의사결정을 위한 확률모형

- 2장 지수분포와 포아송 과정(2) -

김 지 혜 (jihye@pel.sejong.ac.kr)

세종대학교 프로토콜공학연구실

목 차

- 개요
- 복합 포아송 과정
- 비정상 포아송 과정

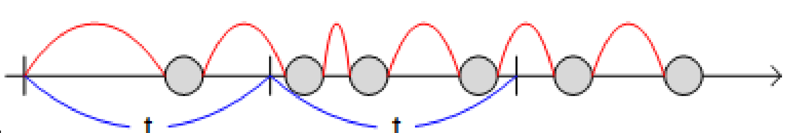
목 차

- 개요
- 복합 포아송 과정
- 비정상 포아송 과정

개 요

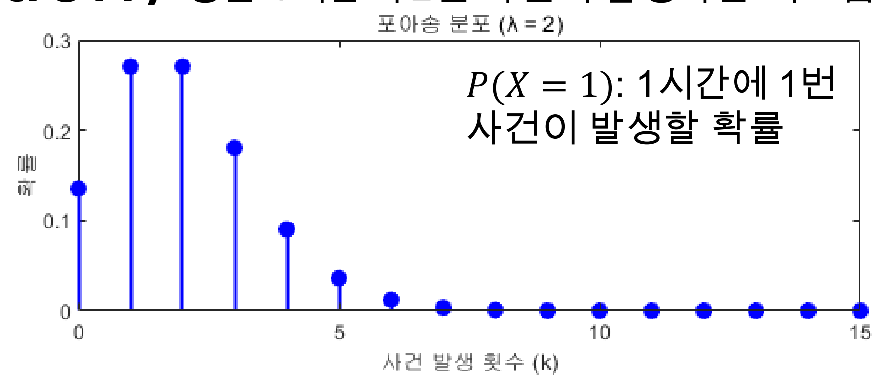
- 포아송분포(Poisson Distribution)
  - 단위시간 당 평균 사건 발생횟수가 λ인 사건 **<u>발생횟수(k)</u>**의 분포
  - $f(k) = \frac{e^{-\lambda}\lambda^{k}}{k!}$, $\left(k = \text{발생횟수}, \lambda = \frac{\text{발생횟수}}{\text{단위시간}}\right)$
- 지수분포(Exponential Distribution)
  - 단위시간 당 평균 사건 발생 횟수가 λ인 사건이 처음 발생할 때까지의 **<u>대기시간(x)</u>**의 분포
  - $F(x) = 1 - e^{-\lambda x}$
  - $f(x) = \lambda e^{-\lambda x}$, $\left(\text{평균시간간격} = \frac{1}{\lambda}\right)$

회색 동그라미: 사건
파랑: 단위시간 t
빨강: 사건 간 대기시간



<그림 2.1> 포아송분포와 지수분포

평균 1시간에 2번 사건이 발생하는 시스템



목 차

- 개요
- 복합 포아송 과정
- 비정상 포아송 과정

복합 포아송 과정

- 정의

정의 2.5

$N(t) \sim PP(\lambda)$라고 하고, iid인 확률변수 $G_1, G_2, \ldots$의 상태공간을 $\{1,2,\ldots\}$라고 하자.
$M(t) = G_1 + G_2 + \cdots + G_{N(t)}$라고 할 때, $\{M(t), t \geq 0\}$을 복합(Compound) 포아송 과정이라고 한다. 즉, 사건 발생은 포아송 과정을 따르지만 각 사건마다 iid 확률변수 만큼의 증가가 이루어지는 경우의 총 증분을 나타내고 있다.

*iid: independent, identically distribution(독립적이고 동일한 분포)
G: 발생한 사건에 대한 증가량(크기)

=> 복합 포아송 분포의 확률변수는 $M(t)$이다. 이 $M(t)$는 포아송분포의 확률변수인 $N(t)$의 사건 크기인 G의 합과 동일하다.

- 특징
  - 독립적인 포아송 사건들의 크기를 합산하여 생성됨
  - 포아송 과정의 단일증분을 제외하고, 2개 이상의 사건에 대한 동시 발생을 허용하는 경우를 포함함

*단일증분: 충분히 작은 구간 내 2개 이상의 결과가 동시 발생할 확률은 0에 가깝다는 가정

복합 포아송 과정

- 일반 포아송 과정과의 비교

구분	일반 포아송 과정	복합 포아송 과정
확률변수	$N(t) \sim PP(\lambda t)$	$M(t) = G_1 + G_2 + \cdots + G_{N(t)}$
증가량	사건마다 1씩 증가	사건마다 임의의 크기씩 증가
기댓값	λt (λ: 사건 발생률, t: 시간)	$\lambda t E(G)$ (G: 발생한 사건에 대한 크기(증가량))
분산		$\lambda t E(G^2)$
단일증분 충족 여부	**충분히 작은 구간에서 2개 이상의 사건이 발생할 확률은 0에 가까움**	**단일증분을 이완하여 충분히 작은 구간에서 2개 이상의 사건발생을 허용하는 경우를 포함**
모델링 대상	사건 발생횟수	사건 발생횟수 및 크기
응용 분야	e.g., 고객 도착 횟수, 고장 횟수 등	e.g., 거래 횟수 및 거래 금액, 자연재해 발생 횟수 및 피해 규모 등

- e.g., 네트워크 서버에 도착하는 패킷 모델링 관련 예시
  - 일반 포아송 과정: 네트워크 서버에 도착하는 패킷의 수를 모델링
  - 복합 포아송 과정: 네트워크 서버에 도착하는 패킷의 수와 각 패킷의 크기를 고려하여 전체 데이터 전송량을 모델링

복합 포아송 과정

- 예제 2.18

네트워크 서버에 패킷이 도착하는 횟수 $N(t)$가 $PP(\lambda)$라고 한다. 해당 서버에는 평균적으로 초당 1,000개의 패킷이 도착하고 각 패킷의 크기가 평균 500Byte일 때, 5초 동안 서버에 도착하는 패킷의 수와 각 패킷의 총 전송량을 구하라.

- 5초 동안 서버에 도착하는 패킷의 수는 <u>포아송 분포</u>를 활용하고, 각 패킷의 크기를 고려하여 총 데이터 전송량을 계산하기 위해서는 <u>복합 포아송 분포</u>를 활용함

- 포아송 분포를 활용한 패킷 도착 횟수는 $t = 5,\ N(5)$
  - $N(5) \sim PP(1{,}000 \times 5) = PP(5{,}000)$, 평균 5,000개를 의미함

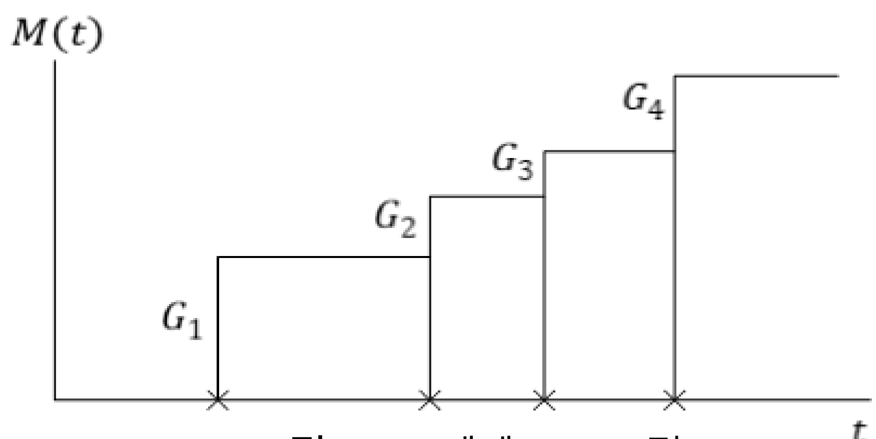
- 각 패킷 크기$(G_1, \ldots, G_5)$의 평균 G는 500Byte임
- 복합 포아송 분포를 활용한 총 패킷 크기는 $\mathrm{t} = 5,\ M(5)$
  - $M(5) = \sum_{i=1}^{N(5)} G_i$
  - 이때 5초 동안의 평균 데이터 전송량을 구하고자 하는 경우, 기댓값을 활용
  - $E\big(M(5)\big) = \lambda t E(G) = 1000 \times 5 \times 500 = 2{,}500{,}000 Byte$

복합 포아송 과정

- 예제 2.19

시점 t까지 고속도로 톨게이트를 통과한 차량의 대수 $N(t)$가 $PP(\lambda)$라고 하자.
또한 확률변수 $G_k(k = 1,2, \ldots)$는 k번째 통과한 차량 내에 탑승한 승객의 수하고 하면, 시점 t까지 톨게이트를 통과한 승객의 수 $M(t) = G_1 + G_2 + \cdots + G_{N(t)}$는 복합 포아송 과정이다.

- 차량 통과 과정: 포아송 과정
  - 시점 t까지 톨게이트를 통과한 차량의 수: $N(t)$
- k번째 차량에 탑승한 승객 수: 확률변수 G_k
- 승객 수의 총합: 복합 포아송 과정 $M(t) = G_1 + G_2 + \cdots + G_{N(t)}$



X 부분: 차량이 톨게이트를 통과하는 순간

<그림 2.22> 예제 2.19 그림

목 차

- 개요
- 복합 포아송 과정
- 비정상 포아송 과정

비정상 포아송 과정

- 정의

정의 2.6

다음과 같은 가정을 만족하는 확률과정 $\{N_1(t), t \geq 0\}$은 비정상 포아송 과정이다.
(i) (독립증분) 겹치지 않는 두 구간에서 각각 발생한 사건의 개수들은 서로 독립이다.
(ii) (단일증분) 충분히 작은 값 dt(시간변화량)에 대하여 다음이 성립한다.
  (a) $P(N(t+dt) - N(t) = 0) \approx 1 - \lambda(t)dt$
  (b) $P(N(t+dt) - N(t) = 1) \approx \lambda(t)dt$
  (c) $P(N(t+dt) - N(t) \geq 2) \approx 0$
이때 비정상 포아송 과정을 따르는 확률변수는 $N(t) \sim NSPP\big(\lambda(t)\big)$로 표현한다.

- 특징
  - 포아송 과정의 정상증분을 제외한 경우를 포함함

*정상증분: 특정 구간 내 사건 발생횟수는 그 구간의 시간적 위치와 무관하다는 가정

비정상 포아송 과정

- 일반 포아송 과정과의 비교

<table>
<tr><th>구분</th><th>일반 포아송 과정</th><th>비정상 포아송 과정</th></tr>
<tr><td>확률변수</td><td>$N(t) \sim PP(\lambda t)$</td><td>$N(t) \sim PP(\lambda t)$</td></tr>
<tr><td>시간 의존성</td><td>시간과 무관하게 사건 발생</td><td>시간에 따라 발생률($\lambda(t)$)이 다를 수 있음</td></tr>
<tr><td>기댓값</td><td rowspan="2">λt
(λ: 사건 발생률, t: 시간)</td><td rowspan="2">$\int_0^t \lambda(s)\, ds$
($\lambda(s)$: 사건 발생률, t: 시간,
즉, 시간에 따른 누적 발생률에 의해 결정)</td></tr>
<tr><td>분산</td></tr>
<tr><td>정상증분
충족 여부</td><td>정상증분에 의해 시간적 위치와 무관하게
특정 구간의 사건 발생횟수는 동일함</td><td>정상증분을 이완하여 시간에 따라 특정
구간의 사건 발생횟수가 달라질 수 있음</td></tr>
<tr><td>모델링 대상</td><td>사건 발생률이 일정한 시스템</td><td>사건 발생률이 시간에 따라 변하는 시스템</td></tr>
<tr><td>응용 분야</td><td>e.g., 공장 기계의 고장 횟수 등</td><td>e.g., 특정 시간대에 몰리는 문의전화 수 등</td></tr>
</table>

- e.g., 네트워크 서버에 도착하는 패킷 모델링 관련 예시
  - 일반 포아송 과정: 네트워크 서버에 일정하게 도착하는 패킷의 도착률을 모델링
  - 비정상 포아송 과정: 네트워크 서버에 시간에 따라 다르게 도착하는 패킷의 도착률을 모델링

비정상 포아송 과정

- 예제 2.20

네트워크 서버 A와 B에 패킷이 도착하는 횟수 $N(t)$가 $PP(\lambda)$라고 한다.

(a) 서버 A에는 패킷이 일정하게 도착한다고 가정할 때, 5초 동안의 패킷 도착률은?

(b) 서버 B에는 오전 9시부터 오후 6시까지는 패킷 도착률이 초당 평균 2,000개로 증가하고, 오후 6시 이후에는 초당 평균 500개로 감소한다고 가정할 때, 5초 동안의 패킷 도착률은?

(a) 풀이

- 서버 A에는 패킷이 일정하게 도착하므로, 시간적 위치와 무관하여 일반 포아송 분포로 모델링할 수 있음
- $N(5) \sim PP(1{,}000 \times 5) = PP(5{,}000)$, 5초 동안의 패킷 평균 도착률은 5,000임

(b) 풀이

- 서버 B에는 시간에 따라 패킷 도착률이 다르므로, 시간에 따른 누적 도착률을 활용해야 하며 비정상 포아송 분포로 모델링할 수 있음
- 오전 9시부터 오후 6시까지의 도착률: $\int_0^5 \lambda(t)\,dt = 2{,}000 \times 5 = 10{,}000$
- 오후 6시 이후의 도착률: $\int_0^5 \lambda(t)\,dt = 500 \times 5 = 2{,}500$

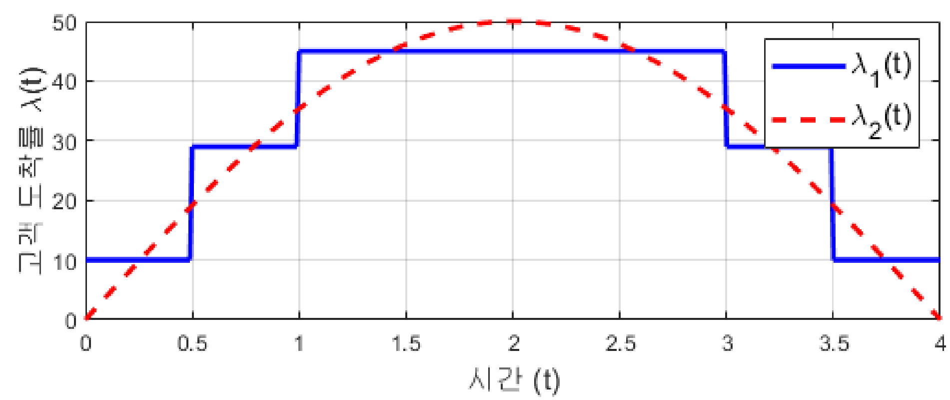
비정상 포아송 과정

- 예제 2.21

11:30부터 13:30까지 식당에 도착하는 고객들을 비정상 포아송 과정으로 모형화할 수 있다. 11:30을 시점 0으로 하고 30분 단위로 1씩 늘게 하여 13:30을 시점 4로 표현한다. 고객들의 도착률은 비정상의 시간 의존형이며 다음 $\lambda_2(t)$와 같은 곡선으로 표시할 수도 있으나 이를 근사하여 $\lambda_1(t)$와 같이 부분선형 함수로 나타낼 수도 있다.

$$\lambda_1(t) = \begin{cases} 10, 0 \le t < 0.5, \\ 29, 0.5 \le t < 1, \\ 45, 1 \le t < 3, \\ 29, 3 \le t < 3.5, \\ 10, 3.5 \le t < 4, \end{cases} \quad \text{또는 } \lambda_2(t) = 50 \sin\left(\frac{\pi t}{4}\right), \ 0 \le t \le 4.$$

- 정상증분이 제외됨에 따라, 네 개의 시점에 따른 구간별 고객 도착률이 다양함
- $\lambda_1(t)$는 실제 시간에 따른 도착률의 변화를 반영한 함수
- $\lambda_2(t)$는 시간에 따른 도착률 변화를 선형으로 근사한 함수



<그림 2.23> 예제 2.21 그림

비정상 포아송 과정

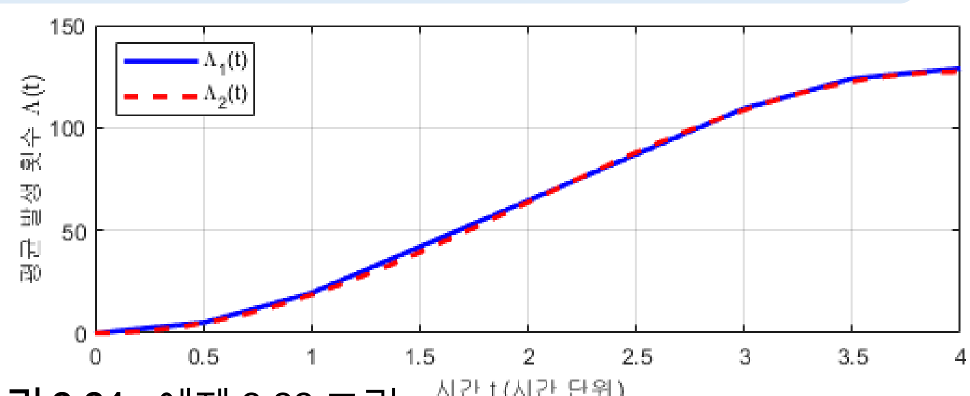
- $\lambda(s)$는 시점 s에서의 순간 발생률이므로, 시간 변화량에 따라 이를 적분하는 경우에 사건 발생횟수의 평균이 됨
  - $\Lambda(t) = \int_0^t \lambda(s)ds = E\big(N(t)\big)$는 구간 $[0, t]$에서 사건 발생 횟수의 평균

- 예제 2.22

예제 2.21에서, $\Lambda_1(t), \Lambda_2(t)$는 시점 t까지의 평균 발생횟수에 해당한다.

$$\Lambda_1(t) = \begin{cases} 10t, & 0 \le t < 0.5, \\ 5 + 29(t - 0.5), & 0.5 \le t < 1, \\ 19.5 + 45(t - 1), & 1 \le t < 3, \\ 109.5 + 29(t - 3), & 3 \le t < 3.5, \\ 124 + 10(t - 3.5), & 3.5 \le t < 4, \end{cases} \quad , \quad \Lambda_2(t) = \int_0^t 50\sin(\pi x/4)dx = \frac{200}{\pi}\left(1 - \cos\left(\frac{\pi t}{4}\right)\right), 0 \le t \le 4.$$

- $\Lambda_1(t), \Lambda_2(t)$에 의하면, $\Lambda_1(t)$의 직선이 $\Lambda_2(t)$의 곡선에 매우 근사하므로, 예제 2.21과 같이 도착률 함수가 $\lambda_2(t)$ 같은 곡선인 경우, 구간을 구분하여 부분선형 함수 $\lambda_1(t)$처럼 표현할 수 있음



<그림 2.24> 예제 2.22 그림

Thanks!

김 지 혜 (jihye@pel.sejong.ac.kr)

- MATLAB 코드 – 예제 2.21

```
t = 0:0.01:4;
lambda1 = zeros(size(t));
lambda1(t >= 0 & t < 0.5) = 10;
lambda1(t >= 0.5 & t < 1) = 29;
lambda1(t >= 1 & t < 3) = 45;
lambda1(t >= 3 & t < 3.5) = 29;
lambda1(t >= 3.5 & t <= 4) = 10;
lambda2 = 50 * sin(pi * t / 4);

fig = figure;
set(fig, 'Position', [100, 100, 600, 220]);
plot(t, lambda1, 'b', 'LineWidth', 2); hold on;
plot(t, lambda2, 'r--', 'LineWidth', 2);

xlabel('시간 (t)', 'FontSize', 12);
ylabel('고객 도착률 \lambda(t)', 'FontSize', 12);
legend('\lambda_1(t)', '\lambda_2(t)', 'FontSize', 12);
grid on;
hold off;
```

부록

- MATLAB 코드 – 예제 2.22

```
t = 0:0.01:4;

Lambda_1 = zeros(size(t));
for i = 1:length(t)
    if t(i) < 0.5
        Lambda_1(i) = 10 * t(i);
    elseif t(i) < 1
        Lambda_1(i) = 5 + 29 * (t(i) - 0.5);
    elseif t(i) < 3
        Lambda_1(i) = 19.5 + 45 * (t(i) - 1);
    elseif t(i) < 3.5
        Lambda_1(i) = 109.5 + 29 * (t(i) - 3);
    else
        Lambda_1(i) = 124 + 10 * (t(i) - 3.5);
    end
End

Lambda_2 = (200/pi) * (1 - cos(pi * t / 4));
fig = figure;
```

```
%Continue

set(fig, 'Position', [100, 100, 600, 220]);
plot(t, Lambda_1, 'b-', 'LineWidth', 2); hold on;
plot(t, Lambda_2, 'r--', 'LineWidth', 2);

xlabel('시간 t (시간 단위)');
ylabel('평균 발생 횟수 \Lambda(t)');
legend('\Lambda_1(t)', '\Lambda_2(t)', 'Location',
'NorthWest’);

grid on;
hold off;
```