Final Paper for Course
Advanced Quantitative Methods in Political Science

On Using Metropolis-Hasting to Analyze Democratic Support

Tobias Stenzel
tobias.stenzel@students.uni-mannheim.de

Prof. Thomas Gschwend, Ph.D.

Submission Date: Mai 22, 2022

Contents

1   Introduction

In recent years scholars have found declining support for democracy in long-established democracies. (Denemark et al. 2016; Foa and Mounk 2016, 2017; Norris 2017; Voeten 2016). This finding raises two important questions: First, what are the reasons for this decline, and second, what are its implications, i.e., does a decline in democratic support endanger the survival of these democracies?

A series of recent articles (Claassen 2019, 2020a, 2020b) researches these questions and presents novel results. Regarding the first question, Claassen (2020a) finds – in contrary to the the widely held belief of self-reinforcing democracies – that democratic support naturally fluctuates over time. The reasons is that increases in democracy levels lead to decreases in public support and vice versa. Regarding question two, so far the results about the relationship between democratic support and its survival have been mixed. Claassen (2020a), however, finds supporting evidence for the natural theory that democratic support plays a positive role in the system's survival.

A main obstacle for researching democratic support is that current panel data contains a large number of missing values. Claassen (2019) finds an approach to simulate dense panel data from the actual fractured data. The two most recent studies both rely on this data. His approach consists of three steps: First, assume a probabilistic structural model of democratic support. Second, estimate its parameters via the Metropolis Hastings algorithm using the fractured data. Third, simulate the data using the model and the large number of parameter estimates.

The objective of this work is to evaluate the robustness of Claassen's findings towards changes in the method's hyperparameters. In other words, I incorparate the uncertainty about the right choice of hyperparameters into Claassen's results to test their reliability. In particular, I propagate a random set of hyperparameters through Claassen's estimation procedure. This is important because arbitrary hyperparameter choices can potentially lead to false or at least random results. Therefore, the respective parametric uncertainty should always be reported along the point estimate.

I come to the following results:

The final paper is structured as follows: Section 2 reviews the literature of

public support with a focus on Claassen's papers and defines the concept. Section 3 explains Claassen's model of public support, its estimation and my approach for the uncertainty propagation. Section 4 presents the results and Section 5 discusses the findings. Section 6 concludes.

2 Literature Review and Data

2.1 The Concept of Democratic Support

There are two major conceptualizations of public support for democracy (PSD). First, the "implicit" approach that requires the support for broader sociopolitical values like post-materialism and egalitarianism. Here, people support democracy implicitly if they support the values that are framed as particularly democratic. Second, the "explicit" approach that requires both an appraisal of democracy and a rejection of autocratic alternatives. Some studies also refer to PSD measured on the national level as "mood."

2.2 Drivers of Democratic Support

The main theory about why citizen and societies begin to support democracy are called *generational socialization* and *instrumental regime performance*. The first theory assumes that individuals are taught to support the regime under which they are socialized during late adolescence (Mannheim 1970; Niemi 1974). One implication is that after a shift to democracy, the support for it will grow over time (Denemark et al. (2016)). Indeed, several single-country studies have found evidence for this claim, for example for 1970s Germany (Baker et al. (1981)), 1980s Spain (Montero, Gunther, and Torcal (1997)), and 1990s Russia (Mishler and Rose (2007)). One the other hand, other studies do not find such an effect analyzing other Central and Eastern European countries (Mishler and Rose (2002)). Furthermore, more recent studies even find a decline in PSD over the last years (Foa and Mounk 2016, 2017).

Regarding the second theory, instrumental regime performance, PSD rises if the system performs well in terms of instrumental benefits such as economic

2

growth and it declines if the opposite is the case (Dalton 1994; Magalhães 2014). Consequently, the theory suggests that during PSD declines during economic crises. However, there are case-studies that do (Dalton 1994; Magalhães 2014; Mishler and Rose 1996) and that do not (Graham and Sukhtankar (2004)) find this relationship in the data.

Claassen (2020b) offers an alternative theory in transferring the thermostatic model of public opinion and policy (Wlezien (1995)) to democracy and democratic support. In particular, he suggests that there is a negative feedback loop between PSD and democracy so that PSD decreases if democracy supply increases and vice versa. In short, the reasoning is that the output of democratic rights and institutions overshoots the initial desire for them which causes another overcompensation in favor of lower levels of democracy. Moreover, Claassen (2020b) finds different causal channels for electoral and minoritarian democracies. These two sub-types differ in the degree to which the majority holds juridical power. Therefore, he tests the following hypotheses: First, increases in democracy have a negative effect on PSD (H1), Second, he specifically looks at electoral (H1-elec) and minoritarian democracies (H1-min). In his analysis he finds evidence for H1 and H1-min but not for H1-elec.

2.3 Democratic Support and Survival of Democracy

Claassen (2020a) distinguishes between two types of PSD: diffuse and specific. Specific PSD is instrumental and focuses on regime outputs (similar to the instrumental regime performance theory), whereas diffuse support is normative and focuses on the principles of the regime. Therefore, principled PSD is more durable than specific support and helps cushioning regimes in times of political or economic crises. It is thus principled PSD that helps to ensure the survival of democracy. Although the theory is widely accepted (e.g., Norris (2017), Booth and Seligson (2009), Mattes and Bratton (2007)), the findings so far have also been mixed with supporting contributions (Inglehart 2003; Welzel and Inglehart 2005) and contributions against the theory (Fails and Pierce 2010; Hadenius and Teorell 2005; Qi and Shin 2011; Welzel 2007). Astonishingly, these studies essentially analyze the same data from the World Value Survey starting at wave 3

where respective PSD items are included.

Claassen (2020a) does not only look at the relationship between PSD and democractic survival but also between PSD and democratic emergence. He includes the argument made by Qi and Shin (2011) that PSD may also function as democratic demand, thus increasing the probability of transitioning from autocracy to democracy. Claassens hypotheses thus are: First, PSD is positively associated with subsequent change in democracy regardless of the initial level of democracy. Second, he specifically looks at PSD in already-existing democracies (H2-dem) and at PSD in autocracies (H2-aut). He finds evidence in support of H2-dem, mixed evidence for H2 and no evidence for H2-aut.

2.4   Measuring Democratic Support and the Data Challenge

We measure PSD with existing survey data on the national level. In particular, we focus on principled support. The questions ask for the respondent's opinion on the appropriateness or desirability of democracy and undemocratic alternatives and comparisons between both.[1]

Such items are available in 14 survey projects,[2] for 150 countries, and from 1988 to 2017. Combined, the dataset includes 3,765 nationally aggregated binary responses from 1,390 nationally representative survey samples.

The data, however, poses the following challenge: It is highly fractured across time and space, with many – oftentimes large – gaps along the time dimension for all countries. Although we look at a 30 years time span, the average number of covered years is slightly below 8. For instance, China is the country with most participants over all years, namely 20895. However, from 1988 to 2020, the data available covers only 8 years: 2001-2003, 2006-2009, and 2011/2012. Many countries are not covered at all or only surveyed once. The South American countries, however, tend to have the best coverage. For instance, Argentina tops the list with 23 years included. Germany and the US are available for 12 and

---

[1]The 8 included survey items and their different variants can be found on the author's webpage.

[2]World and European Values Surveys, the Afrobarometer, Arab Barometer, Latinobarometer, Asiabarometer, Asian Barometer, South Asia Barometer, New Europe Barometer, Latin American Public Opinion Project, Eurobarometer, European Social Survey, pew Global Attitudes Project, and the Comparative Study of Electoral Systems

13 years, respectively.

Figure 1 gives an overview of the sparseness of the data. The first panel shows the availability of at least one survey item across a selection of eight countries from 1990 to 2017. The other three panels focus on the three most common question themes. We see that the single item data is indeed sparse and that even the aggregated data contains lots of gaps.
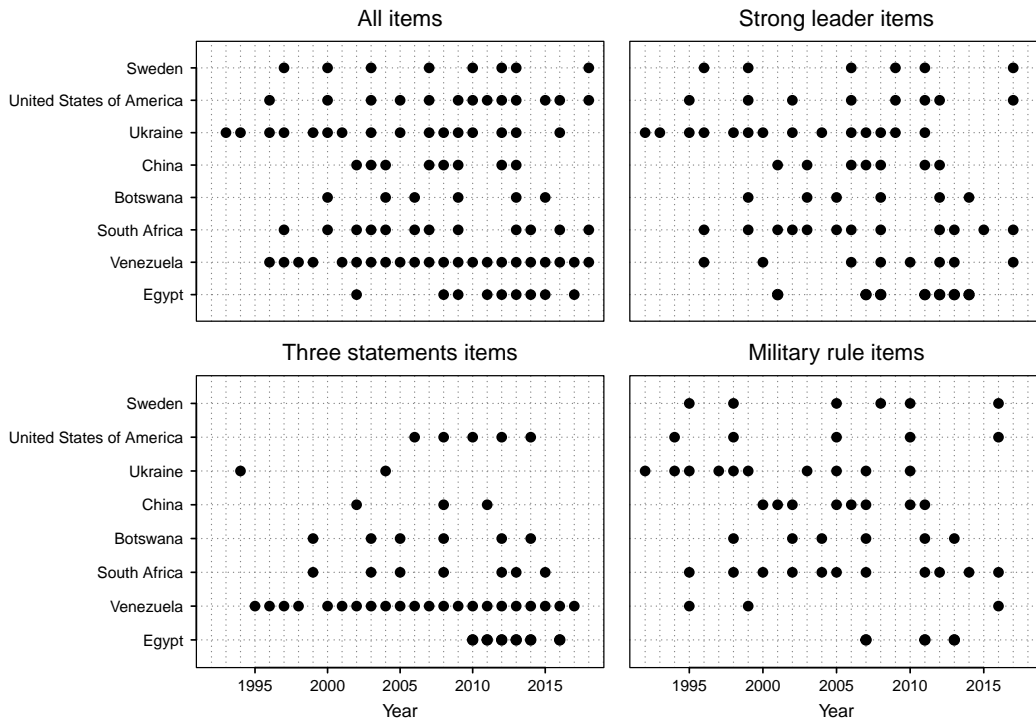

Figure 1: Sparseness of Aggregate Support for Democacy By Country, Year, and Survey Item. The figure is an update from Claassen (2019).

So far, past research has responded to this issue by discarding most of the available data and using only small cross-sectional data sets consisting of data from only one survey project for one year (Inglehart 2003; Qi and Shin 2011; Welzel 2007). Of course, this leads to a disregard not only of all other countries but also of additional items. Claassen (2019) describes how to use most available data for simulating a dataset that is dense across the yearly time dimension. This dataset can then be used for subsequent statistical analyses (Claassen 2020a, 2020b). The next section describes Claassen (2019)'s approach.

3   Model and Estimation

Claassen (2019)'s approach for simulating dense panel data for PSD has the following three main steps: Step 1 is to define a sensible model of PSD item responses as a function of latent public opinion. Step 2 is to estimate the model parameters with the Metropolis Hastings Algorithm. Step 3 is to simulate the the data from the estimated model.

3.1   The Latent Variable Model

Claassen (2019) draws four principles from the literature to model cross-national timeseries for PSD: First, public opinion is an unobserved, latent trait that differs for each year and country. Then, each observed item response is a function of the latent trait. The function should therefore contain item-specific parameters to disaggregate the latent trait into multiple item responses and to account for heterogenuous item functioning. Second, estimating the latent variable from item-specific responses can be though of as *smoothing* the opinion estimate over items since the latent variable does not contain this dimension. One should also smooth over the time dimension by not estimating the latent traits for each time period but rather by estimating the parameters that define a transitional model that holds for all periods. Third, we should not model the percentage of positive item responses but the number of positive and negative responses directly. With that, we can model the problem of smaller response samples. In the following I describe how Claassen (2019) incorporates these principles in the definition of his main model[3] that we will call f.

For each country i, year t, survey item k, the number of positive answers is distrubted binomially with s as the number of total respondands and π as the probability of responding with yes.

$$y_{ikt} \sim \text{Binomial}(s_{ikt}, \pi_{ikt}) \tag{1}$$

---

[3]The main model is called model 5 in Claassen (2019). It achieves lowest discrepancy between simulated and actual data and has the fifth highest complexity of six different specifications.

We could now model π_{ikt} directly as a function of country-year and item-specific effects, θ_{it} and λ_k, respectively. However, we introduce additional dispersion by using another link function, the beta distribution to introduce additional dispersion. Survey data on public opinion are subject to various kinds of errors, for instance, translation, selection and interviewer mistakes. We model this error with the additional dispersion introduce by the beta distribution in Equation (2).

$$\pi_{ikt} \sim \text{Beta}(\alpha_{ikt}, \pi_{ikt}) \tag{2}$$

We reparametrize the two shape parameters α and β to an expectation parameter η and a disperion paramter ϕ in Equation (3) and (4).

$$\alpha_{ikt} = \phi \eta_{ikt} \tag{3}$$

$$\beta_{ikt} = \phi(1 - \eta_{ikt}) \tag{4}$$

Now, we define the expectation of the number of positive responses per year, item and country as a function of item bias λ, country-specific item bias δ and latent, dynamic, country-specific PSD θ.

$$\eta_{ikt} = \text{logit}^{-1}(\lambda_k + \delta_{ik} + \theta_{it}) \tag{5}$$

The item bias effect is distributed normally with expectation μ_λ and variance σ_λ^2.

$$\lambda_k = \mathcal{N}(\mu_\lambda, \sigma_\lambda^2) \tag{6}$$

To model the heterogeneity of item country across item bias across countries (Stegmueller (2011)), we introduce the set of item by country effects δ.

$$\delta_k = \mathcal{N}(0, \sigma_\delta^2) \tag{7}$$

Our main parameter set, the latent parameters θ, capture the underlying time- and country-specific support for democracy. We fully capture the time dimension of the model by modeling the dynamics of θ as an AR(1) process with normally distributed error term with variance σ_θ^2, zero intercept, and zero covariance

as implied by Equation (8).

$$\theta_{it} = \mathcal{N}(\theta_{i,t-1}, \sigma_\theta^2) \tag{8}$$

3.2   Model Estimation with Metropolis-Hastings

One main application for the MH algorithm is Bayesian inference. Specifically, we want to estimate parameters Θ of some probabilistic model f. We have only limited prior knowledge of the distribution of Θ, for instance about its domain, that we use to define prior distributions $p(\Theta)$. And we have a likelihood sample of f given the unknown Θ, namely $p(y|\Theta)$. This would be our PSD data y_{ikt}. We want use both, our prior knowledge and our data, to get our posterior distribution $p(\Theta|y)$ in Equation (9). In this equation, posterior and likelihood are scaled by $\frac{1}{p(y)}$. Without closed-form distributions, $p(y)$ is usually unknown. This also implies that the posterior is proportional to the product of likelihood and prior.

$$p(\Theta|y) = \frac{\mathcal{L}(y|\Theta)p(\Theta)}{p(y)} \propto \mathcal{L}(y|\Theta)p(\Theta) \tag{9}$$

An important insight is that our parameter vector Θ does not only include parameters defined as distributional means, for instance θ, but also standard deviations like σ_θ. If our problem would not include those variation parameters, a more direct option would be to solve $\mathcal{L}(y|\Theta)p(\Theta)$ with the usual optimization algorithms. Those would propose new values of Θ, evaluate its prior probability and the likelihood of the data given Θ until convergence and return the posterior mode or maximum a posteriori probability (MAP) estimate. Yet, since we need to estimate standard deviations, we have to generate the whole parameter distributions to compute mean and variation. To this end, we use a Markov chain Monte Carlo (MCMC) method, namely the Metropolis Hastings algorithm.

Markov chains are stochastic processes that define distributions dependent on the values from only the previous period. The MH algorithm uses Markov chains to sample candidate parameters Θ_i^* depending on Θ_{i-1}. For each candidate Θ_i^*, the algorithm uses the unscaled posterior, $\mathcal{L}(y|\Theta_i)p(\Theta_i)$, and compares it with the posterior of Θ_{i-1} (not requiring constant $p(y)$) to accept or reject new candidates. Comparing relative posterior probabilities thus allows us to sample

from the posterior without knowing y. The algorithm thus explores the domain proportionally to the posterior probability after a number of initial iterations. Besides the model specification using the MH algorithm requires the following choices: the choice of the prior distribution(s) $p(\Theta)$, the choice of the proposal distribution $g(y_i, y_{i-1})$, and the number of initial parameters to drop, i.e. the length of the burnin-period. I include a more technical explanation of the MH algorithm with more general notation in the Appendix.

4   Analysis and Results

4.1   Latent Opinion as Quantity of Interest

Our goal is to analyze the sensitivity of the MH algorithm towards changes in its hyperparameters. Since Claassen (2019) already tested different priors, we are particularly interested in the effect of different choices for the proposal distribution $g(y_i, y_{i-1})$ and the length of the burnin period. Although choosing parameters representing the main results in Claassen (2020b) and Claassen (2020a) would be most illustrative in terms of substantial effects, this choice has the two following disadvantages. First, because these statistical parameters are estimated from simulated data, they are influenced by randomness from the data simulation that could not be distinguished from the estimation-inherent randomness. Second, the additional steps of simulating data and estimating statistical models would potentially double the computation time. Therefore, I choose the latent, dynamic, country-specific PSD θ as the Quantity of Interest.

The results for countries blabla by Classen2020 are shown in Figure blabla...

5   Discussion

6   Conclusion

7   Appendix: Mathematical Background

7.1   Introduction

The Metropolis-Hastings (MH) algorithm is a method for sampling data points from a probability distribution from which direct sampling is difficult. It places among the top 10 algorithms with the greatest influence on science and engineering in the 20th century (Beichl and Sullivan 2000). The MH algorithm belongs to the class of Markov chain Monte Carlo (MCMC) methods. In my explanation I assume prior knowledge on Monte Carlo sampling. However, I will describe the basics of Markov Chains. This section is structured as follows. First, I motivate the usage of the MH algorithm. Second, I explain the basics of Markov Chains. Third, I derive the algorithm and make clear why it works.

7.2   Motivation

One main application for the MH algorithm is Bayesian inference. Specifically, we want to estimate parameters θ of some probabilistic model f. We have only limited prior knowledge of the distribution of θ, $p(\theta)$, and we have a likelihood sample of f given the unknown θ, namely $p(X|\theta)$. The goal is to estimate the posterior distribution of θ, $p(\theta|X)$, given all information that we have. In practice, we do not have a formal definition of the likelihood but only observations. Therefore, we can only approximate the posterior by numerical integration, i.e., we need to sample many points from the posterior to describe it. We can then use the posterior sample to estimate θ with the maximum a posteriori probability estimate.

Claassen (2019) uses the MH algorithm to impute gaps in a panel data set. His approach consists of four steps: First, assume a data generating process

f parameterized by θ. Second, provide the algorithm with the incomplete data X as likelihood and select priors for θ to obtain the posterior distribution $p(\theta|X) = \frac{p(X|\theta)p(\theta)}{P(X)} \propto p(X|\theta)p(\theta)$. Third, use the values for θ with the highest posterior probability as estimates for θ. Finally, insert these estimates into the assumed probabilistic, data generating model f and use it to sample the missing data.

In general and abstracting from Bayesian inference, the MH algorithm generates a sample of random states according to the desired probability distribution $P(X)$. For this purpose, the algorithm employs a Markov process that converges to a unique stationary distribution $\pi(x)$ with $\pi(x) = P(X)$. This distribution can then be used for further steps as previously described. The next section explains the conceptual basics.

7.3 Markov Chains

A Markov chain $(X_t)_{t\in\mathbb{N}}$ is a stochastic process (over time) with the property that the probability of the realization in the next period depends solely on the realization in the current state and not the complete history. This is called the Markov property. Because Markov chains with a countable, or discrete, state space are much more accessible than their continuous variant, in this chapter we will look at the discrete case. Formally, the Markov property writes

$$P(X_{t+1}|X_t, X_{t-1}, ..., X_0) = P(X_{t+1}|X_t). \tag{10}$$

Under some conditions, the stochastic process described by a Markov chain converges to a time-invariant probability distribution, i.e. $P(X_{t+k}|X_{t+k-1}) = P(X_t|X_{t-1}), \forall k > 0$. The crucial step for understanding the MH is to see how it samples a Markov Chain that is certain to converge to a stable posterior distribution. Before exploring how the MH algorithm achieves this result, however, it is necessary to understand its conditions conceptually. To this end, we will use the example depicted by the following graph in Figure 1 that shows the intertemporal transition probabilities between three states representing random events.
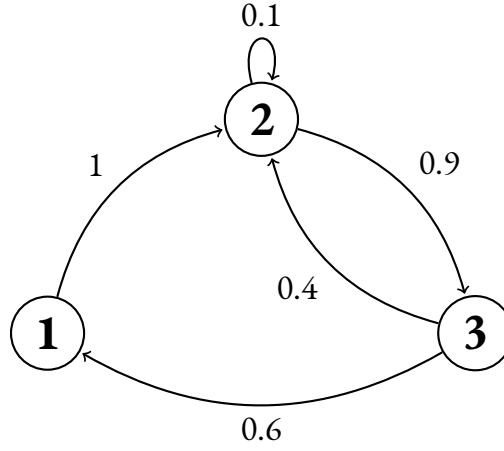
Figure 2: Transition Graph for Markov Chain with 3 states.

This transition graph can be summarized by the $n \times n$ transition matrix T where each element (i, j) represents the probability of moving from state i in period t to state k in period $t + 1$, and where n represents the number of states, i.e $T_{i,j} = P(X_{t+1} = j | X_t = i)$. For our example, we have

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix}. \tag{11}$$

7.3.1 Limit Distribution

As touched upon in the previous subsection, interesting questions can be what the probabilities of each state $j \in \{1, ..., s\}$ are after a finite number or infinitely many steps. For this purpose let $\pi_t(j) = P(X_t = j)$ denote the probability of being in state j in period t. Of course, the probabilities in $t > 0$ depend on the probabilities for the the initial state π_0. We can use the law of total probability to calculate the probability of each state for the next period $t = 1$ by

$$P(X_1 = j) = \sum_{i=1}^{3} P(X_1 = j | X_0 = i)\pi_0(i). \tag{12}$$

I.e., to compute the probability of being in state j in $t = 1$, for each initial state i, we multiply its probability $\pi_0(i)$ by the probability of moving from i to state j.

This is equivalent to $\pi_1 = \pi_0 T$ in vector notation. Further, we can compute the distributions in an arbitrary future period by repeating the matrix multiplication, e.g, $\pi_2 = \pi_0 TT$, or in general, $\pi_t = \pi_0 T^t$.

Now we are ready to define the limit distribution that describes the probability distribution after infinitely many periods by

$$\pi_\infty = lim_{t\to\infty}\pi_t = lim_{t\to\infty}\pi_0 T^t. \tag{13}$$

We can further ask two additional important questions. First, does a limit distribution exist? And second, is it unique, or in other word, do we have the same limit distribution independent from the realization of the initial state X_0? In our example, there does not only exist a limit distribution with $\pi_\infty = (0.2, 0.4, 0.4)$, it is even unique regardless of start distribution π_0. This means that independent of the start state, the probability of each state converges to the same number. For the context of the MH algorithm, this is an important property because we always want to compute the same estimates for our parameters θ, regardless of the starting values of our simulation. In the next section, we introduce and simplify conditions that guarantee a unique limit distribution.

7.3.2 Irreducibility, Periodicity and Stationarity

Definition 7.1. A Markov chain is called *irreducible* if each state is reachable from any other state in a finite number of steps.

Figure 2 shows a Markov chain represented by a bipartite graph. This graph is composed by two times the graph in Figure 1. Obviously, this chain is not irreducible because the initial state impacts all future distributions. More precisely, starting in one subgraph sets the probability of reaching states in the other subgraph to zero. We see that a Markov Chain is only irreducible if there is at least an indirect link between every pair of states. We also observe that if the Markov Chain is not irreducible there can be no limit distribution.
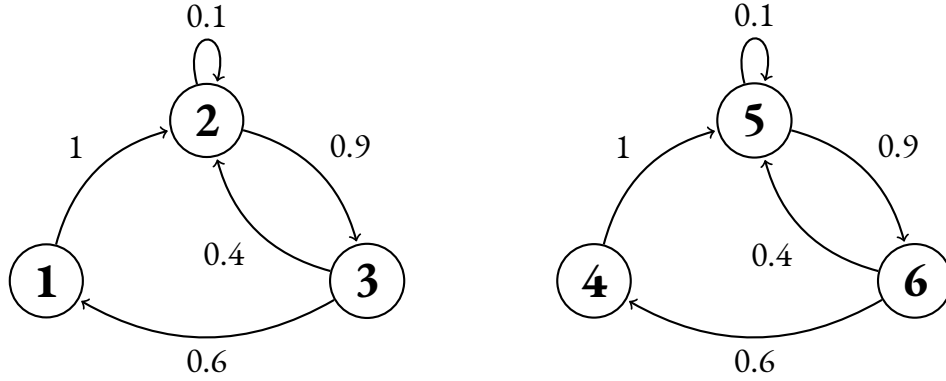
Figure 3: Transition Graph for Irreducible Markov Chain.

Definition 7.2. A state i has a period k if the greatest common denominator of possible revisits is k. A Markov chain is *aperiodic* if the period of all its states is 1.

Consider the five-state Markov chain in Figure 3 as an illustration for the above definition and suppose we start in state 1. Observe that, independent of the random draw for next period, we will arrive again in state 1 after two or four steps. Therefore, state 1 has a period of 2. If a state is revisited in random rather than a fixed time period then the state has period 1. This is automatically the case if a state has a positive edge with itself.
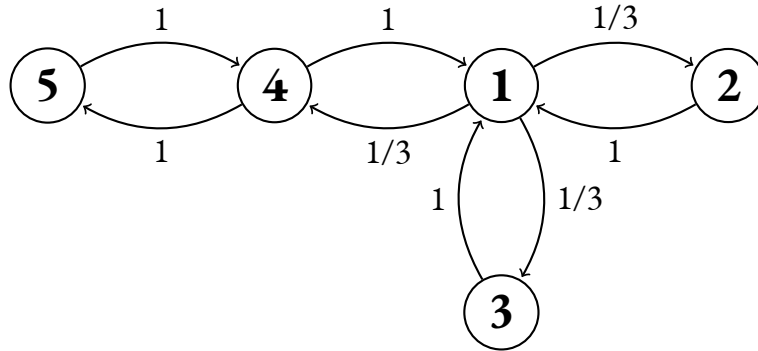

Figure 4: Markov Chain with 2-periodic State 1

Definition 7.3. π^* is the *stationary distribution* of a Markov Chain with Transition matrix T if $\pi^* = \pi^* T$ and π^* is a probability vector.

Verbally, this means that the probability distribution π^* does not change anymore over time. If π^* is also unique, then π^* is our aim, the limit distribution introduces in section 1.3.1, i.e., $\pi^* = \pi_\infty$.

These three definitions are enough to understand the next fundamental theorem.

7.3.3 The Fundamental Theorem of Markov Chains

The next theorem defines formally the condition when a Markov Chain converges to a unique distribution, i.e. the limit distribution.

Theorem 7.1. *(Fundamental Theorem of Markov Chains) If a Markov chain is irreducible and aperiodic (called ergodic) then it has a stationary distribution π^* that is unique (*$\lim_{t\to\infty} P(X_t = i) = \pi_i^*, \forall i$*).*

Therefore, if we want to construct a stable distribution $P(X)$ via Markov chains, we need to ensure that it is irreducible and aperiodic with stationary distribution $\pi^* = P(X)$. In the next subsection, we substitute the stationarity condition by a stronger one before we finally derive the MH algorithm.

7.3.4 Reversibility

Definition 7.4. A Markov chain is *reversible* if there is a probability distribution π over its states such that $\pi(i)T_{ij} = \pi(j)T_{j,i}, \forall i, j$ (reversibility condition).

Theorem 7.2. *A sufficient condition for distribution π^* to be a stationary distribution of a Markov chain with transition matrix T is that it fullfills the reversibility condition.*

Proof. $\sum_i \pi(i)T_{i,j} = \sum_i \pi(j)T_{j,i} = \pi(j) \sum_i T_{j,i} = \pi(j) \implies \pi T = \pi$ $\qquad \square$

Reversibility is a stronger condition than stationarity because it requires that the probability flux from i to j is equal to the one from j to i for each possible pair of states. Recall, that stationarity only requires that the probability flux to one state is equal on aggregate and not that it is symmetric between each pair of states over time. Therefore, if we want to achieve a stationary distribution it is enough to ensure that it is reversible.

7.4   The Algorithm

Recall that we want to generate a sample of a desired distribution $P(X)$. For this purpose, we use a Markov process that is uniquely defined by its transition probabilities $P(X_{t+1}|X)$ with limit distribution π so that $\pi = P(X)$. As explained in the previous section, a Markov process has a limit distribution if each transition $X_t \rightarrow X_{t+1}$ is reversible and if the stationary distribution π is ergodic. With the MH algorithm, we construct such a Markov process with stationary distribution $\pi = P(X)$. The derivation starts with another way of writing reversibility[4]:

$$P(x'|x)P(x) = P(x|x')P(x') \iff \frac{P(x'|x)}{P(x|x')} = \frac{P(x')}{P(x)} \qquad (14)$$

The main idea is to separate transition $P(x'|x)$ in two steps: the proposal step and the acceptance-or-rejection step. Let $g(x')$ be the proposal distribution, i.e., the conditional probability of proposing state x' given x. And let $A(x'|x)$ be the probability of accepting proposed state X'. Formally, we have $P(x'|x) = g(x'|x)A(x'|x')$. Inserting this in Equation (14) gives

$$\frac{P(x')}{P(x)} = \frac{g(x'|x)A(x',x)}{g(x|x')A(x',x)} \iff \frac{A(x',x)}{A(x,x')} = \frac{P(x')}{P(x)}\frac{g(x|x')}{g(x'|x)}. \qquad (15)$$

The following choice, termed the Metropolis choice, is commonly used as an acceptance ratio for sampling x' from $P(x')$ that fulfills the above reversibility condition:

$$A(x',x) = \min\left(1, \frac{P(x')}{P(x)}\frac{g(x|x')}{g(x'|x)}\right) \qquad (16)$$

Note that the minimizer in $A(x',x)$ enforces that the probability is below 1. The MH algorithm writes as follows:

---

[4]We simplify our notation by using x' and x instead of X_{t+1} and X_t.

Algorithm 1 Metropolis-Hastings algorithm

Initialize X_0
for $t \leftarrow 0$ to $T - 1$ **do**
    Draw $u \sim \mathcal{U}_{[0,1]}$
    Draw candidate $X^* \sim P(X^*|X_{t-1})$
    if $u < \min\{1, \frac{p(X^*)g(X_t|X^*)}{p(X_t)g(X^*|X_t)}\}$ **then**
        $X_{t+1} \leftarrow X^*$
    else
        $X_{t+1} \leftarrow X_t$
    end if
end for

Obviously, the construction of the acceptance ratio ensures reversibility. Ergodicity is ensured by the random nature with which we accept proposed states: First, the chain is irreducible because each state is reachable from any other state with positive probability at every single step. Second, for each state x, $P(x' = x)$ is always positive and therefore the Markov chain is aperiodic.

In a general setting, the choice for transition distribution $g(x'|x)$ and the number of iterations until the limit distribution is reached are unclear. These two choices are the hyperparameters of the MH algorithm. In the Bayesian inference application in the article series staring from Claassen (2019), additional choices are the prior distribution $p(\theta)$ and the model choice f.

References

Baker, Kendall L, Russell J Dalton, Kai Hildebrandt, et al. 1981. *Germany Transformed: Political Culture and the New Politics*. Harvard University Press.

Beichl, Isabel, and Francis Sullivan. 2000. "The Metropolis Algorithm." *Computing in Science & Engineering* 2(1): 65–69.

Booth, John A, and Mitchell A Seligson. 2009. *The Legitimacy Puzzle in Latin America: Political Support and Democracy in Eight Nations*. Cambridge University Press.

Claassen, Christopher. 2019. "Estimating Smooth Country–Year Panels of Public Opinion." *Political Analysis* 27(1): 1–20.

———. 2020a. "Does Public Support Help Democracy Survive?" *American Journal of Political Science* 64(1): 118–34.

———. 2020b. "In the Mood for Democracy? Democratic Support as Thermostatic Opinion." *American Political Science Review* 114(1): 36–53.

Dalton, Russell J. 1994. "Communists and Democrats: Democratic Attitudes in the Two Germanies." *British Journal of Political Science* 24(4): 469–93.

Denemark, David, Todd Donovan, Richard G Niemi, and Robert Mattes. 2016. "The Advanced Democracies: The Erosion of Traditional Democratic Citizenship." *Growing up democratic: Does it make a difference*: 181–206.

Fails, Matthew D, and Heather Nicole Pierce. 2010. "Changing Mass Attitudes and Democratic Deepening." *Political Research Quarterly* 63(1): 174–87.

Foa, Roberto Stefan, and Yascha Mounk. 2016. "The Danger of Deconsolidation: The Democratic Disconnect." *Journal of democracy* 27(3): 5–17.

———. 2017. "The Signs of Deconsolidation." *Journal of democracy* 28(1): 5–15.

Graham, Carol, and Sandip Sukhtankar. 2004. "Does Economic Crisis Reduce Support for Markets and Democracy in Latin America? Some Evidence from Surveys of Public Opinion and Well Being." *Journal of Latin American Studies* 36(2): 349–77.

Hadenius, Axel, and Jan Teorell. 2005. "Cultural and Economic Prerequisites of Democracy: Reassessing Recent Evidence." *Studies in comparative international development* 39(4): 87–106.

Inglehart, Ronald. 2003. "How Solid Is Mass Support for Democracy—and How Can We Measure It?" *PS: Political Science & Politics* 36(1): 51–57.

Magalhães, Pedro C. 2014. "Government Effectiveness and Support for Democracy." *European Journal of Political Research* 53(1): 77–97.

Mannheim, Karl. 1970. "The Problem of Generations." *Psychoanalytic review* 57(3): 378–404.

Mattes, Robert, and Michael Bratton. 2007. "Learning about Democracy in Africa: Awareness, Performance, and Experience." *American Journal of Political Science* 51(1): 192–217.

Mishler, William, and Richard Rose. 1996. "Trajectories of Fear and Hope: Support for Democracy in Post-Communist Europe." *Comparative political studies* 28(4): 553–81.

———. 2002. "Learning and Re-Learning Regime Support: The Dynamics of Post-Communist Regimes." *European Journal of Political Research* 41(1): 5–36.

———. 2007. "Generation, Age, and Time: The Dynamics of Political Learning During Russia's Transformation." *American journal of political science* 51(4): 822–34.

Montero, José Ramón, Richard Gunther, and Mariano Torcal. 1997. "Democracy in Spain: Legitimacy, Discontent, and Disaffection." *Studies in comparative international development* 32(3): 124–60.

Niemi, Richard G. 1974. *The Political Character of Adolescence: The Influence of Families and Schools [by] m. Kent Jennings and Richard g. Niemi.* [Princeton, NJ]: Princeton University Press.

Norris, Pippa. 2017. "Is Western Democracy Backsliding? Diagnosing the Risks." *Forthcoming, The Journal of Democracy, April.*

Qi, Lingling, and Doh Chull Shin. 2011. "How Mass Political Attitudes Affect Democratization: Exploring the Facilitating Role Critical Democrats Play in the Process." *International Political Science Review* 32(3): 245–62.

Stegmueller, Daniel. 2011. "Apples and Oranges? The Problem of Equivalence in Comparative Research." *Political Analysis* 19(4): 471–87.

Voeten, Erik. 2016. "Are People Really Turning Away from Democracy?" *Available at SSRN 2882878.*

Welzel, Christian. 2007. "Are Levels of Democracy Affected by Mass Attitudes? Testing Attainment and Sustainment Effects on Democracy." *International Political Science Review* 28(4): 397–424.

Welzel, Christian, and Ronald Inglehart. 2005. "Liberalism, Postmaterialism, and the Growth of Freedom." *International Review of Sociology* 15(1): 81–108.

Wlezien, Christopher. 1995. "The Public as Thermostat: Dynamics of Preferences for Spending." *American journal of political science*: 981–1000.

Statutory Declaration

Hiermit versichere ich, dass diese Arbeit von mir persönlich verfasst ist und dass ich keinerlei fremde Hilfe in Anspruch genommen habe. Ebenso versichere ich, dass diese Arbeit oder Teile daraus weder von mir selbst noch von anderen als Leistungsnachweise andernorts eingereicht wurden. Wörtliche oder sinngemäße Übernahmen aus anderen Schriften und Veröffentlichungen in gedruckter oder elektronischer Form sind gekennzeichnet. Sämtliche Sekundärliteratur und sonstige Quellen sind nachgewiesen und in der Bibliographie aufgeführt. Das Gleiche gilt für graphische Darstellungen und Bilder sowie für alle Internet-Quellen. Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann. Mir ist bekannt, dass von der Korrektur der Arbeit abgesehen werden kann, wenn die Erklärung nicht erteilt wird.

Mannheim, den _____ _____
                              Name und Unterschrift

I hereby declare that the paper presented is my own work and that I have not called upon the help of a third party. In addition, I affirm that neither I nor anybody else has submitted this paper or parts of it to obtain credits elsewhere before. I have clearly marked and acknowledged all quotations or references that have been taken from the works of other. All secondary literature and other sources are marked and listed in the bibliography. The same applies to all charts, diagrams and illustrations as well as to all Internet sources. Moreover, I consent to my paper being electronically stores and sent anonymously in order to be checked for plagiarism. I am aware that the paper cannot be evaluated and may be graded "failed" ("nicht ausreichend") if the declaration is not made.

Mannheim, _____ _____
                         Name and Signature