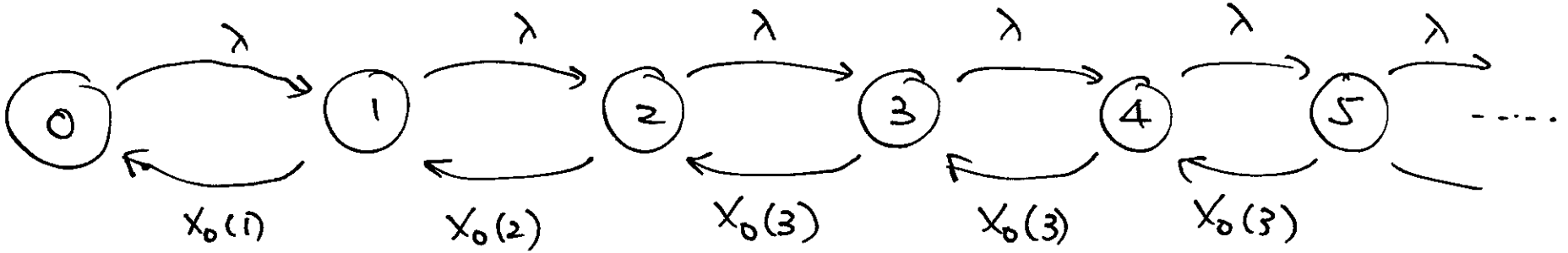

* The transition rate from State k to State (k+1) (for k=0,1,...) is the arival rate of the request.

* The transition rate from State (k+1) to State k (for k=0,...) is the rate at which requests are completed.
  - For State 1 to State 0, this is the same as the throughput of the web server when there is only one client. (Note that throughput is effectively the number of requests completed in an unit time.)
  - For State 2 to State 1, the request completion rate is $X_0(2)$.

- For state 3 to state 2, the request completion rate is $X_0(3)$
- For state (k+1) to state k (where $k \geq 3$), the request completion rate is always $X_0(3)$ because only 3 requests are being processed by the server. The others requests are waiting in the queue.

In order to find the response time, we need to solve the model.

Using the tick given in the notes, we know that

$$P(1) \quad X_0(1) = \lambda P(0)$$

$$P(2) \quad X_0(2) = \lambda P(1)$$

$$P(3) \quad X_0(3) = \lambda P(2)$$

$$P(4) \quad X_0(3) = \lambda P(3)$$

$$P(5) \quad X_0(3) = \lambda P(4)$$

- - - .