Discussion of ‘A Gibbs sampler for a class of random convex polytopes’

Guanyang Wang (Rutgers Stats)

joint work with

Persi Diaconis (Stanford Math & Stats)

Joint Statistical Meetings (JSM) 2021

Outline

A 50-year-old story

The Donkey walk

A 50-year-old story

“When one of the two discussants was a beginning graduate student at Harvard (1971), Art Dempster called him in to suggest a possible thesis topic: ‘Find ways to do the computations required for upper and lower probabilities’. This did not work out at the time but triggered a lifetime's interest. It is inspiring to have tracked his efforts over a 50 year period.”

Outline

A 50-year-old story

The Donkey walk

From Dempster–Shafer theory to the Donkey walk

The Jacob-Gong-Edlefsen-Dempster (JGED) algorithm when $K = 2$:

$$Y^{(t)} = \text{Beta}(N_0, 1) Z^{(t-1)} \tag{1}$$

$$Z^{(t)} = Y^{(t)} + \text{Beta}(1, N_1)(1 - Y^{(t)}) \tag{2}$$

where N_1, N_2 are two fixed positive integers.

From Dempster–Shafer theory to the Donkey walk

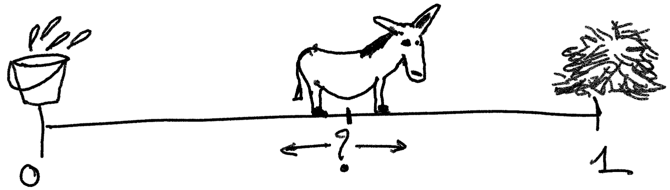
The Jacob-Gong-Edlefsen-Dempster (JGED) algorithm when $K = 2$:

$$Y^{(t)} = \text{Beta}(N_0, 1) Z^{(t-1)} \tag{3}$$

$$Z^{(t)} = Y^{(t)} + \text{Beta}(1, N_1)(1 - Y^{(t)}) \tag{4}$$

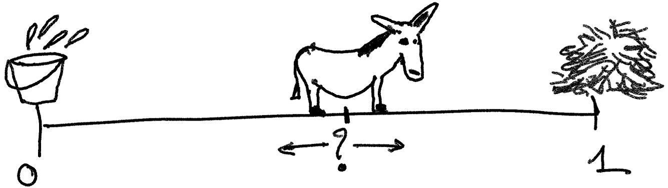
where N_1, N_2 are two fixed positive integers.

The Buridan's Donkey model [4][2] (drawing credit to Pierre Jacob):



The Donkey walk

The Buridan's Donkey model:



- The donkey moves alternatively in the direction of 0 (water), 1 (grass), 0 (water), 1 (grass) ...
- If the donkey choose a uniformly random point on the left, and then a uniformly random point on the left, then the walk is equivalent to the JGED algorithm when $N_0 = N_1 = 1$.
- In general, if the donkey chooses a Beta distribution, then it is precisely the JGED algorithm for two categories.

The Donkey's position after infinitely many(?) steps

- It is not too hard to show $Z^{(\infty)} \sim \text{Beta}(N_0 + 1, N_1)$ distribution.

The Donkey's position after infinitely many(?) steps

- It is not too hard to show $Z^{(\infty)} \sim \text{Beta}(N_0 + 1, N_1)$ distribution.
- An Upper Bound: In the JGED [1] paper, the authors prove

$$\mathcal{W}_1(P^t(z, \cdot), \text{Beta}(N_0 + 1, N_1)) \leq \left(\frac{N_0}{N_0 + 1} \cdot \frac{N_1}{N_1 + 1} \right)^t \cdot \mathbb{E}|Z - z|,$$

where $Z \sim \text{Beta}(N_0 + 1, N_1)$.

The Donkey's position after infinitely many(?) steps

- It is not too hard to show $Z^{(\infty)} \sim \mathsf{Beta}(N_0+1, N_1)$ distribution.
- An Upper Bound: In the JGED [1] paper, the authors prove

$$\mathcal{W}_1(P^t(z,\cdot), \mathsf{Beta}(N_0+1, N_1)) \leq \left(\frac{N_0}{N_0+1} \cdot \frac{N_1}{N_1+1}\right)^t \cdot \mathbb{E}|Z-z|,$$

  where $Z \sim \mathsf{Beta}(N_0+1, N_1)$.
- A Lower Bound: In view of the Kantorovich-Rubinstein Duality, we have a lower bound which matches the upper bound's convergence rate:

$$\mathcal{W}_1(P^t(z,\cdot), \mathsf{Beta}(N_0+1, N_1)) \geq \left(\frac{N_0}{N_0+1} \cdot \frac{N_1}{N_1+1}\right)^t \cdot |\mathbb{E}Z-z|.$$

The Donkey's position after infinitely many(?) steps

- The convergence rate of the Donkey walk (in $\mathcal{W}_1$ distance) is precisely $(\frac{N_0}{N_0+1} \cdot \frac{N_1}{N_1+1})$.
- With a little extra effort, we derive the exact convergence speed for the Donkey walk under the worst-case sceneario:

$$\sup_{z\in[0,1]} \mathcal{W}_1(P^t(z,\cdot), \mathsf{Beta}(N_0+1, N_1)) = \frac{\max\{N_0+1, N_1\}}{N_0+N_1+1}\left(\frac{N_0}{N_0+1} \cdot \frac{N_1}{N_1+1}\right)^t.$$

The Donkey's position after infinitely many(?) steps

- The convergence rate of the Donkey walk (in $\mathcal{W}_1$ distance) is precisely $(\frac{N_0}{N_0+1} \cdot \frac{N_1}{N_1+1})$.
- With a little extra effort, we derive the exact convergence speed for the Donkey walk under the worst-case sceneario:

$$\sup_{z\in[0,1]} \mathcal{W}_1(P^t(z,\cdot),\mathrm{Beta}(N_0+1,N_1)) = \frac{\max\{N_0+1,N_1\}}{N_0+N_1+1}\left(\frac{N_0}{N_0+1}\cdot\frac{N_1}{N_1+1}\right)^t.$$

- Follow-up question: How about the total-variation distance?

The Donkey's position after infinitely many(?) steps

- For simplicity assuming $N_0 = N_1 = 1$, then we have:

$$\left|\frac{2}{3} - z\right| \left(\frac{1}{4}\right)^t \le d_{\mathsf{TV}}(P^t(z, \cdot), \mathsf{Beta}(2,1)) \le 12 \times \left(\frac{1}{4}\right)^t$$

  - The TV distance converges at the same rate as the Wasserstein distance.
- The constant on the RHS can be improved to $4 \times (\frac{1}{4})^t$, pointed out by Yucong Ma (Harvard Stats).
- The idea is based on the one-shot coupling [3].

Thanks & Congrats!

- Check out Pierre's blog post on Dempster's analysis and donkeys. "https://statisfaction.wordpress.com/2021/03/23/dempsters-analysis-and-donkeys"

Pierre E. Jacob, Ruobin Gong, Paul T. Edlefsen, and Arthur P. Dempster.
A Gibbs Sampler for a Class of Random Convex Polytopes.
Journal of the American Statistical Association, 0(0):1–12, 2021.

Gérard Letac.
Donkey walk and Dirichlet distributions.
Statistics & probability letters, 57(1):17–22, 2002.

Neal Madras and Deniz Sezer.
Quantitative bounds for Markov chain convergence: Wasserstein and total variation distances.
Bernoulli, 16(3):882–908, 2010.

Jordan Stoyanov and Christo Pirinsky.
Random motions, classes of ergodic Markov chains and beta distributions.
Statistics & probability letters, 50(3):293–304, 2000.