A cautionary note on the discrete uniform prior for the binomial N: comment

Author(s): Cristiano Villa and Stephen G. Walker

Source: *Ecology* , September 2014, Vol. 95, No. 9 (September 2014), pp. 2674-2677

Published by: Wiley on behalf of the Ecological Society of America

Stable URL: http://www.jstor.com/stable/43494497

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Wiley and *Ecological Society of America* are collaborating with JSTOR to digitize, preserve and extend access to *Ecology*

This content downloaded from
152.3.34.254 on Thu, 02 Jan 2025 02:08:42 UTC
All use subject to https://about.jstor.org/terms

temporal patterns of perceived predation risk ..." (Beschta and Ripple 2013).

Unfortunately, the interpretation of the scientific method presented by Beschta and Ripple (2013) and extended here by Beschta et al. (2014), casts serious doubts on the body of work they are defending. These authors advocate considering only those data that appear to show that which they hypothesize to exist, a BMTC. The dismissal of data from places, "without confirmation of a trophic cascade ..." (Beschta and Ripple 2013), automatically leads to ignoring evidence contradicting the hypothesis. This tautology results in the championing of the BMTC hypothesis with biased, hypothesis-supporting data, rather than rigorous testing with data capable of refuting the hypothesis. When others test the hypothesis in Yellowstone with measured independent variables, either variables assumed to be important by BMTC proponents (Kimble et al. 2011, Winnie 2012), or with field data (Romme et al. 2005, Bilyeu et al. 2008, Kauffman et al. 2010), the hypothesis fails to predict differences in browsing, recruitment and growth.

Literature cited

Bergman, E. J., R. A. Garrott, S. Creel, J. J. Borkowski, R. Jaffe, and G. R. Watson. 2006. Assessment of prey vulnerability through analysis of wolf movements and kill sites. Ecological Applications 16:273–284.

Beschta, R. L., C. Eisenberg, J. W. Laundré, W. J. Ripple, and T. P. Rooney. 2014. Predation risk, elk, and aspen: comment. Ecology 95:2669–2671.

Beschta, R. L., and W. J. Ripple. 2013. Are wolves saving Yellowstone's aspen? A landscape-level test of a behaviorally mediated trophic cascade: comment. Ecology 94:1420–1425.

Bilyeu, D. M., D. J. Cooper, and N. T. Hobbs. 2008. Water tables constrain height recovery of willow on Yellowstone's Northern Range. Ecological Applications 18:80–92.

Halofsky, J. S., and W. J. Ripple. 2008*a*. Fine-scale predation risk on elk after wolf reintroduction in Yellowstone National Park, USA. Oecologia 155:869–877.

Halofsky, J. S., and W. J. Ripple. 2008*b*. Linkages between wolf presence and aspen recruitment in the Gallatin elk winter range of southwestern Montana, USA. Forestry 81:195–207.

Kauffman, M., J. F. Brodie, and E. S. Jules. 2010. Are wolves saving Yellowstone's aspen? A landscape-level test of a behaviorally mediated trophic cascade. Ecology 91:2742–2755.

Kay, C. 1997. Is aspen doomed? Journal of Forestry 95:4–11.

Kay, C. E., and D. L. Bartos. 2000. Ungulate herbivory on Utah aspen: assessment of long-term enclosures. Journal of Range Management 53:145–153.

Kimble, D. S., D. B. Tyers, J. Robinson-Cox, and B. F. Sowell. 2011. Aspen recovery since wolf reintroduction on the Northern Yellowstone winter range. Rangeland Ecology and Management 64(2):119–130.

Ripple, W. J., and R. L. Beschta. 2003. Wolf reintroduction, predation risk and cottonwood recovery in Yellowstone National Park. Forest Ecology and Management 184:299–313.

Ripple, W. J., and R. L. Beschta. 2004. Wolves, elk, willows and trophic cascades in the upper Gallatin Range of southwestern Montana, USA. Forest Ecology and Management 200:161–181.

Ripple, W. J., and R. L. Beschta. 2006. Linking wolves to willows via risk-sensitive foraging by ungulates in the Northern Yellowstone Ecosystem. Forest Ecology and Management 230:96–106.

Ripple, W. J., and E. J. Larsen. 2000. Historic aspen recruitment, elk and wolves in northern Yellowstone National Park. Biological Conservation 95:361–370.

Ripple, W. J., and E. J. Larsen. 2001. The role of postfire woody debris in aspen regeneration. Western Journal of Applied Forestry 16(2):61–64.

Romme, W. H., M. G. Turner, G. A. Tuskan, and R. A. Reed. 2005. Establishment, persistence, and growth of aspen (*Populus tremuloides*) seedlings in Yellowstone National Park. Ecology 86:404–418.

Winnie, J., Jr. 2012. Predation risk, elk, and aspen: tests of a behaviorally mediated trophic cascade in the Greater Yellowstone Ecosystem. Ecology 93(12):2600–2614.

Ecology, 95(9), 2014, pp. 2674–2677
© 2014 by the Ecological Society of America

A cautionary note on the discrete uniform prior for the binomial N: comment

Cristiano Villa[1,3] and Stephen G. Walker[2]

Introduction

In this paper, we discuss objective prior distributions for the discrete parameter N of a binomial distribution, with specific applications to estimation of population or species sizes. In particular, we argue that in statistical *applications* the scale prior $\pi(N) \propto 1/N$ should not be employed due to the lack of a probabilistic interpretation. Hence, this paper can be seen as a comment on the recent paper of Link (2013).

In the statistical analysis of ecological data, it is frequent to deal with data that comes from binomial outcomes, such as the size of a population or the number of species. The capture–recapture models for a closed population introduced by Otis et al. (1978) represent an example of how the estimation of N proceeds in wildlife data analysis.

A common choice of objective prior for N is the scale prior; that is $\pi(N) \propto 1/N$. Recently, Link (2013) has provided support to the scale prior for N on the basis of its better performance in comparison to the uniform prior and, in addition, that it has been proposed by

Manuscript received 24 February 2014; accepted 25 March 2014. Corresponding Editor: E. G. Cooch.

[1] School of Mathematics, Statistics and Actuarial Science, University of Kent, United Kingdom.

[2] Department of Mathematics and Division of Statistics and Scientific Computing, University of Texas at Austin, Texas, USA.

[3] E-mail: cv60@kent.ac.uk

This content downloaded from 152.3.34.254 on Thu, 02 Jan 2025 02:08:42 UTC
All use subject to https://about.jstor.org/terms

Berger et al. (2012). We argue that there is no real motivation in the use of the scale prior; on the contrary, it is an ad hoc solution rather than the result of specific objective probabilistic considerations. In other words, $\pi(N) \propto 1/N$ has no "meaning." We argue that, for a prior to be used in real applications, it needs to be fully motivated.

We believe an objective approach for such a parameter is to consider the *worth* or utility of each value in the parameter space $\mathcal{N} = \{1, 2, \ldots\}$. This idea is thoroughly discussed in Villa and Walker (2014) and Villa and Walker (2013); for convenience, we present it in *Objective priors for N*. It is noteworthy to point out that the scale prior has been used by Wang et al. (2007) and by King and Brooks (2008) for applications in ecology, and by Basu and Ebrahimi (2001) for an example of an application in capture–recapture models in software reliability.

Objective Priors for N

Consider observation $X \sim \text{Bin}(N, p)$, where $N \in \mathcal{N} = \{1, 2, \ldots\}$ represents the number of independent Bernoulli trials, and $p \in (0, 1)$ the probability of success at each trial. The aim is to make inference on the discrete parameter N, assuming p is unknown.

The task of assigning an objective prior to a discrete parameter is not trivial and, in the past, has represented an interesting challenge. The main reason comes from the fact that common objective approaches such as Jeffreys' rule (Jeffreys 1961) and reference analysis (Berger et al. 2009) are not suitable for discrete parameters and, when they are, they do not provide sensible results. Note that the uniform prior $\pi(N) \propto 1$, which may appear to be a natural choice to represent ignorance about N, is not suitable for inference as, for when p is unknown, leads to an improper posterior (Berger et al. 1999, 2012).

A motivation behind the choice of $1/N$ is that, although Jeffreys himself never discussed the prior for N when p is unknown, the choice of $\pi(N) \propto 1/N$ is assumed as natural (Berger et al. 2012), as it is the prior Jeffreys recommends for (continuous) scale parameters. Link (2013), in addition to the above motivation, recommends the scale prior as it solves estimation problems related to the use of the uniform prior (when $\mathcal{N}$ is finite).

The choice of $1/N$ as an objective prior for N is questionable for the following reasons. The motivation for Jeffreys prior in a discrete setting is obsolete. Jeffreys rule is based on invariance property under one-to-one transformations of the parameter of interest, and this notion has no meaning for a discrete parameter space. Furthermore, Kahn (1987), shows that if we assign a Beta prior to p, $\pi(p) \sim \text{Beta}(a, b)$, and assume the parameters of the binomial independent a prior, then $\pi(N) \propto 1/N^c$ yields a proper posterior for N if $a + c > 1$. It is, therefore, legitimate to wonder why c has to be chosen as equal to one. Why not, for example, $\pi(N) \propto 1/N^2$ or $\pi(N) \propto 1/N^3$? This fact adds a level of subjectivity and arbitrariness to the whole procedure, making the process not as objective as intended.

One may argue that the scale prior is the result of a different objective procedure as well. Berger et al. (2012) use an approach that consists of embedding the discrete problem into a continuous one and then apply reference analysis. However, as there exist more than one embedding procedure, they obtain two different priors: the scale prior and

$$\pi(N) \propto \sqrt{N\{N + 4/(n+3)\}}$$

where n is the size of independent and identically distributed random variables: $X_i \sim \text{Bin}(N, p)$, $i = 1, \ldots, n$. As both priors have similar properties, the recommendation of $1/N$ lies in the simplicity of its functional form. Again, the choice of the scale prior does not appear to be truly objective.

It is fundamental to highlight that in an applied (statistical) setting, such as in ecology, an objective prior needs an idea that is well supported. Unlike academic statisticians, who can discourse on objective priors on theoretical grounds, applied statisticians have to put the motive first: an objective prior needs to have an applied meaning. In fact, the derivation of $\pi(N)$ should be the result of a process where there is a clear explanation on why a particular prior is chosen and what it represents; we find, for example, that in Link (2013) this explanation is missing.

The prior we propose is based on the idea of assigning a *worth* to each element $N \in \mathcal{N}$. The worth is objectively measured by assessing what is lost if that parameter value is removed from $\mathcal{N}$ and it is the true one. Once the worth has been determined, this will be linked to the prior probability by means of the self-information loss function (Merhav and Feder 1998), $-\log \pi(N)$. A detailed illustration of the idea can be found in Villa and Walker (2014) and Villa and Walker (2013), but here is an overview.

Let us indicate by f_N the binomial distribution with parameters N, give p (for the moment assumed to be known). The utility (i.e., worth) to be assigned to f_N is a function of the Kullback-Leibler divergence (Kullback and Leibler 1951) measured from the model to the nearest one; where the nearest model is the one defined by $N' \neq N$ such that $D_{\text{KL}}(f_N \,||\, f_{N'})$ is minimized. In fact (see Berk 1966), N' is where the posterior asymptotically accumulates if N is excluded from $\mathcal{N}$. The objectivity of how the utility of f_N is measured is obvious, as it depends on the choice of the model only.

Let us now write $u_1(N) = \log\pi(N)$ and let the minimum divergence from f_N be represented by $u_2(N)$. Note that $u_1(N)$ is the utility associated with the prior

This content downloaded from
152.3.34.254 on Thu, 02 Jan 2025 02:08:42 UTC
All use subject to https://about.jstor.org/terms

probability for model f_N, and $u_2(N)$ is the utility in keeping N in $\mathcal{N}$. We want $u_1(N)$ and $u_2(N)$ to be matching utility functions, as they are two different ways to measure the same utility in N. As it stands, $-\infty < u_1 \leq 0$ and $0 \leq u_2 < \infty$, while we actually want $u_1 = -\infty$ when $u_2 = 0$. The scales are matched by taking exponential transformations; so $\exp(u_1)$ and $\exp(u_2) - 1$ are on the same scale. Hence, we have

$$e^{u_1(N)} = \pi(N) \propto e^{g\{u_2(N)\}} \tag{1}$$

where

$$g(u) = \log(e^u - 1). \tag{2}$$

By setting the functional form of g in Eq. 1, as it is defined in Eq. 2, we derive the proposed objective prior for the discrete parameter N

$$\pi(N) \propto \exp\left\{ \min_{N \neq N' \in \mathcal{N}} D_{\mathrm{KL}}(f_N \,||\, f_{N'}) \right\} - 1.$$

We note that in this way the Bayesian approach is conceptually consistent, as we update a prior utility assigned to N, through the application of Bayes theorem, to obtain the resulting posterior utility expressed by $\log \pi(N \,|\, x)$. Indeed, there is an elegant procedure akin to Bayes, which works from a utility point of view, namely that

$$\log \pi(N \,|\, x) = K + \log f_N(x \,|\, N) + \log \pi(N)$$

which has the interpretation of

$$\text{Utility}(N \,|\, x, \pi) = K + \text{Utility}(N \,|\, x) + \text{Utility}(N \,|\, \pi)$$

where K does not depend on N. There is then a retention of meaning between the prior and the posterior information (here represented as utilities). This property is not shared by the usual interpretation of Bayes theorem when priors are objectively obtained; in fact, the prior would usually be improper, hence not representing probabilities, while the posterior is (and has to be) a proper probability distribution.

In Villa and Walker (2014), we show that the nearest model to f_N is at $N' = N + 1$. Thus, the prior for N is given by

$$\pi(N) \propto \frac{1}{(N+1)(1-p)} \exp\left\{ \sum_{x=0}^{N} \log(N+1-x) \binom{N}{x} p^x (1-p)^{N-x} \right\} - 1. \tag{4}$$

The prior in Eq. 4 is improper but, with just one observation, yields a proper posterior. While looking complicated, prior 4 is fully motivated from an applied perspective.

If p is unknown, the joint prior distribution for the parameters of the binomial is given by

$$\pi(N, p) = \pi(N \,|\, p)\pi(p) \tag{5}$$

where $\pi(N \,|\, p)$ is the prior in Eq. 4 and $\pi(p)$ a suitable prior for the probability of success at each trial.

Snowshoes hares analysis

To illustrate the objective prior we propose, and to compare it with the scale prior, we analyze a popular capture–recapture data set of snowshoes hares. The problem has been originally discussed in Otis et al. (1978) and, from a Bayesian perspective, for example in Royle et al. (2007) and Link (2013). In particular, Link (2013) has analyzed the data using a scale prior for N (although using a data augmentation approach).

The data consists of a sample of $n = 68$ hares captured and recaptured, over $T = 6$ days. The encounter frequencies, over the 6 days, gives the set $\{25, 22, 13, 5, 1, 2\}$, that is for a total of 145 capture–recapture occurrences. For this illustration, we consider model for closed populations M_0, as defined in Otis et al. (1978), which assumes that the capture–recapture probabilities are constant for all the animals and across the 6 days. Thus, indicating by y_i the detection frequency of animal i, with $i = 1, \ldots, N$, the likelihood function is given by

$$L(N, p \,|\, y) \propto \frac{N!}{(N-n)!} p^{\sum_i y_i} (1-p)^{T \cdot N - \sum_i y_i}. \tag{6}$$

We analyze the data by considering both the scale prior and our prior for N. For the scale prior, we have $\pi(N, p) = \pi(N)\pi(p)$, assuming prior independence of the parameters. When we use the prior in Eq. 4, the joint prior has the form of Eq. 5. In both circumstances we set $\pi(p) \sim \text{Beta}(1/2, 1/2)$, that is, Jeffreys' prior. As the posterior distributions are analytically intractable, we obtain the marginal distribution for N through Markov chain Monte Carlo (MCMC) methods.

The histogram of the posterior distributions are plotted in Fig. 1. The posterior for N obtained by applying the scale prior $\pi(N) \propto 1/N$ is shown in panel (a), while the posterior obtained by applying the prior we propose in Eq. 4 is shown in panel (b). Both distributions are positively skewed and accumulate on the same values of N. When the scale prior is used, the median is $N = 81.5$, with 95% credible interval (68.7, 94.3). When our prior is used, we have a median of $N = 81.0$ and 95% credible interval (68.7, 93.4); note that prior 4 gives a smaller credible interval than the one obtained by adopting the scale prior.

For completeness, we note that for p we have medians $p = 0.33$ in both cases, with 95% credible intervals (0.24, 0.43) and (0.24, 0.42), for the scale and our prior respectively.

This content downloaded from 152.3.34.254 on Thu, 02 Jan 2025 02:08:42 UTC
All use subject to https://about.jstor.org/terms

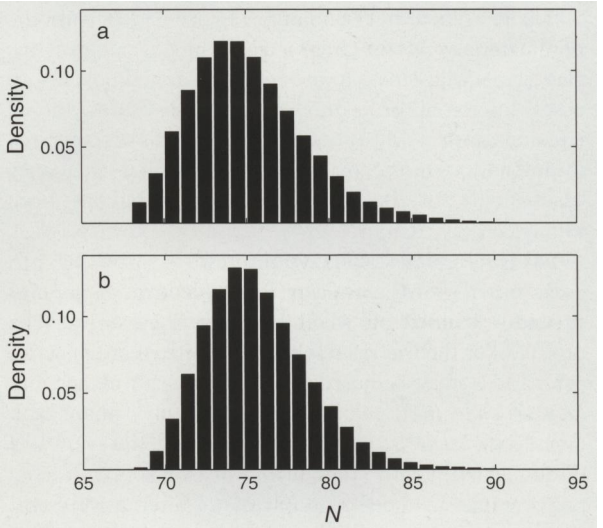

FIG. 1. Histogram of the posterior distribution for the parameter N for the hare data using (a) scale prior and (b) our prior.

Discussion

The choice of an objective prior for N must be based not only on performance, but also on solid motivation; particularly if the prior is intended to be used in real applications. If this criterion is not met, we argue the prior is motiveless. In the example of the hare data, we have shown that the prior based on worth results in a credible interval that is narrower that the one obtained by applying the scale prior for N. Additionally, while the latter prior has no probabilistic justification, the former one is the result of a clear objective motivation.

The prior for N can be applied to any of the remaining capture–recapture models (Otis et al. 1978), that is, when either one or more effects (time effects, behavioral effects, heterogeneity effects) are considered. We have not included any example, either simulated or based on real data, for models including time, behavioral, or heterogeneity effects. However, the implementation is similar to the one outlined.

Literature cited

Basu, S., and N. Ebrahimi. 2001. Bayesian capture–recapture methods for error detection and estimation of population size: heterogeneity and dependence. Biometrika 88:269–279.

Berger, J. O., J. M. Bernardo, and D. Sun. 2009. The formal definition of reference priors. Annals of Statistics 37:905–938.

Berger, J. O., J. M. Bernardo, and D. Sun. 2012. Objective priors for discrete parameter spaces. Journal of the American Statistical Association 107:636–648.

Berger, J. O., B. Liseo, and R. L. Wolpert. 1999. Integrated likelihood methods for eliminating nuisance parameters. Statistical Science 18:1–28.

Berk, R. H. 1966. Limiting behaviour of posterior distributions when the model is incorrect. Annals of Mathematical Statistics 37:51–58.

Jeffreys, H. 1961. Theory of probability. University Press, Oxford, UK.

Kahn, W. D. 1987. A cautionary note for Bayesian estimation of the binomial parameter *n*. American Statistician 41:38–39.

King, R., and S. P. Brooks. 2008. On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. Biometrics 64:816–824.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of Mathematical Statistics 22:79–86.

Link, W. A. 2013. A cautionary note on the discrete uniform prior for the binomial *N*. Ecology 94:2173–2179.

Merhav, N., and M. Feder. 1998. Universal prediction. IEEE Transactions on Information Theory 44:2124–2147.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62. 135 pages.

Royle, J. A., R. M. Dorazio, and W. A. Link. 2007. Analysis of multinomial models with unknown index using data augmentation. Journal of Computational and Graphical Statistics 16:67–85.

Villa, C., and S. G. Walker. 2013. Objective prior for the number of degrees of freedom of a *t* distribution. Bayesian Analysis 9:197–220.

Villa, C., and S. G. Walker. 2014. An objective approach to prior mass functions for discrete parameter spaces. Journal of the American Statistical Association, *in press*.

Wang, X., C. Z. He, and D. Sun. 2007. Bayesian population estimation for small sample capture-recapture data using noninformative priors. Journal of Statistical Planning and Inference 137:1099–1118.

Ecology, 95(9), 2014, pp. 2677–2679
© 2014 by the Ecological Society of America

A cautionary note on the discrete uniform prior for the binomial *N*: reply

WILLIAM A. LINK[1]

The rats were devouring the house, but instead of examining the cat's teeth and claws they only concerned themselves to find out if it was a holy cat. If it was a pious cat, a moral cat, all right, never mind about the other capacities, they were of no consequence.

—Mark Twain, Personal Recollections of Joan of Arc

I published "*A cautionary note* ..." this past year in response to widespread use of the discrete uniform (DU) prior for N in wildlife studies (Link 2013). I addressed a specific failure of the DU prior: Suppose $X_i \sim B(N_i, p)$ are independent for $i = 1, 2, \ldots k$ with N_i's and p

Manuscript received 5 May 2014; accepted 8 May 2014. Corresponding Editor: E. G. Cooch.

[1] USGS Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA. E-mail: wlink@usgs.gov

This content downloaded from
152.3.34.254 on Thu, 02 Jan 2025 02:08:42 UTC
All use subject to https://about.jstor.org/terms