# 확률통계 및 프로그래밍

## 5장 확률분포

: 연속확률분포(지수분포)

# 지수 분포 (exponential distribution)

포아송 분포는 "**단위시간 동안 발생하는 사건의 횟수**를 확률변수 $X$로 정의한 이산확률분포"이다. 반면 지수분포는 첫 번째 사건이 일어날 때까지의 걸리는 대기시간을 확률변수 $X$로 정의한 연속확률분포"이다.

그러면 **단위구간에서 평균 발생건수가** $\lambda$(모수)인 포아송 분포에서 단위구간을 확장하자.

$$P(X=x)=e^{-\lambda}\frac{\lambda^{x}}{x!}\quad \text{(확률질량함수)}$$

구간을 $[0,t]$로 확장을 하면 이 구간에서는 평균 발생건수는 단위시간당 평균 $\lambda$이므로 $\lambda t$이다. 그럼 확률분포함수는

$$P(X=x)=e^{-\lambda t}\frac{(\lambda t)^{x}}{x!}$$

여기서 $x=0$일 때

$$P(X=0)=e^{-\lambda t}\frac{(\lambda t)^0}{0!}=e^{-\lambda t}$$

이것의 의미는 구간 $[0,t]$, $t$ **시간 동안 사건이 한 번도 일어나지 않을 확률**이다.

여기에서 "**최초 사건이 발생하기까지의 걸린 시간**"을 나타내는 확률변수를 시간 $T$라고 하자.

$T>t$일 확률은 관측을 시작한 이후로 $t$시간이 지난 이후에 사건이 발생함을 의미한다. 즉, $[0,t]$에서 사건이 전혀 발생하지 않음을 의미한다. 따라서 $T>t$와 $X=0$은 동치이다.

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

한편, 구간 $t$ 시간 동안 사건이 적어도 한 번은 일어날 확률은 사건이 발생하기까지의 걸린 시간 $T$가 $t$ 보다 작을 확률과 같다. 즉

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

이고 누적확률분포함수이다. 따라서 $\dfrac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t} = f(t)$

지수분포의 확률밀도함수는

$$f(x) = \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \ge 0 \end{cases}$$

## [정의] 지수분포(exponential distribution)

지정된 시점으로부터 어떤 사건이 일어날 때까지 걸리는 시간(X)을 측정하는 확률분포

$$X \sim Exp(\lambda) \qquad f(x) = \lambda e^{-\lambda x}, \; x > 0$$

- **누적분포함수** $F(x) = P(X \le x) = \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}, \;\; x > 0$

- **기댓값과 분산**

$$E(X) = \frac{1}{\lambda} \qquad E(X^2) = \frac{2}{\lambda^2}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}$$

**# 지수분포의 R 함수**

# 확률밀도함수 f(x) (rate=λ), 초기치 (rate = 1)

```
dexp(x, rate = 1)
```

# 누적분포함수 (q=분위수)

```
pexp(q, rate = 1, lower.tail = T)
```

# 분위수 (p=누적확률)

```
qexp(p, rate = 1, lower.tail = T)
```

# 지수 확률변수 (n=난수의 개수)

```
rexp(n, rate = 1)
```

[예제1] 어느 기업의 부품은 1개월에 3번씩 고장을 일으킨다. 부품이 고장 나서 고친 후에 다시 고장이 발생할 때까지 걸리는 시간(X)을 측정하는 분포를 구하고, 2개월이내에는 고장 나지 않을 확률을 구하여라.

$\lambda = 3 \qquad f(x) = 3e^{-3x}, x > 0$

```
# 2개월 이내에 고장 날 누적확률
pexp(2,3)
# 2개월 이내에는 고장 나지 않을 확률 P(X>2)
1-pexp(2,3)
```

```
# λ=3일 때, P(1<X<2) 구하기
pexp(q=2, rate=3)-pexp(q=1, rate=3)
#지수확률분포의 95% 분위수 구하기
qexp(p=0.95, rate=3)
pexp(0.9985774,3)

# λ=3인 지수함수의 plot
x <- seq(0, 5, length.out=101)
lambda <- 3
y <- lambda * exp(-lambda * x)
plot(x, y)

x <- seq(0, 5, length=101) # length를 통해 벡터의 길이를 조정할 수 있다.
plot(x, dexp(x,3), col='red')
plot(seq(0, 5, length=101), dexp(x,3), col='red')
plot(x, dexp(x,3), col='red', type='l')
```

[예제2] A 사의 전구수명은 평균 1000 시간 의 지수분포를 따른다. 100개의 전구 가운데 1000 시간 이상 사용할 수 있는 것의 평균개수를 구하여라.

**풀이** 전구수명을 확률변수 $X$라고 했을 때 $\lambda = \dfrac{1}{1000}$이다.

$$P(X \geq 1000) = 1 - P(X < 1000) = 1 - F(1000) = e^{-1} = 0.3679$$

이므로 구하고자 하는 평균은 $100 \times 0.368$, 약 37개가 된다.

```
# pexp(확률변수 값, lambda=1/평균)를 이용하여, 지수분포로부터 P(X>1000)를 구하자.
pexp(1000, 1/1000)       # 1000시간까지의 누적확률
1-pexp(1000, 1/1000)   # 1000시간 이상 사용할 수 있을 확률
```