

Physics 210B Non-equilibrium Fall 2025

Assignment 2 – Solutions

Due 11:59pm Monday, October 13, 2025

1. Brain-warmer: Optimal selection.

Take a gazillion random humans and break them into groups of n . Make each group of n race some fixed distance, say 26.2 miles, and record their average speed. Normalize these results by the world record, which is (amazingly) $\frac{26.2 \text{ miles}}{2:00:35}$ (held by the late Kelvin Kiptum since 2023), so $r = \frac{\text{speed}}{\text{world record speed}} \in [0, 1]$. (For people who don't make it the whole distance (for example because they are infants and can't walk yet) the result is $r = 0$.) Now suppose we select the largest result from each group of n , i.e. the average speed of the winner of each race, and call it x . What will be the distribution of values of x ? If we pick n big enough, so that we find an elite marathon runner in each group, we might expect that they are going to cluster pretty close together.

For small n , the distribution of x will be broad, but as n increases, the distribution will become more concentrated near 1, reflecting the selection of the fastest runners.

- (a) Let $\{r_\alpha\}$ be n random numbers, each independently chosen from a probability density $p(r)$, with $r \in [0, 1]$. Find the probability density $p_n(x)$ for the *largest* value in the set, i.e. for $x = \max\{r_1, \dots, r_n\}$.

$$p_n(x) = np(x) \left(\int_0^x p(r') dr' \right)^{n-1}. \quad (1)$$

- (b) Consider the special case where $p(r)$ is uniform on $[0, 1]$. Find $\langle x \rangle$ and $\text{Var}(x)$ for this case and think about what happens at large n . Does it agree with the expectation mentioned above?

$$\langle x \rangle = \frac{n}{n+1}, \quad \text{Var}(x) = \frac{n}{(n+1)^2(n+2)}. \quad (2)$$

As n increases, $\langle x \rangle$ approaches 1, and $\text{Var}(x)$ decreases, indicating that the distribution becomes more concentrated near 1. This agrees with the expectation that elite runners cluster closely together.

2. **Large deviation principle.** In this problem we will see whether the rate function $I(S)$ ($S \equiv \sum_{i=1}^N x_i^N / N$) computed by saddle point (i.e. the large deviation principle) agrees with the exact calculation for two examples of random walks.

(a) **Fixed-step-length with bias.** Consider a 1d random walk where each step has magnitude 1 with probability p for $x_i = +1$ and probability $1 - p$ for $x_i = -1$. Find the exact distribution $P_N(S)$ of the average displacement $S = \sum_{i=1}^N x_i / N$ after N steps and use Stirling's formula to write it in the form $P_N(S) = f(S)e^{-NI(S)}$.

$$S = \frac{1}{N} \sum_{i=1}^N x_i, \quad x_i = \pm 1, \quad (3)$$

Let K be the number of $+1$ steps, then the number of -1 steps is $N-K$. Thus,

$$S = \frac{K - (N - K)}{N} = \frac{2K - N}{N} \implies K = \frac{N(1 + S)}{2}. \quad (4)$$

The probability of getting K steps of $+1$ and $N-K$ steps of -1 is given by the binomial distribution:

$$P_N(S) = \binom{N}{K} p^K (1 - p)^{N-K} = \binom{N}{\frac{N(1+S)}{2}} p^{\frac{N(1+S)}{2}} (1 - p)^{\frac{N(1-S)}{2}}. \quad (5)$$

Using Stirling's approximation, we can express this in the form:

$$P_N(S) \approx f(S)e^{-NI(S)}, \quad (6)$$

where

$$I(S) = \frac{1 + S}{2} \ln \left(\frac{1 + S}{2p} \right) + \frac{1 - S}{2} \ln \left(\frac{1 - S}{2(1 - p)} \right). \quad (7)$$

Now find the rate function by the method of large deviations and check that you get the same answer.

$$S = \frac{1}{N} \sum_{i=1}^N x_i \quad (8)$$

$$I(s) = \sup_{t \in \mathbb{R}} [ts - \lambda(t)] \quad (9)$$

$$\lambda(t) = \log \mathbb{E} [e^{tx_1}] = \log (pe^t + (1 - p)e^{-t}) \quad (10)$$

$$\frac{d\lambda}{dt} = s \quad (11)$$

$$\frac{pe^t - (1-p)e^{-t}}{pe^t + (1-p)e^{-t}} = s \quad (12)$$

$$s = \frac{py - (1-p)}{py + (1-p)} \quad (13)$$

let $y = e^{2t}$

$$y = \frac{(1-p)(1+s)}{p(1-s)} \quad (14)$$

$$t^*(s) = \frac{1}{2} \log \left[\frac{(1-p)(1+s)}{p(1-s)} \right] \quad (15)$$

$$I(s) = t^*(s) s - \lambda(t^*(s)) \quad (16)$$

$$q = \frac{1+s}{2}, \quad s = 2q - 1 \quad (17)$$

$$I(s) = q \ln \frac{q}{p} + (1-q) \ln \frac{1-q}{1-p} \quad (18)$$

$$I(s) = \frac{(s - (2p-1))^2}{8p(1-p)} + \mathcal{O}((s - (2p-1))^3) \quad (19)$$

(b) **Exponential distribution.** Suppose now that each step is sampled from the exponential distribution,

$$P(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0. \quad (20)$$

Find the exact distribution for the average displacement $S = \sum_{i=1}^N x_i / N$ after N steps and use Stirling's formula to write it in the form $P_N(S) = f(S)e^{-NI(S)}$.

$$P(x) = \frac{1}{\mu} e^{-x/\mu}, \quad x \geq 0 \quad (21)$$

$$S = \frac{1}{N} \sum_{i=1}^N x_i \quad (22)$$

$$f_X(X) = \frac{1}{\mu^N \Gamma(N)} X^{N-1} e^{-X/\mu}, \quad X \geq 0 \quad (23)$$

$$P_N(S) = \frac{N^N}{\mu^N \Gamma(N)} S^{N-1} e^{-NS/\mu}, \quad S \geq 0 \quad (24)$$

$$\Gamma(N) \approx \sqrt{2\pi} N^{N-\frac{1}{2}} e^{-N} \quad (25)$$

$$P_N(S) \approx \frac{\sqrt{N}}{\sqrt{2\pi} S} \exp \left\{ -N \left(\frac{S}{\mu} - 1 - \ln \frac{S}{\mu} \right) \right\} \quad (26)$$

$$I(S) = \frac{S}{\mu} - 1 - \ln \frac{S}{\mu}, \quad S \geq 0 \quad (27)$$

Now find the rate function by the method of large deviations. Do you get the same answer?

$$\lambda(t) = \log \mathbb{E}[e^{tx}] = -\log(1 - \mu t), \quad t < \frac{1}{\mu} \quad (28)$$

$$I(S) = \sup_{t < 1/\mu} [tS - \lambda(t)] = \sup_t [tS + \log(1 - \mu t)] \quad (29)$$

$$S - \frac{\mu}{1 - \mu t} = 0 \implies t^* = \frac{1}{\mu} - \frac{1}{S} \quad (30)$$

$$I(S) = t^* S + \log(1 - \mu t^*) = \left(\frac{1}{\mu} - \frac{1}{S} \right) S + \log \frac{\mu}{S} = \frac{S}{\mu} - 1 - \log \frac{S}{\mu} \quad (31)$$

$$I(S) = \frac{(S - \mu)^2}{2\mu^2} + O((S - \mu)^3), \quad \text{Var}(S) \approx \frac{\mu^2}{N} \quad (32)$$

3. Tracy-Widom distribution. [Bonus problem]

Use the method of large deviations to derive the Tracy-Widom distribution for the largest eigenvalue of a (Gaussian) random (real, symmetric) matrix.

4. **Gillespie algorithm or continuous-time Monte Carlo, and demographic stochasticity.** Consider a ‘dimerization reaction’: a molecule M (‘monomer’) joins with another monomer to form a dimer $2M \leftrightarrow D$. The forward reaction rate is k_d and the backwards rate is k_u . That is, the backwards unbinding rate per unit volume is $k_u[D]$ (where $[D]$ is the concentration of dimers), while the forward binding rate per unit volume is $k_b[M]^2$ (since the binding requires a collision of two monomers). [actually $k_b M(M-1)$.] For definiteness, let’s take all N monomers to be initially unbound, and $k_u = 2k_b$.

(a) Write an ODE for $\partial_t M$ treating M and D as continuous variables.

Find the equilibrium values of $[M]$ and $[D]$ for the given values of k_u and k_b for $N = 2, 90, 10000$. Solve the ODE numerically and verify that the late time behavior agrees with the equilibrium solutions you found.

$$\dot{M} = -2k_b M^2 + 2k_u D \quad (33)$$

$$M + 2D = N \quad (34)$$

$$-2k_b M^2 - k_u M + k_u N = 0 \quad (35)$$

$$M^* = \frac{-k_u + \sqrt{k_u^2 + 8k_b k_u N}}{4k_b} \quad (36)$$

$$M^* = \frac{-1 + \sqrt{1 + 8rN}}{4r}, \quad r \equiv \frac{k_b}{k_u} \quad (37)$$

$$D^* = \frac{N - M^*}{2} = \frac{k_b}{k_u} (M^*)^2 \quad (38)$$

$$\text{For } k_b = 1, k_u = 2 : \quad M^* = \frac{-1 + \sqrt{1 + 4N}}{2} \quad (39)$$

$$D^* = \frac{N - M^*}{2} \quad (40)$$

$$\text{For } N = 2 : \quad M^* = 1, \quad D^* = 0.5 \quad (41)$$

$$\text{For } N = 90 : \quad M^* = 9, \quad D^* = 40.5 \quad (42)$$

$$\text{For } N = 10000 : \quad M^* \approx 99.50125, \quad D^* \approx 4950.2494 \quad (43)$$

When the number of molecules is small, we expect that the continuum approximation above may fail. Here is an efficient Monte Carlo algorithm to simulate the individual molecule evolution.

Suppose that the reactions have rates Γ_i with $\Gamma_{\text{tot}} = \sum_i \Gamma_i$, so the probability that the next reaction that happens is reaction j is $\Gamma_j/\Gamma_{\text{tot}}$. To simulate until a final time t_f : Pick a random time t_{wait} with probability distribution $p(t) = \Gamma_{\text{tot}} e^{-\Gamma_{\text{tot}} t}$.

If the current time t plus t_{wait} is bigger than t_f , you are done.

Otherwise, increment t by t_{wait} , pick a random number r uniformly in the interval $[0, \Gamma_{\text{tot}})$, and execute the reaction j for which $\sum_{i < j} \Gamma_i \leq r < \sum_{i < j+1} \Gamma_i$, i.e. such that r lands in the j th interval of the sum forming Γ_{tot} .

Repeat. (If some reaction becomes impossible, you may need to recompute Γ_{tot} .)

- (b) Implement this algorithm for the dimerization reaction, and simulate for $N = 2, 90, 10000$ and compare individual trajectories with the continuum solution. For what N are the fluctuations at late times less than $\pm 20\%$?

For $N=10000$, the fluctuations at late times are less than ± 20 (See figure 1,2,3).

- (c) Find the average of many realizations of the discrete evolution for $N = 2, 90$ and compare with the ODE result. How much is the long-term average shifted by the demographic stochasticity? How large does N need to be for the ensemble average of $M(t)$ to be well-described by the ODE?

For $N=90$, the ensemble average of $M(t)$ is well-described by the ODE (See figure 4,5).

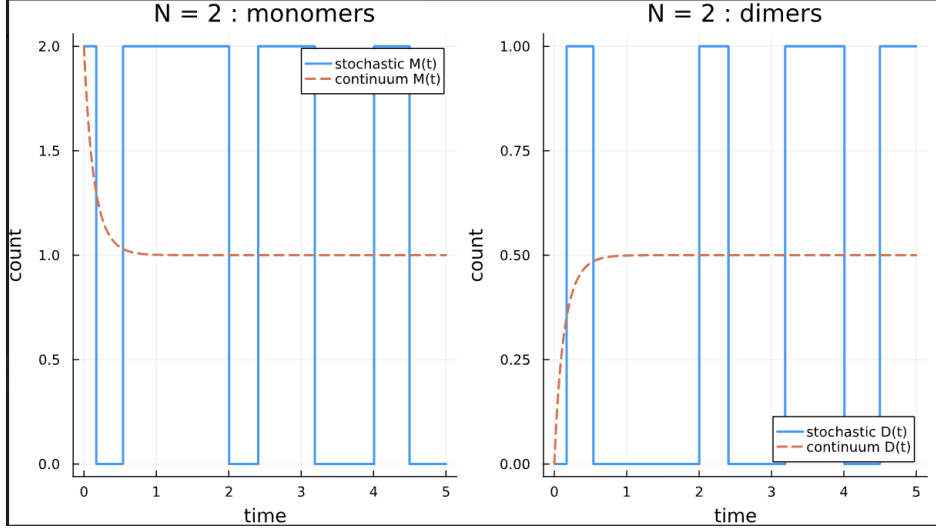


Figure 1: Stochastic trajectory for $N=2$

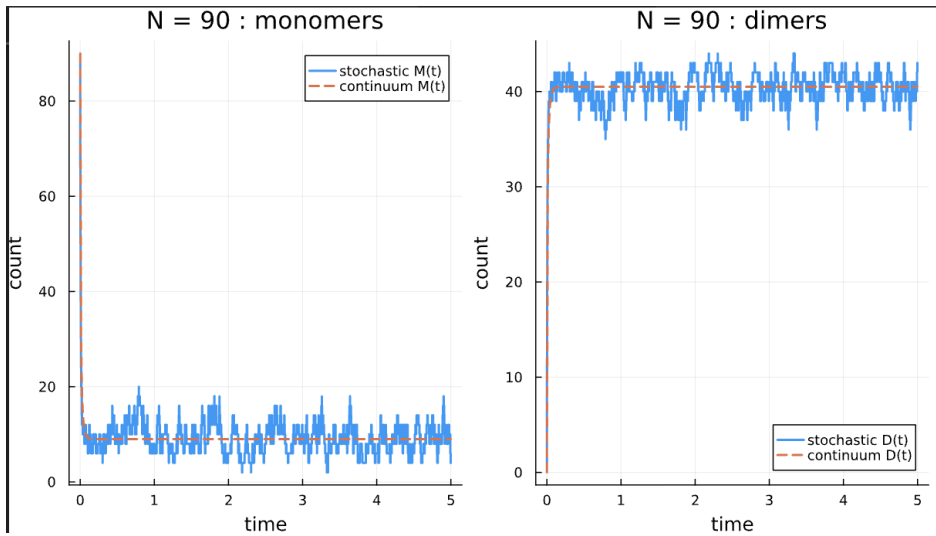


Figure 2: Stochastic trajectory for $N=90$

- (d) Use these ideas to simulate a 1d continuous-time random walk (CTRW) (say for a distribution $p_t(t)$ with a long tail) without actually waiting for the possibly-long waiting times.

Please see figure 6.

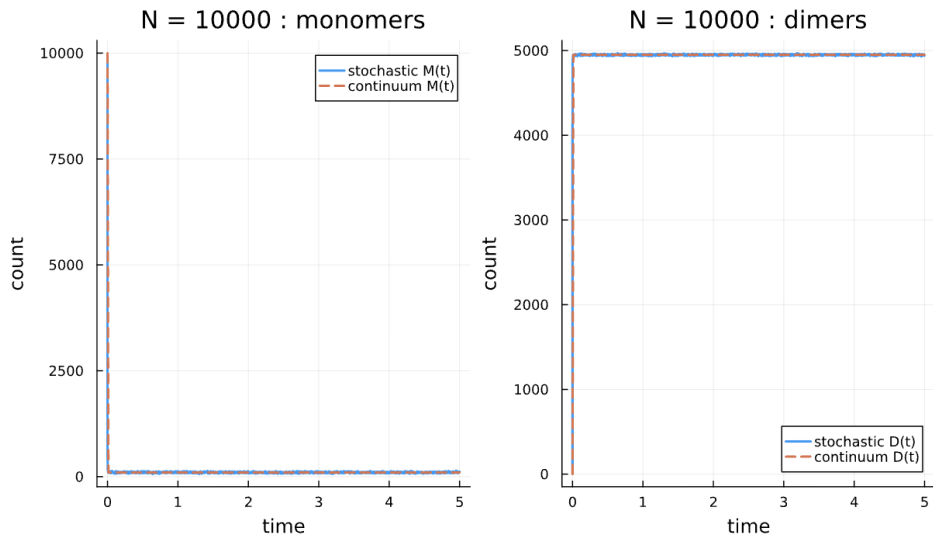


Figure 3: Stochastic trajectory for $N=10000$

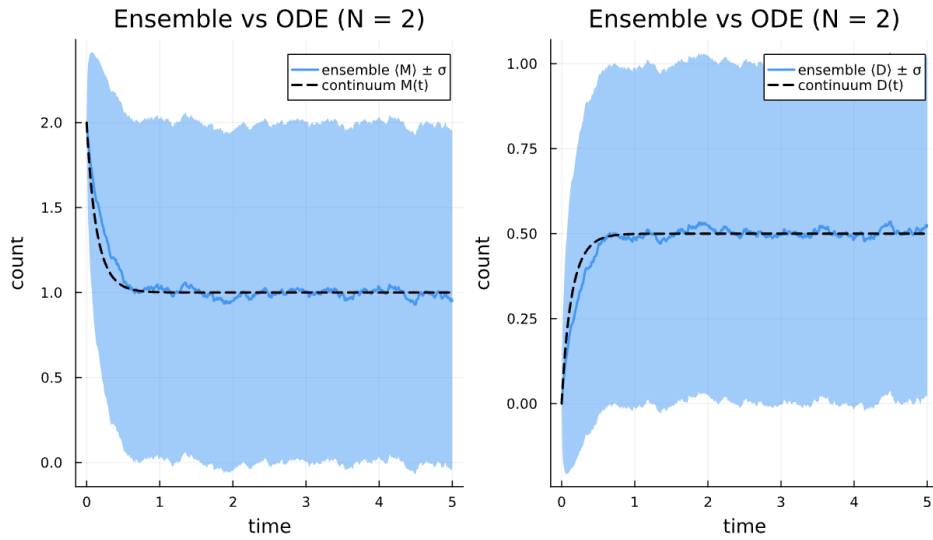


Figure 4: Average of many realizations for $N=2$, compared with ODE result

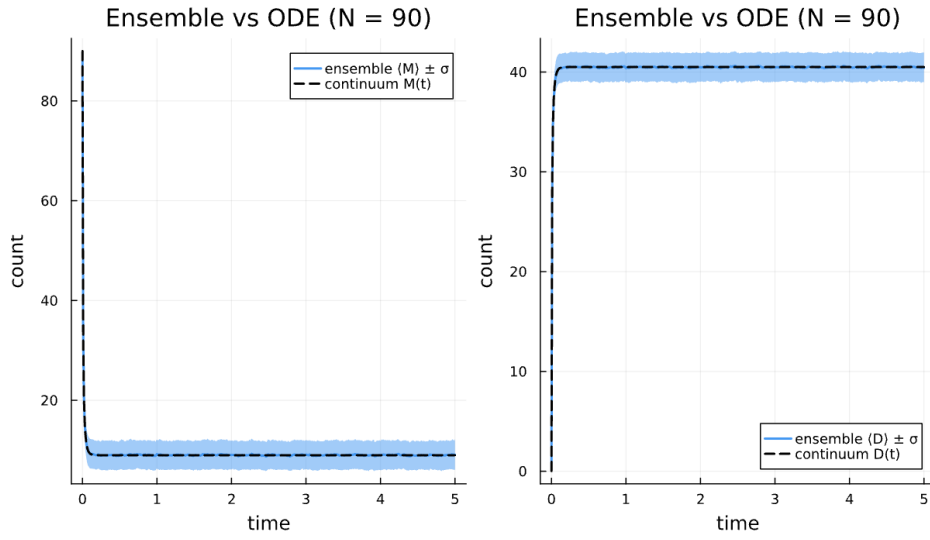


Figure 5: Average of many realizations for $N=90$, compared with ODE result

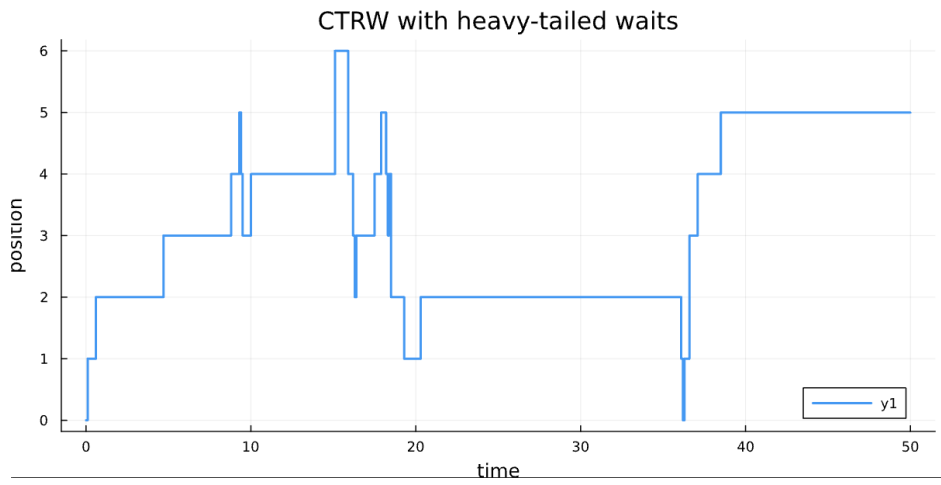


Figure 6: CTRW with heavy-tailed waits