This report explores tumour growth and movement simulations, demonstrating how simple computational models can represent complex biological processes. The first part examines random movement on a grid, beginning with four directions and later expanding to eight, illustrating how increased trials reduce bias and lead to uniform distributions, as described by the law of large numbers. The second part focuses on modelling tumour growth using the Gompertz equation and Euler’s method, analysing how step size influences computational complexity and error, and highlighting the trade-offs between precision and efficiency in predicting steady states. The final part integrates movement and growth, simulating tumour propagation across a grid under constraints such as non-overlapping cells, to explore the interaction between spatial expansion and localised growth. Overall, this study demonstrates how basic simulations provide insights into the balance between computational accuracy and complexity, laying the groundwork for more sophisticated modelling approaches.  
  
  
This task focuses on understanding how distributions, the law of large numbers, computational complexity, and accuracy interplay in simulating random processes, specifically cell movements on a grid. It aims to demonstrate how bias emerges and diminishes in random processes and how uniformity develops with increasing iterations. Depending on how the randomness is implemented, it could lead to a variety of distributions, such as the normal distribution, which describes data using a probability density function with mean μ and standard deviation σ and is often used in random number generation (Shiflet & Shiflet, 2014). The uniform distribution ensures that all outcomes have an equal probability in discrete cases, or that equal-length intervals have an equal chance of occurring in continuous cases (Shiflet & Shiflet, 2014). The Bernoulli distribution arises when modelling binary outcomes, and its conditional version, the conditional Bernoulli distribution, is useful when several Bernoulli random variables are conditioned on their sum equalling a specific value, with applications in sampling and hypothesis testing (Chen and Liu, 1997). The beta-binomial distribution applies when the probability of success in a binomial process itself follows a beta distribution, offering a model for over-dispersed data, common in biological and reliability studies (Skellam, 1948; Altham, 1978). Lastly, the Poisson distribution generalizes the binomial distribution by allowing different probabilities of success in each trial, which is useful in survey sampling and logistic regression (Chen and Liu, 1997). These varying distributions help explain how randomness in cell movement can lead to different patterns, and how bias and uniformity evolve as the process progresses. In practice however never can a distribution truly be split uniformly however due to the law of large numbers the outcomes will be slowly converging towards an even probability; so, when trying to find if the distribution is uniform a chi-square test can be implemented the purpose of the chi-square test is to compare the expected and observed outcomes. The chi-square test does not confirm whether the distribution is random, but it does give an indication as the chi-square test only looks at frequency a test that looks at the shape of the frequency which allows for detection of small deviation even when looking at large sample sizes, that test is the Kolmogorov-Smirnov test.

|  |
| --- |
| Pseudocode |

Task 1

The simulation results in the form of Chi-Squared and Kolmogorov-Smirnov (KS) tests offer valuable insights into the uniformity of the cell movement over different steps. Initially, the observed frequencies deviate from the expected uniform distribution at the early checkpoints (e.g., 10 and 25 steps), with Chi-Squared and KS tests showing p-values above 0.05 (indicating no significant deviation from uniformity). However, as the number of steps increases, the results tend to approach uniformity, and p-values generally move closer to 0.05, suggesting that with larger sample sizes, the randomness of the movement tends to become more evenly distributed.

In early steps of the simulation, such as at checkpoint 10, observed frequencies show some deviation from the expected values (e.g., at Run 1, checkpoint 10, we have observed frequencies [4, 4, 1, 1] instead of the expected [2.5, 2.5, 2.5, 2.5]). This is expected due to the limited number of steps involved. The Chi-Squared statistic (e.g., 3.6 for checkpoint 10) and the KS test statistic (e.g., 0.6) suggest that randomness in small samples can cause fluctuations and result in frequencies that do not perfectly match the expected uniform distribution. As the number of steps increases, however, the number of observations improves, reducing the influence of these fluctuations.

This behavior is an example of the **Law of Large Numbers** (LLN), which states that as the sample size grows, the sample mean (or in this case, the direction distribution) converges to the expected value. At lower step counts, there is insufficient data for the distribution to settle, and the results may be biased or skewed. At larger step counts, such as 50 or 100, we see that the data becomes closer to the expected distribution, though some discrepancies still exist.

With more steps (e.g., 50, 100), observed frequencies move closer to the expected uniform distribution. For example, at checkpoint 50, the frequencies are [12, 14, 12, 12], which is much closer to the expected [12.5, 12.5, 12.5, 12.5]. This is in line with the expected behavior as the number of trials increases. The Chi-Squared statistic (e.g., 0.24) and the p-value (e.g., 0.9709) for checkpoint 50 confirm that the observed distribution is very close to the expected uniform distribution, supporting the idea that larger datasets tend to smooth out irregularities and lead to more uniform distributions.

The results indicate that the distribution doesn’t perfectly match the expected uniformity, especially in the smaller runs. A potential reason for the bias in the movement could be the **initialization of the random number generation**. Each run starts with a fresh random seed, and this can lead to minor biases in the movement during the early steps. As the number of steps increases (i.e., as the simulation runs for more trials), the bias tends to diminish, and the distribution becomes more uniform. However, if each simulation run is treated separately, the bias may not fully disappear and will likely remain present in small-scale simulations.

To further mitigate the bias, if we were to consolidate all the steps across multiple runs into one larger simulation (e.g., combining 100, 500, and 10000 steps into a single simulation), and then extract the data at checkpoints (100, 500, 1000, 5000), we would see a more uniform distribution. The distribution would gradually approach uniformity as the number of steps grows, aligning with expectations based on the law of large numbers.

The complexity of the simulation grows with the number of steps, and the increased computational effort required for larger numbers of steps reflects the trade-off between **accuracy and efficiency**. As the sample size increases, we get a better approximation of the uniform distribution, but the computational time also grows. For example, with larger simulations (e.g., 10,000 steps), we see that the direction distribution converges to near-uniformity, but the simulation takes longer to complete. This represents a compromise between achieving more accurate results and the computational cost associated with simulating more steps.

The complexity of the model also introduces challenges in terms of the randomness and the simulation’s behavior over time. Each random step contributes to the overall distribution, and in smaller simulations, this randomness can lead to irregularities that prevent the distribution from being uniform. However, with sufficiently large datasets, the irregularities diminish, and the movement behavior becomes more predictable, demonstrating the balance between randomness, complexity, and accuracy in simulations.

The results of the Chi-Squared and Kolmogorov-Smirnov tests show that the distribution of directions in the early stages of the simulation deviates from uniformity due to the small sample size. This bias reduces as the number of steps increases, confirming the importance of the Law of Large Numbers in simulations. The initialization of the random number generator is a key factor that influences the observed bias, and by consolidating multiple runs into a single extended simulation, we can observe the trend toward uniformity more clearly. The complexity of the simulation, while requiring more computational effort for larger step counts, enables a more accurate assessment of the distribution, balancing between efficiency and accuracy. Ultimately, larger sample sizes lead to more uniform distributions, aligning with theoretical expectations, while smaller samples may show biases that dissipate with increased trials.  
  
Task 2   
The simulation of tumor growth using the Gompertz model provides a comprehensive illustration of how distributions, the law of large numbers, complexity, and accuracy interact in computational modeling. The tumor growth curve exhibits a sigmoidal distribution, characterized by an initial phase of exponential growth when the tumor size (*NN*N) is far smaller than the carrying capacity (*MM*M), followed by a gradual slowdown as resources become limited. This behavior reflects biological constraints, with the growth rate reducing over time and eventually reaching a steady state where *N≈MN \approx M*N≈M. At this steady state, the term *ln⁡(M/N)\ln(M/N)*ln(M/N) becomes zero, leading to *dN/dt=0dN/dt = 0*dN/dt=0. The steady-state value *MM*M, which represents the maximum achievable tumor size, is a critical parameter in understanding the system's dynamics and is prominently displayed in the graphs. This transition from rapid growth to saturation is a hallmark of many biological systems and is captured effectively by the model.

The law of large numbers plays a vital role in ensuring the reliability of the numerical solution, especially when smaller step sizes (*hh*h) are used. Smaller *hh*h reduces the discretization error in approximating the differential equation, allowing the numerical solution to closely match the theoretical steady-state value *MM*M. However, this improvement in accuracy comes at a cost: the computational complexity of the simulation increases as the number of steps required scales with *O(T/h)O(T/h)*O(T/h), where *TT*T is the total simulation duration. Conversely, larger step sizes reduce the number of steps and computational time but introduce greater errors, which can manifest as deviations from the expected steady-state value. This trade-off between accuracy and computational efficiency is a central consideration in numerical modeling and must be carefully balanced.

The errors in the simulation are influenced by both the choice of *hh*h and the limitations of machine precision. Discretization error decreases with smaller *hh*h, improving the accuracy of the simulation, but this comes at the expense of higher cumulative rounding errors caused by finite-precision arithmetic. These rounding errors scale with *ϵ/h\epsilon/h*ϵ/h, where *ϵ\epsilon*ϵ is the machine's precision. This dual dependence of error on *hh*h highlights the compromises inherent in such simulations. While a smaller *hh*h increases accuracy by reducing discretization error, it amplifies the cumulative effect of machine precision error and increases the computational cost due to a greater number of steps. On the other hand, larger *hh*h reduces computational time and minimizes the accumulation of machine precision errors but sacrifices accuracy in capturing the system's dynamics, particularly near the steady state.

Including the carrying capacity *MM*M as a reference on the graph underscores the biological significance of the steady-state value and helps visualize how *NN*N approaches *MM*M over time. This visual representation complements the theoretical understanding of the system, allowing for an intuitive grasp of the steady-state behavior. Furthermore, the simulation demonstrates how the time to reach steady state depends on both the initial tumor size (*N0N\_0*N0 ) and the growth rate constant (*kk*k). Larger *kk*k values accelerate convergence to steady state, while smaller *kk*k values extend the time required. The choice of step size (*hh*h) also directly affects the precision with which this convergence is captured, with smaller *hh*h providing finer granularity and more accurate tracking of the growth curve.

Overall, the simulation effectively balances biological realism, numerical accuracy, and computational complexity. By systematically exploring the trade-offs between step size, error, and computational effort, the model provides valuable insights into tumor growth dynamics. It highlights the importance of choosing appropriate numerical methods and step sizes to achieve a balance between accuracy and efficiency, ensuring the results are both biologically meaningful and computationally feasible. This simulation serves as an example of how complex systems can be modeled and analyzed, demonstrating the power of mathematical and computational tools in understanding real-world phenomena.

### **References**

* Chen, S.X. and Liu, J.S., 1997. Statistical applications of the Poisson-binomial and conditional Bernoulli distributions. *Statistica Sinica, 7*(4), pp.875–892.
* Altham, P.M.E., 1978. Two generalizations of the binomial distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 27*(2), pp.162–167.
* Skellam, J.G., 1948. A probability distribution derived from the binomial distribution by regarding the probability of success as variable. *Journal of the Royal Statistical Society: Series B (Methodological), 10*(2), pp.257–261.
* Shiflet, A.B. & Shiflet, G.W., 2014. *Introduction to computational science: Modeling and simulation for the sciences*. Princeton University Press.

Appendicies  
*A group of graphs with different colored lines

Description automatically generated*

Figure 1.1.1

*A group of graphs with different colored lines

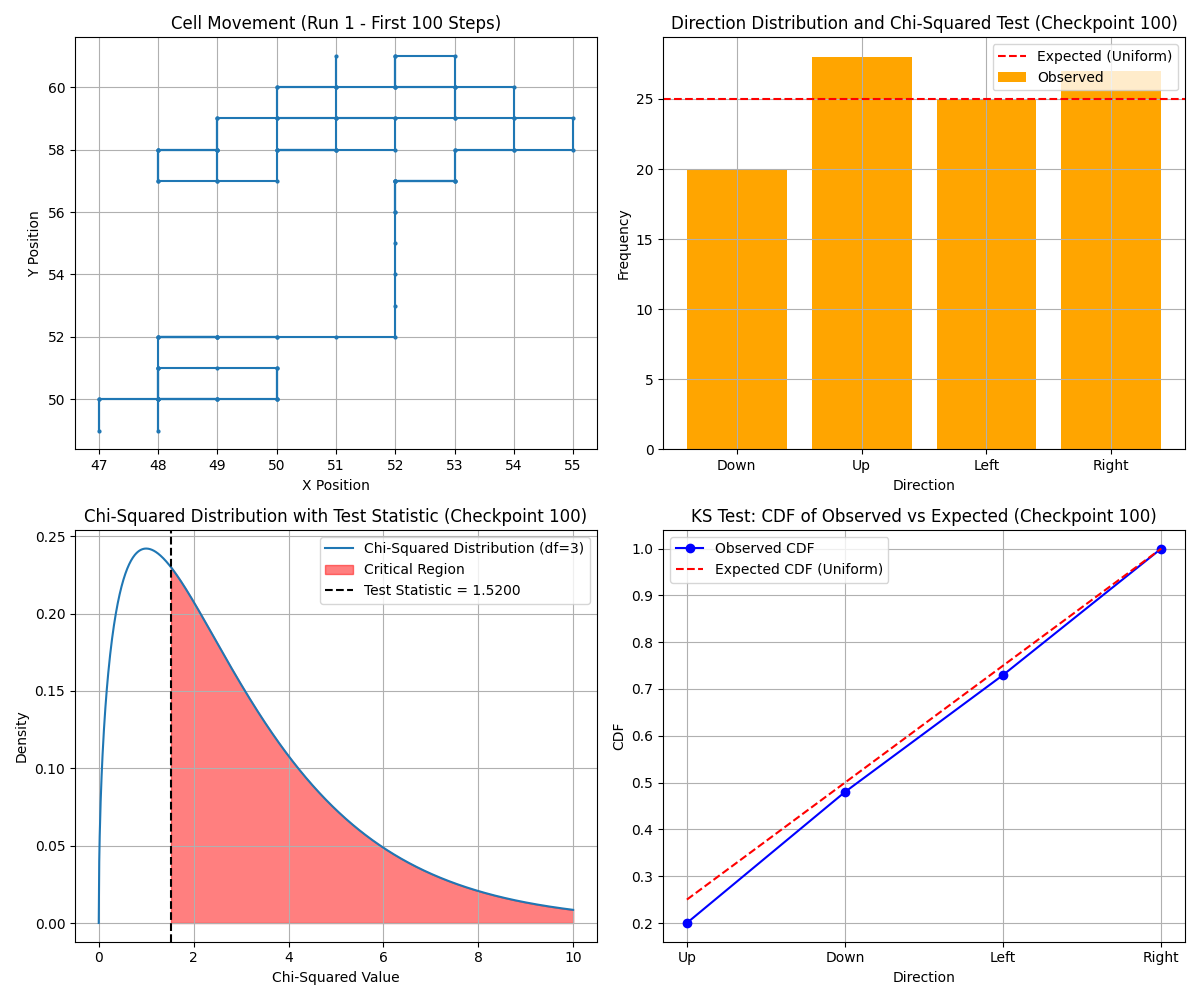
Description automatically generated with medium confidence*Figure 1.1.2

Figure 1.1.3*A group of graphs and diagrams

Description automatically generated*

Figure 1.1.4*A group of graphs and diagrams

Description automatically generated*

Figure 1.2.1

A group of graphs with different colored lines

Description automatically generated with medium confidence

Figure 1.2.2

Figure 1.2.2

A group of graphs and diagrams

Description automatically generated

Figure 1.2.3

A group of graphs and diagrams

Description automatically generated

Figure 1.3.1

A group of graphs and diagrams

Description automatically generated

Figure 1.3.1

A group of graphs and diagrams

Description automatically generated

Figure 1.3.2

A group of graphs and diagrams

Description automatically generated

Figure 1.3.3

A group of graphs and diagrams

Description automatically generated

Figure 1.3.4

A collage of graphs and diagrams

Description automatically generated

Figure 2.1.1

A collage of graphs and diagrams

Description automatically generated

Figure 2.1.2

A collage of graphs and diagrams

Description automatically generated

Figure 2.2.1

A collage of graphs

Description automatically generated

Figure 2.2.2

A screenshot of a graph

Description automatically generated

Figure 2.3.1

A collage of graphs and diagrams

Description automatically generated

Figure 2.3.2

A graph of a model and a log-turd

Description automatically generated with medium confidence

Figure 3

Figure 4.1

Figure 4.2