Prof. Dr. Joachim Krug
Alexander Klug
Benjamin Schmiegelt
Suman G. Das

Summer term 2021

Biological Physics II

Problem Set 5

Please hand in your solutions before 12:00 noon on Wednesday, June 23, 2021.

1. Moment generating function

3+4+5= 12 pts

Let $t \geq 0$ be a continuous random variable with probability density function (PDF) $P_t(t)$. Then its moment generating function (MGF) is given by $G_t(z) = \langle e^{tz} \rangle$, where $\langle \cdot \rangle$ denotes average with respect to the distribution $P_t(t)$.

a) Let the Taylor series of the MGF be

$$G_t(z) = \sum_{k=0}^{\infty} \frac{a_k}{k!} z^k. \tag{1.1}$$

The k-th moment of t is defined as $\langle t^k \rangle$. Show that it is equal to a_k.

b) Let $t_1,\ t_2, \ldots,\ t_n$ be n independent random variables, with the respective MGFs $G_{t_1}(z_1),\ G_{t_2}(z_2), \ldots, G_{t_n}(z_n)$. Let $T_n = t_1 + t_2 + \cdots + t_n$. Show that the MGF of T_n is

$$G_{T_n}(z) = \prod_{j=1}^{n} G_{t_j}(z). \tag{1.2}$$

Now let the variables $t_j,\ j = 1, 2, \ldots, n$ be consecutive time intervals in a Poisson process, where each waiting time is identically distributed with rate constant r. Then T_n is the time required for n events to occur. Express the MGF of T_n. Determine the first and second moments of T_n by using the result of part a).

c) Numerically simulate a Poisson process with rate constant $r = 1$ and find the distribution of T_n for $n = 5$, $n = 10$ and $n = 20$. Use a bin size of about 0.05 together with 10^6 - 10^7 realizations of the process for a smooth histogram. Argue that the PDF of T_n for large n should approach a Gaussian distribution. Analytically calculate the mean and variance of T_{20} from the formulas in the previous part, and use these to plot the Gaussian approximation to the distribution of T_{20} and plot it in the same figure as the numerical result.

d) Bonus: Here is a puzzle, related to the waiting time paradox discussed in the lectures. Suppose you have two friends Alice and Bob, and you can visit either by going to the station and taking the train going west (for Alice) or east (for Bob). The trains in both directions come at equally spaced intervals of 10 minutes (not randomly!). You want to visit each of them with equal frequency, and so decide on the following strategy. Every Friday you go to the station at some random time and take whichever train (east or west) comes first. But after some time you realize that you are visiting Alice about 4 times as frequently as Bob by following this protocol. How do you explain this? Hint: Think of a schedule for the trains that will lead to this effect.

2. The Gillespie algorithm

3+4+4+7= 18 pts

In this problem we will define and use the Gillespie algorithm. It is used to simulate a discrete memoryless random variable $x(t)$ which is a function of continuous time t. From any state $x(t) = i$, the system can transition to any of a set of ν_i states. We thus have to consider ν_i waiting times τ_{ji}, $j = 1, 2, \ldots, \nu_i$ which are independently and exponentially distributed with rates r_{ji}. The transition happens after a waiting time τ, where $\tau = \min\{\tau_{1i}, \tau_{2i}, \ldots, \tau_{\nu_i i}\}$, and to the state k for which $\tau_{ki} = \tau$.

a) Show that the variable τ is distributed as an exponential with rate $R_i = \sum_{j=1}^{\nu_i} r_{ji}$. Hint: It helps to consider the cumulative distribution function of τ.

b) Let the distribution of τ be denoted by $P_\tau(\tau) = R_i e^{-R_i \tau}$. In the **Gillespie algorithm**, the **first step** is to draw a number τ from this distribution, and update the time to $t \to t + \tau$. The **second step** is to choose a state k from the set of ν_i target states according to the probabilities $\frac{r_{ki}}{R_i}$, and set this as the state at the updated time. After this, one repeats the same steps starting from the new state at the updated time, and so on. We apply this algorithm to a simple model of a bacterial population growing by cell division. A population of size N can transition to $N + 1$ cells at rate Nb or $N - 1$ cells at rate Nd, where b is the division rate and d the death rate per cell. Simulate the process using the Gillespie algorithm, with parameters $b = 1$ and $d = 0$, and starting with $N = 1$ cell at $t = 0$. Plot the distribution of the number of cells at time $t = 4$. You will need about 10^7 realizations for a smooth histogram.

c) Consider the same system and compute and plot the distribution of the time taken for the population to reach 100 cells.

d) Now consider the same system but with $d = 0.5$ and other parameters and initial condition unchanged. Numerically compute and plot the mean number of cells as a function of time between $t = 0$ and $t = 4$. Now compute and plot the mean number of cells as a function of time *conditioned* on the population not going extinct by $t = 4$ in the same figure. Plot at least 5 sample trajectories each of $N(t)$ for the cases where the population does and does not go extinct by $t = 4$. Also, simulate the same system and numerically determine the probability that the population does not go extinct by time T for $T = 1, 3, 5, 7, 9$. Use this to estimate the probability that the population never goes extinct.