Numerical Methods in Scientific Computing, spring 2023

Problem sheet 10: statistics

Return code and solutions by Monday 3.4.2023 23:59 to Moodle.

Bonus exercises are voluntary: they don't count towards the maximum for the exercises. They do, however, give you extra points in the grading of the course – it's possible to score over 100% of the points on the course with bonus exercises. No solutions will be provided for the bonus exercises.

1. (pen and paper, 6 points) A colleague of yours is running an experiment that samples a Gaussian random process x with $\mu = 3.7$ and $\sigma^2 = 76.6$. The experiment is painstaking, and it takes a long time to run and extract the results. They have measured the following set of samples of x

21.9514	28.1882	28.7511	31.1946	37.3700	20.5656	31.2570	31.6869	37.4599	26.0999
29.9478	36.6224	34.1511	25.8640	33.2204	29.4145	24.8233	23.7741	25.2026	26.4967

but are not sure whether their equipment is working properly. Based on these results, can you help them figure this out?

2. (computer, 6 points) Study the central limit theorem numerically by computing histograms of the average of K random numbers x_i

$$z = \frac{1}{K}\sum_{i=1}^{K} x_i$$

and showing that the histogram of z approach a normal distribution for large values of K. (Because x_i are random numbers, also z is a random number.)

Use $K = 3, 10, 30, 100$. For each value of K, generate 10 000 random numbers z, generate a histogram with 100 bins, and plot it against a "best-fit" normal distribution (you don't need to optimize μ and σ tightly, getting the idea is enough). Study the following cases

a. uniformly distributed random numbers, $x \in \left[-\frac{1}{2}, \frac{1}{2}\right]$

b. exponentially distributed random numbers, $x \in [-\infty, \infty]$

$$p(x) = \frac{\gamma}{2} e^{-\gamma|x|}$$

c. Lorentz distributed random numbers, $x \in [-\infty, \infty]$

$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

Use inverse transform sampling to generate the random numbers. You can use a standard library implementation to get the uniform random numbers.

Hint: the lecture notes and the cumulants derived in the following problem might be of assistance in figuring out what the normal distribution should look like, but this is totally optional.

3. (pen and paper, 6 points) Derive the cumulants $c_1(x)$, $c_2(x)$, and $c_3(x)$ defined in the lecture notes for

a) the uniform distribution

$$p(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

b) the exponential distribution

$$p(x) = \frac{1}{2}e^{-|x|},$$

c) the normal distribution

$$p(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$

How do the values of the cumulants compare to the moments?

4. (pen and paper, 6 points) Antti is running Monte Carlo simulations of a random walk on a one-dimensional lattice. The *walkers* in such simulations are a bit like people you might see late at night in Helsinki (or mid-day in Kallio): you cannot tell which direction the walker is going to turn next, based on their earlier history i.e. walk path. The random walk on the one-dimensional lattice can thus be modeled with a binomial distribution, with the outcomes being determined by $p =$ take a step to the left and $q =$ take a step to the right.

   Assuming a set of walkers leave the pub at the origin $x = 0$ (at a time $t = 0$), derive the mean and the standard deviation of the distribution of walkers as a function of the total number of steps taken N. You can assume that it is so late at night that the probabilities for both outcomes are $p = \frac{1}{2} = q$.

5. (BONUS, computer, 6 points) Arska, who lives in Kallio, was enjoying a few pints at a bar on Helsinginkatu with his friends, until the bouncer threw them out since they were being disruptive. On the street, Arska's friend Pena invites Arska and the rest of the bunch to his house, which is very close to the bar – only 200 m away. Arska's home is also on Helsinginkatu, just a short walking distance (0.25 miles accoding to US research studies, i.e. 400 m) away from the pub; only, in the wrong direction from Pena's place.

   After leaving the bar, Arska has conflicting ideas whether to go home, try again to get back into the bar, or to join Pena and friends, and may change direction at the spur of a moment at each step, assumed to be 78 cm long. However, despite his drunken stupor, Arska as a native Helsinkian knows not to stray from Helsinginkatu, since all three destinations are along the same street.

   Calculate the probabilities for Arska to get home to bed, p_{bed}, to get back into the bar to continue drinking, p_{bar}, to join Pena and the others, p_{Pena}, or to get picked up by the police and thrown in the drunk tank, $p_{\text{drunk tank}}$, as a function of the number of steps Arska has taken.

   To estimate the probabilities, you can assume that Arska goes out every night. To make sure you have collected statistics, run nightly simulations for a time period covering 10 years.

   Take into account the following aspects in your simulations.

   (a) the computer says that the distance between Arska's home and the bar is exactly $d_1 = 512$ steps. Similarly, the distance between the bar and Pena's place is exactly $d_2 = 256$ steps. Use these values in your simulation. You can assume Arska lives to the "left" of the bar, and Pena lives to the "right".

   (b) if Arska finds himself at the bar, he tries to get back in. However, the probability that the bouncer throws him out again is

   $$p_{\text{bouncer}} = 2^{-n/d_b},$$

   where n is the number of steps Arska has taken so far and $d_b = 2048$. You can assume that if Arska gets in, he stays in the bar, since the rest of the bunch aren't there to cause trouble. If Arska can't get in, he continues his walk.

   (c) if Arska finds himself at his house, he goes to sleep until the morning. However, there's a chance of

   $$p_{\text{miss}} = \frac{1}{10}$$

   that Arska doesn't notice that he is in front of his house and instead continues walking.

   (d) if Arska finds himself at Pena's place, he rings the bell, but the probability that someone lets him in depends on two independent factors.

   First, Pena has to have arrived for Arska to get in, the probability for which is

   $$p_{\text{Pena at home}} = 1 - 2^{-n/d_2}.$$

Second, Pena was recently on a trip to Latvia, meaning he has a stocked liquor cabinet, which he and his friends are taking liberties with. This means that the probability that the party has passed out by the time Arska rings the bell is

$$p_{\text{passed out}} = 1 - 2^{-n/d_L},$$

where $d_L = 512$. If the party has passed out, Arska can't get in and continues his walk.

(e) Helsinginkatu is 2.2 km long, according to Google Maps. Assuming Arska lives at Sällikoti, the distance from Arska's home to the end of Helsinginkatu is 900 m, while the distance from Sällikoti, past the bar and Pena's place to the other end of Helsinginkatu at Kurvi (where the Sörnäinen metro station is) is 1.3 km.

Translated into steps, the one end of Helsinginkatu is 1154 steps from Arska's house to the "left", while the other end of Helsinginkatu – past the pub and Pena's place – is 1666 steps to the "right" from Arska's house.

If Arska goes to Kurvi, he realizes that he has gone too far, and will take a step back. (Reflective boundary condition)

The Kisahalli drunk tank, however, can be found at the other end of Helsinginkatu *i.e.* close to the the opera house. Because the police know Arska well, if Arska makes it to this end of Helsinginkatu, the police take him into the drunk tank for the night, because the nights are cold in Helsinki.

Do the probabilities sum up to one?