3장 연속형분포

1. 연속형 확률분포
2. 균일 분포
3. 지수 분포
4. 감마 함수와 감마 분포
5. 베타 분포와 카이제곱 분포
6. 정규 분포

1. 연속형 확률분포

[정의 3.1-1] 연속형 확률변수의 밀도함수 는 다음의 성질을 갖는 함수이다.
(a) $0 \le f(x, y) \le 1$
(b) $\sum\sum_{(x,y)\in\mathbb{R}} f(x, y) = 1$
(c) $P[(X, Y) \in A] = \sum\sum_{(x,y)\in A \subset \mathbb{R}} f(x, y)$
확률 $P(a < X < b)$는 밀도함수 $f(x)$의 그래프, x축, $x=a$, $x=b$로 둘러싸인 면적이다.

[정의 3.1-2] 연속인 확률변수의 분포함수는 다음과 같이 정의된다.

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt.$$

[연속형 확률변수의 성질]
(1) $P(X=b)=0$; 이산형 확률변수와 연속형 확률변수와의 차이점
(2) $F'(x) = f(x)$
(3) $P(a < X < b) = P(a < X \le b) = P(a \le X < b) = P(a \le X \le b) = F(b) - F(a)$

[정의 3.1-3]
(a) 평균: $\mu = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$
(b) 분산: $\sigma^2 = Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)dx = E(X^2) - \mu^2$
(c) 적률생성함수: $M(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx$, $-h < t < h$
(d) 누가적률함수: $R(t) = \ln M(t)$

$$\mu = R'(0),\ \sigma^2 = R''(0)$$

[예 3.1-1] 확률변수 X를 pdf가 $g(x)=2x$, $0<x<1$인 연속확률변수라고 하면 X의 분포함수는 아래와 같다.

$$F(x)=\begin{cases}0, & x<0,\\ \int_0^x 2t\,dt = x^2, & 0\le x<1,\\ 1, & 1\le x\end{cases}$$

$$P\left(\frac{1}{2}\le X\le \frac{3}{4}\right)=F\left(\frac{3}{4}\right)-F\left(\frac{1}{2}\right)=\left(\frac{3}{4}\right)^2-\left(\frac{1}{2}\right)^2=\frac{5}{16},$$

$$P\left(\frac{1}{4}\le X< 2\right)=F(2)-F\left(\frac{1}{4}\right)=1-\left(\frac{1}{4}\right)^2=\frac{15}{16}.$$

[예 3.1-2] 확률변수 X를 pdf가 다음과 같을 때 X의 평균과 분산을 구하라.

$$f(x)=\begin{cases}xe^{-x}, & 0\le x<\infty\\ 0, & o/w\end{cases}$$

풀이> 확률변수 X의 적률생성함수

$$M(t)=\int_0^\infty e^{tx}\cdot xe^{-x}dx=\int_0^\infty x\cdot e^{-(1-t)x}dx$$

$$=\left[-\frac{xe^{-(1-t)x}}{1-t}-\frac{e^{-(1-t)x}}{(1-t)^2}\right]_0^\infty$$

$$=\frac{1}{(1-t)^2},\quad t<1.$$

누가적률 $R(t)=\ln M(t)=-2\ln(1-t)$,

$$R'(t)=\frac{2}{1-t},\ R''(t)=\frac{2}{(1-t)^2},$$

평균과 분산 $\mu=R'(0)=2$, $\sigma^2=R''(0)=2$.

2. 균일분포

[정의 3.2-1] 확률변수 X를 pdf가 구간 $[a,b]$에서

$$f(x)=\frac{1}{b-a},\ a\le x\le b$$

로 주어졌을 때 X는 균일분포(Uniform distribution)를 갖는다: $X\sim U(a,b)$

확률변수 X의 한 점을 랜덤하게 구간 $[a,b]$, $-\infty<a<b<\infty$에서 뽑았을 때 그 점이 구간 $[a,x]$, $a\le x<b$에 들어 있을 확률은 $(x-a)/(b-a)$이다.

$$X\text{의 분포함수 } F(x)=\begin{cases}0, & x<a,\\ \frac{x-a}{b-a}, & a\le x<b,\\ 1, & b\le x.\end{cases}$$

률변수 X는 연속확률변수이므로 $F'(x)$가 존재한다면 $F'(x)=f(x)=1/(b-a)$.

[정리 3.2-1] 확률변수 X가 균일분포를 가질 때, X의 적률생성함수, 평균과 분산

$$M(t)=\begin{cases}\frac{e^{tb}-e^{ta}}{t(b-a)}, & t\ne 0,\\ 1, & t=0.\end{cases}\quad E(X)=\frac{a+b}{2},\ Var(X)=\frac{1}{12}(b-a)^2.$$

증명> $t\ne 0$일 때 $M(t)=E[e^{tX}]=\int_a^b e^{tX}\cdot\frac{1}{b-a}dx=\left[\frac{e^{tx}}{t(b-a)}\right]_a^b=\frac{e^{tb}-e^{ta}}{t(b-a)}$

$t=0$일 때 $M(0)=E[e^{0\cdot X}]=E[1]=1$

평균과 분산은 적률생성함수를 이용하는 것보다는 정의에서 직접구하는 것이 편리

$$E(X)=\int_a^b x\cdot\frac{1}{(b-a)}dx=\frac{a+b}{2},$$

$$Var(X)=E(X^2)-[E(X)]^2=\int_a^b x^2\cdot\frac{1}{(b-a)}dx-\left[\frac{(a+b)}{2}\right]^2=\frac{(b-a)^2}{12}.$$

[정의 3.2-2] 확률변수 X를 pdf가 아래와 같이 주어질 때, X는 표준균일분포

$$f(x)=\begin{cases}1, & 0<x<1,\\ 0, & o/w.\end{cases}\quad X\sim U(0,1)$$

[참고] 확률변수 U가 균일분포 $U(0,1)$을 갖는다면 U의 일차변환

$$X=a+(b-a)U$$

는 균일분포 $U(a,b)$를 갖게 되며, 역으로 확률변수 X가 균일분포 $U(0,1)$을 갖는다면 X의 일차변환

$$U=\frac{X-a}{b-a}$$

는 균일분포 $U(0,1)$을 갖게 된다.

[정리 3.2-2] (확률 적분 변환(Probability integral transformation))

(1) $F(x)$가 연속형 확률변수의 누적분포함수일 때, 확률변수 $U=F(X)$는 표준균일분포 $U(0,1)$를 따르게 된다.
(2) 역으로 확률변수 U가 표준균일분포 $U(0,1)$를 따를 때, 확률변수 $X=F^{-1}(U)$의 누적분포함수는 $F(X)$이다.

증명> (1) $P[U \le u] = P[F(X) \le u] = P[X \le F^{-1}(u)] = FF^{-1}(u) = u$

$$P[X \le x] = P[F^{-1}(U) \le x] = P[U \le F(x)] = F(x),$$

확률변수 X의 누적분포함수가 $F(x)$임을 보여준다.

[예 3.2-1] 확률변수 X의 pdf가 $f(x)=1/2\sqrt{x}, 0<x<1$로 주어진다면 X의 누적분포함수는 $F(x)=\sqrt{x}$, $0<x<1$이 된다. 지금 Y를 표준균일분포 $U(0,1)$을 따르는 확률변수라고 하자. 위의 정리를 이용하기 위해 $Y=F(x)=\sqrt{x}$라 두면 $Y^2=x$의 밀도함수와 분포함수는 앞에서 주어진 바와 같다. 부연하여 설명하자면, y가 구간 $(0,1)$에서 균일하게 생성되는 난수 (Random number)라고 하면 생성된 난수의 제곱인 y^2은 앞에서 주어진 분포에서의 난수가 된다.

[예 3.2-2] 확률변수 X의 분포함수가 $F(x)=\int_0^x \lambda e^{-\lambda t}dt = 1-e^{-\lambda x}$, $x>0$로 주어졌다고 하자. $U=1-e^{-\lambda X}$로 두면

$$X=-\frac{1}{\lambda}\ln(1-U) \approx -\frac{1}{\lambda}\ln U.$$

• $U=F(X)$는 [정리 3.2-2]에서 표준균일분포 $U(0,1)$을 따르게 되므로 표준균일분포에서 생성된 난수의 $-\log$ 값을 λ로 나누어서 만들어진 난수는 확률변수 X의 분포 (모수가 λ인 지수분포)에서 생성된 난수로 보면 된다.

• U가 표준균일분포를 따르면 $1-U$도 표준균일분포 $U(0,1)$을 따르게 된다.

3. 지수분포

X를 평균이 λt인 포아송 분포를 따르는 확률변수라고 하면 시간 $(0,t)$에서 사건이 전혀 일어나자 않을 확률

$$P[X=0] = e^{-\lambda t}$$

$$P[T>t] = P[\text{시간 } (0,\ t)\text{에서 사건이 발생치 않음}] = P[X=0] = e^{-\lambda t}, \quad t>0$$

확률변수 T는 연속확률변수로 볼 수 있으며 T의 분포함수는

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

[정의 3.3-1] 확률변수 X를 pdf가

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

으로 주어질 때, X는 모수가 λ인 지수분포(Exponential distribution)를 갖는다.

[정리 3.3-1] 확률변수 X는 모수가 λ인 지수 분포를 갖는다고 하자. X의 적률생성함수, 평균, 분산

$$M(t) = \frac{\lambda}{\lambda-t}, \quad t<\lambda, \quad E(X)=\frac{1}{\lambda}, \quad Var(X)=\frac{1}{\lambda^2}.$$

증명>

$$M(t) = \int_0^\infty e^{tx}\cdot f(x)dx = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \int_0^\infty \lambda e^{(t-\lambda)x}dx$$

$$= \left[\frac{\lambda}{t-\lambda}e^{(t-\lambda)x}\right]_0^\infty = \frac{\lambda}{\lambda-t}$$

따라서, $M'(t)=\dfrac{1/\lambda}{(1-t/\lambda)^2}$, $M''(t)=\dfrac{2/\lambda^2}{(1-t/\lambda)^3}$이므로 $E(X)=M'(0)=\dfrac{1}{\lambda}$,

$E(X^2)=M''(0)=\dfrac{2}{\lambda^2}$, $Var(X)=M''(0)-[E(X)]^2=\dfrac{1}{\lambda^2}$이다.

단위시간에 발생하는 평균사건수 λ로 두었을 때, 하나의 사건이 일어날 때까지 기다리는 시간의 평균은 $1/\lambda$이고 분산은 $1/\lambda^2$

[예 3.3-1] 시간당 평균 20명의 손님이 노래방에 들어온다고 했을 때, 첫 손님이 오는데 주인이 5분 이상 기다릴 확률은 얼마인가?
풀이>

확률변수 X를 첫 손님이 들어올 때까지 주인이 기다리는 시간(단위 : 분)이라고 하면 $\lambda = 1/3$은 분당 들어오는 손님의 기대값이다. 따라서 X의 밀도함수가

$$f(x) = \frac{1}{3} e^{-\frac{1}{3}x}, \quad 0 \le x < \infty$$

이므로 구하는 확률은 아래와 같다.

$$P(X > 5) = \int_5^\infty \frac{1}{3} e^{-\frac{1}{3}x} dx = e^{-\frac{5}{3}} = 0.189.$$

[정리 3.3-2] (지수분포의 건망성(혹은 망각성, Memoryless property))
확률변수 X가 모수 λ인 지수분포를 따를 때,

$$P[X > y + x \mid X > x] = P[X > y], \quad x, \ y > 0$$

이 성립하며 이 성질은 지수분포의 건망성이라고 한다.

증명> $P[X > y + x \mid X > x] = \dfrac{P[X > y + x]}{P[X > x]} = \dfrac{e^{-\lambda(y+x)}}{e^{-\lambda x}} = e^{-\lambda y} = P[X > y].$

[예 3.3-2] 전구의 수명은 평균 사용시간이 500시간인 지수분포를 따른다고 한다. 확률변수 X를 전구의 수명이라고 하면

$$P[X > t] = \int_t^\infty \frac{1}{500} e^{-\frac{x}{500}} dx = e^{-\frac{t}{500}}$$

를 얻는다. 어떤 전구가 현재까지 300시간 작동하였다면 앞으로 600시간을 더 사용할 수 있는 확률은 얼마인가?
풀이> 구하는 확률은 다음과 같다.

$$P[X > 600 + 300 \mid X > 300] = \frac{P[\{X > 900\} \cap \{X > 300\}]}{P[X > 300]} = \frac{P[X > 900]}{P[X > 300]}$$
$$= \frac{e^{-9/5}}{e^{-3/5}} = e^{-6/5}.$$

[참고] 건망성은 지수분포뿐만 아니라 기하분포도 갖고 있는 성질이다.