Dirichlet draws are sparse with high probability

Matus Telgarsky

Abstract

This note provides an elementary proof of the folklore fact that draws from a Dirichlet distribution (with parameters less than 1) are typically sparse (most coordinates are small).

1 Bounds

Let $\mathrm{Dir}(\alpha)$ denote a Dirichlet distribution with all parameters equal to α.

Theorem 1.1. *Suppose $n \geq 2$ and $(X_1, \ldots, X_n) \sim \mathrm{Dir}(1/n)$. Then, for any $c_0 \geq 1$ satisfying $6c_0 \ln(n) + 1 < 3n$,*

$$\Pr\left[\left|\left\{i : X_i \geq \frac{1}{n^{c_0}}\right\}\right| \leq 6c_0 \ln(n)\right] \geq 1 - \frac{1}{n^{c_0}}.$$

The parameter is taken to be $1/n$, which is standard in machine learning. The above theorem states that (with high probability) as the exponent on the sparsity threshold grows linearly $(n^{-1}, n^{-2}, n^{-3}, \ldots)$, the number of coordinates above the threshold cannot grow faster than linearly $(6\ln(n), 12\ln(n), 18\ln(n), \ldots)$.

The above statement can be parameterized slightly more finely, exposing more tradeoffs than just the threshold and number of coordinates.

Theorem 1.2. *Suppose $n \geq 1$ and $c_1, c_2, c_3 > 0$ with $c_2 \ln(n) + 1 < 3n$, and $(X_1, \ldots, X_n) \sim \mathrm{Dir}(c_1/n)$; then*

$$\Pr\big[|\{i : X_i \geq n^{-c_3}\}| \leq c_2 \ln(n)\big] \geq 1 - \frac{1}{e^{1/3}}\left(\frac{1}{n}\right)^{\frac{c_2}{3} - c_1 c_3} - \frac{1}{e^{4/9}}\left(\frac{1}{n}\right)^{\frac{4c_2}{9}}.$$

The natural question is whether the factor $\ln(n)$ is an artifact of the analysis; simulation experiments with Dirichlet parameter $\alpha = 1/n$, summarized in Figure 1a, exhibit both the $\ln(n)$ term, and the linear relationship between sparsity threshold and number of coordinates exceeding it.

The techniques here are loose when applied to the case $\alpha = o(1/n)$. In particular, Figure 1b suggests $\alpha = 1/n^2$ leads to a single nonsmall coordinate with high probability, which is stronger than what is captured by the following theorem.

Theorem 1.3. *Suppose $n \geq 3$ and $(X_1, \ldots, X_n) \sim \mathrm{Dir}(1/n^2)$; then*

$$\Pr\big[|\{i : X_i \geq n^{-2}\}| \leq 5\big] \geq 1 - e^{2/e-2} - e^{-8/3} \geq 0.64.$$

Moreover, for any function $g : \mathbb{Z}_{++} \to \mathbb{R}_{++}$ and any n satisfying $1 \leq \ln(g(n)) < 3n - 1$,

$$\Pr\big[|\{i : X_i \geq n^{-2}\}| \leq \ln(g(n))\big] \geq 1 - e^{2/e-1/3}\left(\frac{1}{g(n)}\right)^{1/3} - e^{-4/9}\left(\frac{1}{g(n)}\right)^{4/9}.$$

(Take for instance g to be the inverse Ackermann function.)

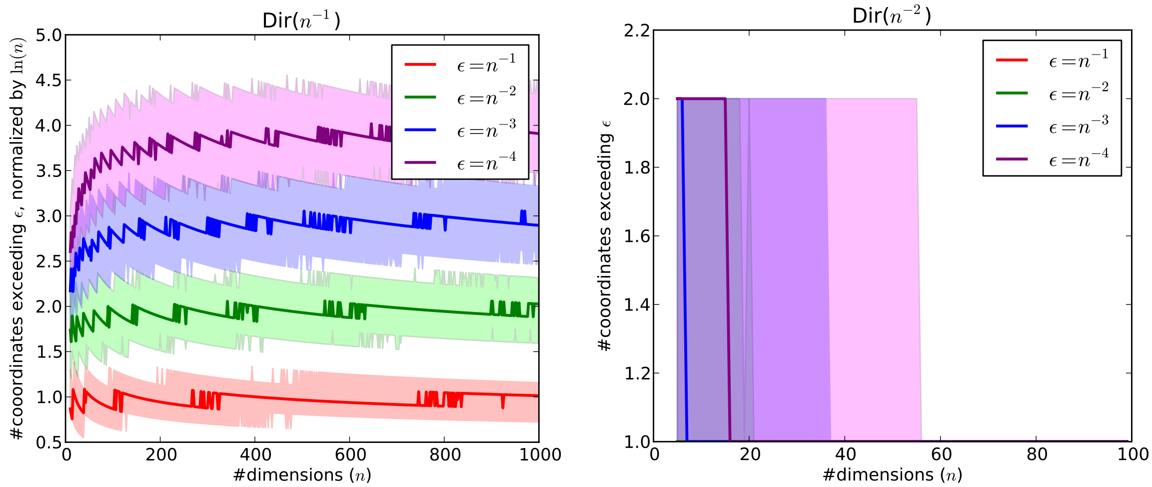

Figure 1: For each Dirichlet parameter choice $\alpha \in \{n^{-1}, n^{-2}\}$ and each number of dimensions n (horizontal axis), 1000 Dirichlet distributions were sampled. For each trial, the number of coordinates exceeding each of 4 choices of threshold were computed. In the case of $\alpha = n^{-1}$, these counts were then scaled by $\ln(n)$ to better coordinate with the suggested trends in Theorems 1.1 and 1.2. Finally, these counts values (for each (n, ϵ)) were converted into quantile curves (25%–75%).

2 Proofs

Theorems 1.1 to 1.3 are established via the following lemma.

Lemma 2.1. *Let reals $\epsilon \in (0, 1]$ and $\alpha > 0$ and positive integers k, n be given with $k + 1 < 3n$. Let $(X_i, \ldots, X_n) \sim \text{Dir}(\alpha)$. Then*

$$\Pr\Big[|\{i : X_i \geq \epsilon\}| \leq k\Big] \geq 1 - \epsilon^{-n\alpha} e^{-(k+1)/3} - e^{-4(k+1)/9}.$$

The proof avoids dependencies between the coordinates of a Dirichlet draw via the following alternate representation. Throughout the rest of this section, let Gamma(α) denote a Gamma distribution with parameter α.

Lemma 2.2. *(See for instance Balakrishnan and Nevzorov (2003, Equation 27.17).) Let $\alpha > 0$ and $n \geq 1$ be given. If $(X_1, \ldots, X_n) \sim \text{Dir}(\alpha)$ and $\{Y_i\}_{i=1}^n$ are n i.i.d. copies of* Gamma(α)*, then*

$$(X_1, \ldots, X_n) \stackrel{d}{=} \left\{ \frac{Y_i}{\sum_{i=1}^n Y_i} \right\}.$$

Before turning to the proof of Lemma 2.1, one more lemma is useful, which will allow a control of the Gamma distribution's cdf.

Lemma 2.3. *For any $\alpha > 0$, $c \geq 0$, and $z \geq 1$,*

$$\Pr[\text{Gamma}(\alpha) \leq zc] \leq z^\alpha \Pr[\text{Gamma}(\alpha) \leq c].$$

Proof. Since $e^{-zx} \le e^{-x}$ for every $x \ge 0$ and $z \ge 1$,

$$\begin{aligned}
\Pr[\mathrm{Gamma}(\alpha) \le zc] &= \frac{1}{\Gamma(\alpha)} \int_0^{zc} e^{-x} x^{\alpha-1} dx \\
&= \frac{1}{\Gamma(\alpha)} \int_0^{c} e^{-zx} (zx)^{\alpha-1} z dx \\
&\le \frac{z^\alpha}{\Gamma(\alpha)} \int_0^{c} e^{-x} x^{\alpha-1} dx \\
&= z^\alpha \Pr[\mathrm{Gamma}(\alpha) \le c]. \qquad \square
\end{aligned}$$

Proof of Lemma 2.1. Since $z \mapsto \Pr[\mathrm{Gamma}(\alpha) \ge z]$ is continuous and has range $[0,1]$, choose $c \ge 0$ so that

$$\Pr[\mathrm{Gamma}(\alpha) > c] = \Pr[\mathrm{Gamma}(\alpha) \ge c] = \frac{k+1}{3n}, \tag{2.4}$$

where $(k+1)/(3n) < 1$. By this choice and Lemma 2.3,

$$\Pr[\mathrm{Gamma}(\alpha) \le c/\epsilon] \le \epsilon^{-\alpha} \Pr[\mathrm{Gamma}(\alpha) \le c] = \epsilon^{-\alpha}\left(1 - \frac{k+1}{3n}\right) \le \epsilon^{-\alpha} e^{-(k+1)/(3n)}. \tag{2.5}$$

Now let $\{Y_i\}_{i=1}^n$ be n i.i.d. copies of Gamma(α). Define the events

$$A := [\exists i \in [n] \centerdot Y_i \ge c/\epsilon] \qquad \text{and} \qquad B := [|\{i \in [n] : Y_i \le c\}| \ge n-k]\,.$$

The remainder of the proof will establish a lower bound on $\Pr(A \wedge B)$. To see that this finishes the proof, define $S := \sum_i Y_i$; since event A implies that $S \ge c/\epsilon$, it follows that $Y_i \le c$ implies $Y_i/S \le \epsilon$. Consequently, events A and B together imply that $Y_i/S \le \epsilon$ for at least $n-k$ choices of i. By Lemma 2.2, it follows that $\Pr(A \wedge B)$ is a lower bound on the event that a draw from Dir(α) has at least $n-k$ coordinates which are at most ϵ.

Returning to task, note that

$$\Pr(A \wedge B) = 1 - \Pr(\neg A \vee \neg B) \ge 1 - \Pr(\neg A) - \Pr(\neg B). \tag{2.6}$$

To bound the first term, by eq. (2.5),

$$\Pr(\neg A) = \Pr[\forall i \in [n] \centerdot Y_i < c/\epsilon] = \Pr[Y_1 \le c/\epsilon]^n \le \epsilon^{-n\alpha} e^{-(k+1)/3}. \tag{2.7}$$

For the second term, define indicator random variables $Z_i := [Y_i > c]$, whereby

$$\mathbb{E}(Z_i) = \Pr[Z_i = 1] = \Pr[Y_i > c] = \Pr[Y_i \ge c] = \frac{k+1}{3n}.$$

Then, by a multiplicative Chernoff bound (Kearns and Vazirani, 1994, Theorem 9.2),

$$\Pr(\neg B) = \Pr[|\{i \in [n] : Y_i > c\}| \ge k+1] = \Pr\left[\sum_i Z_i \ge 3n\mathbb{E}(Z_i)\right] \le \exp(-4n\mathbb{E}(Z_i)/3). \tag{2.8}$$

Inserting (2.7) and (2.8) into the lower bound on $\Pr(A \wedge B)$ in (2.6),

$$\Pr(A \wedge B) \ge 1 - \epsilon^{-n\alpha} e^{-(k+1)/3} - e^{-4(k+1)/9}. \qquad \square$$

Proof of Theorem 1.2. Instantiate Lemma 2.1 with $k = c_2 \ln(n)$, $\alpha = c_1/n$, and $\epsilon = n^{-c_3}$. $\square$

Proof of Theorem 1.1. Instantiate Theorem 1.2 with $c_1 = 1$, $c_2 = 6c_0$, $c_3 = c_0$, and note

$$\frac{1}{e^{1/3}}\left(\frac{1}{n}\right)^{c_0} + \frac{1}{e^{4/9}}\left(\frac{1}{n}\right)^{\frac{8c_0}{3}} \leq \frac{1}{n^{c_0}}\left(\frac{1}{e^{1/3}} + \frac{1}{e^{4/9}}\left(\frac{1}{2}\right)^{\frac{5c_0}{3}}\right) \leq \frac{1}{n^{c_0}}. \qquad \square$$

Proof of Theorem 1.3. Define the function $f(z) := z^{-z}$ over $(0,\infty)$. Note that $f'(z) = -(\ln(z)+1)z^{-z}$, which is positive for $z < 1/e$, zero at $z = 1/e$, and negative thereafter; consequently, $\sup_{z\in(0,\infty)} f(z) = f(1/e) = e^{1/e}$. As such, instantiating Lemma 2.1 with $\epsilon = n^{-2}$, $\alpha = n^{-3}$, and any $k < 3n-1$ gives

$$\begin{aligned}\Pr\big[|\{i : X_i \geq n^{-2}\}| \leq k\big] &\geq 1 - n^{2/n}e^{-(k+1)/3} - e^{-4(k+1)/9} \\ &\geq 1 - e^{2/e}e^{-(k+1)/3} - e^{-4(k+1)/9}.\end{aligned}$$

Plugging in $k \in \{5, \ln(g(n))\}$ gives the two bounds. $\square$

Acknowledgement

The author thanks Anima Anandkumar and Daniel Hsu for relevant discussions.

References

N. Balakrishnan and V. B. Nevzorov. *A primer on statistical distributions.* Wiley-Interscience, 2003.

M. J. Kearns and U. V. Vazirani. *An introduction to computational learning theory.* MIT Press, Cambridge, MA, USA, 1994.