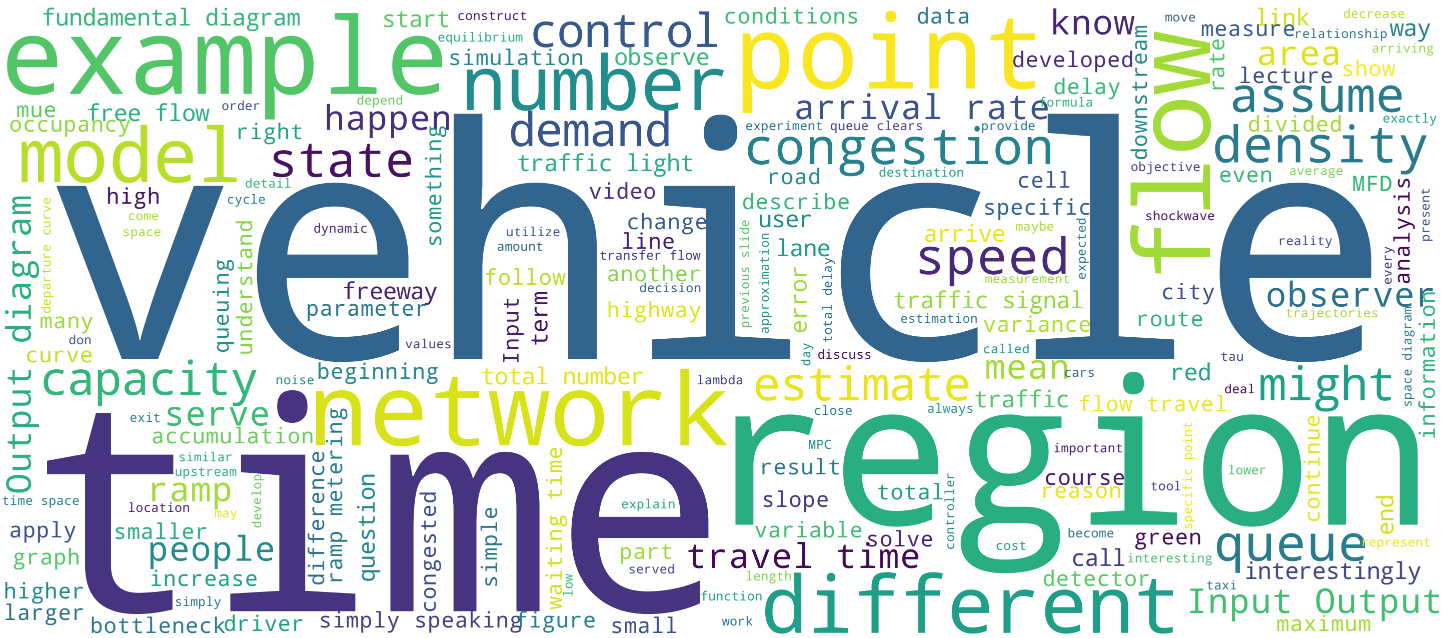
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Input – Output Diagrams

Intro to traffic flow modeling and ITS

Prof. Nikolas Geroliminis

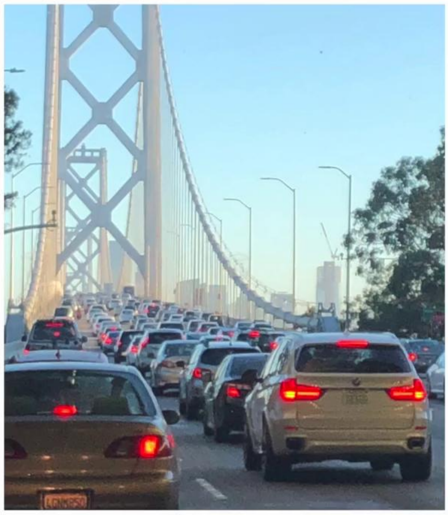


Search MOOC

Video

EPFL

Welcome



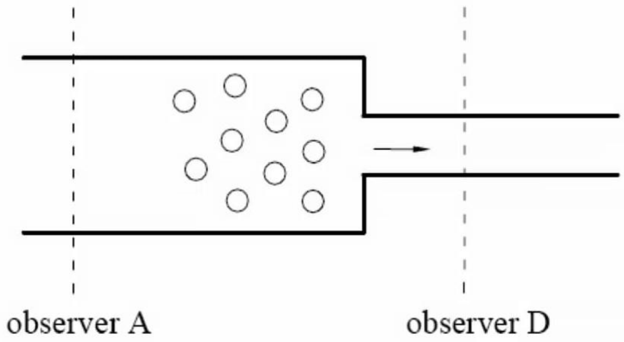
Welcome. My name is Nikolas Geroliminis. In the picture on my left we see a number of vehicles crossing the Bay Bridge in San Francisco under heavily congested conditions. One interesting question we might ask is how we can quantify and we can estimate the collective amount of congestion and delays that these vehicles cause when they were crossing the bridge. And to answer this question we will use a tool which is called Input-Output diagram and we will describe in more detail during this lecture.

Notes

Summary

0m 05s

A queueing system



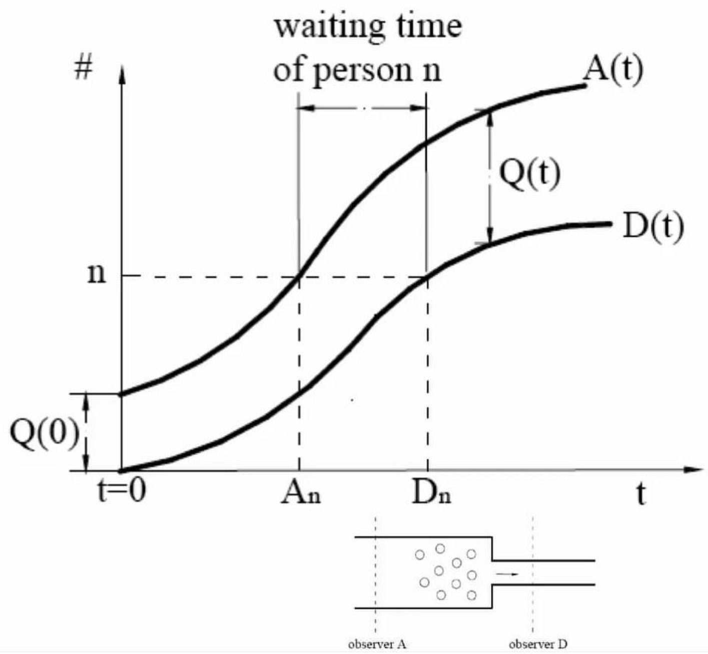
Let's start from the basics. Here we present what is a queuing system. A queuing system is a system where people or vehicles arrive to get some service. This, for example, can be as in the previous picture that vehicles want to cross a bridge or it can be people that arrive in a ticket office to buy some tickets or it can be people that wait in a bar to get a beer. And the question is how we can estimate the amount of time it takes to serve these vehicles or people. And to do that we might not have information what exactly is happening inside the system but we can, let's say, install two observers; one observer upstream that measures the arrivals in the system, call it A and one observer downstream that measures the departure for the system, call it D.

Notes

Summary

0m 43s

Input-Output Diagrams



- $Q(0)$ = number of customers in queue at time 0
- $Q(t)$ = number of customers in queue at time t
- A_n = time of arrival of the n^{th} customer
- D_n = time of departure of the n^{th} customer

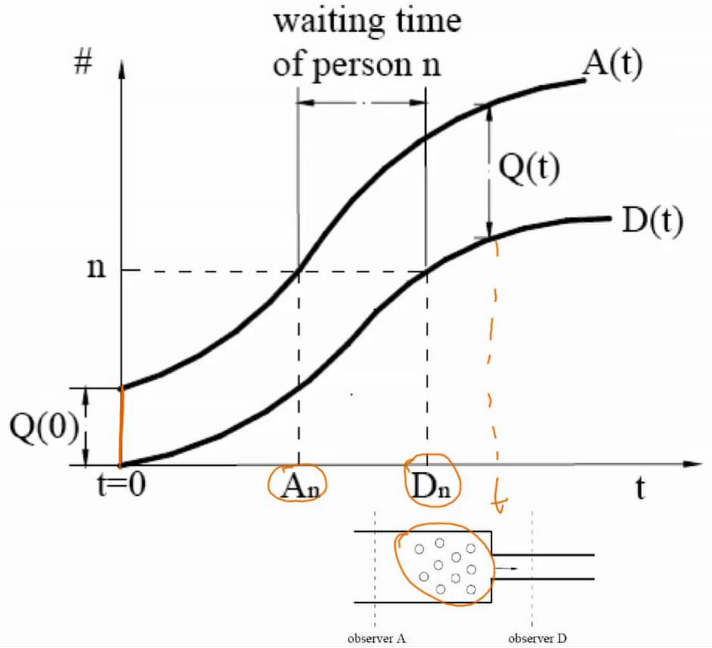
And with these two observers we can measure the total number of vehicles that are entering and exiting the system. So let's present this through an example. Here you see an example of vehicles travelling through a route and in the upper figure you see a time-space diagram and we see the trajectories of four vehicles while in the bottom figure, we see a cumulative plot that actually shows the total number of vehicles measured by observers A and D since the beginning of the experiment. And actually what we see is that when the first vehicle crosses observer A at time 't1', this is the time where the cumulative input increases from zero to one and the same we can say for the remaining vehicles and also for the other observer. Interestingly these cumulative plots of input and output are in principal step functions because vehicles are discrete quantities. But as step functions are challenging to meet to deal with from a mathematical point of view, for the remaining of this lecture we will assume that we will use some approximation that might be like this and it looks like a continuous function. And interestingly given that we have a large number of vehicles, this approximation is very reasonable to describe the phenomenon.

Notes

Summary

1m 47s

Input-Output Diagrams



- $Q(0)$ = number of customers in queue at time 0
- $Q(t)$ = number of customers in queue at time t
- A_n = time of arrival of the n[th] customer
- D_n = time of departure of the n[th] customer

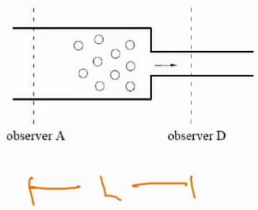
Let's now look more carefully and define exactly what is an Input-Output diagram. Let's follow again the same queuing system as we defined in the beginning and let's assume that we have two observers that observer A measures the cumulative number of arrivals while observer D measures the cumulative number of departures. What we also need to know to do our analysis, we need to know how many vehicles are in the system in the beginning of the experiment at time 't zero', let's call 'Q zero'. So let's try to see what we can observe in this graph. So first of all if we choose a specific vehicle, let's say with an id 'n' which is the nth vehicle observed by the observer. This vehicle entered the system at time 'An' and exited the system at time 'Dn'. So simply speaking, the difference between these two times is the waiting time or the total time spent by these vehicles in this queuing system. And this is the horizontal distance between cumulative input and cumulative output. Let's now look at the vertical distance between these two curves, let's say, at a specific time t. So what we see over here, we see the total number of vehicles that are in the system at time t and actually this is a very interesting quantity because if we sum up all over the time, we can estimate the total vehicle hours spent in the system to serve a number of customers or people or vehicles.

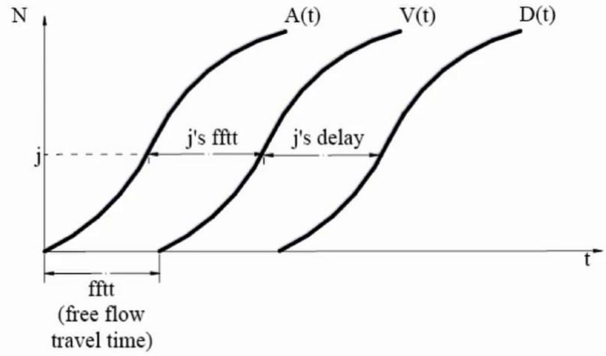
Notes

Summary

3m 29s

A(t), V (t) and D (t) curves





- $V(t)$ represents virtual departures: the time you would have departed the downstream end if there was no queue.
- If L ~ small $\Rightarrow$ $A(t) \approx V(t)$
- If L ~ large $\Rightarrow$ $A(t) \neq V(t)$

So when would [inaudible 5:23:00] traffic systems and for example, consider the vehicles that are transversing one link. The travel time of these vehicles consist of two components. We have the free flow travel time which is the travel time the vehicles would need if they were travelling alone without any effect of congestion and interaction with other vehicles. And we have an additional component which is the delay. So if we consider the cumulative arrival curve A of t and the cumulative departure curve D of t and we take the arrival curve A of t and we shift it by the free flow travel time then what we get, we get what we call virtual departure curve, so virtual departure curve represents the times that these vehicles would exit the system if they were travelling all at the same free flow travel time with zero delay. So if the distance travelled, which in this case is ‘L’, is null then what we can understand that the free flow travel time is almost zero so curves A and V of t coincide. But if A is larger then we have to do this horizontal shift of the curve A to the curve V of t, so it’s something like we subtract the free flow travel time from all the vehicles and then the remaining analysis we do is to estimate the delays of these vehicles in the system.

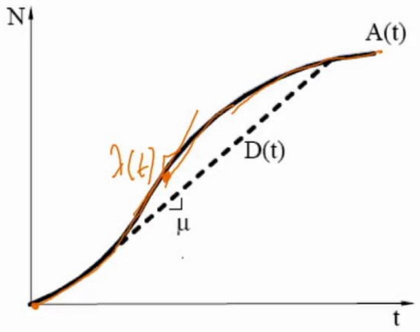
Notes

Summary

5m 22s

Construct I-O curves

- Given $A(t)$, L, v_f and μ (capacity or service rate)



So now I would like to explain to you how we can construct the Input-Output curves if we have some information about the system. So let's assume that we have a queuing system like a road and let's assume that the vehicles are arriving in this system with time-dependent arrival rate. So the A of t curve here represents the cumulative arrivals and if I look at a specific point in time and I take the tangent, the slope of this line is, simply speaking, the arrival rate at time t. So the question is given that we know the arrival rate at time t, the length of the link, the free flow speed and also mue which is the capacity or the service rate, the question is if we can construct the output cumulative curve. So the reason we need to know the length and the free flow is as we discussed in the previous slide, so to shift the arrival curve to the virtual departure curve and subtract the free flow travel time. So let's assume that we have done this already, so what we see in this system is that in the beginning this slope is small which means that the arrival rate is small and then it increases and after some point in time it starts decreasing again and we try to understand how congestion is developed.

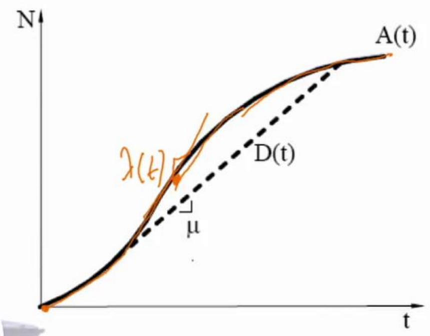
Notes

Summary

6m 56s

Construct I-O curves

- Given $A(t)$, L, v_f and μ (capacity or service rate)



	$Q(t)=0$	$Q(t)>0$
$\lambda(t)>\mu$	μ	μ
$\lambda(t)\leq\mu$	$\lambda(t)$	μ

So here we consider two different cases. The one case is that the arrival rate is larger than the capacity which is the maximum flow that a road or a traffic signal or a bottleneck can serve vehicles and the case where the arrival rate is smaller or equal than the capacity. And what it also matters, it matters what is the state of the system at time t and by the state of the system I mean if at time t there is a queue of vehicles or not in the system of analysis, so if Q of t is zero or if Q of t is larger than zero. So here what we can see, we can see that we have four different combinations and we will explain how we construct the output curve. So in the case that the demand is smaller than the capacity and in the beginning at this specific time we don't have any queues, we serve the vehicles in the rate they arrive. But in case that the demand is higher than capacity, we cannot serve vehicles with higher rate than mue which is the maximum, so we serve with mue. When during initial conditions we have some queues independently if the demand is low or high, we will always serve the system at the maximum possible rate until the queue clears. And actually this is what we see in this figure.

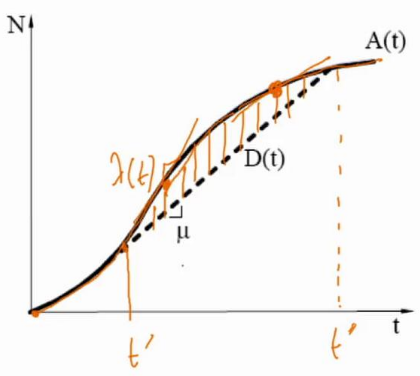
Notes

Summary

8m 39s

Construct I-O curves

- Given $A(t)$, L, v_f and μ (capacity or service rate)



	$Q(t)=0$	$Q(t)>0$
$\lambda(t)>\mu$	μ	μ
$\lambda(t)\leq\mu$	$\lambda(t)$	μ

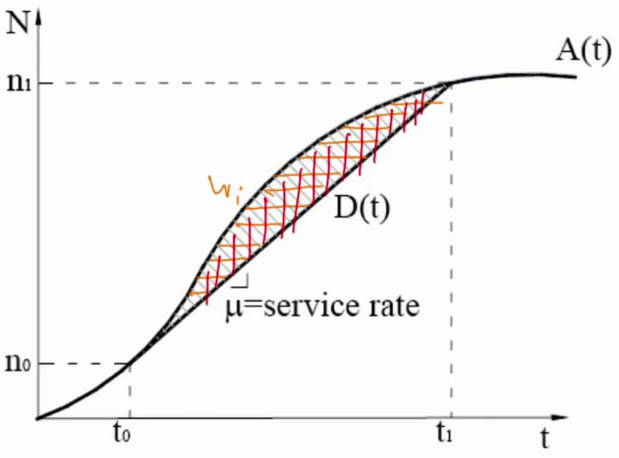
So when the demand is low and there are no queues, the input curve and the output curve coincide where after this point in time where demand starts to get higher, so the slope is higher than mue then what we see, we see that some queue is developed and this queue creates additional delays and even if the demand is lower than the capacity, let's say at this point, we continue to serve at the maximum possible rate. But one where its point in time t double prime which is the time when the queue clears then we return in this scenario and we continue the serve vehicles at the rate they arrive.

Notes

Summary

10m 24s

Little's Formula



Area $= (n_1 - n_0) \cdot \bar{w}$
$= (t_1 - t_0) \cdot \bar{Q}$

$\bar{w}$:average waiting time

$\bar{Q}$:average queue length

λ :average arrival time

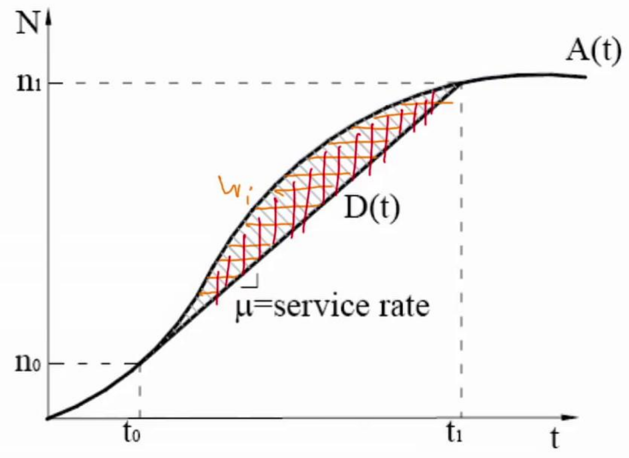
So now we would like to describe the famous and simple formula in queuing theory which is called Little's formula and exists and applies under steady stay conditions. So let's use the example as in the previous slide with time-dependent arrival rate at the bottleneck where flow is small in the beginning then it increases and then it decreases again and we already constructed the cumulative output so the area between these two lines is the total vehicle hours travelled by all vehicles that experience some congestion. And if we want to calculate this area, we have two ways to do it. The one way to do it is to sum horizontally the delays or the waiting times of all the vehicles. So in that case if 'wi' is the waiting time of vehicle 'i', this area is simply speaking the total number of vehicles that experience delay multiplied by the average waiting time. Another way is to look vertically and if we look vertically what we can estimate, we can estimate at every point in time the number of vehicles that are queued in the system. So in that case the same area will be simply speaking the amount of time that we experience congestion multiplied by the average queue length.

Notes

Summary

11m 19s

Little's Formula



$$\text{Area} = (n_1 - n_0) \cdot \bar{w}$$
$$= (t_1 - t_0) \cdot \bar{Q}$$

$$\frac{n_1 - n_0}{t_1 - t_0} = \frac{\bar{Q}}{\bar{w}} = \lambda$$

$\bar{w}$:average waiting time

$\bar{Q}$:average queue length

λ :average arrival time

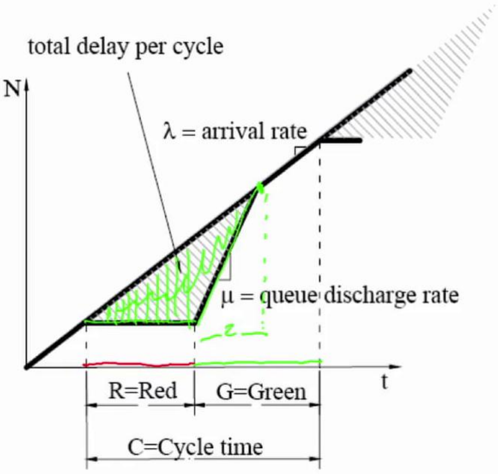
So given that this area is the same, these two quantities have to be the same and if we do some simple calculations what we have, we have that the number of vehicles that queued divided by the time queue was present is equal to average waiting time divided, so as you know that they had a typo here, so is equal to average queue length which is Q bar divided by the average waiting time which is W bar. And what is this term over here? Actually this term over here is the average arrival rate.

Notes

Summary

13m 04s

I-O curves (Example of a traffic signal)



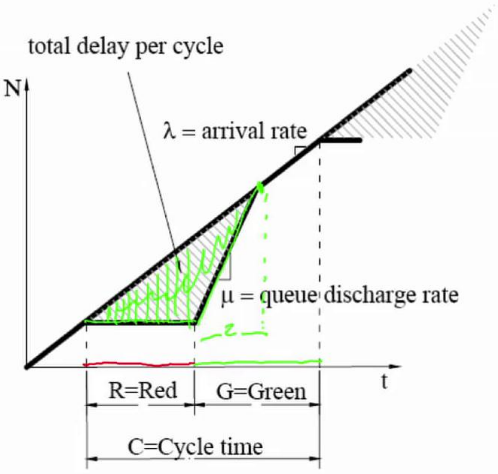
Let's look now at an example of a traffic signal where vehicles are arriving at a constant arrival rate and the traffic signal for some amount of time is red while for another amount of time is green. And for simplicity we can consider that the yellow phase is part of red. So what we see here that when the traffic signal is red, you cannot serve any vehicles. So the service rate is equal to zero and this is the region we see output of the system to be exactly zero. But when the signal turns green given that there is some queue which has been developed here, what it happens is that the traffic signal operates at the maximum discharge rate which in that case is mue and vehicles are serving until the queue clears. So when the queue clears at this specific point in time then the arrival rate is smaller than mue and the vehicles are served in the right they arrive without any delay. And the question is how can we estimate the total delay experienced by all these vehicles which simply speaking is the area of this triangle. So an important variable that I would like to show you here is time tau which is simply the time that the queue is present.

Notes

Summary

13m 50s

I-O curves (Example of a traffic signal)



$(R+\tau)\lambda = \mu\cdot\tau$

$R\lambda = (\mu-\lambda)\tau$

$\tau = \frac{R\lambda}{(\mu-\lambda)}$

$Area = \frac{1}{2} R\mu\cdot\tau = \frac{1}{2}\cdot\lambda\mu\frac{R^2}{\mu-\lambda}$

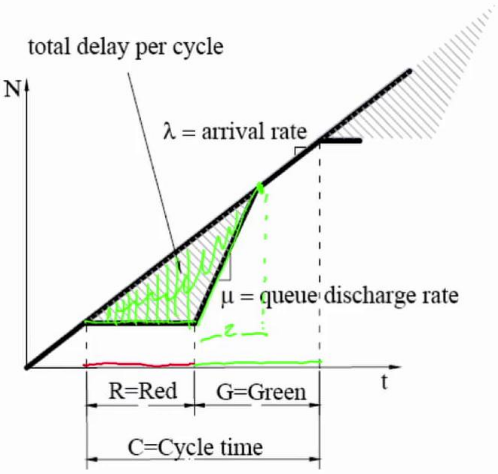
So I would like to focus between the time the green starts and the time R plus tau which is the time needed to clear the queue. The total number of vehicles that arrived in this traffic light during this interval is simply speaking R plus tau multiplied by lambda. But this quantity is equal with the queue discharge rate mue multiplied by tau. So if we do some algebra here, we have that R lambda is mue minus lambda times tau and in that case we can solve for tau and get this specific equation. And now we can write down what is the area of this specific triangle, so this area is one half the base of the triangle which is R multiplied by the height which is mue times tau. So the total area is one half lambda times mue R square divided by mue times lambda. And interestingly what we can observe that as lambda increases, the denominator becomes really small and actually for lambda closer to mue, this area is almost infinity. And actually there is on this means that conditions are over-saturated or otherwise we can call them congested and it means that a traffic light it cannot serve all the vehicles during the cycle they arrive.

Notes

Summary

15m 34s

I-O curves (Example of a traffic signal)



$(R+\tau)\lambda = \mu\cdot\tau$

$R\lambda = (\mu-\lambda)\tau$

$\tau = \dfrac{R\lambda}{(\mu-\lambda)}$

$\text{Area} = \dfrac{1}{2}R\mu\cdot\tau = \dfrac{1}{2}\cdot\lambda\mu\dfrac{R^2}{\mu-\lambda}$

$\lambda\cdot C \leq \mu\cdot G$

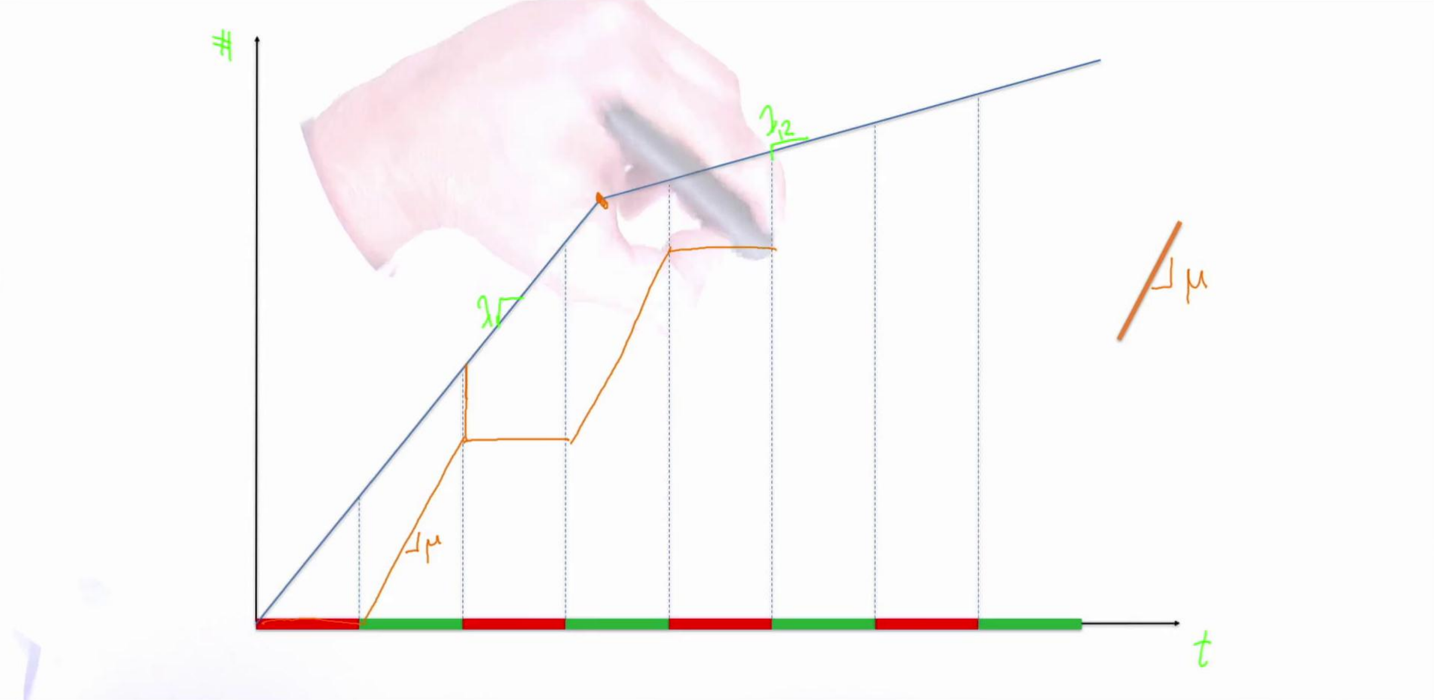
So this formula is valid for under-saturated conditions and these are the conditions where the total number of vehicles that arrive during one cycle which is lambda times C is smaller or equal than the total number of vehicles that can be served which is simply mue times the green time.

Notes

Summary

17m 38s

An oversaturated traffic signal



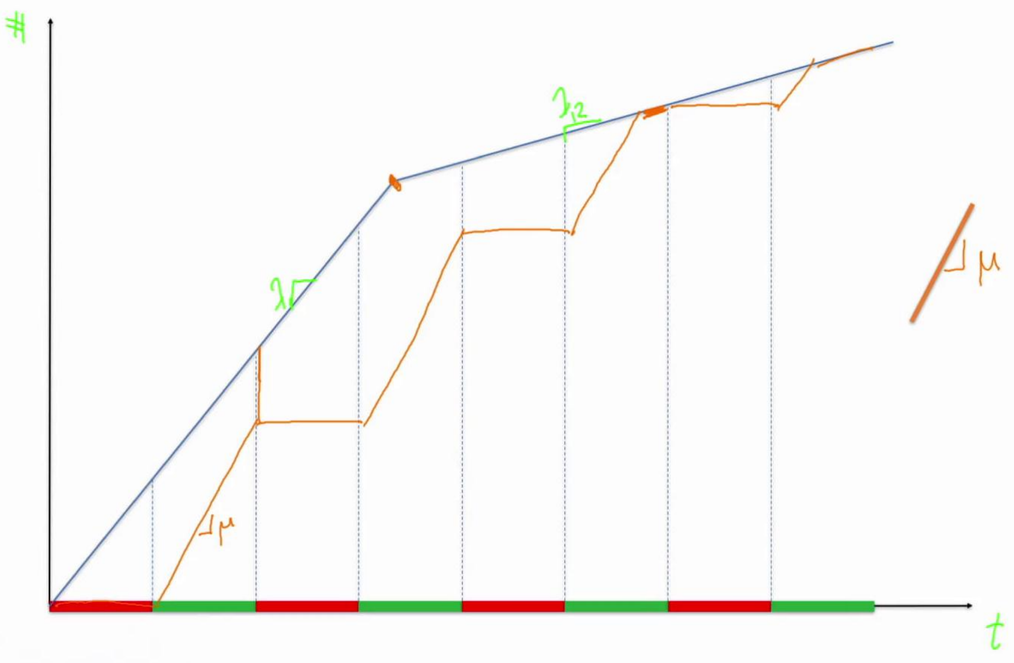
Let's now look at an example of another saturated traffic signal. So what we see here, we see a cumulative plot, so the vertical axis is total number of vehicles the horizontal axis is time, and here we see the signal phases that alternate between red and green and what we have, let's consider that during the peak hour we have a high arrival rate which is lambda one while after the end of the peak we have a lower arrival rate which is lambda two. So now let's assume that the slope of this orange line is the maximum discharge rate of the traffic light and let's try to draw the Input-Output diagram. So during red we cannot serve any vehicles. A queue is developed, so as the queue is developed when the traffic signal turns green, we serve in the maximum rate we can which is mue and what we see, we see that by the end of green there is a remaining queue of vehicles which is a clear example of over-saturated conditions. But we cannot serve these vehicles anymore because the traffic signal turns red and when and this means that by the end of the next red we have a longer queue and now again we serve at the maximum possible until the end of the green but what we have noticed, we have noticed that during this period the arrival flow has decreased.

Notes

Summary

18m 00s

An oversaturated traffic signal



So if we continue like that we are able to clear all the queues and at this point the vehicle will be served at the right they arrive and then it's again red and again green and we return in the end to under-saturated conditions.

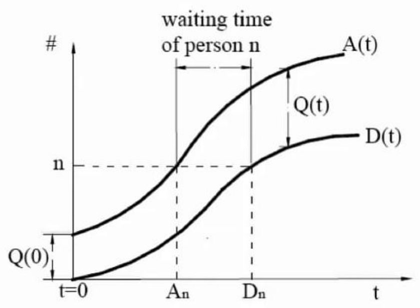
Notes

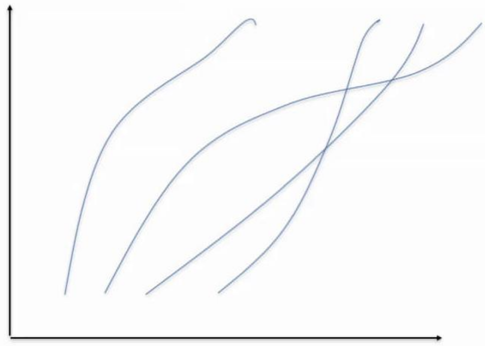
Summary

19m 38s

Discussion (I) - FIFO

- Violation of First in – First out (FIFO) Principle
  - Problematic for individual vehicle delay
  - Indifferent for estimation of aggregated variables (e.g. delays or accumulations)





I would like to finish this lecture by discussing two very interesting points about Input-Output diagrams. So the Input-Output diagrams are fully valid when we have what is called in queuing theory FIFO principle. So the first vehicle that comes in is also the first vehicle that goes out. So it’s something like we don’t have any overtaking.

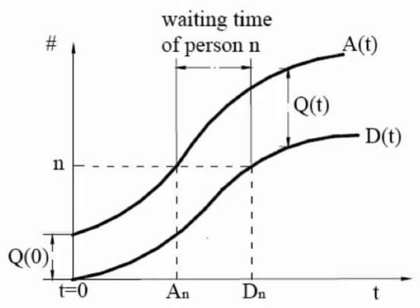
Notes

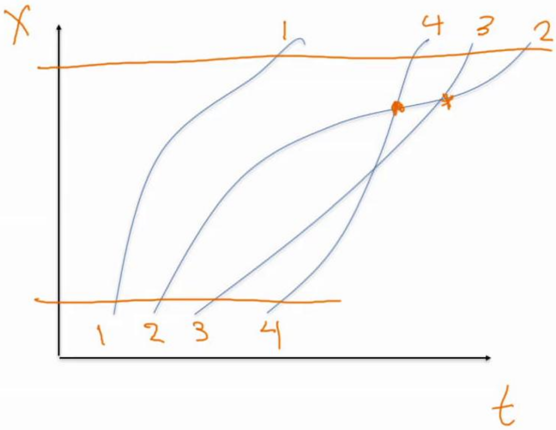
Summary

19m 55s

Discussion (I) - FIFO

- Violation of First in – First out (FIFO) Principle
  - Problematic for individual vehicle delay
  - Indifferent for estimation of aggregated variables (e.g. delays or accumulations)





But in reality what we know in transportation systems is that we really have overtaking because vehicles do not have all the same speed. And the question is does this creates any issues when we do our analysis. So with respect to individual vehicle delays, yes indeed, this can have a problem but given that our objective is to estimate aggregated variables like the total delays or the accumulations of the vehicles at a specific point in time, actually our estimation in terms of aggregated quantities is exact even if vehicles are over passing each other. And a simple intuitive way to understand this is to look at a time-space diagram with some trajectories of vehicles where actually we have some overpassing. So here let's assume that we have as usual two observers; one observer upstream and one observer downstream, so upstream, this is the sequence that vehicles arrive while downstream, the vehicles arrive in a different sequence. But what we could assume is that when a vehicle overtakes another vehicle they simply switch their ids. So in that case at this point vehicle two and four switch their labels and during this time vehicle three and four now switch their labels again.

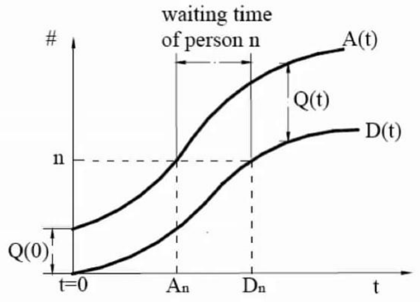
Notes

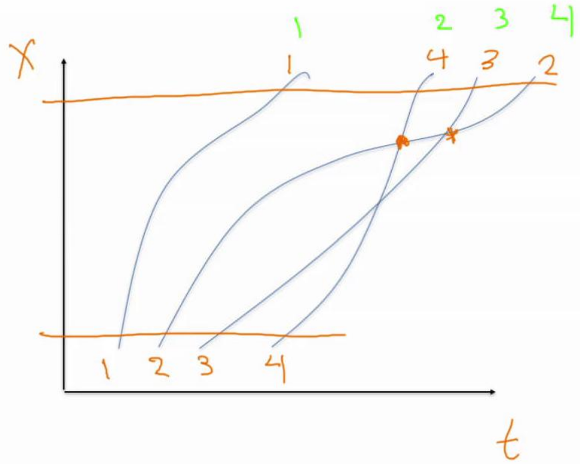
Summary

20m 20s

Discussion (I) - FIFO

- Violation of First in – First out (FIFO) Principle
  - Problematic for individual vehicle delay
  - Indifferent for estimation of aggregated variables (e.g. delays or accumulations)





So if you consider that what will happen, we will have again that the vehicles are ranked as one, two, three, four and this is the reason that even if for individual vehicles the Input-Output diagrams can have errors for aggregated variables so our estimation is exact. So FIFO is not a very important principle.

Notes

Summary

21m 56s

Discussion (II) - Noise

- Effect of measurement errors and noise
  - Theory of random walk

$$Y_t = Y_{t-1} + \varepsilon_t, Y_0 = 0 \qquad \varepsilon_t \sim N(0, \sigma^2)$$

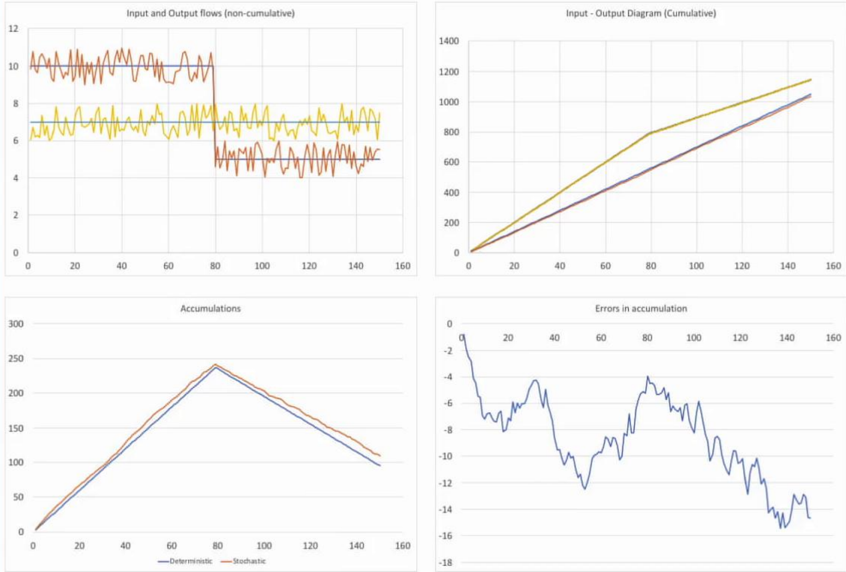
$$E(Y_t) = E(Y_{t-1}) + E(\varepsilon_t) = \cdots =$$
$$= E(Y_0) + E(\varepsilon_1) + E(\varepsilon_2) + \cdots = 0$$

$$Var(Y_t) = Var(Y_{t-1}) + Var(\varepsilon_t) = \cdots =$$
$$= Var(Y_0) + Var(\varepsilon_1) + Var(\varepsilon_2) + \cdots = t\sigma^2$$

Y_t: Position of a walker at time t

ε_t: Random noise

- An example (Errors increase with time)



So far we assumed that the observers that take measurements make perfect measurements without any errors. But as we know the reality is more complicated. So observers and also sensors that we install in transport networks make measurement errors and this might influence our analysis and our results. So to better understand this, I would like to introduce a simple and elegant concept from statistics and this is called random walk. So what is a random walk? Consider a walker that at time zero is at location zero and then at time t, he moves based on his previous position plus some random term E of t. So we assume that this random term E of t is what we call in statistics, random noise which is an random variable that can follow a normal distribution with a mean zero and some standard deviation, let's call sigma. So if we want to estimate what is the expected distance of this walker from the point he was at time zero, what we really need to consider is that the expected value of variable 'Yt' is the sum of the expected value of the variable 'Y t minus 1' plus expected value of the error.

Notes

Summary

22m 19s

Discussion (II) - Noise

- Effect of measurement errors and noise
  - Theory of random walk

$$Y_t = Y_{t-1} + \varepsilon_t, Y_0 = 0 \qquad \varepsilon_t \sim N(0, \sigma^2)$$

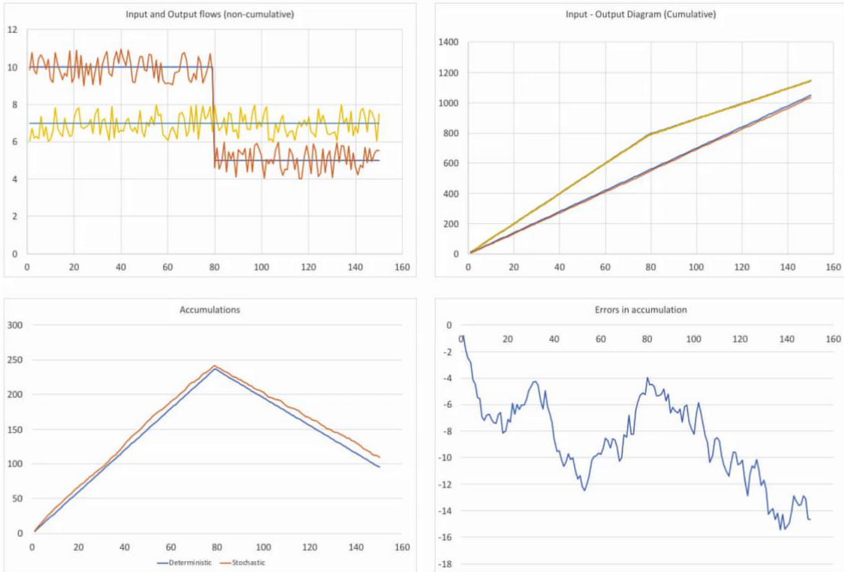
$$E(Y_t) = E(Y_{t-1}) + E(\varepsilon_t) = \cdots = \\ = E(Y_0) + E(\varepsilon_1) + E(\varepsilon_2) + \cdots = 0$$

$$Var(Y_t) = Var(Y_{t-1}) + Var(\varepsilon_t) = \cdots = \\ = Var(Y_0) + Var(\varepsilon_1) + Var(\varepsilon_2) + \cdots = t\sigma^2$$

Y_t: Position of a walker at time t

ε_t: Random noise

- An example (Errors increase with time)



And if we continue to utilize this equation we come to compose this in what is written here but please pay attention that each of these terms is equal to zero because what we have, we have random noises. If now we want to look at variance of the position of the walker at time t, we can only say that the variance of ‘Yt’ is the sum of the variances of ‘Y t minus 1’ and epsilon of t only if these variables are independent which is actually the case because the walker makes a decision at a specific interval without considering what he decided at the previous time interval. So if we do here some simple derivations what we obtain, we obtain that variance of this random walk is t times sigma square which means that the variance increases with time. So if now we want to utilize this theory and make a connection with Input-Output diagrams, Input-Output diagrams utilize cumulative curves so this means that if we have some measurement errors or some noise in the input and in the output measurements, these errors will accumulate with time and actually our estimation can be erroneous. And let’s look at a very simple example.

Notes

Summary

23m 56s

Discussion (II) - Noise

- Effect of measurement errors and noise
  - Theory of random walk

$$Y_t = Y_{t-1} + \varepsilon_t, Y_0 = 0 \qquad \varepsilon_t \sim N(0, \sigma^2)$$

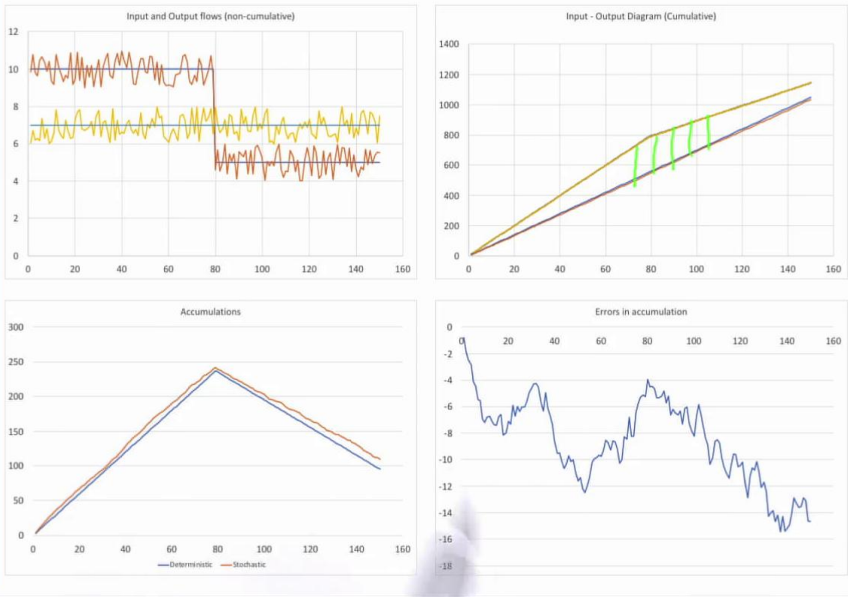
$$E(Y_t) = E(Y_{t-1}) + E(\varepsilon_t) = \cdots =$$
$$= E(Y_0) + E(\varepsilon_1) + \mathrm{E}(\varepsilon_2) + \cdots = 0$$

$$Var(Y_t) = Var(Y_{t-1}) + Var(\varepsilon_t) = \cdots =$$
$$= Var(Y_0) + Var(\varepsilon_1) + Var(\varepsilon_2) + \cdots = t\sigma^2$$

Y_t: Position of a walker at time t

ε_t: Random noise

- An example (Errors increase with time)



So let's consider an example where vehicles are arriving at a specific bottleneck and between time zero and 80 they arrive at a rate 10 where between time 80 and later they arrive at a smaller rate 5. And let's assume that the capacity of this bottleneck is always equal to 7. So let's also assume that what we have is that both the input and the output have a random error which is uniformly distributed between -1 and +1. So what we see in the first figure in the bottom left, we see the input and output flows are non-cumulative ones. And what we learn in this lecture, we can construct the Input-Output diagrams. So what we see here in the second graph in the top right, we see the input and the output and what we also see, we see that vehicles are queued and some congestion is developed but what we observe is that the curves without the noise and with a noise are very close to each other. If now we look at the accumulation in the deterministic and in the stochastic case what we see, we see actually that these two curves are not identical but they start to having some error which interestingly, this error grows with time.

Notes

Summary

25m 28s

Discussion (II) - Noise

- Effect of measurement errors and noise
  - Theory of random walk

$$Y_t = Y_{t-1} + \varepsilon_t, Y_0 = 0 \qquad \varepsilon_t \sim N(0, \sigma^2)$$

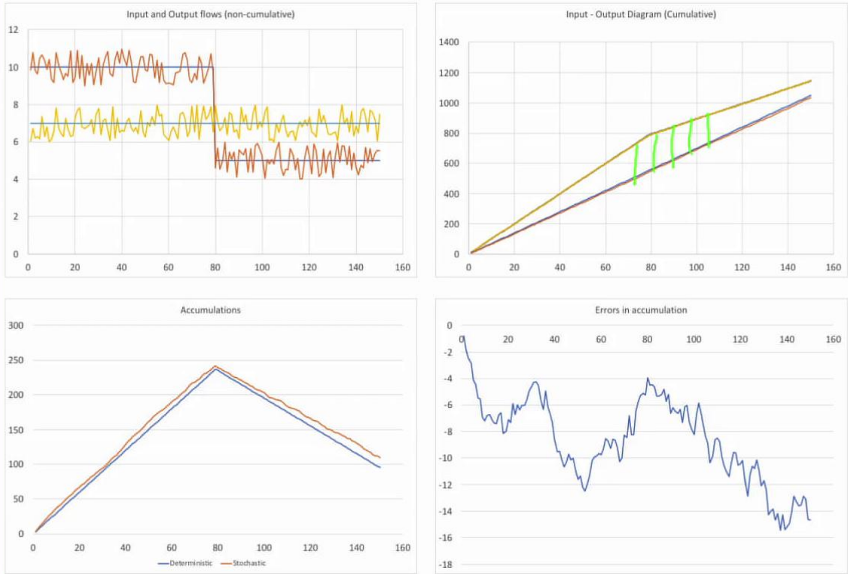
$$E(Y_t) = E(Y_{t-1}) + E(\varepsilon_t) = \cdots = \\ = E(Y_0) + E(\varepsilon_1) + E(\varepsilon_2) + \cdots = 0$$

$$Var(Y_t) = Var(Y_{t-1}) + Var(\varepsilon_t) = \cdots = \\ = Var(Y_0) + Var(\varepsilon_1) + Var(\varepsilon_2) + \cdots = t\sigma^2$$

Y_t: Position of a walker at time t

ε_t: Random noise

- An example (Errors increase with time)



And the way we can see that is in the last figure in the bottom right where we see the error in the accumulation by considering the stochastic arrivals and the deterministic arrivals and as you see at the end of our experiment the error is significantly larger. So this explains that when we deal with Input-Output diagrams, we should be careful when we do our calculations and maybe one solution is that frequently every, let's say, some intervals of time we try to develop some methodology and some estimation method to know what is the exact number of vehicles that are waiting in the system and in this case our approach will be much more accurate.

Notes

Summary

27m 03s

Input/Output Diagram - Summary

So what we learn in this video is how to construct Input-Output diagrams and to utilize them to estimate the congestion and the time delays occurred in a queuing system or in a traffic road and also we learn Little's formula and we saw some examples under when the Input-Output diagrams might not be very accurate. So would like to refer you to have a look at the class material where we also have some additional exercises which I believe will be very interesting for you to solve. See you next week.

Notes

Summary

27m 51s