Sampling on the simplex

P.G.L. Porta Mana
Kavli Institute, Trondheim, Norway <piero.mana@ntnu.no>

14 July 2019; updated 25 July 2019

Two simple approaches to sampling general probability distributions defined over the simplex, especially with Monte Carlo algorithms and in high dimensions, are presented.

Note: Dear Reader & Peer, this manuscript is being peer-reviewed by you. Thank you.

Dedicated to Nerone. We still miss you every day.

The need to sample from a general distribution of probability over a simplex, that is, over a space of distributions, is more common than one might think. For example, it's necessary whenever we use exchangeable probability models (Dawid 2013; de Finetti 1930; 1937; Bernardo et al. 2000 § 4.2), ubiquitous in science owing to their intimate connection with reproducibility (JCGM 2012 §§ B.2.15–16; ISO 2006 §§ 3.3.6, 3.3.11; Poincaré 1992 ch. IX; Duhem 1914 ch. VI; de Finetti 1937; Jeffreys 1973 ch. IV; Popper 2005 ch. 1).

Sampling on an n-dimensional simplex, for example by Monte Carlo methods (Neal 1993; MacKay 2003 chs 29, 30; Hall 2014), is difficult because the allowed ranges of the sampled quantities are interdependent: denoting by $\boldsymbol{f} \coloneqq (f_0, \dots, f_n)$ a sampled distribution, we must have

$$\sum_{i=0}^{n} f_i = 1, \qquad 0 \leqslant f_i \leqslant 1, \quad \text{all } i, \tag{1}$$

so that n independent components, for example $\boldsymbol{F} \coloneqq (f_1, \dots, f_n)$, must satisfy

$$\sum_{j=1}^{n} f_j \leqslant 1. \tag{2}$$

The individual bounds $\{0 \leqslant f_i \leqslant 1\}$ are not difficult to handle with a change of variable, using a logistic function for example, because they are independent. It's the joint constraint (2) that's the problem. It needs to be tested for any proposed sample, and in high dimensions it leads to high rejection rates, because the n-dimensional simplex only occupies $1/n!$ of the volume of the unit hypercube $\{0 \leqslant f_j \leqslant 1\}$.

I didn't find solutions to this problem in the literature, except for an ingenious change of variables $\boldsymbol{z} \mapsto \boldsymbol{f}$, discussed by Betancourt (2012; Altmann et al. 2014), which maps the unit hypercube onto (but not into) the simplex:

$$\begin{aligned} f_1 &= 1 - z_1 \\ f_2 &= (1 - z_2)\, z_1 \\ f_3 &= (1 - z_3)\, z_2\, z_1 \\ &\cdots \\ f_n &= (1 - z_n)\, z_{n-1} \cdots z_1 \\ f_0 &\equiv z_n \cdots z_1. \end{aligned} \tag{3}$$

You can easily check that the quantities $\boldsymbol{z} \coloneqq (z_1, \dots, z_n)$ have independent bounds $\{0 \leqslant z_j \leqslant 1\}$. The Jacobian determinant of this transformation can also be computed. This transformation has a couple of practical disadvantages, however, owing to its recursive structure. The numerical relative imprecision of the z_j accumulates in the f_j as j increases, leading to an additional variability in the f_j with larger j, which manifests as an artefactual broadening of their marginal densities. The non-linear transformation $\boldsymbol{z} \mapsto \boldsymbol{f}$ moreover destroys the convex structure of the simplex; it heavily bends regions that are linear in $\boldsymbol{f}$, and can thus lead to a slower exploration of bent regions of high probability for some Markov-chain methods.

I want to propose two alternative approaches to the problem of interdependent bounds (2). These approaches also work very well in high dimensions. They are so simple – the first in particular – that they have probably been used before, so I'm not making any claim of originality. But I haven't seen them explicitly discussed in the literature, so it may be useful to advertise them. The main idea behind them is to give an additional dimension to the simplex.

Suppose we want to sample from the probability distribution

$$\begin{gathered} G(\boldsymbol{f})\, \mathrm{d}\boldsymbol{F} \\ f_0 \equiv 1 - \textstyle\sum_{j=1}^{n} f_j, \quad f_j \geqslant 0, \quad \sum_{j=1}^{n} f_j \leqslant 1, \end{gathered} \tag{4}$$

defined over the n-dimensional simplex. This distribution can be considered as the marginal distribution of the joint, separable distribution

$$L(r)\, G(\boldsymbol{f})\, \mathrm{d}r\, \mathrm{d}\boldsymbol{F}, \qquad r \geqslant 0, \quad f_j \geqslant 0, \quad \textstyle\sum_{j=1}^{n} f_j \leqslant 1, \tag{5}$$

where r is a positive quantity with a (regular) distribution of probability $L(r)\,\mathrm{d}r$. If we sample $(r, \boldsymbol{F})$ from this joint distribution, the samples of $\boldsymbol{F}$ will obviously come from $G(\boldsymbol{f})\,\mathrm{d}\boldsymbol{F}$.

First approach Now consider the quantities $\boldsymbol{x} \coloneqq (x_0, \dots, x_n)$ in the positive $(n+1)$-dimensional orthant $\{x_i \geqslant 0\}$, and the following transformation $\boldsymbol{x} \mapsto (r, \boldsymbol{F})$:

$$r = x_0 + \dots + x_n \tag{6a}$$

$$f_i = \frac{x_i}{x_0 + \dots + x_n}. \tag{6b}$$

You can prove, for example using the inverse transformation $(r, \boldsymbol{F}) \mapsto \boldsymbol{x}$, that the Jacobian determinant of $\boldsymbol{x} \mapsto (r, \boldsymbol{F})$ is r^{-n}. The joint probability distribution (5) can thus be written as

$$M[r(\boldsymbol{x})]\, G[\boldsymbol{f}(\boldsymbol{x})]\, \mathrm{d}\boldsymbol{x}, \qquad x_i \geqslant 0, \tag{7}$$

where the Jacobian determinant has been absorbed into $M(r) \coloneqq L(r)/r^n$. Most important, *the x_i have independent bounds*.

We can therefore easily sample from the distribution (7) for the quantities $\boldsymbol{x}$, on a space with no interdependent bounds, and then construct samples of $\boldsymbol{f}$ from $\boldsymbol{x}$ via the transformation (6b). The resulting $\boldsymbol{f}$ samples are drawn from our distribution of interest (4). The r samples are drawn from $L(r)\,\mathrm{d}r$ but they aren't calculated (unless this quantity is of interest for some reason). Note that L and therefore M are completely arbitrary; we can choose a normal distribution as the latter, for example. Also, the numerical relative imprecision of every f_i is roughly twice that of x_i: no accumulation of imprecision occurs.

If our probability distribution (4) has its main mass enough far away from the faces of the simplex, it's practical to sample the $\boldsymbol{x}$ directly – despite their lower bound – by using a density function $M(r)$ with a sharp peak at high values of r, say $r \approx n$. This way, even if some typical f_i may be close to zero, the corresponding x_i will be of order unity, enough far away from its lower bound to be sampled directly without worrying about high rejection rates.

Second approach If we prefer to have completely unbounded quantities, we can still consider the probability distribution (5) extended with r, but use the following transformation $\boldsymbol{t} \mapsto (r, \boldsymbol{F})$ (cf. MacKay 2003 § 23.5, p. 316):

$$r = t_0 + \cdots + t_n \tag{8a}$$

$$f_i = \frac{\exp(t_i)}{\exp(t_0) + \cdots + \exp(t_n)}. \tag{8b}$$

You can prove, for example by induction on n, that the Jacobian determinant of the transformation above is

$$\left|\det \frac{\partial(r, \boldsymbol{F})}{\partial \boldsymbol{t}}\right| = (n+1) \prod_{i=0}^{n} f_i \equiv (n+1)\ \exp(r)\ \left[\textstyle\sum_i \exp(t_i)\right]^{-(n+1)}. \tag{9}$$

The joint probability distribution (5) can thus be written as

$$(n+1)\ L[r(\boldsymbol{t})]\ G[\boldsymbol{f}(\boldsymbol{t})]\ \prod_{i=0}^{n} f_i(\boldsymbol{t})\ \mathrm{d}\boldsymbol{t}, \qquad \boldsymbol{t} \in \mathbf{R}^{n+1}. \tag{10}$$

We can therefore sample from the distribution (10), defined over $\mathbf{R}^{n+1}$, and then construct samples for $\boldsymbol{f}$ from $\boldsymbol{t}$ via the transformation (8b). Also in this case the resulting $\boldsymbol{f}$ samples are drawn from our distribution of interest (4). The probability density function $L(r)$ is again completely arbitrary, and can be conveniently chosen as a normal sharply peaked at $r \approx 0$; it's also possible to include the Jacobian term $\exp(r)$ from (9) in it.

A final remark: instead of the transformation (8a) we could also have considered $r = \sum_i \exp(\boldsymbol{t}_i)$. This transformation is inconvenient, however: surfaces of constant r would then be heavily bent in $\boldsymbol{t}$ coordinates, leading to slower convergence for some Markov-chain methods. It's also possible to use the transformation (8b) without r, by using the independent quantities $(t_1, \dots, t_n)$ and defining $t_0 \coloneqq -\sum_{j=1}^{n} t_j$. The Jacobian determinant in this case is still (9); the proof is left again as an exercise for you.

Thanks

. . . to the Kavli Foundation and the Centre of Excellence scheme of the Research Council of Norway (Yasser Roudi group) for financial support. To Iván Davidovich for a useful discussion. To Mari & Miri for continuous

encouragement and affection. To Buster Keaton and Saitama for filling life with awe and inspiration. To the developers and maintainers of LaTeX, Emacs, AUCTeX, Open Science Framework, Python, Inkscape, Sci-Hub for making a free and unfiltered scientific exchange possible.

Bibliography

('de *X*' is listed under D, 'van *X*' under V, and so on, regardless of national conventions.)

Altmann, Y., McLaughlin, S., Dobigeon, N. (2014): *Sampling from a multivariate Gaussian distribution truncated on a simplex: a review*. IEEE Workshop Stat. Signal Process. **2014**, 113–116. http://dobigeon.perso.enseeiht.fr/papers/Altmann_IEEE_SSP_2014.pdf. Note: the 'covariance' matrices used in § 5 of this paper are not positive-definite.

Bernardo, J.-M., Smith, A. F. (2000): *Bayesian Theory*, reprint. (Wiley, New York). First publ. 1994.

Betancourt, M. (2012): *Cruising the simplex: Hamiltonian Monte Carlo and the Dirichlet distribution*. Am. Inst. Phys. Conf. Proc. **1443**, 157–164.

Damien, P., Dellaportas, P., Polson, N. G., Stephens, D. A., eds. (2013): *Bayesian Theory and Applications*. (Oxford University Press, Oxford).

Dawid, A. P. (2013): *Exchangeability and its ramifications*. In: Damien, Dellaportas, Polson, Stephens (2013), ch. 2, 19–29.

de Finetti, B. (1930): *Funzione caratteristica di un fenomeno aleatorio*. Atti Accad. Lincei: Sc. Fis. Mat. Nat. **IV**^{5}, 86–133. http://www.brunodefinetti.it/Opere.htm.

— (1937): *La prévision : ses lois logiques, ses sources subjectives*. Ann. Inst. Henri Poincaré **7**^{1}, 1–68. Transl. in Kyburg, Smokler (1980), pp. 53–118, by Henry E. Kyburg, Jr.

Duhem, P. (1914): *La Théorie Physique : son objet – sa structure*, 2nd ed. (Marcel Rivière, Éditeur, Paris). http://virtualbooks.terra.com.br/freebook/fran/la_theorie_physique.htm. First publ. 1906. Transl. as Duhem (1991).

— (1991): *The Aim and Structure of Physical Theory*, transl. of the 2nd ed. (Princeton University Press, Princeton). Transl. of Duhem (1914) by P. P. Wiener. First publ. in French 1906.

Hall, B. (2014): *MCMC algorithms*. https://m-clark.github.io/docs/ld_mcmc/. See also https://github.com/LaplacesDemonR/LaplacesDemon, https://web.archive.org/web/20141224051720/http://www.bayesian-inference.com/index.

ISO (2006): *ISO 3534-2:2006: Statistics – Vocabulary and symbols – Part 2: Applied statistics*. International Organization for Standardization.

JCGM (2012): *Guide to the Expression of Uncertainty in Measurement*, corr. version. Joint Committee for Guides in Metrology (JCGM), BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML. http://www.bipm.org/en/publications/guides/gum.html. Includes various supplements. First publ. 1993.

Jeffreys, H. (1973): *Scientific Inference*, 3rd ed. (Cambridge University Press, Cambridge). First publ. 1931.

Kyburg Jr., H. E., Smokler, H. E., eds. (1980): *Studies in Subjective Probability*, 2nd ed. (Robert E. Krieger, Huntington, USA). First publ. 1964.

MacKay, D. J. C. (2003): *Information Theory, Inference, and Learning Algorithms*. (Cambridge University Press, Cambridge). http://www.inference.phy.cam.ac.uk/mackay/itila/. First publ. 1995.

Neal, R. M. (1993): *Probabilistic inference using Markov chain Monte Carlo methods*. Tech. rep. CRG-TR-93-1. (University of Toronto, Toronto). http://www.cs.utoronto.ca/~radford/review.abstract.html, http://omega.albany.edu:8008/neal.pdf.

Poincaré, H. (1905): *Science and Hypothesis*. (Walter Scott, London). Transl. of Poincaré (1992) by W. J. Greenstreet; with a Preface by J. Larmor. First publ. 1902. Partly repr. in Poincaré (1958).

— (1958): *The Value of Science*. (Dover, New York). Authorized transl. with an introduction by G. B. Halsted. First publ. 1913.

— (1992): *La science et l'hypothèse*. (Éditions de la Bohème, Rueil-Malmaison, France). http://gallica.bnf.fr/document?O=N026745. First publ. 1902; transl. as Poincaré (1905).

Popper, K. (2005): *The Logic of Scientific Discovery*, reprint. (Routledge, London). First publ. in German 1935.