# Bottom rank statistics for brownian reshuffling

Ayan Shah supervised by A/Prof Mario Kieburg

The University of Melbourne

## Motivation

Rankings are ubiquitous tools for analysing extremes, from the largest companies to the best universities. However, when the underlying data is stochastic in nature, these hierarchies are not fixed but evolve continuously.
Recent empirical work [1] on the top-100 wealth lists of various countries and the world found that the overlap ratio collapses onto a single curve, suggesting universal behaviour. This phenomenon was theoretically modelled in [2] using a Brownian motion process on the positive real line. In contrast, we are interested in the order statistics of the smallest participants and derive an analytical curve for the bottom-n order statistics.

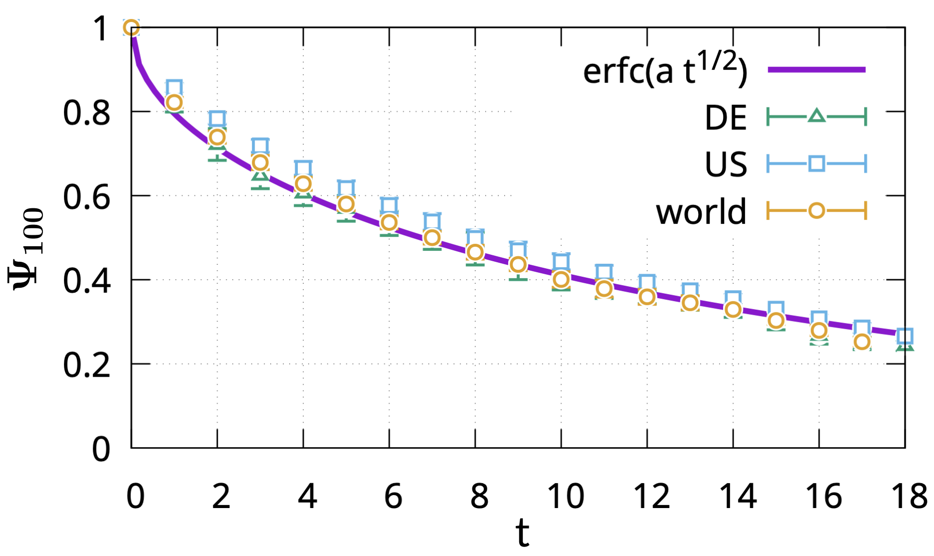


Figure 1. Data points represent empirical data presented in [1]. The time t is measured in years. The solid line represents the theoretical curve for a model of Brownian reshuffling derived in [2].

## Setting

### Overlap Ratio

For a system with $N$ ranked elements the overlap ratio for bottom rank statistics $\Omega_n$, is the ratio of elements that remain in the bottom-n ranks after some time $t$.

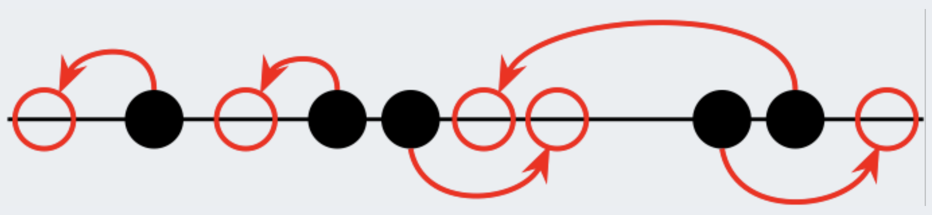

Figure 2. Schematic illustration of particles moving stochastically along a line, depicting how order statistics change over time. The black filled circles represent the initial positions, while the red circles represent the position after time t. The overlap ratios in this example are $\Omega_1 = 1$, $\Omega_2 = 1$, $\Omega_3 = 2/3$, $\Omega_4 = 3/4$ and $\Omega_5 = 1$. Image taken from [2].

The average overlap ratio is explicitly given by

$$\mathbb{E}(\Omega_n) = \frac{1}{n}\sum_{j,k=1}^{n} \mathbb{P}(k,j,t) \tag{1}$$

where $\mathbb{P}(k,j,t)$ is the probability the $j^{th}$ smallest particle will be the $k^{th}$ smallest particle after time t. Hence, the average overlap ratio is the probability that the $n$ lowest ranked elements remain the $n$ lowest ranked after time t.

### Brownian Motion on $\mathbb{R}^+$

We assume that each particle performs an independent Brownian motion with diffusion constant $\sigma^2/2$ and drift $\mu < 0$ towards a reflective hard wall at the origin $x = 0$. The probability density $p(x,t)$ evolves according to the Fokker-Planck equation

$$\partial_t p(x,t) = -\mu\partial_x p(x,t) + \frac{\sigma^2}{2}\partial_x^2 p(x,t). \tag{2}$$

The reflective hard wall at $x = 0$ is enforced by the Robin boundary condition

$$-\mu p(0,t) + \frac{\sigma^2}{2}\partial_x p(0,t) = 0. \tag{3}$$

## Bottom Rank Statistics with Overlap

We consider $N$ i.i.d. particles on the real line with initial positions $\vec{y}' = (y_1', y_2', .., y_N') \in \mathbb{R}_+^N$ and final positions $\vec{x}' = (x_1', x_2', .., x_N') \in \mathbb{R}_+^N$ after time t. The dynamics of each particle is determined by an arbitrary probability density given by $f'(x_i', y_i')$. The resulting transition probabilities are

$$\mathbb{P}(k,j) = N\int_{-\infty}^{\infty} dy' \int_{-\infty}^{\infty} dx'\, f'(x',y') \oint_{|z|=1} \frac{dz}{2\pi i z^k} \oint_{|w|=1} \frac{dw}{2\pi i w^j} \times \left(zwM_{00}'(x',y') + wM_{10}'(x',y') + zM_{01}'(x',y') + M_{11}'(x',y')\right)^{N-1} \tag{4}$$

with

$$M_{ij}'(x',y') = \int_{I_i} d\tilde{x} \int_{J_j} d\tilde{y}\, f'(\tilde{x},\tilde{y}) \tag{5}$$

where $I_0 = [0, x']$, $I_1 = (x', \infty)$, $J_0 = [0, y']$ and $J_1 = (y', \infty)$.

## Probability Integral Transform

To simplify the probabilities (4) and (5) we unfold the random variables $\vec{x}'$ and $\vec{y}'$,

$$x = \int_0^{x'} d\tilde{x}\, p_1(\tilde{x}) \quad \text{and} \quad y = \int_0^{y'} d\tilde{y}\, p_0(\tilde{y}) \tag{6}$$

where $p_1(x') = M_{00}'(x',y') + M_{01}'(x',y')$ and $p_0(y') = M_{00}'(x',y') + M_{10}'(x',y')$ are the marginal distributions of $x'$ and $y'$, respectively. This transformation simplifies the average overlap ratio to

$$\mathbb{E}(\Omega_n) = \frac{N}{n}\int_0^1 dy \int_0^1 dx \oint_{|z|=\frac{1}{2}} \frac{dz}{2\pi i(z-1)z^n} \oint_{|w|=\frac{1}{2}} \frac{dw}{2\pi i(w-1)w^n} \partial_x\partial_y M_{00}\Big((1-z)(1-w)M_{00}(x,y) + y(w-1) + x(z-1) + 1\Big)^{N-1} \tag{7}$$

where $M_{00}(x,y) = M_{00}'(x',y')$ and the Fundamental Theorem of Calculus tells us $f(x,y) = \partial_x\partial_y M_{00}(x,y)$

### Scale of Positions

For large $N$, the smallest particles concentrate at the origin on the scale $1/N$ so that we rescale $x \to x/N$ and $y \to y/N$ and obtain

$$\mathbb{E}(\Omega_n) = \frac{1}{n}\int_0^N dy \int_0^N dx \oint_{|z|=1/2} \frac{dz}{2\pi i(z-1)z^n} \times \oint_{|w|=1/2} \frac{dw}{2\pi i(w-1)w^n} \partial_x\partial_y \left[NM_{00}\left(\frac{x}{N},\frac{y}{N}\right)\right] \times \left(1 + \frac{1}{N}\left[(z-1)x + (w-1)y + (1-z)(1-w)NM_{00}\left(\tfrac{x}{N},\tfrac{y}{N}\right)\right]\right)^{N-1} \tag{8}$$

## Time Scale

The solution of the boundary value problem (2) and (3) has been solved in [3, 4] and yields the following heat kernel by setting $\alpha = -2\mu/\sigma^2 > 0$

$$W(x,y,t) = \frac{1}{\sqrt{2\pi t}}\exp\left(-\left[\frac{2(y-x)-\alpha t}{\sqrt{8t}}\right]^2\right) + \frac{\alpha e^{-\alpha x}}{2}\mathrm{erfc}\left[\frac{2(y+x)-\alpha t}{\sqrt{8t}}\right] + \frac{e^{-\alpha x}}{\sqrt{2\pi t}}\exp\left(-\left[\frac{2(y+x)-\alpha t}{\sqrt{8t}}\right]^2\right). \tag{9}$$

The first term describes the Brownian motion towards the wall and the other two terms represent the reflection.

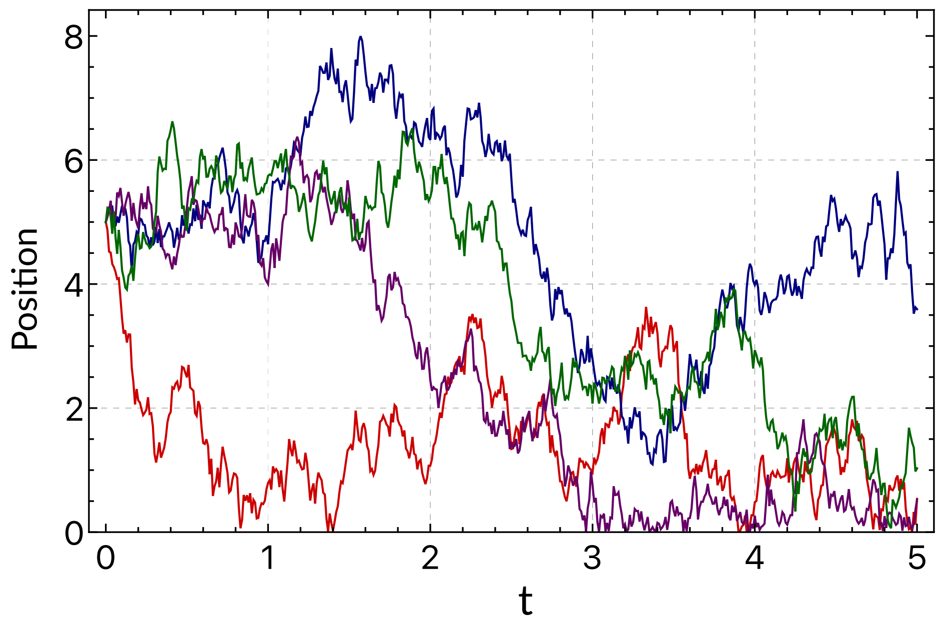


Figure 3. Simulated sample paths of Brownian motion with drift towards a reflective hard wall at the origin.

Applying (8) to (9) we arrive at

$$\mathbb{E}(\Omega_n) \approx \frac{n}{2N}\left(1 + \mathrm{erf}\left[\alpha\sqrt{\frac{t}{8}}\right] + \frac{8}{\sqrt{2\pi t}\alpha}e^{-\frac{\alpha^2 t}{8}}\right) + \mathcal{O}(N^{-2}) \tag{10}$$

as $N \to \infty$. Hence, as the number of particles in the system approaches infinity, the probability of remaining in the bottom-$n$ ranks vanishes.

In contrast to top rank statistics, the ordering of the smallest particles changes much more frequently due to the drift towards the reflective hard wall at the origin, see Figs. 3 and 4.

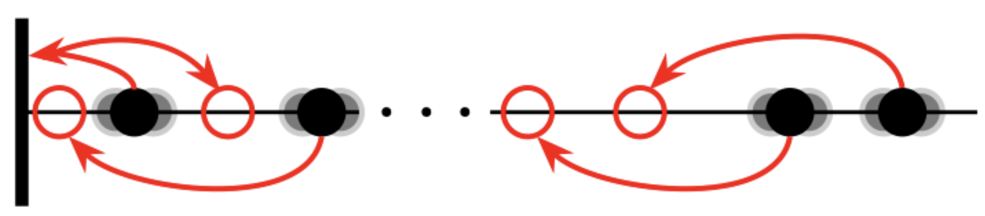

Figure 4. Schematic Illustration of particles performing Brownian motion on $\mathbb{R}_+$ with drift towards reflective wall at origin. The combined effects of drift and the reflective wall induce frequent reshuffling of bottom-rank statistics, whereas top-rank statistics exhibit greater stability.

The scale where the overlap ratio drops to zero is of the order $1/N$. Thus, we rescale the time as follows

$$t = \frac{n^2}{N^2\alpha^2}\tau. \tag{11}$$

The new time scale is dependent on the number of particles in the system changing the large $N$ expansion.

## Asymptotic limit

The limit $\lim_{N\to\infty}\mathbb{E}(\Omega_n)$ can be performed as follows:

1. **Change of Variables.** We change to centre-of-mass and rescaled (with respect to the centre of mass) relative coordinates

$$R = \frac{x+y}{n\sqrt{2\tau}} \quad \text{and} \quad r = \frac{x-y}{x+y}. \tag{12}$$

2. **Calculating Limits.** The limits of the density and the corresponding probability are

$$F(R,r) = n\tau R \lim_{N\to\infty} \partial_x\partial_y NM_{00}\left(\frac{x}{N},\frac{y}{N}\right) = \frac{R\sqrt{\tau}}{\sqrt{2\pi}}\left(e^{-R^2} + e^{-R^2r^2}\right), \tag{13}$$

$$G(R,r) = \lim_{N\to\infty} NM_{00}\left(\frac{x}{N},\frac{y}{N}\right) = n\sqrt{\frac{\tau}{2}}\left(R\,\mathrm{erf}[R] - Rr\,\mathrm{erf}[Rr] + \frac{e^{-R^2} - e^{-R^2r^2}}{\sqrt{\pi}}\right) \tag{14}$$

3. **Short Lists.** The average overlap ratio is then

$$\lim_{N\to\infty}\mathbb{E}(\Omega_n) = \int_0^\infty dR \int_{-1}^1 dr \oint_{|z|=1/2} \frac{dz}{2\pi i(z-1)z^n} \times \oint_{|w|=1/2} \frac{dw}{2\pi i(w-1)w^n} F(R,r)\, e^{(1-z)(1-w)G(R,r)} \times \exp\left(nR\sqrt{\frac{\tau}{2}}\Big[z + w - 2 + r(z-w)\Big]\right) \tag{15}$$

This expression can be evaluated numerically for small $n$ as seen in Fig. 5 for $n = 1$.

4. **Long Lists.** For very long lists ($n \gg 1$), we need to change the contour variables,

$$z = 1 - \frac{1 - i\varphi}{n} \quad \text{and} \quad w = 1 - \frac{1 - i\delta}{n}. \tag{16}$$

After taking the limit $n \to \infty$, first the integrals over $\varphi$ and $\delta$ can be carried out and then the remaining integrals over $R$ and $r$. This gives the result

$$\lim_{n\to\infty}\lim_{N\to\infty}\mathbb{E}(\Omega_n) = \mathrm{erf}\left[\sqrt{\frac{2}{\tau}}\right] - \sqrt{\frac{\tau}{2\pi}}\left[1 - e^{-2/\tau}\right]. \tag{17}$$

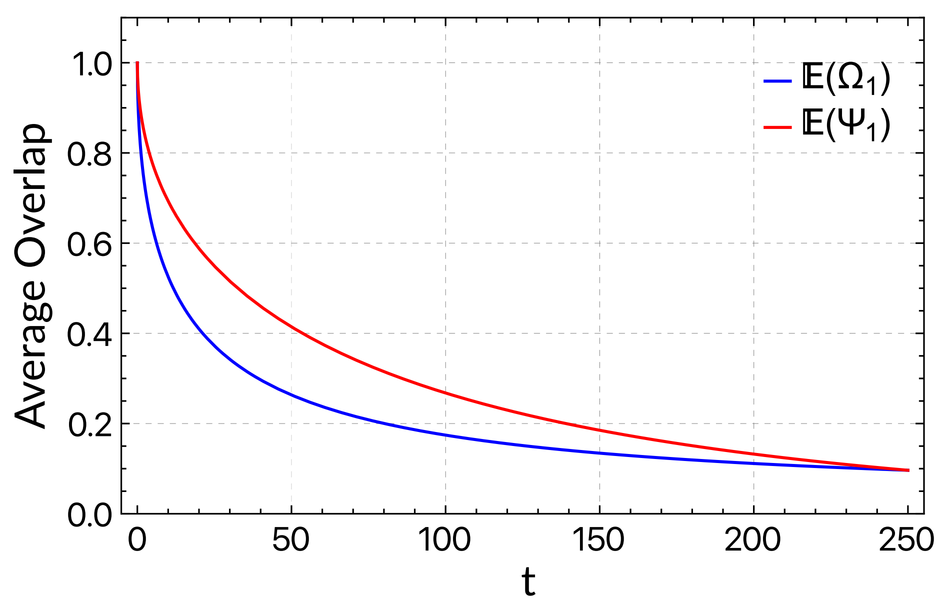


Figure 5. Comparison of decay rates for the average overlap ratio for bottom rank statistics $\mathbb{E}(\Omega_1)$ and top rank statistics $\mathbb{E}(\Psi_1)$. The time scales of the top and bottom rank statistics differ, and the normalisation is done to ensure that they have the same probability at the end points.

## Summary

In summary, we have derived expressions for the average overlap ratio for bottom rank statistics when the number of particles $N$ goes to infinity, for both short lists (15) and long lists (17). Furthermore, our results demonstrate that the bottom ranks are subject to much greater instability than the top ranks.

### Decay Rates

In contrast to the average overlap ratio for top rank statistics, which decays exponentially as seen in [2],

$$\lim_{n\to\infty}\lim_{N\to\infty}\mathbb{E}(\Psi_n) = \mathrm{erfc}[\sqrt{\tau}], \tag{18}$$

the ratio for bottom rank statistics decays algebraically. The difference can be traced back to the dominant mechanisms of the reordering. At the bottom the reshuffling at the wall is the primary driver while at the top it is mostly the diffusion as illustrated in Fig. (4). This also causes the difference in time scales which is of order $\mathcal{O}(1)$ for the top rank statistics and of order $\mathcal{O}(1/N^2)$ for the bottom rank statistics.

## References

[1] Z. Burda, M. J. Krawczyk, K. Malarz, and M. Snarska, "Wealth rheology," *Entropy*, 2021.

[2] Z. Burda and M. Kieburg, "Top rank statistics for brownian reshuffling," *Physical Review E*, 2025.

[3] J. M. Harrison, "Brownian motion and stochastic flow systems," *Wiley, New York*, 1988.

[4] J. Abate and W. Whitt, "Transient behavior of regulated brownian motion, i: Starting at the origin," *Advances in Applied Probability*, 1987.

Mathematics and Statistics Vacation Scholarship Program
ayashah@student.unimelb.edu.au