1. Introduction and Examples

Models in Science and their relation to Mathematical Models

1.1 What is a stochastic process?

1.2 Topics to be discussed

1.3 Examples

- Waiting for a bus
- Play the winner rule
- Prevalence and Incidence Models
- Clinical Trials
- Early Detection of Disease
- Planning Early Detection Trials
- Cell Kinetics
- Models for Prevention Trials

Mathematical Models

- A model can be used to vary the inputs and make predictions of the outputs.
- Mathematical models can be used to guide experiments.
- Nearly all mathematical models are simplifications of the phenomenon being observed. However a useful model captures the important features of the phenomenon.

Models in Science

MODELS: Science has many different kinds of models.

Animal Models: To study human diseases.

Physical Models: To demonstrate the apparent truth of a principle.

Mathematical Models: To gain insight and understanding of a phenomenon which is being observed.

$$\text{inputs} \Longrightarrow \boxed{\text{MODEL}} \Longrightarrow \text{Outputs}$$

A Probability Model is a mathematical model which incorporates the probability aspects of the observed phenomenon. If the observed process has a time parameter (or something equivalent) we refer to the probability model as a stochastic process.

The model can be used to:

- Understand the phenomenon better
- Make predicitons of how outcome change with inputs

1.1 What is a Stochastic Process ?

A Stochastic Process is a collection of random variables $\{X(t), t \in T\}$ indexed by the quantity t taking on values in the set T. Generally t may be time.

$X(t)$ takes on values in a set S called State Space.

Two Major Cases (Time Domain)

Discrete Time: $T = \{0, 1, 2, \dots\}$, then the process is written as $\{X_n, n \geq 0\}$

Continuous Time: $T = [0, \infty)$ we represent the process by $\{X(t), t \geq 0\}$.

Two Major Cases (State Space)

Discrete or Countable: $S : 0, 1, 2, \ldots$

Continuous: $S : (-\infty, \infty)$

If $\{X(t), t \in T\}$ is an observed outcome of a stochastic process it is referred to as a sample path.

The set of all sample paths is the sample space.

Classification of Processes

		Time	
		Discrete	Continuous
State Space	Discrete	—	—
	Continuous	—	—

Goals of Analysis

- Find distribution of $X(t)$
- Find low order moments: mean, variance
- Find behavior at $t \to 0$ or $t \to \infty$
- Find special characteristics of process; e.g. First passage times (time to reach a state for first time).

1.2 Topics to be Discussed

- Review of Laplace Transforms
- Relations between Incidence and Prevalence of a disease
- Simple Time Dependent Models
  - Poisson Processes
  - Birth and Death Processes
- Markov Chains
- Renewal Processes
- Semi-Markov Process
  - Relation to Markov Chains and Processes

Topics to be Discussed (Cont.)

- Application to Clinical Trials
- Applications to Screening Problems
- Applications to Planning Disease Prevention Trials
- Miscellaneous Applications
  - Cell Kinetics
  - Family History Problems

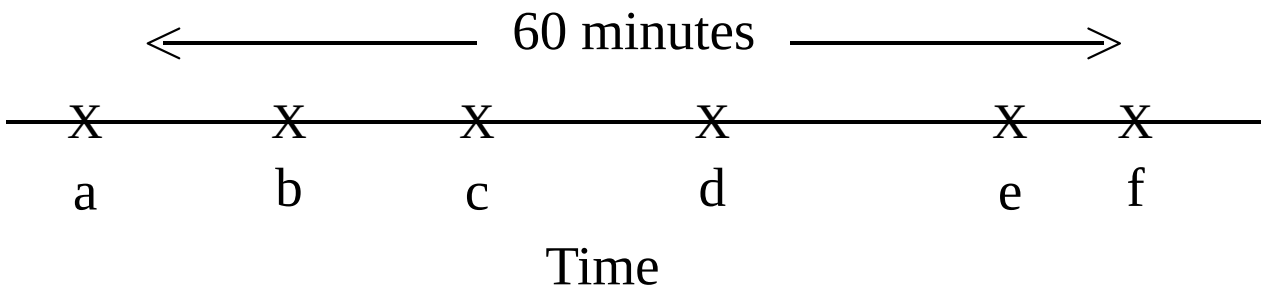
1.3 Examples

Waiting for a BUS

Problem: Why is it when one waits for a bus it always seems to take an unusually long time?

Hypothetical situation

Observed time between bus arrivals

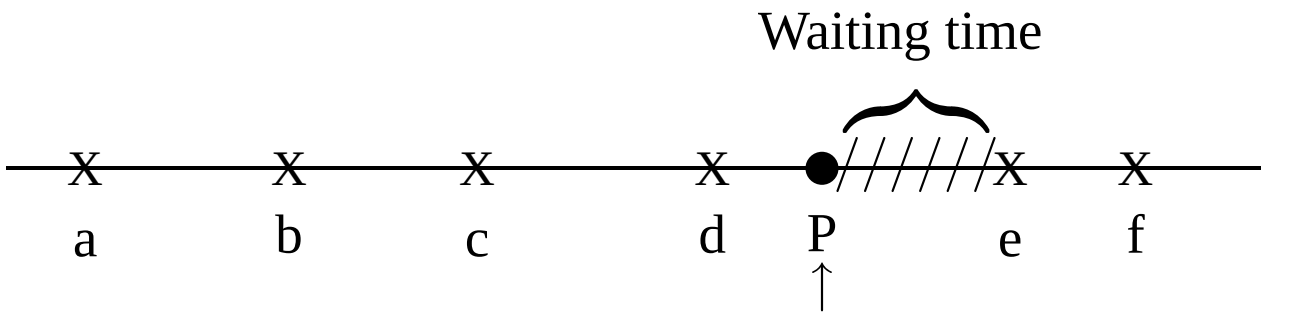


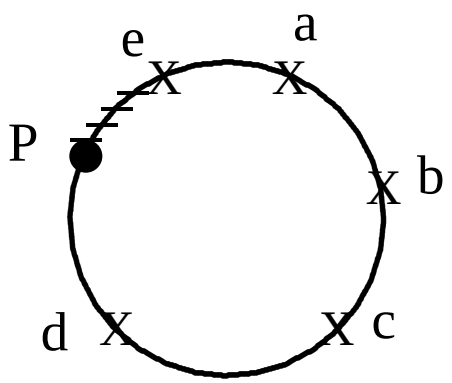
5 Intervals in 60 minutes

$$\text{MTBA} = \text{Mean time between arrivals}$$

$$= \frac{\text{Total observed time}}{\text{No. of intervals}} = \frac{60}{5} = 12 \text{ mintes}$$

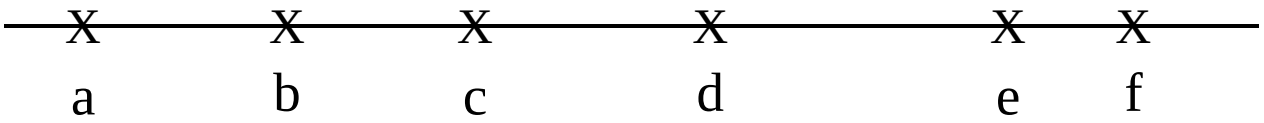
Suppose a friend decides to catch a bus sometime during the observational period. How long does she have to wait for the bus to arrive (on the average)? What is her expected mean waiting time if MTBA = 12 minutes?

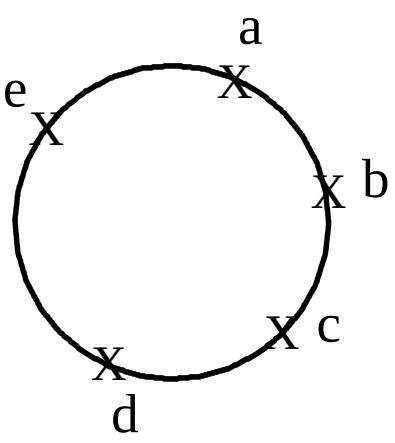




Find Mean distance $\overline{Pe}$

MTBA = 12 minutes

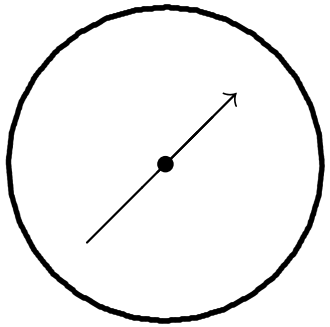




60 minutes clock

5 intervals

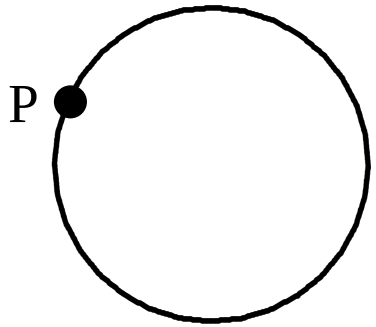
Equivilant to placing 5 points at random on clock; i.e.



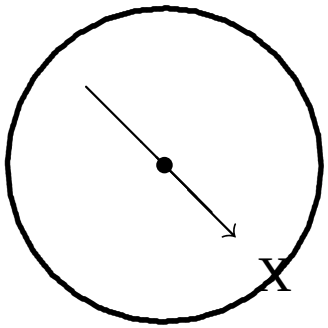
Flip point 5 times

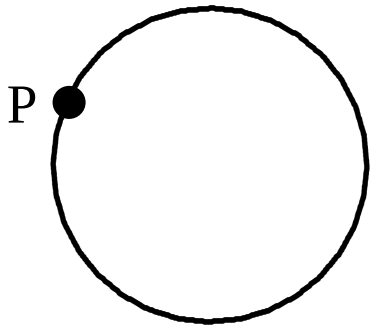
Informal Analysis

Suppose the arrival point is placed on the clock initially



or equivilantly spin the pointer once and move x to P.





Conclusion:

Placing a point on a clock is the same as spinning a pointer once. But we have 5 spins and one placed point. This is equivilant to 6 spins and results in 6 intervals. Hence average time between points is

$$\frac{60 \text{ minutes}}{6} = 10 \text{ minutes}$$

Therefore average waiting time = 10 minutes.

Reason for the unexpected result is that the arrival of friend at the bus stop is equivalent to choosing an inter-arrival interval. The probability of choosing an interval is proportional to the length of the interval.

b. Play the Winner Rule (Two State Markov Chain)

Two treatments are to be evaluated. Outcome of individual observation is Success or Failure.

$$p_i = P\{\text{Success on Treatment } i\} \quad q_i = 1 - p_i, \quad i = 0, 1$$

$$X_n = \begin{cases} 1 & \text{if } \mathrm{n}^{th} \text{ patient is on } 1 \\ 0 & \text{if } \mathrm{n}^{th} \text{ patient is on } 0 \end{cases}$$

PWR: If $(n-1)^{st}$ patient is on i and outcome is Success $\Rightarrow$ assign n^{th} patient to i. If outcome is Failure $\Rightarrow$ assign n^{th} patient to other treatment. Choose $n = 1$ patient to be on treatment i with $prob. = 1/2$.

$$P\{X_1 = 1\} = 1/2$$
$$P\{X_n = 1 | X_{n-1} = 1\} = p_1, \quad P\{X_n = 1 | X_{n-1} = 0\} = q_0$$

$S_n = X_1 + X_2 + \ldots + X_n =$ Number on treatment 1 for first n patients.

Sample Path: $x_1, x_2, \ldots, x_n$

Joint Distribution: $P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$

$$P\{X_1, X_2, \ldots, X_n\} = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)\cdot\cdot P(x_n|x_1, \ldots x_{n-1})$$

But $P(x_r|x_1, x_2, \ldots, x_{r-1}) = P(x_r|x_{r-1})$

$$\therefore P(X_1, X_2, \ldots, X_n) = P(x_1)P(x_2|x_1)(x_3|x_2)\ldots P(x_n|x_{n-1})$$

where the probabilities can be expressed as

$$P = \begin{bmatrix} P(X_n = 0|X_{n-1} = 0) & P(X_n = 1|X_{n-1} = 0) \\ P(X_n = 0|X_{n-1} = 1) & P(X_n = 1|X_{n-1} = 1) \end{bmatrix} = \begin{bmatrix} p_0 & q_0 \\ q_1 & p_1 \end{bmatrix}$$

c. Prevalence and Incidence Models

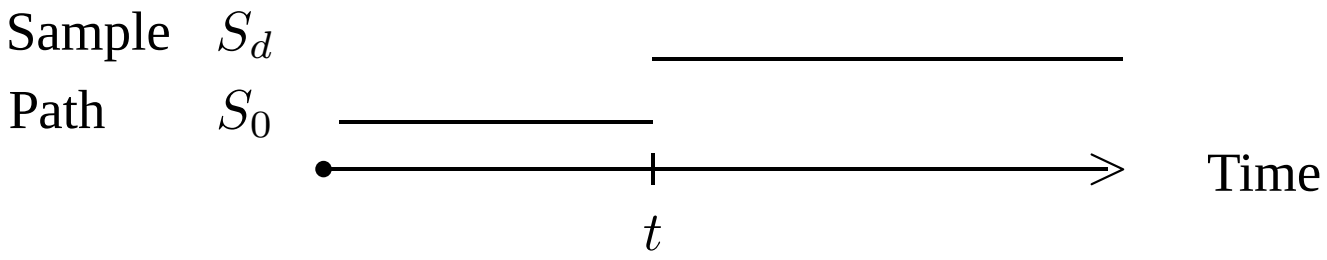
$I(t)$: Incidence of disease at time t

$P(t)$: Prevalence of disease at time t

t: may refer to age or chronological time

States: S_0 : disease free

S_d : disease



$I(t)dt = P\{S_0 \rightarrow S_d \quad \text{during} \quad t, t+dt\}$

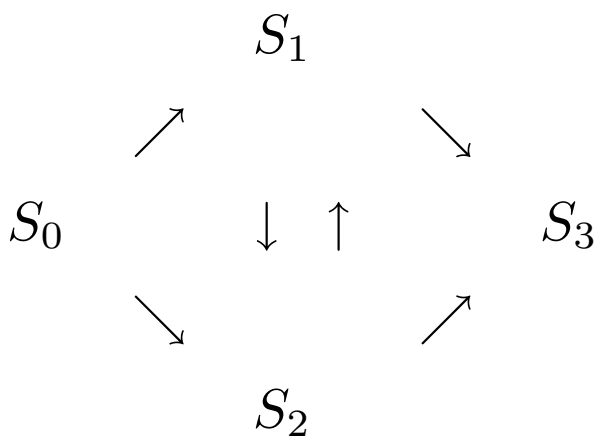
What are relations between $I(t)$ and $P(t)$? Are there differences when t is chronological time or age?

d. Clinical Trials

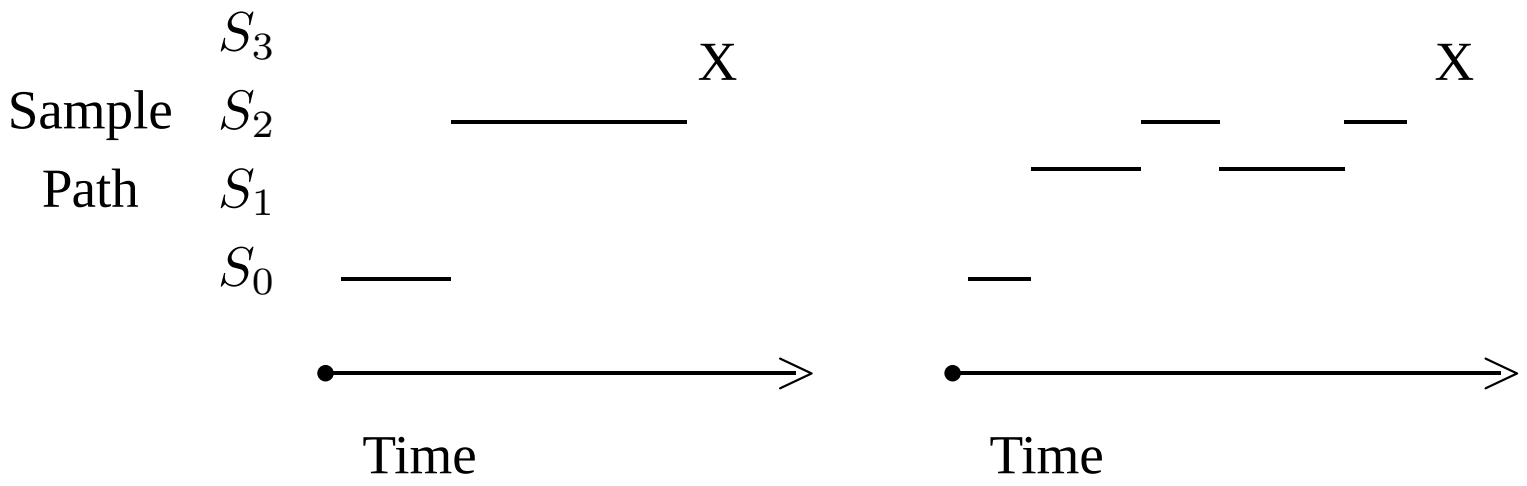
Patients are entered in a clinical trial at time $t = 0$. State of a patient at time $t = 0$ is S_0.

As time goes on, possible other states are:

S_1 : response, S_2 : relapse, S_3 : death (absorbing state)



S_1 : Response, S_2 : Relapse, S_3 : Death



Note: S_3 is absorbing state.

- Survival Time is time from $S_0 \rightarrow S_3$
- Does time in state depend on next state?

e. Early Detection of Disease

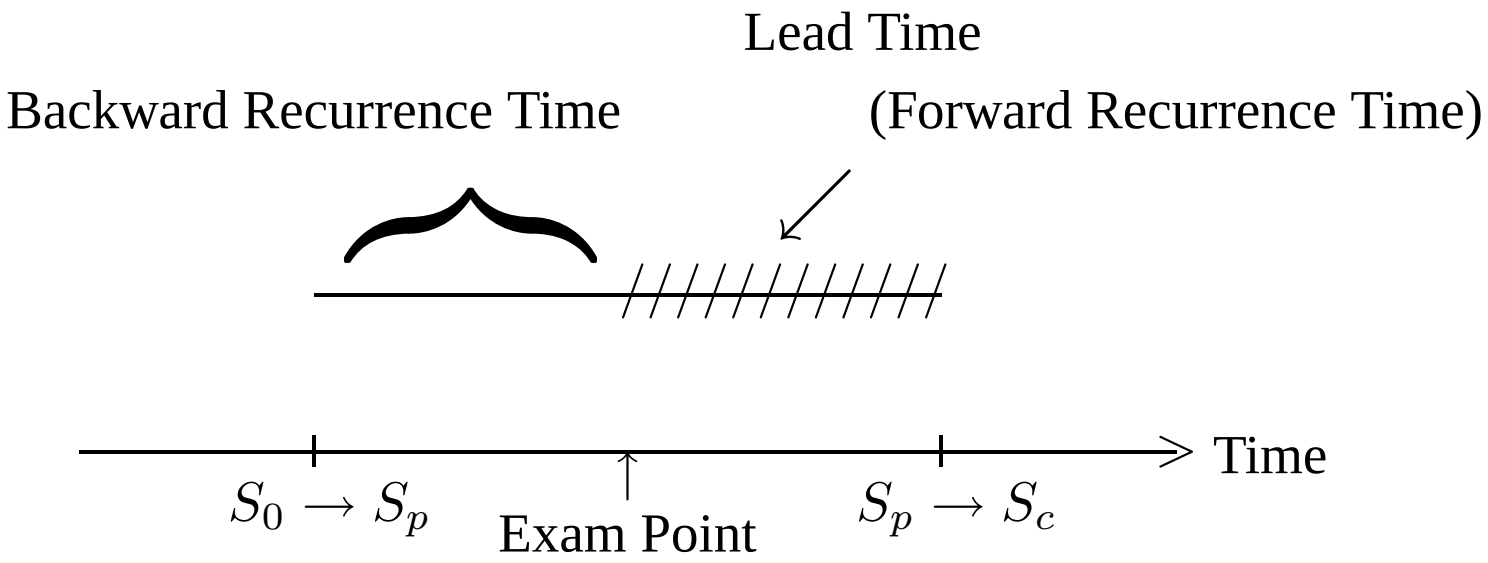
Suppose a person can be in one of three states; i.e.

S_0: Disease Free

S_p: Pre-clinical State (Individual has disease, but is unaware of it. Disease can be diagnosed by special exam)

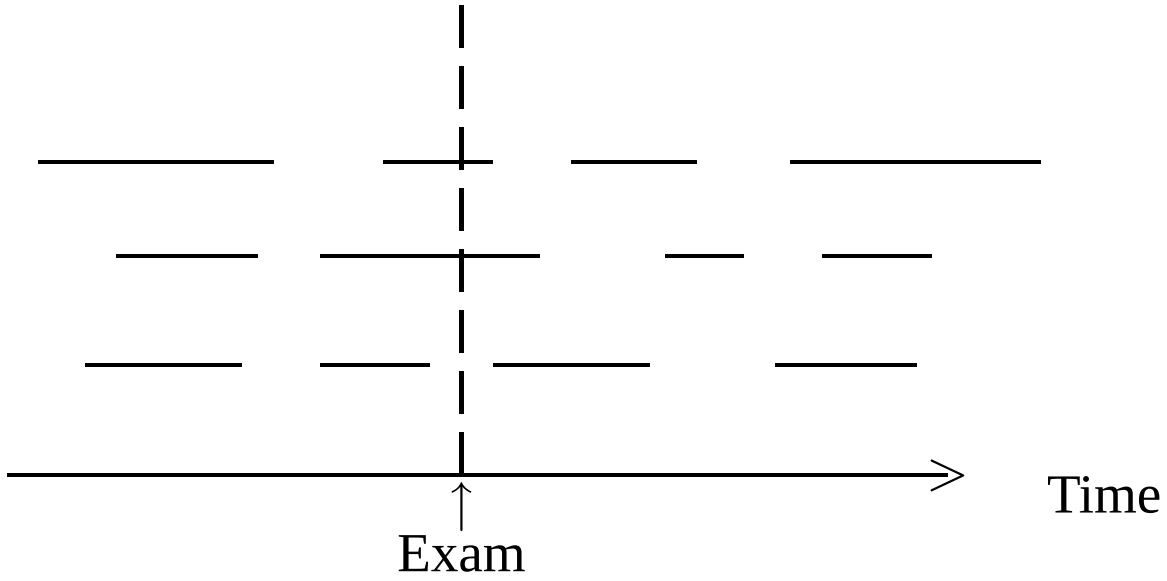
S_c: Clinical Diagnosis

- Progressive Disease Model: $S_0 \rightarrow S_p \rightarrow S_c$
- Goal of Early Detection Program is to find people in S_p



Problems What is distribution of Lead Time? Note lead time is not observed.

What is distribution of backward recurrence time? (Time in S_p before diagnosis). Note that the time when the transition $S_0 \to S_p$ takes place is not observed.



At time of exam, those in S_p are diagnosed (Assuming exam sensitivity is unity). Case finding is “throwing” a random vertical line on figure. Diagnosis is when vertical line intersects horizontal line.

Note larger horizontal lines have a greater chance of being intersected.

Longer horizontal line $\Longrightarrow$ Longer time in S_p

Longer Time in S_p $\Longrightarrow$ slower evolving disease.

Slower evolving disease $\Longrightarrow$ live longer.

Problems

- What is distribution of time in S_p for those diagnosed by special exam compared to those not diagnosed by special exam?
- What is probability of detection by exam if disease incidence depends on age?

f. Planning Early Detection Trials

Eligible $\Longrightarrow$ | RANDOMIZE |

- Study Group (k special exams)

- Control Group (Usual care)

End-Point: Disease Mortality

Issues

How many exams (k)?

Spacing between exams (staggered, equidistant)

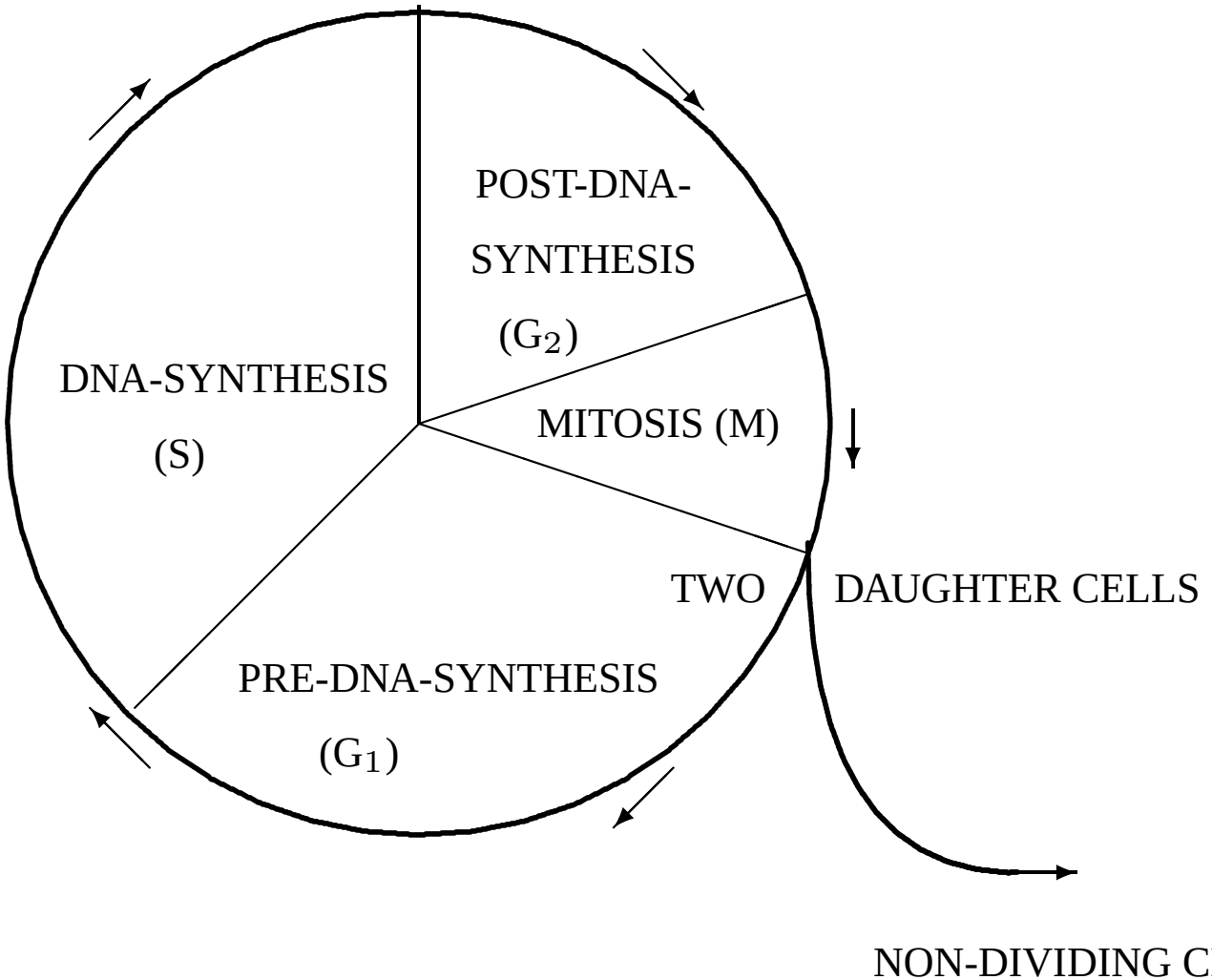
Optimal time for analysis?

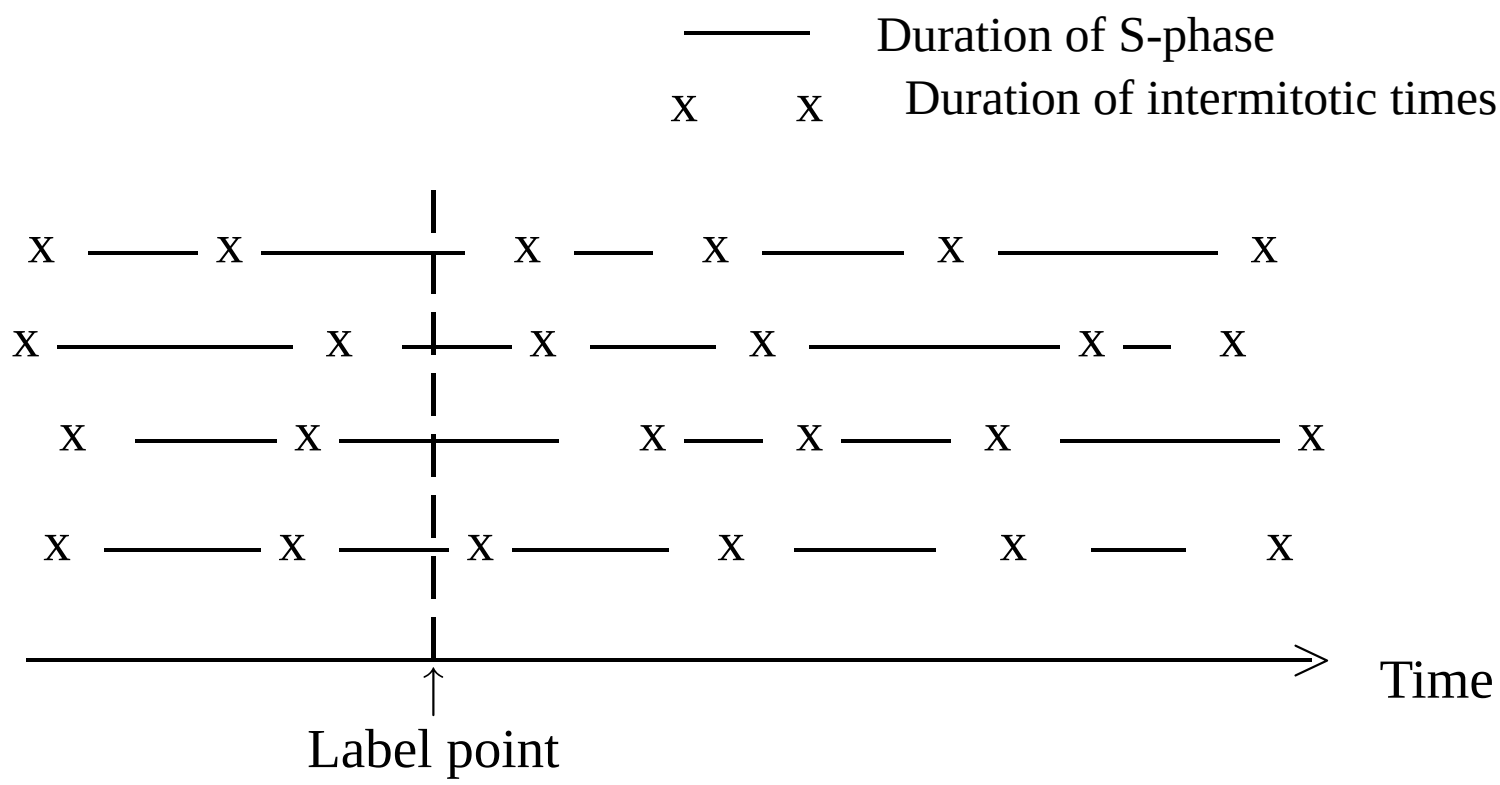
g. Cell Kinetics

Life cycle of cell consists of 4 states; i.e.

$$G_1 \rightarrow S \rightarrow G_2 \rightarrow \text{Mitosis} \rightarrow \text{Daughter Cell}$$

- Interest is in measuring total cell cycle time and time in S-state (DNA Synthesis State)
- Technique is to expose cells to a radio active label (Tridiated Thymadine, HTDR)
- Cells which are in S state at time of labeling incorporate HTDR. Samples of cells are taken at subsequent times.
- Label Index = Proportion of labeled cells = $L(t)$.


POST-DNA-
SYNTHESIS
(G2)
DNA-SYNTHESIS
(S)
MITOSIS (M)
TWO
DAUGHTER CELLS
PRE-DNA-SYNTHESIS
(G1)
NON-DIVIDING CELLS



Population of asynchronous proliferating cells. Labeled cells are those in S-phase at time label is given to cell population.

A cell with long S-phase relative to cell cycle time has a higher probability of being labelled. Therefore population of labelled cells is not the same population as the initial cell population. The sample of cells selected by labelling is neither a representative or random sample of cells.

Problem: How to estimate cell cycle time?

h. Models for Prevention Trials (Cancer)

Define:

S_0: Disease free state

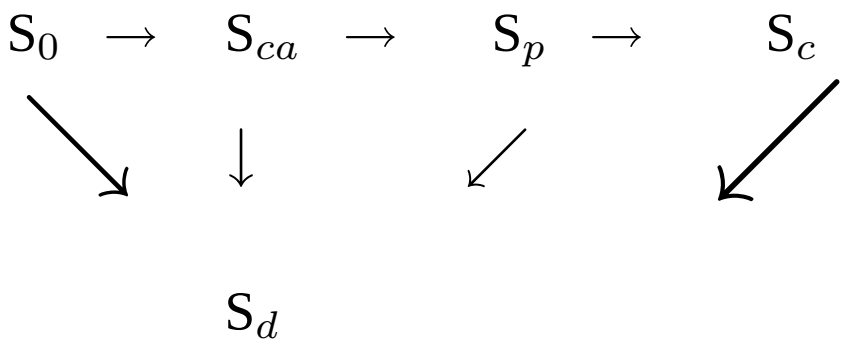
S_{ca}: Carcinogenic state

S_p: Pre-clinical state

S_c: Clinical state

S_d: Death

$S_p \rightarrow S_c$ and transitions $S_0 \rightarrow S_d,\ S_c \rightarrow S_d$ are observable



To prevent disease, it is necessary to either reduce transition $S_0 \rightarrow S_{ca}$ and/or $S_{ca} \rightarrow S_c$.

How does one model process?