \exp\left(\sum_{i=1}^{p} \alpha_i^* \frac{1}{2}(y_i + 1) + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* \frac{1}{2}(y_i + 1)\frac{1}{2}(y_j + 1)\right)} = p(y),$$

$$(7)$$

and observe that this transformation affects both main effects and pairwise interactions. Working out the sum over pairs of variables, we find

$$\sum_{i=1}^{p-1} \sum_{j>i}^{p} \beta_{ij}^* \frac{1}{2}(y_i + 1)\frac{1}{2}(y_j + 1)$$

$$= \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* (y_i y_j + y_i + y_j + 1)$$

$$= \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* y_i y_j + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* y_i$$

$$+ \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* y_j + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^*$$

$$= \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* y_i y_j + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p}$$

$$\frac{1}{4} \beta_{ij}^* y_i + \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* = \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^* y_i y_j + \sum_{i=1}^{p} \frac{1}{4} \beta_{i+}^* y_i$$

$$+ \sum_{i=1}^{p-1} \sum_{j>i}^{p} \frac{1}{4} \beta_{ij}^*, \quad (8)$$

where the first term reflects pairwise interactions between the variables y, the second term reflects main effects of the

variables with main effect $\beta_{i+}^* = \sum_{j=1}^p \beta_{ij}^*$, and the last term is constant with respect to (w.r.t.) the variables y. Similarly, we can express the sum over the main effects as

$$\sum_{i=1}^p \alpha_i^* \frac{1}{2}(y_i + 1) = \sum_{i=1}^p \alpha_i^* \frac{1}{2} y_i + \sum_{i=1}^p \alpha_i^* \frac{1}{2}, \qquad (9)$$

where the last term is again constant w.r.t. the variables y. Collecting the main effects,

$$\sum_{i=1}^p \frac{1}{2}\alpha_i^* y_i + \sum_{i=1}^p \frac{1}{4}\beta_{i+}^* y_i = \sum_{i=1}^p \left(\frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*\right) y_i, \qquad (10)$$

and constant terms,

$$C = \sum_{i=1}^p \frac{1}{2}\alpha_i^* + \sum_{i=1}^p \sum_{j=1}^p \frac{1}{4}\beta_{ij}^*, \qquad (11)$$

we obtain:

$$p(y) = \frac{\exp\left(\sum_{i=1}^p \left(\frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*\right) y_i + \sum_{i=1}^{p-1}\sum_{j>i}^p \frac{1}{4}\beta_{ij}^* y_i y_j + C\right)}{\sum_y \exp\left(\sum_{i=1}^p \left(\frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*\right) y_i + \sum_{i=1}^{p-1}\sum_{j>i}^p \frac{1}{4}\beta_{ij}^* y_i y_j + C\right)}$$

$$= \frac{\exp\left(\sum_{i=1}^p \left(\frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*\right) y_i + \sum_{i=1}^{p-1}\sum_{j>i}^p \frac{1}{4}\beta_{ij}^* y_i y_j\right)}{\sum_y \exp\left(\sum_{i=1}^p \left(\frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*\right) y_i + \sum_{i=1}^{p-1}\sum_{j>i}^p \frac{1}{4}\beta_{ij}^* y_i y_j\right)}, \qquad (12)$$

which is equal to the Ising model for variables in the $\{-1, 1\}$ domain when we write $\alpha_i = \frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*$ and $\beta_{ij} = \frac{1}{4}\beta_{ij}^*$. In a similar way, one can obtain the parameter values of the binary case from a solution of the Ising model for variables in the $\{-1, 1\}$ domain using $\alpha_i^* = 2\alpha_i - 2\beta_{i+}$ and $\beta_{ij}^* = 4\beta_{ij}$. Thus, we can obtain the binary Ising model parameters α^* and β^* from a simple transformation of the $\{-1, 1\}$ coded Ising model parameters α and β, and vice versa. Table 3 summarizes these transformations:

Table 3. Transformation functions to obtain the threshold and interaction parameters in one parameterization from the threshold and interaction parameters in the other parameterization.

Transformation	α	β
$\{0, 1\} \Rightarrow \{-1, 1\}$	$\alpha_i = \frac{1}{2}\alpha_i^* + \frac{1}{4}\beta_{i+}^*$	$\beta_{ij} = \frac{1}{4}\beta_{ij}^*$
$\{-1, 1\} \Rightarrow \{0, 1\}$	$\alpha_i^* = 2\alpha_i - 2\beta_{i+}$	$\beta_{ij}^* = 4\beta_{ij}$

Parameters with asterisk indicate parameters in the {0, 1} domain.

Appendix D. Model equivalence across domains with penalized estimation

If one estimates the Ising model with an unbiased estimator, one can estimate with domain $\{0, 1\}$ and obtain by transformation the estimates one would have obtained by estimating with domain $\{-1, 1\}$ (and vice versa). In this section we ask whether this is also the case for penalized estimation, which is a popular way to estimate the Ising model (e.g., Ravikumar, Wainwright, & Lafferty, 2010; van Borkulo et al., 2015).

In penalized estimation, the likelihood is maximized with respect to a constraint c, typically on the ℓ_1-norm of the vector of interaction parameters β_{ij}

$$\sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p |\beta_{ij}| < c.$$

Estimation with an ℓ_1-penalty is attractive because it sets small parameter estimates to zero, which makes it easier to interpret the model. The key problem in this setting is selecting an appropriate constraint c. A popular approach is to consider a sequence of candidate constraints $C = \{c_1, ..., c_k\}$ and select the c_i that minimizes the Extended Bayesian Information Criterion (EBIC) (Foygel & Drton, 2010), which extends the BIC (Schwarz, 1978) by an additional penalty (weighted by γ) for the number of nonzero interaction parameters

$$\text{EBIC}_{c_i} = -2LL_{c_i} + s_0 \log n + 4 s_0 \gamma \log p,$$

where LL_{c_i} is the maximized log-likelihood under constraint c_i, s_0 is the number of nonzero interaction parameters, n is the number of observations and p the number of estimated interaction parameters.

We are interested in whether selecting models with this procedure in the two domains, $\{0, 1\}$ and $\{-1, 1\}$, leads to statistically equivalent models. This is indeed the case for the following reason: Assume that c^* minimizes the EBIC for domain $\{0, 1\}$, then from the transformation in Table 2, $\frac{c^*}{4}$ should give the lowest EBIC in domain $\{-1, 1\}$, because the constraint $||\beta^*||_1 < c^*$ on $\{0, 1\}$ is equivalent to the constraint $||\beta^*||_1 < \frac{c^*}{4}$ on Y. Thus, if $\frac{c^*}{4}$ is included in the candidate set C, when estimating in domain $\{-1, 1\}$, two statistically equivalent models should be selected. Note that exactly $\frac{c}{4}$ has to be included, because a slightly larger/smaller constraint can lead to a very different model, if the number of nonzero parameter changes. This nonlinearity arises from the EBIC, in which s_0 decreases by 1 (large change) if some parameter with a tiny value (e.g., 0.0001) is set to zero (small change). Therefore, in order to ensure statistically equivalent models one would need to search a dense sequence C. Clearly, this is unfeasible in practice. This means that, in practice ℓ_1-regularized estimation can return models from domains $\{0, 1\}$ and $\{-1, 1\}$ that are not statistically equivalent. We leave the task of investigating this issue for different estimation algorithms for future research. In what follows we provide an extended version of this argument.

We define:

$$c^* = \arg_{c \in C}\min \text{EBIC}_c$$

$$= \arg_{c \in C}\min -2\log\left[\frac{1}{Z}\prod_{m=1}^n \exp\left\{\sum_{i=1}^p \alpha_i^* X_i + \sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p \beta_{ij}^* X_i X_j\right\}\right]$$

$$+ s_0 \log n + 4 s_0 \gamma \log\left[p(p-1)/2\right],$$

with constraint

$$\sum_{i=1}^p \sum_{\substack{j=1 \\ j \neq i}}^p |\beta_{ij}^*| < c,$$

where s_0 is the number of nonzero interaction parameters, n is the sample size, p is the number of variables $\frac{p(p-1)}{2}$ is the total number of interaction parameters, and γ is a tuning parameter.

Now, we would like to show that if c^* minimizes the EBIC in domain $\{0, 1\}$, then $4c^*$ minimizes the EBIC in $\{-1, 1\}$.

We use the transformation in Table 2 to rewrite the EBIC into the parameterization implied by $\{-1, 1\}$:

$$c^* = \arg_{c \in C} \min \text{EBIC}_c$$

$$= \arg_{c \in C} \min - 2 \log \left[\frac{1}{Z} \prod_{m=1}^{n} \exp \left\{ \sum_{i=1}^{p} \left(\frac{1}{2} \alpha_i^* + \frac{1}{4} \sum_{\substack{j=1 \\ j \neq i}}^{p} \beta_{ij}^* \right) X_i \right. \right.$$

$$\left. \left. + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \frac{1}{4} \beta_{ij}^* X_i X_j \right\} \right] + s_0 \log n + 4 s_0 \gamma \log \left[p(p-1)/2 \right],$$

with constraint

$$\sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} \left| \frac{1}{4} \beta_{ij}^* \right| < c^*.$$

We can rewrite the constraint into

$$\sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} |\beta_{ij}^*| < 4c^*.$$

The last inequality shows that the constraint is four times larger for the parameterization in domain $\{0, 1\}$. Or the other way around, the constraint is $\frac{1}{4}$ times smaller in $\{-1, 1\}$ compared to $\{0, 1\}$.

We know that the models are statistically equivalent across domains. Therefore, the likelihood of the model with constraint c in domain $\{0, 1\}$ is equal to the likelihood of the model with constraint $\frac{c}{4}$ in domain $\{-1, 1\}$. Now, since the transformation never changes a zero estimate in a nonzero estimate or vice versa with probability 1, also the terms $s_0 \log n + 4 s_0 \gamma \log[p(p-1)/2]$ in the EBIC remain constant across domains. It follows that, if $c^* = \arg_{c \in C} \min \text{EBIC}_c$ in domain $\{0, 1\}$, then $\frac{c^*}{4} = \arg_{c \in C} \min \text{EBIC}_c$ in domain $\{-1, 1\}$.