의사결정을 위한 확률모형

- 2장 지수분포와 포아송 과정 -

김 지 혜 (jihye@pel.sejong.ac.kr)
세종대학교 프로토콜공학연구실

목 차

- 개요
- 지수분포
- 포아송 과정
- 복합 포아송 과정
- 비정상 포아송 과정

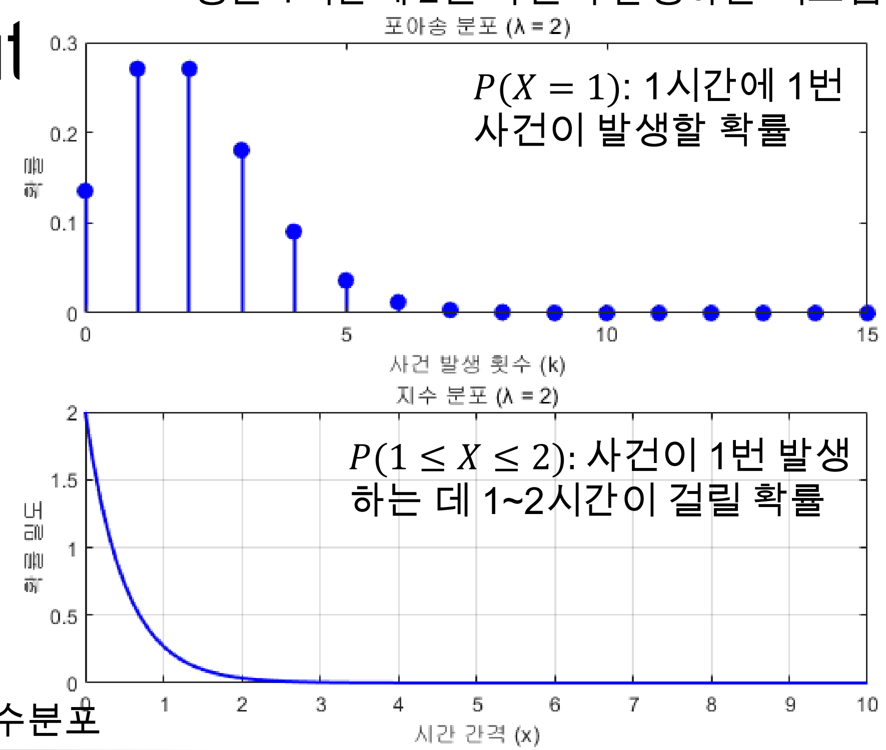
목 차

- 개요
- 지수분포
- 포아송 과정
- 복합 포아송 과정
- 비정상 포아송 과정

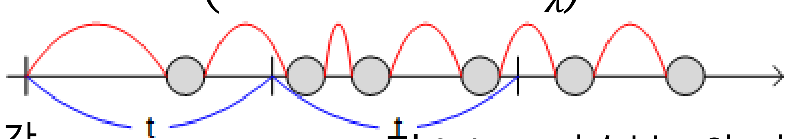
개 요

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

평균 1시간에 2번 사건이 발생하는 시스템



회색 동그라미: 사건
파랑: 단위시간 t
빨강: 사건 간 대기시간



<그림 2.1> 포아송분포와 지수분포

개 요

- 포아송분포와 지수분포
  - 예제 2.1

평균적으로 10초에 2개씩 패킷이 도착하는 네트워크 서버가 있다.
(a) 10초 동안 3개의 패킷이 도착할 확률
(b) 첫 패킷이 5초 이내에 도착할 확률

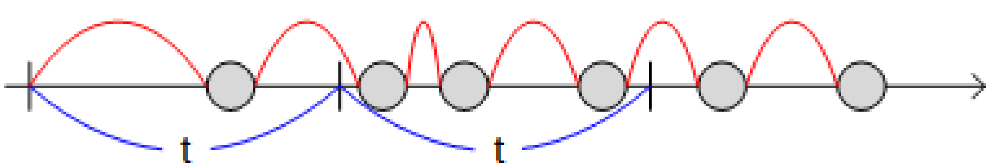
- (a): 포아송분포, (b): 지수분포로 풀이, 이때 $\lambda = 2/10 = 0.2$로 동일
- 단위시간 t(=10초) 동안 평균 2개의 패킷이 도착하는 서버

(a) 그 중 10초 동안 3개의 패킷이 도착할 확률은 두 번째 구간 t 그림과 같은 상황의 확률

$\lambda = 0.2, k = 3$이므로, $P(\mathrm{X} = 3) = f(3) = \dfrac{e^{-0.2}0.2^3}{3!}$

(b) 첫 패킷이 5초 이내에 도착할 확률은 첫 번째 빨간 구간이 5초 이내일 확률

$\lambda = 0.2, x = 5$이므로, $P(X \leq 5) = F(5) = 1 - e^{-0.2 \cdot 5}$

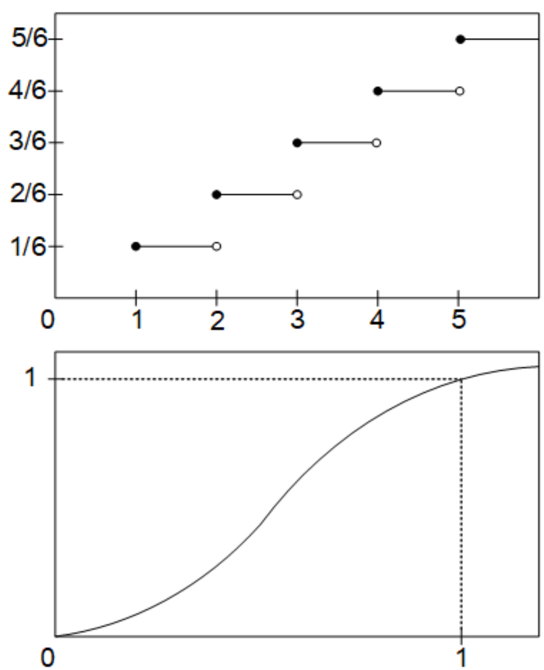


목 차

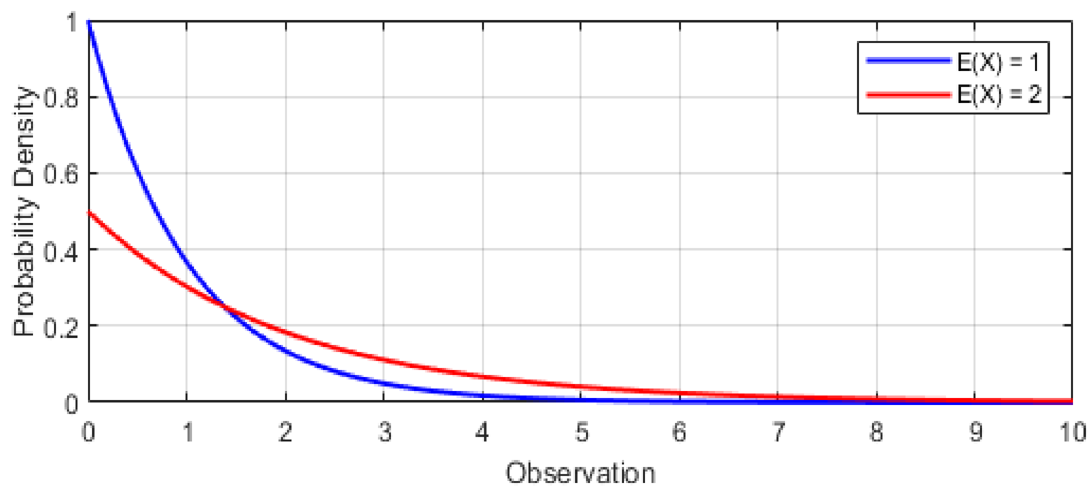
- 개요
- 지수분포
- 포아송 과정
- 복합 포아송 과정
- 비정상 포아송 과정

지수분포

- 정의
  - 평균 사건 발생률이 λ인 사건이 처음 발생할 때까지의 대기 시간에 대한 연속확률분포
    - 연속확률분포: 확률변수가 가질 수 있는 실수 값이 연속적인 분포
    - 모수 λ는 단위시간 당 사건 평균 발생횟수(발생률) 의미



<그림 2.2> 이산확률분포와 연속확률분포



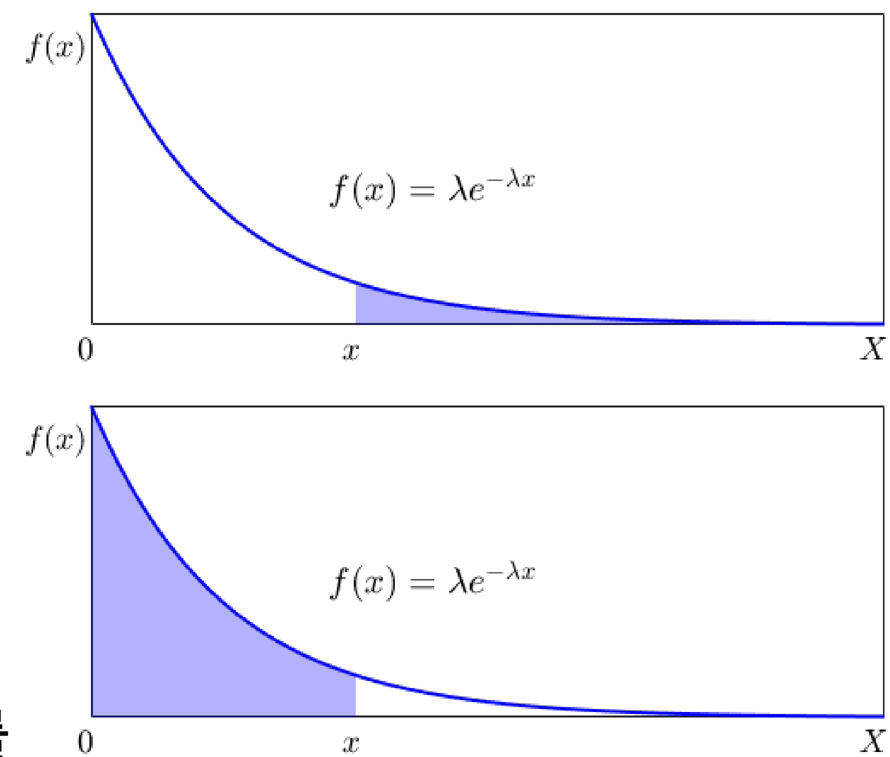
<그림 2.3> 지수분포의 PDF

$x = 2$(2시간)에서 가장 높은 밀도를 가지는 것이 아닌, $x = 0$에서 가장 높은 밀도를 가짐

λ: 단위시간 당 사건 평균 발생횟수, $E(X)$: 평균 사건 발생시간
- 파랑: 특정 사건이 1시간에 1번 발생하는 경우에 대한 분포
- 빨강: 특정 사건이 2시간에 1번 발생하는 경우에 대한 분포

지수분포

- 표현법 (1/2)
  - PDF(Probability Density Function): $f(x) = \lambda \mathrm{e}^{-\lambda x},\ x \geq 0$
  - CDF(Cumulative Distribution Function): $F(x) = 1 - \mathrm{e}^{-\lambda x}$, $x \geq 0$
    - $P(X \geq x) = \int_x^\infty f(x)dx$
      $= \int_x^\infty \lambda e^{-\lambda x} dx = e^{-\lambda x},\ x > 0$
      - 사건이 x시간 동안 한 번도 발생하지 않을 확률
    - $P(0 \leq X \leq x) = \int_0^x f(x)dx$
      $= 1 - \int_x^\infty \lambda e^{-\lambda x} dx$
      $= 1 - e^{-\lambda x},\ x > 0$
      - 사건이 x시간 이내에 발생할 확률



<그림 2.4> 지수분포의 CDF 증명을 위한 PDF

지수분포

- 표현법 (2/2)
  - 평균: $E(x) = 1/\lambda$
    - 특정 사건이 발생하는 평균 대기 시간
  - 분산: $Var(x) = E(X^2) - E(X)^2 = 1/\lambda^2$
    - 특정 사건에 대한 발생 시간의 변동성
      - 분산이 클수록 사건 발생 시간의 변동이 큼
      - 분산이 작을수록 대기 시간은 평균에 가까워짐
      - e.g., $\lambda = 2$이면, 평균 대기시간은 $E(X) = 1/2 = 0.5$이고, 그에 대한 분산은 $Var(X) = 1/4 = 0.25$임. 이는 사건 발생 시간이 평균 대기 시간으로부터 $\sqrt{0.25} = \pm 0.5$(표준편차) 범위의 변동성이 존재함을 의미

지수분포

- 특징
  - 변동계수(CV, Coefficient of Variation)는 1임
    - 표준편차를 평균으로 나눈 값
  - 기억상실성질(Memoryless Property)을 가짐
    - 과거 조건이 현재 확률에 영향을 미치지 않는 특성
  - 특정 사건이 발생하기까지의 대기 시간이나 특정 사건이 발생하는 시간 간격을 모델링하는 데 사용됨
    - e.g., 전자기기의 고장 간격, 다음 고장 발생까지의 대기 시간, 패킷 도착 간격, 콜센터 대기 시간 등

지수분포

- 변동계수(CV, Coefficient of Variation)
  - 표준편차를 평균으로 나눈 값: $CV(X) = \frac{\sqrt{Var(X)}}{E(X)} = \sqrt{\frac{1}{\lambda^2}} / \left(\frac{1}{\lambda}\right) = 1$
    - 평균 대기시간의 크기와 무관하게 상대적으로 일정한 변동성을 가짐을 의미
  - 집단이 가진 데이터의 흩어진 정도를 비교하는 데 사용
  - 변동계수가 작은 집단의 데이터가 더 안정적임

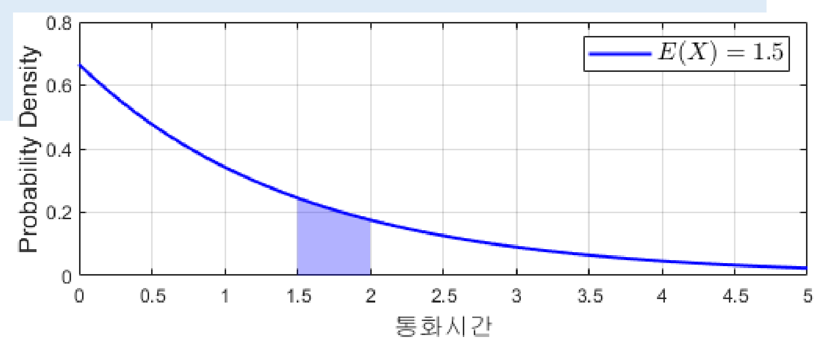
- 예제 2.2

콜센터 문의전화 하나의 통화 시간 X가 평균이 1분 30초인 지수 확률분포를 따른다.
(i) 한 통화 시간이 1분 30초 이상 2분 이내일 확률
(ii) 한 통화 시간이 3분을 넘을 확률

- $\frac{1}{\lambda} = 1.5$분이므로, $\lambda = \frac{2}{3}$

(i) $P(1.5 \leq X \leq 2) = \int_{1.5}^{2} \frac{2}{3} e^{-\left(\frac{2}{3}\right)x}\, dx = 0.1043$

(ii) $P(X \geq 3) = 1 - F(3) = e^{-\left(\frac{2}{3}\right)\times 3} = 0.1353$



<그림 2.5> 예제 2.1(i) 그림

지수분포

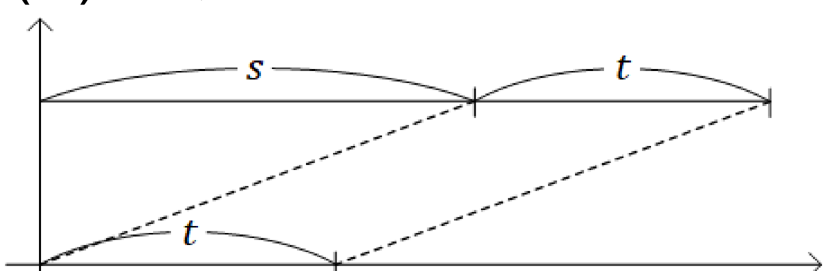
- 기억상실성질(Memoryless Property) (1/3)
  - 과거 조건이 현재 확률에 영향을 미치지 않는 특성

(i) 시간 s동안 통화가 진행 중일 때$(X > s)$,
추가로 시간 t가 지나도 통화가 지속될 확률$(X > t + s)$은
처음 통화를 시작하여 시간 t 이상 지속될 확률$(X > t)$과 같음

(ii) 시점 s에서 통화가 다시 처음 시작된 것과 같은 효과임

(iii) 이는 모든 s, t에 대해 성립하므로, 통화는 매 순간 다시 개시되는 것과 확률적으로 같음

(iv) 즉, 매 순간마다 과거 기억을 상실하고 다시 출발하는 것과 같음



<그림 2.6> 지수 확률변수의 기억상실성질(1)

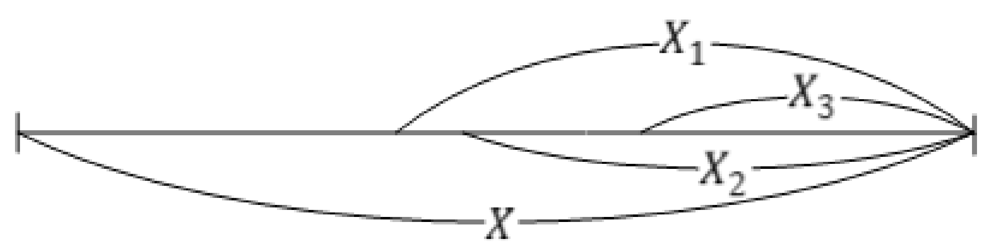
$$F(t) = P(0 \le X \le t) = 1 - e^{-\lambda t}$$
$$1 - F(t) = P(X > t) = e^{-\lambda t}$$

$P(X > t + s | X > s) = P(X > t)$에 대한 증명

$$P(X > t + s | X > s) = \frac{P(X > t + s)}{P(X > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}}$$
$$= \frac{e^{-\lambda t} \cdot e^{-\lambda s}}{e^{-\lambda s}} = \boldsymbol{e^{-\lambda t}}$$
$$P(X > t) = \boldsymbol{e^{-\lambda t}}$$
$$\therefore P(X > t + s \mid X > s) = P(X > t)$$

지수분포

- 기억상실성질(Memoryless Property) (2/3)
  - (v) 과거 경험 또는 기억이 없으므로, 통화시간 길이의 예측이 어려움
    - 완전한 랜덤성을 가진 어떤 사건의 발생간격을 모형화할 수 있음
  - (vi) 역으로, 어떤 연속확률변수가 기억상실성질을 가지는 경우, 해당 확률변수는 지수 확률변수임
    - 지수분포는 연속확률분포 중 유일하게 기억상실성질을 가짐
- $X \sim Exp(\lambda)$인 경우, $X_1, X_2, \ldots, X_i$는 모두 X와 동일한 모수의 지수 확률변수임
  - 지수 확률변수 X는 $X \sim Exp(\lambda)$로 표기

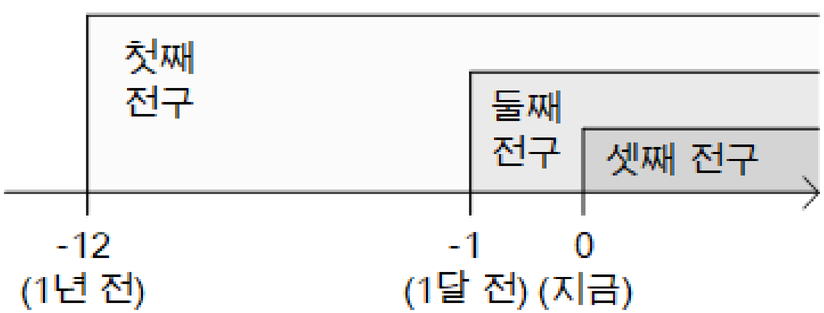


<그림 2.7> 지수 확률변수의 기억상실성질(2)

- 기억상실성질(Memoryless Property) (3/3)
  - 예제 2.3

전구의 수명 X가 $X \sim Exp(\lambda)$이다. 첫째 전구는 1년 전부터 켜두었고, 둘째 전구는 1달 전부터 켜두고 있다. 이제 셋째 전구를 막 켰을 때, 셋째 전구가 가장 먼저 끊어질 확률은 얼마인가?

- 기억상실성질에 의해, 셋째 전구를 막 켠 시점부터 세 전구의 수명은 모두 동일한 지수분포를 따름
- 이러한 대칭성에 의해 가장 먼저 끊어질 확률은 세 전구 모두 동일하게 $1/3$임



<그림 2.8> 예제 2.2 그림

지수분포

- 여러 지수 확률변수들의 확률적 성질 (1/8)
  - $X_i \sim Exp(\lambda_i), i = 1,2, \ldots, n$이고, 서로 독립이라고 가정

1. 서로 독립인 지수 확률변수들의 최소값
   - 여러 독립 사건 중 가장 빨리 발생하는 사건의 시간 모델링에 활용
     - e.g., 여러 장비의 고장 시간 중 가장 빨리 고장나는 장비의 고장 시간 예측
   - 공식: $\min(X_1, X_2) \sim Exp(\lambda_1 + \lambda_2)$
2. 서로 독립인 지수 확률변수들의 비교
   - 여러 독립 사건 발생 시간을 비교할 때 활용
     - e.g., 하나의 장비가 다른 장비보다 먼저 고장날 확률 예측
   - 공식: $P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$
3. 서로 독립인 지수 확률변수들 중 최대-최소 차이의 기댓값
   - 여러 독립 사건의 발생 시간 차이를 비교하여 시스템 유지보수에 활용
     - e.g., 고객이 여러 서비스를 받는 대기시간을 예측하여 시스템 성능 최적화에 활용
   - 공식: $E(\max(X_1, X_2) - \min(X_1, X_2)) = \frac{1}{\lambda_2} \cdot \frac{\lambda_1}{\lambda_1+\lambda_2} + \frac{1}{\lambda_1} \cdot \frac{\lambda_2}{\lambda_1+\lambda_2}$

지수분포

- 여러 지수 확률변수들의 확률적 성질 (2/8)
  - 예제 2.4

만약 3개의 장비가 있고, 각 장비의 고장률이 $\lambda_1, \lambda_2, \lambda_3$라고 할 때, 가장 빨리 고장나는 장비의 고장 시간은 $Exp(\lambda_1 + \lambda_2 + \lambda_3)$의 지수분포를 따른다.

- 서로 독립인 지수 확률변수들의 최소값 공식 활용
- 각 장비의 고장률이 각각 $\lambda_1 = 0.1, \lambda_2 = 0.2, \lambda_3 = 0.3$인 경우, 가장 빨리 고장나는 장비의 고장 시간은 $\min(X_1, X_2, X_3) \sim Exp(0.1 + 0.2 + 0.3) = Exp(0.6)$의 지수분포를 따름
- 이는 평균적으로 $\frac{1}{0.6} \approx 1.67$ 단위시간이 걸림을 의미함

지수분포

- 여러 지수 확률변수들의 확률적 성질 (3/8)
  - 예제 2.5

콜센터에 가입(=1) 및 탈퇴(=2) 요청 등 2가지 종류의 문의전화가 온다.
전화 도착간격이 각각 $X_i \sim Exp(\lambda_i)(i=1,2)$이고 서로 독립일 때, 문의전화의 도착 간격은 $\min(X_1, X_2) \sim Exp(\lambda_1+\lambda_2)$이며, 도착률은 시간당 $\lambda_1+\lambda_2$이다.

- 서로 독립인 지수 확률변수들의 최소값 공식 활용
- 가입 문의전화 도착 간격: $X_1 \sim Exp(\lambda_1)$, 탈퇴 문의전화 도착 간격: $X_2 \sim Exp(\lambda_2)$
- λ_1, λ_2는 각 사건이 발생할 확률 즉, 문의전화 도착률을 의미함
- $\min(X_1, X_2)$: 두 종류의 문의전화 중 더 빨리 도착하는 전화의 도착 시간
- 따라서, 문의전화의 전체 도착 간격은 $\min(X_1, X_2)$가 되고, 이는 지수분포 성질에 의해 $Exp(\lambda_1+\lambda_2)$를 따름
- 즉, 빨리 도착하는 문의전화의 분포는 단위시간 당 $\lambda_1+\lambda_2$인 지수분포를 따른다는 의미

지수분포

- 여러 지수 확률변수들의 확률적 성질 (4/8)
  - 예제 2.6

세 개의 지수 확률변수 X_1, X_2, X_3가 각각 $\lambda_1 = 0.1, \lambda_2 = 0.2, \lambda_3 = 0.3$일 때,
각 확률변수가 가장 빨리 발생할 확률은 무엇인가?

- 서로 독립인 지수 확률변수들의 비교 공식 활용
- $P(X_1 = \min(X_1, X_2, X_3)) = \frac{0.1}{0.1+0.2+0.3} = \frac{0.1}{0.6} \approx 0.167$
- $P(X_2 = \min(X_1, X_2, X_3)) = \frac{0.2}{0.1+0.2+0.3} = \frac{0.2}{0.6} \approx 0.333$
- $P(X_3 = \min(X_1, X_2, X_3)) = \frac{0.3}{0.1+0.2+0.3} = \frac{0.3}{0.6} \approx 0.500$
- 따라서, X_3이 가장 빨리 발생할 것임

지수분포

- 여러 지수 확률변수들의 확률적 성질 (5/8)
  - 예제 2.7

(예제 2.5) 콜센터에 가입(=1) 및 탈퇴(=2) 요청 등 2가지 종류의 문의전화가 온다. 전화 도착간격이 각각 $\mathrm{X_i \sim Exp(\lambda_i)(i = 1,2)}$이고 서로 독립일 때, 문의전화의 도착 간격은 $\mathrm{min(X_1, X_2) \sim Exp(\lambda_1 + \lambda_2)}$이며, 도착률은 시간당 $\lambda_1 + \lambda_2$이다.

예제 2.5에서 첫 번째 문의전화가 가입 요청 문의일 확률은 $\frac{\lambda_1}{\lambda_1+\lambda_2}$이다.

- 서로 독립인 지수 확률변수들의 비교 공식 활용
- 첫 번째 문의전화가 가입 요청 문의이기 위해서, $P(X_1 < X_2)$이어야 함
- 지수분포의 기억상실성질의 특성에 기반하여, 독립적인 확률변수 X_1, X_2에 대한 $P(X_1 < X_2)$는 $\frac{\lambda_1}{\lambda_1+\lambda_2}$ 이므로, 해당 확률 값은 $\frac{\lambda_1}{\lambda_1+\lambda_2}$임

지수분포

- 여러 지수 확률변수들의 확률적 성질 (6/8)
  - 예제 2.8

서버가 두 가지 다른 서비스 요청을 처리할 때, 각 요청의 응답 시간은 지수 분포를 따르며 서로 독립이다.
- 요청 1의 응답 시간 $X_1 \sim Exp(\lambda_1)$의 응답 도착률은 $\lambda_1 = 0.2$
- 요청 2의 응답 시간 $X_2 \sim Exp(\lambda_2)$의 응답 도착률은 $\lambda_2 = 0.3$

- 서로 독립인 지수 확률변수들 중 최대-최소 차이의 기댓값 공식 활용
- $E(Y) = E(Y|X_1 < X_2)P(X_1 < X_2) + E(Y|X_2 \leq X_1)P(X_2 \leq X_1)$

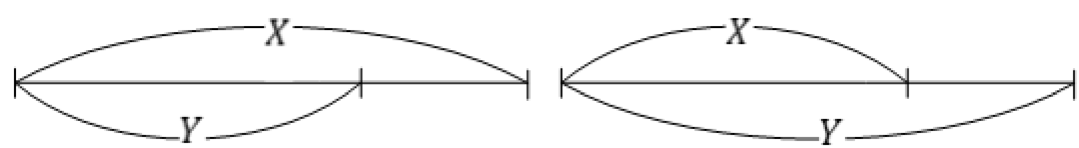
$$= \frac{1}{\lambda_2} \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_1} \cdot \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{1}{0.3} \cdot \frac{0.2}{0.2 + 0.3} + \frac{1}{0.2} \cdot \frac{0.3}{0.2 + 0.3} = 4.33$$

- 따라서, 두 서비스 요청의 응답시간의 차이는 평균적으로 약 4.33초이며, 이 값보다 더 큰 차이를 가지는 경우에는 시스템 성능 개선이 필요한 것으로 간주할 수 있음

- 여러 지수 확률변수들의 확률적 성질 (7/8)
  - 예제 2.9 (1/2)

콜센터에서 하나의 문의전화에 대한 통화시간은 $Y \sim Exp(\mu)$인데, 각 문의전화의 도착간격은 $X \sim Exp(\lambda)$이며, 이 둘은 서로 독립이다.
(a) 직원이 하나의 문의전화를 응대하는 동안 다른 문의전화가 하나도 걸려오지 않았을 확률은 $P(X > Y) = \mu/(\lambda + \mu)$이다.

- $P(X > Y)$: 하나의 문의전화 이후에 다음 전화가 도착할 확률
- $P(X < Y)$: 하나의 문의전화 도중에 다음 전화가 도착할 확률
- (a)는 하나의 문의전화 응대 중 다른 문의전화가 걸려오지 않을 확률이므로 $P(X > Y)$에 해당하고, 이는 서로 독립인 지수 확률변수들의 비교 공식에 의해 $P(X > Y) = \mu/(\lambda + \mu)$가 성립



<그림 2.9> 예제 2.8(a) 그림

지수분포

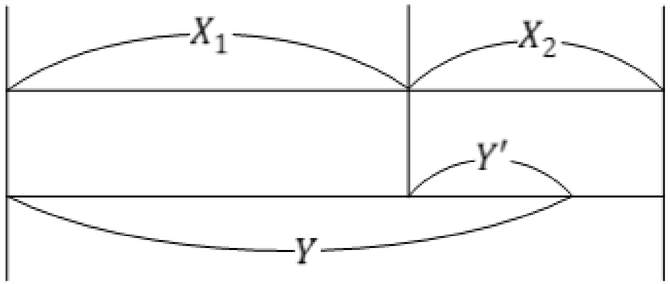
- 여러 지수 확률변수들의 확률적 성질 (8/8)
  - 예제 2.9 (2/2)

(b) 하나의 문의전화를 응대하는 동안 다른 문의전화가 한 개만 걸려왔을 확률을 구하는 것은 다음 그림과 같이 생각하자. 첫 번째 길이다툼에서는 Y가 X_1보다 길었다(문의전화 1회 도착). 두 번째 길이다툼에서는 Y'이 X_2보다 짧아야 한다. 이 두사상은 서로 독립이므로, 구하는 확률은 $\frac{\lambda}{\lambda+\mu} \times \frac{\mu}{\lambda+\mu}$이다.

이와 같은 방법을 적용하면 하나의 문의전화를 응대하는 동안 다른 문의전화가 $k(k = 1,2,\dots)$개 걸려왔을 확률은 $\left(\frac{\lambda}{\lambda+\mu}\right)^k \times \frac{\mu}{\lambda+\mu}$이다.

- 첫 번째 길이: $P(X_1 < Y)$, 나머지: $P(X_2 > Y')$, 이들은 독립임
- 서로 독립인 지수 확률변수들의 비교 공식에 의해 구하고자 하는

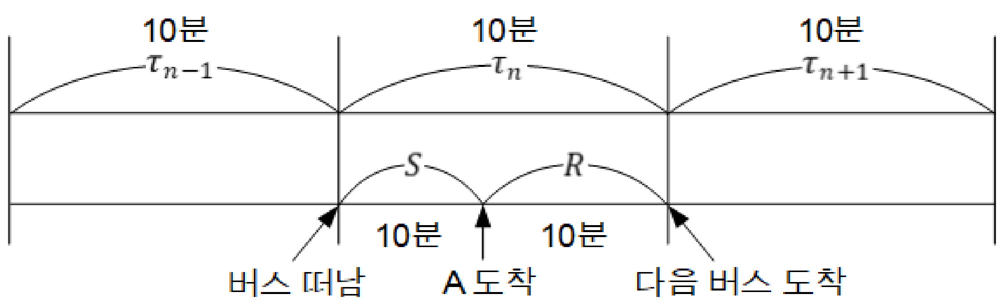
확률: $P(X_1 < Y)P(X_2 > Y') = \frac{\lambda}{\lambda+\mu} \times \frac{\mu}{\lambda+\mu}$



<그림 2.10> 예제 2.8(b) 그림

지수분포

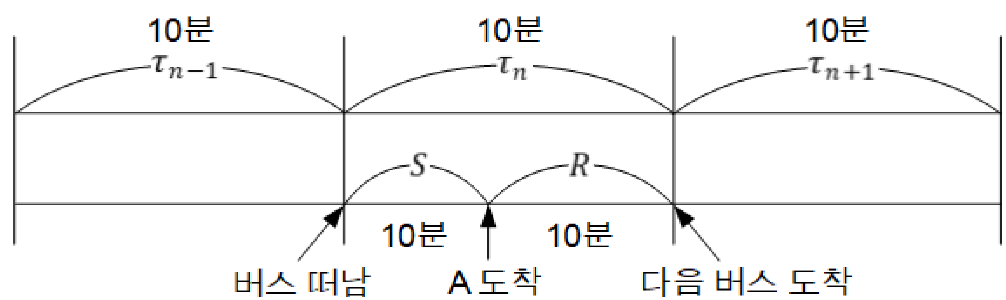
- 검사시점의 역설(Inspection Paradox) (1/4)
  - 임의로 선택된 한 시점에서 발생한 사건의 평균 길이가 실제 사건의 평균 길이보다 더 길게 나타나는 현상
  - 예시 (1/2)
    - 버스 정류장에 도착하는 버스들의 시간간격 $\tau_k, k = 1,2, \ldots$가 평균 10분인 지수분포를 따른다고 가정
    - A가 버스 정류장에 도착한 후, 버스를 탈 때까지의 시간(=잔여시간 R)도 지수 확률변수의 기억상실성질에 의해 평균이 10분인 지수분포이므로 A의 대기시간은 평균 10분



<그림 2.11> 예시 그림

지수분포

- 검사시점의 역설(Inspection Paradox) (2/4)
  - 임의로 선택된 한 시점에서 발생한 사건의 평균 길이가 실제 사건의 평균 길이보다 더 길게 나타나는 현상
  - 예시 (2/2)
    - 그러나, A가 정류장에 도착하기 전, 이전 버스가 정류장에서 떠난 이후 A의 도착 시점까지의 시간(=경과시간 S) 또한 평균 10분인 지수분포를 따름
    - 따라서, (버스 도착간격 τ) = (경과시간 S) + (잔여시간 R)
    - 이는 (평균 10분) = (평균 10분) + (평균 10분)을 의미



<그림 2.11> 예시 그림

지수분포

- 검사시점의 역설(Inspection Paradox) (3/4)
  - 특정 시점에서 특정 구간에 속할 확률은 구간의 길이인 x의 크기에 비례
    - x가 2배인 경우 그 구간이 선택될 가능성도 2배를 의미

예시) 버스 정류장에 버스가 도착하는 간격(x)이 길면, 승객이 정류장에 도착하는 시점이 그 긴 구간 내에 포함될 가능성이 더 커짐

  - 구간의 길이가 x인 구간의 출현횟수는 구간의 길이가 x일 확률($\lambda e^{-\lambda x} dx$)에 비례
    - 긴 구간일수록 사건이 발생할 확률이 낮음(출현횟수가 적음)
    - 지수 확률변수 X의 PDF는 $\lambda e^{-\lambda x}$이고, X가 클수록 낮은 확률을 가짐

예시) 평균적으로 버스 대기 시간이 1~2분으로 짧아서 버스가 자주 도착한다고 할 때, 그 대기 시간은 자주 출현할 것임
그러나, 버스 대기 시간이 10~20분으로 길어진다고 할 때, 긴 구간은 자주 출현할 수 없음

- 검사시점의 역설(Inspection Paradox) (4/4)
  - 도착시점을 포함하는 시간간격 τ의 확률분포와 평균 $E(\tau)$
    - 도착 간격은 $X \sim Exp(\lambda)$, 도착 시점 X_1, X_2는 독립으로 가정
    - $f_\tau(x) = \lambda^2 x e^{-\lambda x}, \quad F_\tau(x) = 1 - e^{-\lambda x} - \lambda x e^{-\lambda x}, x \geq 0$

시간 간격 τ는 $\tau = X_2 - X_1$로 정의됨

$P(X_1 \leq t) = 1 - e^{-\lambda t}$, $P(X_2 \leq X_1 + x) = 1 - e^{-\lambda(X_1 + x)}$

$F_\tau(x) = P(\tau \leq x) = P(X_2 - X_1 \leq x) = P(X_2 \leq X_1 + x) = \int_0^\infty P(X_2 \leq X_1 + x | X_1 = t) f_{X_1}(t) dt$

$= \int_0^\infty \left[1 - e^{\lambda(t+x)}\right] \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} - \lambda x e^{-\lambda x}$

$f_\tau(x) = F_\tau'(x) = \lambda^2 x e^{-\lambda x}$

- $E(\tau) = 2/\lambda$

$E(\tau) = \int_0^\infty x f_\tau(x) dx = \int_0^\infty x(\lambda^2 x e^{-\lambda x}) dx = \lambda^2 \cdot \frac{2}{\lambda^3} = \frac{2}{\lambda}$

- 위 공식을 활용하여 검사시점의 역설을 반영하도록 예측 모델을 조정함으로써, 기존 시스템보다 정확한 예측 및 분석을 수행하고, 효과적으로 개선된 시스템을 제공할 수 있음

목 차

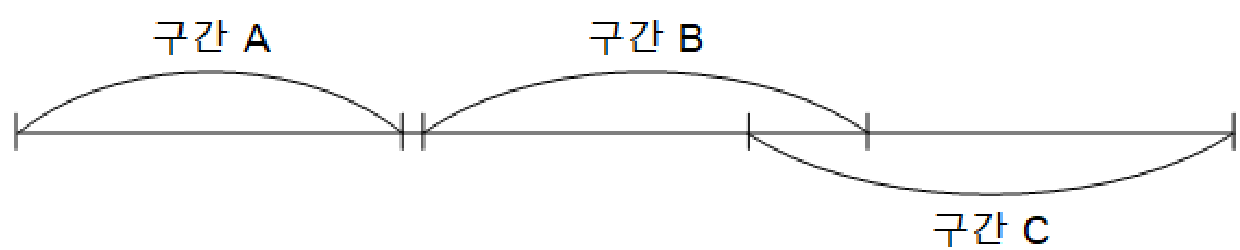
- 지수분포
- 포아송 과정
- 복합 포아송 과정
- 비정상 포아송 과정

포아송 과정

- 정의
  - 시간 t 동안의 사건 발생횟수 $N(t)$를 나타내는 확률과정
    - 확률과정: 시간에 따라 발생하는 사건들이 확률적으로 변화하는 구조
    - $N(t)$: 시점 $t(\geq 0)$까지 발생한 어떤 사건의 누적 횟수이자 확률변수 (단, $N(0) = 0$)
- 특징
  - 시간 t까지의 사건 발생횟수를 모델링하여 확률적 예측 및 분석을 수행할 수 있음
    - e.g., 시점 t까지 콜센터에 걸려온 문의전화 수, 교환기에 도착한 트래픽 수, 온라인 쇼핑사이트에 접속한 고객 수 등 예측
  - 포아송 과정은 3가지 가정을 만족해야 하며, 이를 통해 포아송 분포를 유도할 수 있음
    - 독립증분, 정상증분, 단일증분

포아송 과정

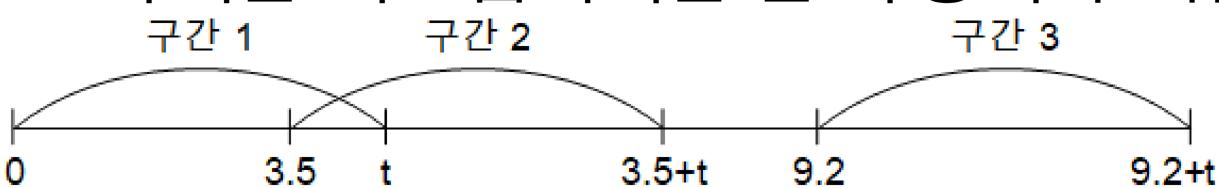
- 가정 1 – 독립증분(Independent Increment)
  - 겹치지 않는 두 구간에서 각각 발생한 사건의 개수들은 서로 독립임
    - e.g., 문의전화 시간에 따른 전화 빈도 수
      - 15:00 ~ 15:30 동안 콜센터에 걸려온 문의전화 개수는 15:30 ~ 15:45 동안의 문의전화 개수와 무관함
  - 포아송 과정에서는 독립증분에 의해, 서로 다른 구간에서 사건이 발생하는 것은 서로 독립적으로 고려됨
    - 구간 A와 구간 B에서의 사건 수는 서로 독립이나, 구간 B와 구간 C에서의 사건 수는 독립이 아님



<그림 2.12> 독립증분의 가정

포아송 과정

- 가정 2 – 정상증분(Stationary Increment)
  - 구간내에서 발생한 사건수는 그 구간의 길이에만 영향을 받고 구간의 시간적 위치와는 무관함
    - e.g., 시간에 따른 불량품 생산 개수
      - 24시간 부품을 생산하는 기계에서 15:00 ~ 15:30의 30분 동안 생산된 불량품 개수에 대한 확률분포는 07:00 ~ 07:30의 30분 동안과 동일
  - 포아송 과정에서는 정상증분에 의해, 사건 발생 빈도가 일정한 사건 발생 확률은 지수분포의 평균 시간에 대한 구간 길이에 비례함
    - 구간 1에서 $k(k = 0,1,2, \dots)$개의 사건이 발생할 확률은 구간 2, 3에서도 모두 동일함
    - 그러나, 콜센터 문의전화 같이 시간대에 따라 다양한 빈도수를 가지는 시스템에서는 본 가정이 부적합할 수 있음



<그림 2.13> 정상증분의 가정

포아송 과정

- 가정 3 – 단일증분(Single Increment)
  - 충분히 짧은 시간이나 작은 공간에서는 2개 이상의 결과가 동시에 발생할 확률이 0에 가까움
    - 충분히 작은 값 Δt에 대해 다음이 성립함

      (a) $P(N(t+\Delta t)-N(t)=0) \approx 1-\lambda\Delta t,$

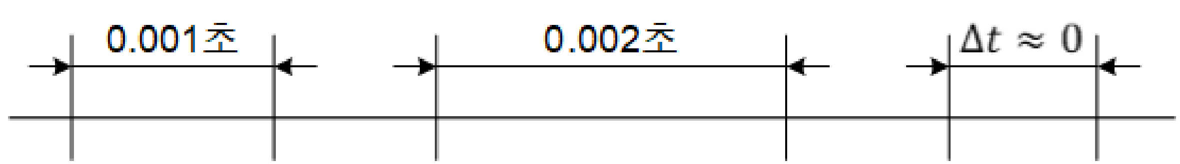
      (b) $P(N(t+\Delta t)-N(t)=1) \approx \lambda\Delta t,$

      (c) $P(N(t+\Delta t)-N(t) \geq 2) \approx 0$

      - Δt가 0에 가까울수록 값이 정확함

Δt: 시간 변화율
λ: 단위시간 당 사건 발생률
$N(t)$: 사건 발생횟수

  - 포아송 과정에서는 단일증분에 의해, 사건은 하나씩만 발생하고, 구간 길이가 충분히 짧은 경우에는 해당 구간에서 하나의 사건이 발생할 확률이 그 구간 길이에 비례함
    - 구간 길이가 짧을수록 사건이 한 번 발생할 확률도 적음을 의미



<그림 2.14> 단일증분의 가정

포아송 과정

- 3가지 가정에 의한 포아송 과정

정의 2.1

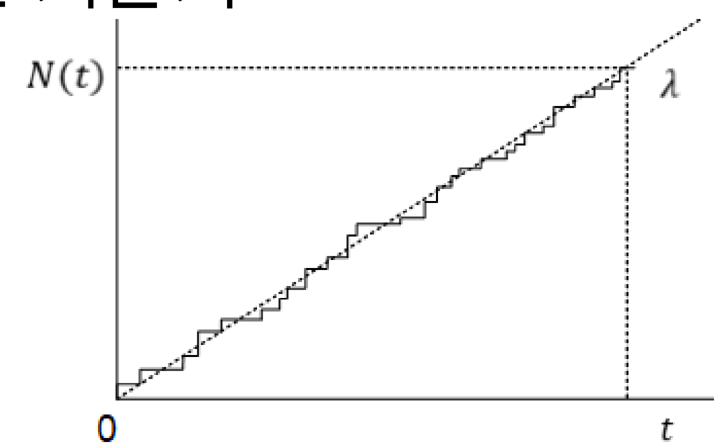
$N(t) \sim PP(\lambda)$이면 $N(t)$는 모수 λt의 포아송 확률변수이다. 즉, $N(t)$는 다음과 같은 확률 분포를 가진다.

$$P(N(t) = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, n = 0,1,2, \dots$$

λ: 단위시간 t 당 사건 발생률
$N(t)$: 사건 발생횟수

그러므로 $E\left(N(t)\right) = \lambda t, Var\left(N(t)\right) = \lambda t$이다.

- 모수 λ의 의미
  - t가 충분히 큰 경우, $N(t)$의 대략적인 기울기
    - $N(t) \approx \lambda t$, 즉 $\lambda \approx N(t)/t$이므로 총 사건 발생횟수를 시간으로 나눈 값
  - 단위시간당 사건 평균 발생횟수, 또는 사건의 의미에 따라 발생률, 도착률, 전이율 등으로 해석 가능



<그림 2.15> 포아송 과정에서 모수 λ 의미

포아송 과정

- 예제 2.10

$N(t) \sim PP(\lambda)$일 때,

$$P(N(t+s)=m+k|N(s)=m) = \boldsymbol{P(N(t+s)-N(s)=k)} = P(N(t)=k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

두 번째 등식은 $N(0)=0$과 정상증분에 의해 성립한다.

- m은 시간 s에 대한 사건 발생횟수, k는 시간 t에 대한 사건 발생횟수
- "$N(0)=0$과 정상증분에 의해 성립한다."는 포아송 과정의 초기 조건인 $N(0)=0$을 따르고, 정상증분에 의해 시간적 위치와 무관하게 구간 길이에만 영향을 받음
- 따라서, $P(N(t+s)-N(s)=k) = P(N(t+0)-N(0)=k)$

포아송 과정

- 예제 2.11

$$
\begin{aligned}
P(N(t+s)=m+k, N(s)=m) &= \boldsymbol{P(N(t+s)-N(s)=k, N(s)=m)} \\
&= \boldsymbol{P(N(t+s)-N(s)=k)P(N(s)=m)} \\
&= P(N(t)=k)P(N(s)=m) \\
&= \frac{e^{-\lambda t}(\lambda t)^k}{k!}\cdot\frac{e^{-\lambda s}(\lambda s)^m}{m!}.
\end{aligned}
$$

첫 번째와 두 번째 등식은 독립증분의 성질을 이용한 것이며, 세 번째 등식은 정상증분에 의한 것이다.

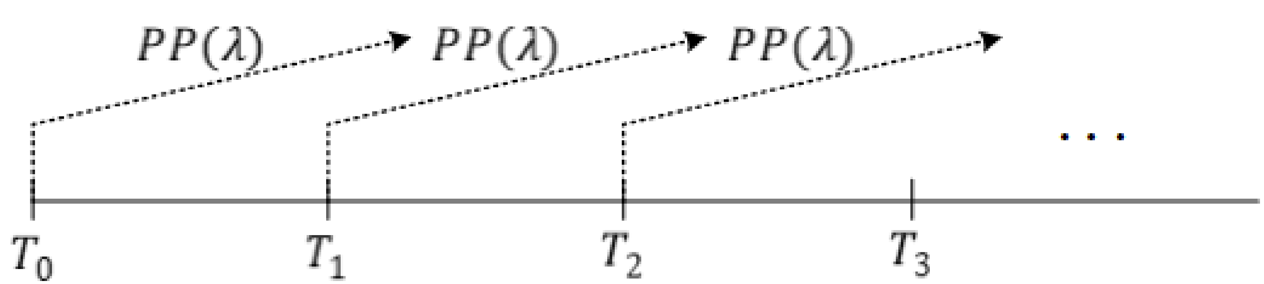
- m은 시간 s에 대한 사건 발생횟수, k는 시간 t에 대한 사건 발생횟수
- $P(N(t+s)=m+k, N(s)=m)$는 시간 s까지 m개의 사건이 발생하고, 이후 t 동안 k개의 사건이 더 발생하여 총 $m+k$개의 사건이 $t+s$까지 발생했을 확률
- 포아송 과정의 독립증분에 의해 $N(t+s)-N(s)$와 $N(s)$는 독립임
- 포아송 과정의 정상증분에 의해 시간적 위치와 무관하게 구간 길이에만 영향을 받으므로 $P(N(t+s)-N(s)=k)=P(N(t)=k)=\frac{e^{-\lambda t}(\lambda t)^k}{k!}$임

포아송 과정

- 포아송 과정의 사건발생 간격
  - 포아송 과정에서 $T_1, T_2, \ldots, T_n$을 각각 $1,2,\ldots,n$번째 사건이 발생한 시점들이라고 가정
  - $N(t) \sim PP(\lambda)$에서, $\{N(x) = 0\}$인 사상은 $\{T_1 > x\}$인 사상과 동치(Equivalent)임
    - 동치는 두 명제가 논리적으로 같음을 의미
    - $N(x) = 0$: 시간 x 까지 사건이 발생하지 않음을 의미
    - $T_1 > x$: 첫 번째 사건이 시간 x 이후에 발생했음을 의미
  - 따라서, $P(N(x) = 0) = e^{-\lambda x} = P(T_1 > x)$이고
    첫 번째 사건이 발생할 때까지의 시간 T_1은 $T_1 \sim Exp(\lambda)$,
    즉 모수 λ의 지수분포를 따름
    - 포아송 과정에서 기억상실성질을 가짐

포아송 과정

- 포아송 과정의 사건발생 간격
  - 포아송 과정의 독립증분에 의해, 구간 $[0, T_1]$에서의 확률은 구간 $[T_1, \infty)$과 독립적이므로, 시점 T_1에서 포아송 과정이 처음부터 다시 시작한다고 생각할 수 있음
  - 구간 $[T_1, \infty)$에서 보면, T_2 또한 첫 번째 사건발생 시점이므로 $T_2 - T_1$도 모수 λ의 지수분포를 따르며, T_1과는 독립임
  - 그러므로 $T_3 - T_2$도 $T_1, T_2 - T_1$과 독립이며, 모수 λ의 지수분포를 따름



<그림 2.16> 포아송 과정의 사건발생 간격

포아송 과정

- 포아송 과정의 사건발생 간격

정의 2.2

$N(t) \sim PP(\lambda)$일 때, 사건의 발생 간격 $T_{k+1} - T_k (k = 0,1,2, \dots)$는 iid*이며, 모수 λ의 지수 분포를 따른다.

*iid(independent and identically distributed): 독립적이고 동일한 분포

- 포아송 과정은 사건발생 간격에 대해 독립적이고 동일한 분포를 가짐
  - $T_{k+1} - T_k$에 대한 동일한 지수분포를 따름
- 포아송 과정은 기억상실성질을 가지는 지수분포를 따름
- 포아송 과정은 완전한 랜덤성을 가진 특정 사건들이 발생하는 것을 표현할 수 있음
  - e.g., 콜센터에 걸려오는 문의전화, 방사능 물질에서 방출되는 방사능 입자 수 등

포아송 과정

- 포아송 과정의 사건발생 간격
  - $T_k = (T_k - T_{k-1}) + (T_{k-1} - T_{k-2}) + \cdots + T_1$이므로 T_k는 모수 (k, λ)를 가지는 얼랑(Erlang) 확률변수임
    - 얼랑 확률변수: k개의 독립적인 지수 확률변수 합
    - 포아송 과정에서 사건 발생까지의 총 시간 모델링에 유용
  - T_k의 PDF: $f_{T_k}(t) = \frac{\lambda(\lambda t)^{k-1} e^{-\lambda t}}{(k-1)!}$
  - T_k의 CDF: $F_{T_k}(t) = P(T_k \leq t) = \int_0^t \frac{\lambda(\lambda x)^{k-1} e^{-\lambda x}}{(k-1)!} dx$
  - T_k의 평균: $E(T_k) = k \times \left(\frac{1}{\lambda}\right) = \frac{k}{\lambda}$

λ: 단위시간 당 사건 발생률

포아송 과정

- 예제 2.12

$N(t) \sim PP(\lambda)$일 때, 다음이 성립한다.
(a) $E(T_4) = \frac{4}{\lambda}$.
(b) $E(T_4 - T_3) = \frac{1}{\lambda}$.
(c) $E(T_7 - T_3) = \frac{4}{\lambda}$.
(d) $E(T_7 - T_3 | T_3 = c) = \frac{4}{\lambda}$. ($c$는 상수)

- (a)는 얼랑분포 평균 공식에 따라 계산함
- (b)는 사건발생 간격이 $T_{k+1} - T_k$는 지수분포를 따르므로 지수분포의 평균 공식을 따름
- (c)는 $T_7 - T_3$는 4개의 지수분포를 따르는 간격의 합이므로 $4 \times 1/\lambda$
- (d)는 기억상실성질에 의해 $T_7 - T_3$가 $T_3 = c$와 무관하게 동일한 분포를 따르므로 (c)와 동일하게 계산됨

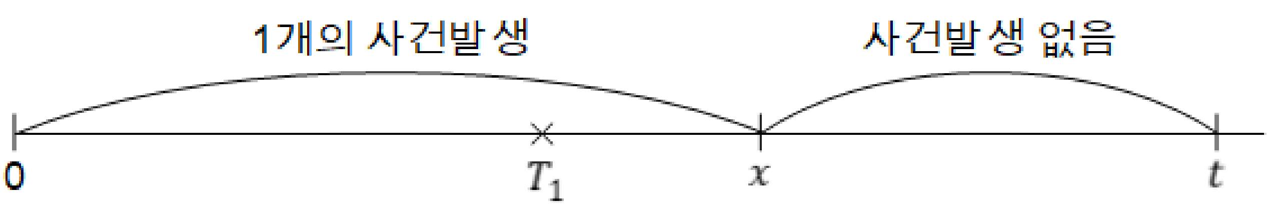
포아송 과정

- 조건부 발생시점
  - 포아송 과정에서 첫 번째 사건의 발생시점 T_1은 지수분포를 따르고, 시간 구간 $[0, t]$ 동안 사건이 한 번 발생했다는 조건 하에 가지는 확률분포
    - 첫 번째 사건 발생시점 T_1의 분포, $T_1|N(t) = 1$

$$P(T_1 \le x|N(t) = 1) = \frac{P(T_1 \le x, N(t) = 1)}{P(N(t) = 1)} = \frac{P([0,x]\text{에서 한 사건 발생}, (x,t]\text{에서 발생 없음})}{P(N(t) = 1)}$$

$$= \frac{P(N(x) = 1)P(N(t) - N(x) = 0)}{P(N(t) = 1)} = \frac{e^{-\lambda x}(\lambda x) \cdot e^{-\lambda(t-x)}}{e^{-\lambda t}(\lambda t)} = \frac{x}{t}, \qquad 0 \le x \le t$$

- $T_1|N(t) = 1$은 $[0, t]$에서 균일분포를 따름
  - e.g., 전화가 10분$(= t)$에 한 번 온다고 할 때, 그 전화가 1분 후에 올 확률이나 5분 후에 올 확률 등은 모두 동일함



<그림 2.17> 조건부 발생시점의 분포

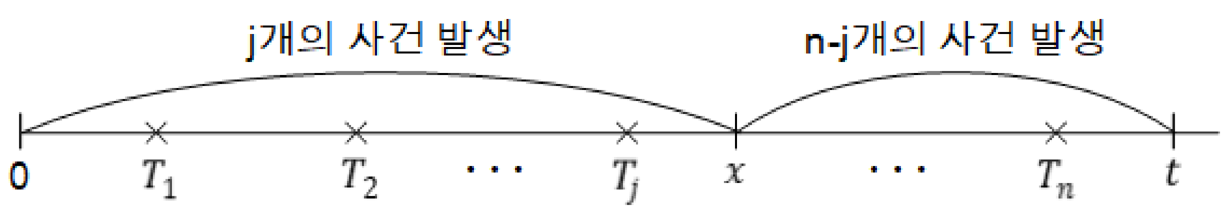
포아송 과정

- 조건부 발생시점
  - 어떤 시간 구간 $[0, t]$ 동안 n개의 사건이 발생했음을 알고 있을 때, k번째 발생한 사건의 발생시점 T_k의 확률분포

$$P(T_k \leq x | N(t) = n) = \frac{P(T_k \leq x, N(t) = n)}{P(N(t) = n)} = \frac{P(N(x) \geq k, N(t) = n)}{P(N(t) = n)}$$

$$= \sum_{j=k}^{n} \frac{P(N(s) = j, (x, t]\text{에서 } n - j\text{개의 사건 발생})}{P(N(t) = n)}$$

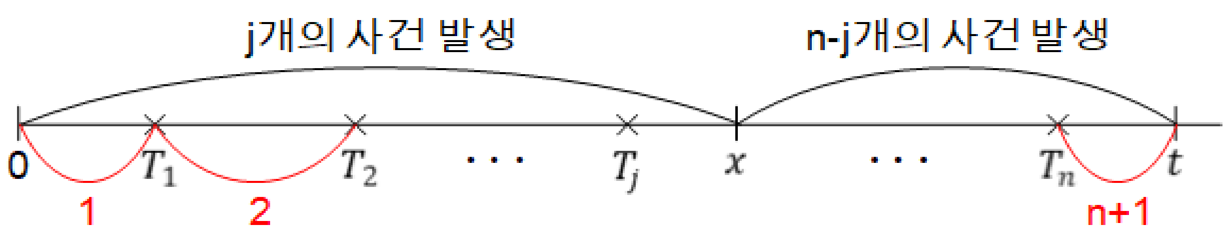
$$= \sum_{j=k}^{n} \frac{e^{-\lambda x}(\lambda x)^j / j! \cdot e^{-\lambda(t-x)}\big(\lambda(t-x)\big)^{n-j} / (n-j)!}{e^{-\lambda t}(\lambda t)^n / n!} = \sum_{j=k}^{n} \binom{n}{j} \left(\frac{x}{t}\right)^j \left(1 - \frac{x}{t}\right)^{n-j}, 0 \leq x \leq t$$



<그림 2.18> n개 사건의 조건부 발생시점의 분포

포아송 과정

- 조건부 발생시점
  - 조건부 확률변수 $T_k|N(t)=n$는 구간 $[0,t]$에서 균일분포를 따르는 확률변수 $T_1,\dots,T_n$ 중 T_k와 동일한 분포를 가짐
  - 이는 n개의 사건들이 발생한 시점인 $T_1,\dots,T_n$이 $T_1<\cdots<T_n$를 유지하며 그 구간 내에서 나타날 가능성이 균일함을 의미
    - $[0,t]$ 동안 n개의 사건이 발생했음을 알고 있을 때, 임의로 선택한 특정 시점에서 사건이 발생할 가능성은 $[0,t]$ 내에서 $n+1$만큼 균일하게 분포되어 있음



<그림 2.19> 조건부 발생시점 분포의 구간 수

포아송 과정

- 예제 2.13

$PP(\lambda)$에서, 구간 $[0, t]$ 동안 다섯 개의 사건이 발생했음을 알고 있을 때, 네 번째 발생한 사건의 발생시점 평균값을 구하라.

- $P(T_4 \leq x | N(t) = 5) = \sum_{j=4}^{5} \left(\frac{x}{t}\right)^j \left(1 - \frac{x}{t}\right)^{5-j} = 5\left(\frac{x}{t}\right)^4 - 4\left(\frac{x}{t}\right)^5$이므로
- $E(T_4 | N(t) = 5) = \int_0^t \left(1 - 5\left(\frac{x}{t}\right)^4 - 4\left(\frac{x}{t}\right)^5\right) dx = \frac{2}{3} t$이다.
- $k(k = 1,2,3,4,5)$번째 사건의 발생 시점의 평균값은 $k \times t/(1+5) = kt/6$이다.
- 이는 구간 $[0, t]$를 6등분한 시점들이다.

- 예제 2.14

10:00~11:00 동안 세 개의 사건이 발생했다. 세 사건 모두 10:45~11:00 사이에 발생했을 확률은 $0.25^3 = 0.015625$이다.

- 총 구간 길이는 60분, 10:45~11:00의 길이는 15분이므로, 확률은 $15/60 = 0.25$
- 세 개의 사건은 모두 독립적으로 발생하므로 확률은 $0.25^3 = 0.015625$

포아송 과정

- 예제 2.15

어떤 버스 정거장에서 버스가 30분마다 한 대씩 떠난다. 승객들은 포아송 과정으로 도착하는데, 30분 사이에 8명의 승객들이 도착했다면 이들의 누적 대기시간의 평균은?

- 8명 고객들의 도착시점의 평균값은 30분을 9등분한 사람들이다.
- 따라서 누적 대기시간은 $\left(\frac{30}{9}\right)\times 8+\cdots+\left(\frac{30}{9}\right)\times 1=\left(\frac{30}{9}\right)\times 36=120$분이다.

- 예제 2.16

시스템에 N명의 고객이 도착하는 순간 시스템이 종료된다. 종료될 때까지 시스템 내에 존재하는 고객 수는 평균적으로 몇 명인가?

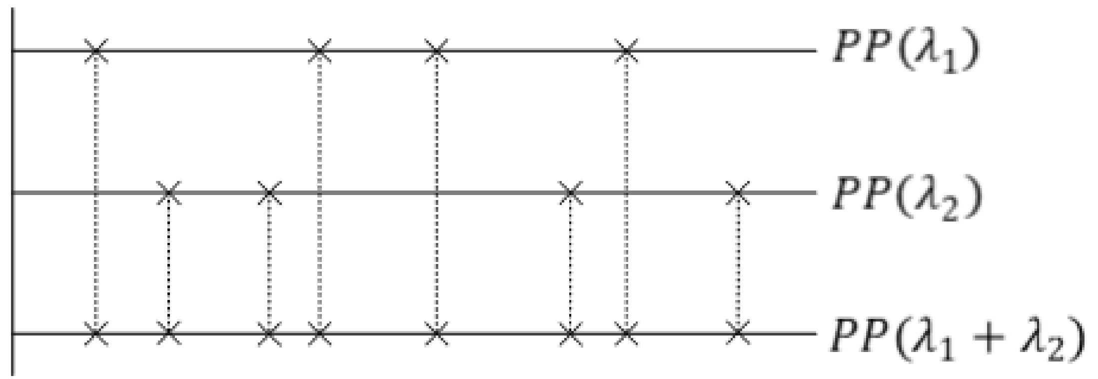
- 시스템이 종료하기 전까지는 총 $N-1$ 명의 포아송 도착고객이 있다.
- 이들의 도착시점들의 평균은 동일한 간격을 가지므로 시스템 내 고객 수가 0명인 기간, 1명인 기간, $\ldots$, $N-1$명인 기간들은 그 길이의 평균이 모두 동일하며 총 N개의 기간이 있다.
- 따라서 평균 고객수는 $(0+1+\cdots+N-1)/N=(N-1)/2$명이다.

포아송 과정

- 포아송 과정의 중첩과 분해
  - 중첩(Superposed) 포아송 과정

정의 2.3

두 개의 서로 독립인 포아송 과정 $\{N_1(t), t \geq 0\}$과 $\{N_2(t), t \geq 0\}$이 있는데, 모수는 각각 λ_1, λ_2이다. 이들의 사건 발생간격은 $Exp(\lambda_1), Exp(\lambda_2)$을 따른다.
이들을 서로 겹쳐서 만든 새로운 확률과정 $\{N_1(t) + N_2(t), t \geq 0\}$은 모수가 $\lambda_1 + \lambda_2$인 포아송 과정이다. 중첩 포아송 과정의 사건발생 간격은 $\min\left(Exp(\lambda_1), Exp(\lambda_2)\right)$이므로 $Exp(\lambda_1 + \lambda_2)$를 따른다. 셋 이상의 독립적인 포아송 과정을 중첩하는 경우도 마찬가지로 확장해서 생각할 수 있다.



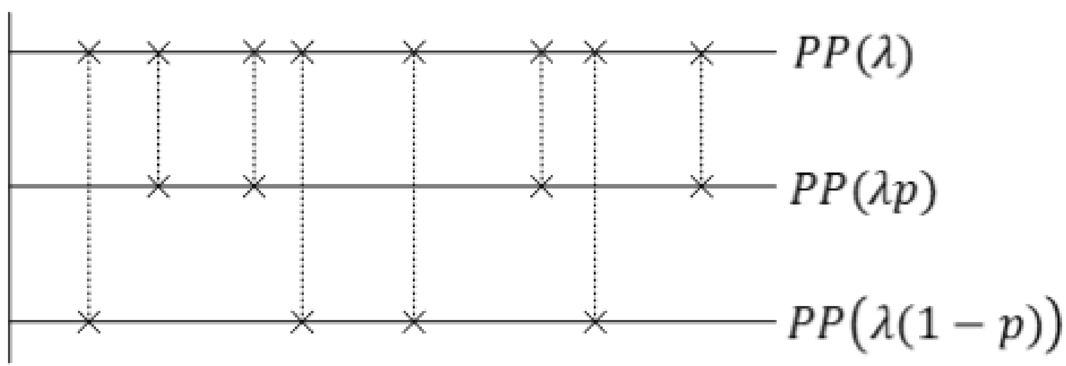
<그림 2.20> 중첩 포아송 과정

포아송 과정

- 포아송 과정의 중첩과 분해
  - 베르누이 분해(Bernoulli Decomposed) 포아송 과정

정의 2.4

모수가 λ인 포아송 과정 $\{N_1(t), t \geq 0\}$에서, 사건이 발생할 때마다 앞면이 나올 확률이 p인 동전을 던져서 앞면이 나오면 '사건 1'로, 뒷면이 나오면 '사건 2'로 기록을 한다. 이렇게 기록된 사건 1들은 모수가 $p\lambda$인 포아송 과정을, 사건 2들은 $(1-p)\lambda$인 포아송 과정을 이룬다. 이를 포아송 과정의 베르누이 분해라고 한다.



<그림 2.21> 베르누이 분해 포아송 과정

포아송 과정

• 예제 2.17

콜센터에 가입(=1) 및 탈퇴(=2) 요청 등 두 가지 종류의 문의전화가 각각 시간당 $\lambda_1 = 10$건, $\lambda_2 = 9$건씩 걸려온다.

(a) 걸려온 문의전화가 가입 요청을 위한 전화일 확률은?

(b) 탈퇴 문의전화의 15%는 콜센터 직원의 설득으로 탈퇴를 철회한다. 하루 8시간 동안 탈퇴요청을 철회하는 횟수 M이 15건 이상이 될 확률은?

(a) 걸려온 문의전화가 가입 요청을 위한 전화일 확률

- 중첩 포아송 과정에 의해, $\lambda_1/(\lambda_1 + \lambda_2)$ 활용
- 전체 문의전화의 도착률은 $10 + 9 = 19$이며, 이 가운데 도착률 10으로 가입 요청이 오므로 구하는 확률은 $10/(10 + 9) = 0.53$임

(b) 하루 8시간 동안 탈퇴요청을 철회하는 횟수 M이 15건 이상이 될 확률

- 베르누이 분해 포아송 과정에 의해, 탈퇴 문의전화의 포아송 과정을 $p = 0.15$의 확률로 분해하면, 탈퇴철회는 평균 $\lambda = 9 \times 0.15 = 1.35$(건/시간)의 포아송 과정임
- $t = 8$(시간)이므로 $\lambda t = 10.8$이고,
  구하는 확률은 $P(M \geq 15) = 1 - \sum_{i=0}^{14} e^{-10.8}(10.8)^i/i! = 1 - 0.8682 = 0.1318$임

목 차

- 지수분포
- 포아송 과정
- 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.

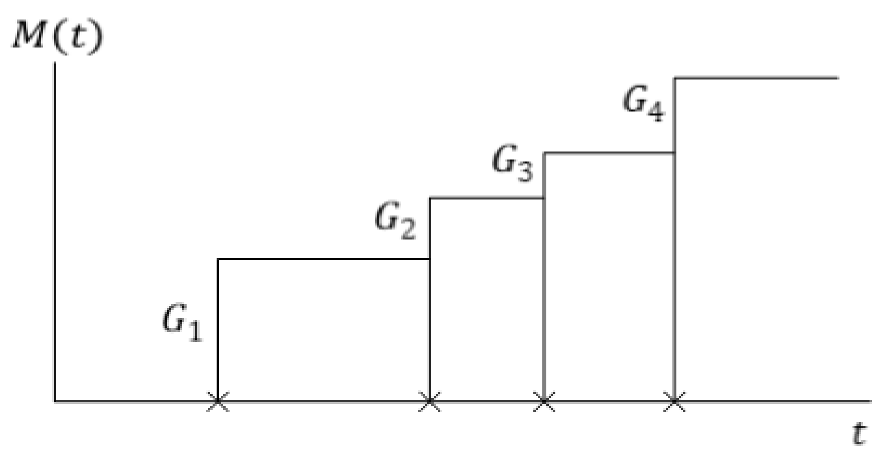
- 특징
  - 복합 포아송 과정은 각 포아송 사건의 크기를 합산하여 생성됨
  - 포아송 과정의 단일증분을 제외하고, 독립증분과 정상증분 중에 하나 이상의 가정을 허용하는 경우에도 복합 포아송 과정이 됨

복합 포아송 과정

• 예제 2.18

시점 t까지 고속도로 톨게이트를 통과한 차량의 대수 $N(t)$가 $PP(\lambda)$라고 하자.
또한 확률변수 $G_k(k = 1,2, \dots)$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.

- 확률과정 $M(t)$의 그래프에서 각 점프의 시점들은 $PP(\lambda)$를 이루고, 점프의 높이는 확률변수 $G_k(k = 1,2, \dots)$임
- 총 승객 수인 $M(t)$는 랜덤합이므로 $E\big(M(t)\big) = E\big(N(t)\big)E(G) = \lambda t E(G)$임



<그림 2.22> 예제 2.18 그림

복합 포아송 과정

- 예제 2.19

베르누이 분해된 포아송 과정은 복합 포아송 과정으로도 생각할 수 있다. $N(t) \sim PP(\lambda)$이고 점프의 높이가 확률 p로 1, 확률 $q = p - 1$로 0인 복합 포아송 과정을 생각하자. 이 때 시점 t까지의 총 도착수가 k일 확률을 구해보자.

- 구하는 확률은 $P(M(t) = k)$이고, $G_i = 1$(확률 p) 또는 0(확률 q)이므로

$$P(Mt) = k) = P\left(G_1 + G_2 + \cdots + G_{N(t)} = k\right)$$

$$= \sum_{n=k}^{\infty} P\left(G_1 + G_2 + \cdots + G_{N(t)} = k | N(t) = n\right) P(N(t) = n)$$

$$= \sum_{n=k}^{\infty} P\left(G_1 + G_2 + \cdots + G_{N(t)} = k\right) \frac{e^{-\lambda t}(\lambda t)^n}{n!} = \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \frac{e^{-\lambda t}(\lambda t)^n}{n!} = \cdots = \frac{e^{-\lambda pt}(\lambda pt)^n}{k!}$$

- 네 번째 등식에서는 점프높이를 나타내는 G_i가 베르누이 확률변수이므로 $G_1 + \cdots + G_n$이 이항확률변수임을 활용함
- 마지막 등식에서는 $\sum_{x=0}^{\infty} \frac{a^x}{x!} = e^a$의 등식을 적용함
- 따라서, $M(t)$는 $N(t)$에서 점프 크기가 1인 경우만 걸러내서 얻은 베르누이 분해된 복합 포아송 과정임

목 차

- 지수분포
- 포아송 과정
- 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 dt에 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) $P(N(t+dt)-N(t)=0) \approx 0$

  - 확률과정 $\{N(t), t \geq 0\}$이 비정상 포아송 과정을 따르는 것은, $N(t) \sim NSPP\big(\lambda(t)\big)$로 표현함

- 특징
  - 포아송 과정에서 정상증분 가정을 제외한 경우에도 비정상 포아송 과정이 됨
    - 정상증분 가정은 어떤 구간내에서의 사건 발생 수는 그 구간의 시간적 위치와 무관함을 정의함

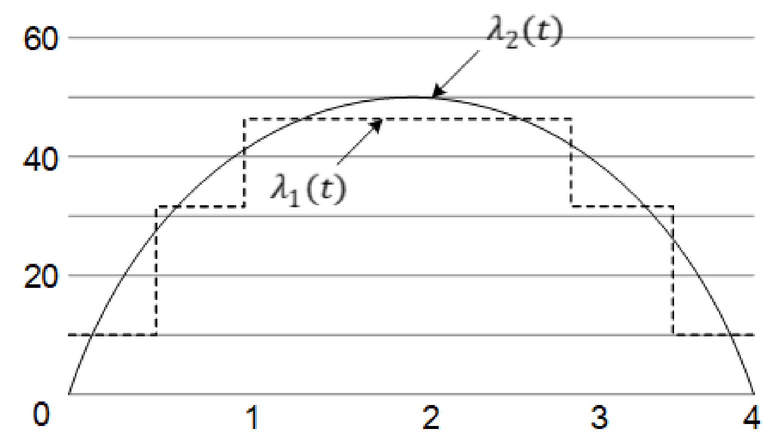
비정상 포아송 과정

- 예제 2.20

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 시간에 따른 도착률 변화를 선형으로 근사한 함수
- $\lambda_2(t)$는 실제 시간에 따른 도착률의 변화를 반영한 함수



<그림 2.23> 예제 2.20 그림

비정상 포아송 과정

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

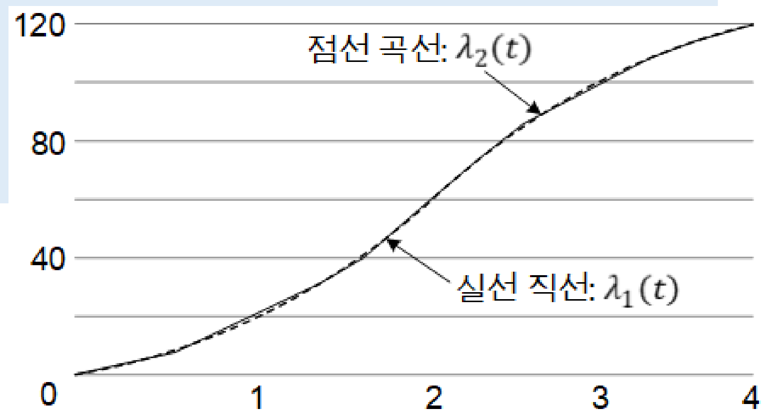
- 예제 2.21

예제 2.20에서, $\Lambda_1(t), \Lambda_2(t)$는 시점 t까지의 평균 발생횟수에 해당한다.

$$\Lambda_1(t) = \begin{cases} 10t, & 0 \le t < 0.5, \\ 5 + 29(t - 0.5), & 0.5 \le t < 1, \\ 19.5 + 45(t-1), & 1 \le t < 3, \\ 109.5 + 29(t-3), & 3 \le t < 3.5, \\ 124 + 10(t - 3.5), & 3.5 \le t < 4, \end{cases}$$

$$\Lambda_2(t) = \int_0^t 50\sin(\pi x/4)dx = \frac{200}{\pi}\left(1 - \cos\left(\frac{\pi t}{4}\right)\right), 0 \le t \le 4.$$

- $\Lambda_1(t), \Lambda_2(t)$에 의하면, $\Lambda_1(t)$의 직선이 $\Lambda_2(t)$의 곡선에 매우 근사하므로, 예제 2.20과 같이 도착률 함수가 $\lambda_2(t)$ 같은 곡선인 경우, 구간을 구분하여 부분선형 함수 $\lambda_1(t)$처럼 표현할 수 있음



<그림 2.24> 예제 2.21 그림

Thanks!

김 지 혜 (jihye@pel.sejong.ac.kr)

- MATLAB 코드 – 그림 2.1a

```
lambda = 2;
k = 0:15;

pmf = poisspdf(k, lambda);

disp('포아송 분포의 PMF:');
disp(table(k', pmf', 'VariableNames', {'사건발생횟수', '확률'}));

fig = figure;
set(fig, 'Position', [100, 100, 600, 220]);
stem(k, pmf, 'filled', 'Color', 'Blue', 'LineWidth', 1.5);
hold on;

xlabel('사건 발생 횟수 (k)');
ylabel('확률');
title(['포아송 분포 (λ = ', num2str(lambda), ')']);
```

- MATLAB 코드 – 그림 2.1b

```
lambda = 2;
x = linspace(0, 10, 100);

pdf = exppdf(x, 1/lambda);

disp('지수 분포의 PDF:');
disp(table(x', pdf', 'VariableNames', {'시간간격', '확률밀도'}));

fig = figure;a
set(fig, 'Position', [100, 100, 600, 220]);
plot(x, pdf, 'b', 'LineWidth', 1.5);
xlabel('시간 간격 (x)');
ylabel('확률 밀도');
title(['지수 분포 (λ = ', num2str(lambda), ')']);
grid on;
```

부록

- MATLAB 코드 – 그림 2.3

```
x = 0:0.1:10;
y1 = exppdf(x,1); %/mu=1
y2 = exppdf(x,2); %/mu=2

fig = figure;
set(fig, 'Position', [100, 100, 600, 220]);
plot(x, y1, 'Color', 'Blue', 'LineWidth', 1.5);
grid on;
hold on;

plot(x, y2, 'Color', 'Red', 'LineWidth', 1.5);
xlabel('Observation')
ylabel('Probability Density')
legend('E(X) = 1', 'E(X) = 2')
```

- MATLAB 코드 – 그림 2.4

```
x = 0:0.1:6;
y = exppdf(x,1);
fig = figure;
set(fig, 'Position', [100, 100, 600, 220]);
plot(x, y, 'Color', 'Blue', 'LineWidth', 1.5);
hold on;
x_fill = [0 x(x <= 2) 2];
y_fill = [0 y(x <= 2) 0];
fill(x_fill, y_fill, 'blue', 'FaceAlpha', 0.3, 'EdgeColor',
'none');
text(1.8, 0.5, '$f(x) = \lambda e^{-\lambda x}$',
'Interpreter', 'latex', 'FontSize', 16, 'VerticalAlignment',
'top');
text(-0.1, -0.02, '0', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
text(1.9, -0.02, '$x$', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
text(5.8, -0.02, '$X$', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
text(-0.5, 0.95, '$f(x)$', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
set(gca, 'XTick', []);
set(gca, 'YTick', []);
```

```
x = 0:0.1:6;
y = exppdf(x,1);
fig = figure;
set(fig, 'Position', [100, 100, 600, 220]);
plot(x, y, 'Color', 'Blue', 'LineWidth', 1.5);
hold on;
x_fill = [2 x(x >= 2) max(x)];
y_fill = [0 y(x >= 2) 0];
fill(x_fill, y_fill, 'blue', 'FaceAlpha', 0.3, 'EdgeColor',
'none');
text(1.8, 0.5, '$f(x) = \lambda e^{-\lambda x}$',
'Interpreter', 'latex', 'FontSize', 16, 'VerticalAlignment',
'top');
text(-0.1, -0.02, '0', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
text(1.9, -0.02, '$x$', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
text(5.8, -0.02, '$X$', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
text(-0.5, 0.95, '$f(x)$', 'Interpreter', 'latex', 'FontSize', 14,
'VerticalAlignment', 'top');
set(gca, 'XTick', []);
set(gca, 'YTick', []);
```

부록

- MATLAB 코드 – 그림 2.5

```
x = 0:0.1:5;
y = exppdf(x,1.5);
fig = figure;

set(fig, 'Position', [100, 100, 600, 220]);
plot(x, y, 'Color', 'Blue', 'LineWidth', 1.5);
grid on;
hold on;

overlap_x = 1.5:0.01:2;
overlap_y = interp1(x, y, overlap_x);

fill([overlap_x, fliplr(overlap_x)], [overlap_y, zeros(size(overlap_y))], 'Blue', 'FaceAlpha', 0.3, 'EdgeColor',
'none');

xlabel('통화시간', 'FontSize', 12)
ylabel('Probability Density', 'FontSize', 12)
legend('$E(X) = 1.5$', 'Interpreter', 'latex', 'FontSize', 12)
```

부록

- 여러 지수 확률변수들의 확률적 성질 (1)

1. 서로 독립인 지수 확률변수들의 최소값
   - 지수 확률변수들의 최솟값 $\min(X_1, X_2, \dots, X_n)$의 확률분포

$P(\min(X_1, X_2) > t) = P\big((X_1 > t) \cap (X_2 > t)\big)$

$= P(X_1 > t)P(X_2 > t) = e^{-(\lambda_1+\lambda_2)t}$

$P(\min(X_1, X_2) \le t) = 1 - e^{-(\lambda_1+\lambda_2)t}, t \ge 0$

$\therefore \min(X_1, X_2) \sim Exp(\lambda_1 + \lambda_2)$

- $Exp(\lambda_1 + \lambda_2)$ 누적분포함수: $F(x) = 1 - e^{-(\lambda_1+\lambda_2)x}$

독립사상의 곱셈법칙
$P(A \cap B) = P(A)P(B)$

- 또한, $\min(X_1, \dots, X_n) = \min(X_1, \min(X_2, \dots, X_n))$임을 활용하여, $\min(X_1, \dots, X_n) \sim Exp(\lambda_1 + \dots + \lambda_n)$로 일반화

부록

- 여러 지수 확률변수들의 확률적 성질 (2)

2. 서로 독립인 지수 확률변수들의 비교
   - 지수 확률변수 X_1, X_2를 비교할 때, X_1이 X_2보다 먼저 발생할 확률을 계산하는 것

$$P(X_1 < X_2) = \int_0^\infty P(X_1 < X_2 | X_1 = x)\lambda_1 e^{-\lambda_1 x} dx$$

$$= \int_0^\infty P(X_2 > x)\lambda_1 e^{-\lambda_1 x} dx = \int_0^\infty e^{-\lambda_2 x}\lambda_1 e^{-\lambda_1 x} dx = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

$$\therefore P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

   - 또한, 이를 확장하여 $P\left(X_j = \min_{k=1,\ldots,n}(X_k)\right) = P\left(X_j < \min_{j \neq k}(X_k)\right) = \frac{\lambda_j}{\lambda_1 + \cdots + \lambda_n}$으로 일반화

부록

- 여러 지수 확률변수들의 확률적 성질 (3)

3. 서로 독립인 지수 확률변수들 중 최대-최소 차이의 기댓값
   - $Y = \max(X_1, X_2) - \min(X_1, X_2)$일 때, 전체확률의 법칙(Rule of the Total Probability) 사용하는 경우

$$E(Y) = E(Y|X_1 < X_2)P(X_1 < X_2) + E(Y|X_2 \le X_1)P(X_2 \le X_1)$$

- $X_1 < X_2$인 경우
  - $Y = X_2 - X_1, E(Y|X_1 < X_2) = E(X_2)$, 기억상실성질에 의해 $E(X_2|X_1 < X_2) = E(X_2) = \frac{1}{\lambda_2}$
  - 이 경우의 확률은 $P(X_1 < X_2) = \frac{\lambda_1}{\lambda_1+\lambda_2}$
- $X_1 \ge X_2$인 경우
  - $Y = X_1 - X_2, E(Y|X_1 \ge X_2) = E(X_1)$, 기억상실성질에 의해 $E(X_1|X_1 \ge X_2) = E(X_1) = \frac{1}{\lambda_1}$
  - 이 경우의 확률은 $P(X_1 \ge X_2) = \frac{\lambda_2}{\lambda_1+\lambda_2}$

부록

- 여러 지수 확률변수들의 확률적 성질 (3)

3. 서로 독립인 지수 확률변수들 중 최대-최소 차이의 기댓값

   - 이때 기억상실성질에 의해, $E(Y|X_1 < X_2) = E(X_2) = 1/\lambda_2$,

   $E(Y|X_2 \le X_1) = E(X_1) = 1/\lambda_1$

   $$\therefore E(Y) = \frac{1}{\lambda_2} \cdot \frac{\lambda_1}{\lambda_1 + \lambda_2} + \frac{1}{\lambda_1} \cdot \frac{\lambda_2}{\lambda_1 + \lambda_2}$$