

# Mathematical Data Science HW4

20180617 You SeungWoo

October 16, 2023

## Problem 1

*Solution.* The equation is:

$$P(n, k+1) = p_r P(n-1, k) + p_l P(n+1, k) + p_s P(n, k) \quad (1)$$

where  $p_r, p_l, p_s$  are probabilities of moving right, left, and the same each. They satisfy  $p_r + p_l + p_s = 1$ . Suppose  $p_r = p_l = p_s = \frac{1}{3}$ . By Taylor expansion,

$$P(n, k+1) \simeq P(n, k) + P_k + \frac{1}{2} P_{kk} + \dots$$

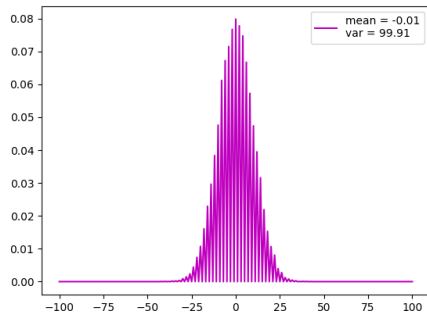
$$P(n-1, k) \simeq P(n, k) - P_n + \frac{1}{2} P_{nn} - \dots$$

$$P(n+1, k) \simeq P(n, k) + P_n + \frac{1}{2} P_{nn} + \dots$$

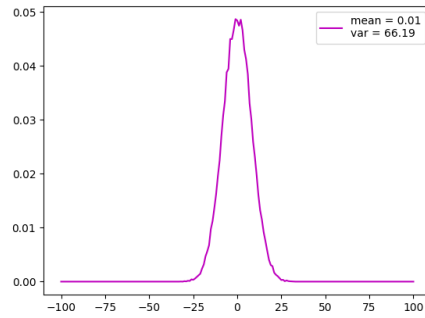
Note that  $\Delta k = \Delta n = 1$  in this discrete situation. By putting them into the equation (1), we get

$$P_k \simeq \frac{1}{3} P_{nn}$$

Therefore, the diffusion coefficient is  $\frac{1}{3}$  if  $\Delta k = \Delta n = 1$ . If not, then the general version is  $\frac{1}{3} \frac{\Delta n^2}{\Delta k}$ . If we apply the similar analysis as we did in class, then we get the same result:  $\mu(t) = 0, \sigma^2(t) = 2Dt$  where  $D$  is a diffusion coefficient. However, since  $D$  has decreased compared to two movement choices,  $\sigma^2$  is further reduced. Here are the results.



(a) 2-choice walk



(b) 3-choice walk

□

## Problem 2

*Proof.* Just put  $n = \frac{x}{\Delta x}, k = \frac{t}{\Delta t}$  into  $p(n, k)$  and run brute-force calculation. Then we get the desired result:

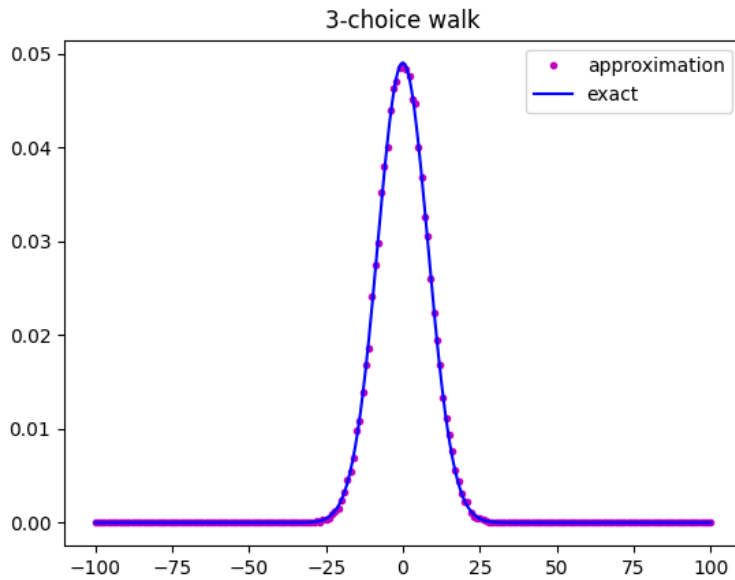
$$p(n, k) = p\left(\frac{x}{\Delta x}, \frac{t}{\Delta t}\right) = 2\Delta x \cdot p(x, t)$$

This model has only 2 choices: left and right. Let  $m_i$  be a position at  $i$ th iteration. Let  $N_l, N_r$  be a number of choices of left and right each. Then  $m_i = m_0 - N_l + N_r = N_l + N_r$  and  $N_l + N_r = i$ , so

$$m_i + i = 2N_r,$$

$$i - m_i = 2N_l$$

This means if  $i$  is even, then  $m_i$  also even. So it is impossible to spread equally in odd places and even places, just one places have both. Therefore, the term  $2\Delta x$  needs to explain this model well. To fix this, we need to add more variety to movement. We can use the 3-choice model instead of the 2-choice model.

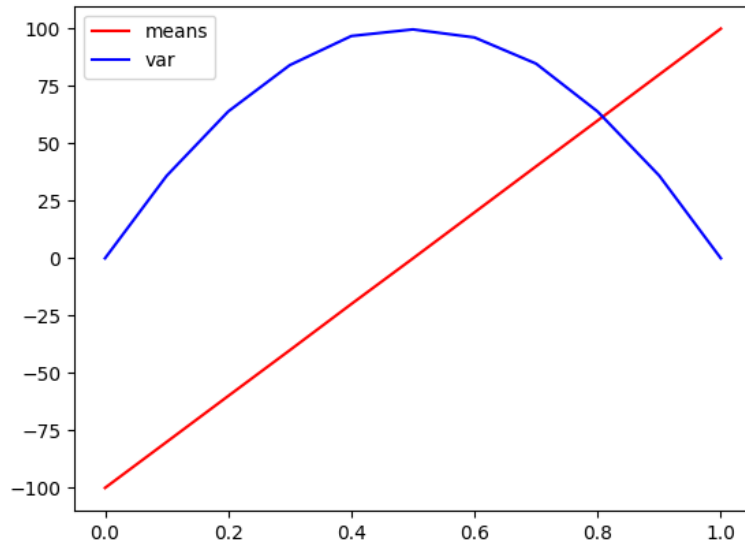


□

### Problem 3

*Solution.* We can guess that if  $p_r$  increases, then  $\mu(t)$  also increases. Increasing  $p_r$  means that each particle can move further to the right. But  $\sigma^2$  is different. It is based on the difference between  $p_r$  and  $p_l$ . Because,  $p_r$  means ‘move to the right’ and  $p_l$  ‘means move to the left’. We can think of them as a force acting in opposite directions. If there is no difference between the two values, the force will be balanced and particles can spread to various places based on the center. However, if the force on one side is strong, particles are concentrated around that side.

The result below is analyzed numerically by varying  $p_r$  from 0 to 1.



□

## Problem 4

*Solution.* The given logistic map is a recurrence relation. Modify it as following:

$$\begin{aligned}x_{k+1} &= rx_k(1 - x_k) \\ \Rightarrow x_{k+1} &= -rx_k^2 + rx_k \\ \Rightarrow rx_k^2 - rx_k + x_{k+1} &= 0\end{aligned}$$

It is hard to solve directly because of the non-linear recurrence form. But since  $x_{k+1}$  will be given, we can consider  $x_{k+1}$  as a constant. Then this is a second-order equation for  $x_k$  if  $r \neq 0$ . The solutions are

$$x_k = \frac{r \pm \sqrt{r^2 - 4rx_{k+1}}}{2r}$$

Note that there are two solutions at most. If we find all  $x_i$  from  $x_0$ , then it takes  $O(n)$  time. However, if starting from the end:  $x_n$ , then it takes  $O(2^n)$  time. Because each  $x_i$  can have 2 states, we should check the validity of all values.

□

## Problem 5

### *Solution.*

질문: 현대의 채용 프로세스는 컴퓨터를 통해 인적성검사를 실시한다. 정해진 답이 있는 것처럼, 문항에 따라 평가 지표가 만들어지고 관리된다. 이것이 사람의 평가보다 공정하다고 할 수 있을까? 평가 알고리즘 또한 사람이 만들었기에 공정하지 않을 수 있고, 오류나 차별이 발생해도 불투명한 과정으로 인해 불이익을 당하는 사람이 있을 수 있다. 현대의 채용 프로세스가 공정하지 못하다면, 어떻게 해결할 수 있을까?

답변: 현대의 채용 프로세스가 공정하지 못할 수 있는 이유에 대해 말씀드리겠습니다.

1. **알고리즘의 편향:** 알고리즘은 그것을 개발한 사람들의 편견이나 선입견을 반영할 수 있습니다. 예를 들어, 특정 인구 그룹에 대한 선입견이나 차별이 반영되어 있을 수 있습니다.
2. **데이터의 편향:** 알고리즘은 학습 데이터를 기반으로 작동합니다. 이 데이터가 특정 인구 그룹을 대표하지 않거나, 특정 그룹에 대한 편향이 있으면, 그 결과도 공정하지 않을 수 있습니다.
3. **투명성의 부족:** 어떤 평가 기준으로 어떻게 평가되는지에 대한 투명성이 부족할 경우, 신청자들은 자신이 왜 불이익을 받게 되었는지 이해하기 어렵습니다.

이러한 문제를 해결하기 위해 몇 가지 방안이 있습니다:

1. **다양한 의견 수렴:** 평가 알고리즘을 개발할 때 다양한 관점을 수렴시키고 다양한 인종, 성별, 백그라운드의 사람들을 참여시켜야 합니다.
2. **데이터의 다양성 보장:** 학습 데이터를 수집할 때, 가능한 모든 인구 그룹을 대표할 수 있는 다양성을 확보해야 합니다.
3. **투명성 확보:** 알고리즘이 어떻게 작동하는지, 어떤 평가 기준으로 평가되는지를 설명하는 투명한 문서를 작성하고, 지원자들에게 공개해야 합니다.
4. **지속적인 감시와 개선:** 채용 프로세스와 알고리즘은 지속적으로 모니터링되어야 하며, 발견된 문제는 빠르게 수정되어야 합니다.
5. **피드백 반영:** 지원자들로부터의 피드백을 수렴하고, 그를 토대로 프로세스를 개선해야 합니다.

이러한 접근 방법들을 통해 공정한 채용 프로세스를 만들 수 있습니다. 또한, 인적성 검사 결과를 토대로 최종 판단을 내리기 전에 다양한 평가 방법을 사용하여 종합적으로 평가하는 것도 중요합니다.

비교: ChatGPT는 상당히 비슷한 해결 방안을 제시했다. 이것이 보편적으로 할 수 있는 방안이기 때문일 것이다. 하지만 ChatGPT는 분석과 개선을 어떻게 해야하는가에 대한 조금 더 구체적인 답변을 제시했다. 많은 사람의 의견을 반영하고, 학습 데이터에 대표성을 보장해야하며, 알고리즘에 대한 문서를 공개할 필요가 있다고 언급했다.

□