MAT2410 (응용수학I)
과제 1-2

서강대학교 컴퓨터공학과 박수현 (20181634)

서강대학교 컴퓨터공학과

4 이산확률분포

4.1 이산균등분포

3. X는 2와 11 사이의 정수들에 대하여 이산균등분포를 이룬다고 한다.

(1) X의 확률질량함수를 구하라.
(2) $Y = X - 1$의 확률질량함수를 구하라.
(3) $Y = X - 1$의 평균과 분산을 구하라.
(4) X의 평균과 분산을 구하라.

Solution.

(1)
$$f(x) = \frac{1}{11-2+1} = \frac{1}{10} \qquad 2 \le x \le 11, x \in \mathbb{Z}$$

(2)
$$f(y) = \frac{1}{10} \qquad 1 \le y \le 10, y \in \mathbb{Z}$$

(3) $Y \sim \mathrm{DU}(10)$이므로 기댓값은
$$\frac{1+10}{2} = \frac{11}{2}$$
이고, 분산은
$$\frac{10^2-1}{12} = \frac{33}{4}$$
이다.

(4)
$$\mathrm{E}(X) = \mathrm{E}(Y+1) = \frac{13}{2}$$
$$\mathrm{Var}(X) = \mathrm{Var}(Y+1) = \frac{33}{4}$$

4.2 초기하분포

3. 비복원추출에 의하여 52장의 카드가 들어 있는 주머니에서 임의로 세 장의 카드를 꺼낼 때, 세 장의 카드 안에 포함된 하트의 수를 확률변수 X라 한다.

(1) X의 확률질량함수를 구하라.
(2) 평균과 표준편차를 구하라.

Solution.

(1) 카드 52장 중 3장을 뽑는 경우의 수는 $\binom{52}{3}$이고, 3장 중 하트가 x장 나오는 경우의 수는 $\binom{13}{x}\binom{39}{3-x}$이다. 따라서 확률질량함수는

$$\frac{\binom{13}{x}\binom{39}{3-x}}{\binom{52}{3}} \qquad x = 0, 1, 2, 3$$

이며, $X \sim \mathrm{H}(52, 13, 3)$이다.

(2)

$$\begin{aligned}
\mu &= n\frac{M}{N} \\
&= 3 \times \frac{13}{52} \\
&= \frac{3}{4} \\
\sigma^2 &= n\frac{M}{N}\frac{N-M}{N}\frac{N-n}{N-1} \\
&= 3 \times \frac{13}{52}\frac{39}{52}\frac{49}{51} \\
&= \frac{147}{272} \\
\therefore \sigma &= \sqrt{\frac{147}{212}}
\end{aligned}$$

6. 지하수 오염 실태를 조사하기 위하여 30곳의 구멍을 뚫어 수질을 조사하였다. 그 걀과 19곳은 오염이 매우 심각하였고, 6곳은 약간 오염되었다고 보고하였다. 그러나 채취한 지하수 병들이 섞여 있어 어느 지역이 깨끗한 지하수를 갖고 있는지 모르는 상황에서 5곳을 선정하였을 때, 다음을 구하라.

(1) 오염 정도에 따른 확률분포
(2) 선정된 5곳 중에서 매우 심각하게 오염된 지역이 3곳, 약간 오염된 지역이 1곳일 확률
(3) 5곳 중에서 적어도 4곳에서 심각하게 오염되었을 확률

Solution. 선정한 5곳 중 매우 심각하게 오염된 지역의 수를 X, 약간 오염된 지역의 수를 Y라 하자.

(1) 30곳 중 5곳을 선정했는데 매우 심각하게 오염된 지역을 x곳, 약간 오염된 지역을 y곳 선정했을 확률을 $f(x, y)$라 하면

$$
\begin{aligned}
f(x, y) &= \frac{\binom{19}{x}\binom{6}{y}\binom{30-19-6}{5-x-y}}{\binom{30}{5}} \\
&= \frac{\binom{19}{x}\binom{6}{y}\binom{5}{5-x-y}}{\binom{30}{5}} \qquad 0 \le x+y \le 5,\ 0 \le x, y \le 5,\ x, y \in \mathbb{Z}
\end{aligned}
$$

이며, 확률분포는 확률질량함수 f를 갖는다.

(2)

$$
\begin{aligned}
f(3, 1) &= \frac{\binom{19}{3}\binom{6}{1}\binom{5}{1}}{\binom{30}{5}} \\
&= \frac{1615}{7917} \approx 0.2040
\end{aligned}
$$

(3)

$$
\begin{aligned}
P(X \ge 4) &= f(4, 0) + f(4, 1) + f(5, 0) \\
&= \frac{\binom{19}{4}\binom{6}{0}\binom{5}{1}}{\binom{30}{5}} + \frac{\binom{19}{4}\binom{6}{1}\binom{5}{0}}{\binom{30}{5}} + \frac{\binom{19}{5}\binom{6}{0}\binom{5}{0}}{\binom{30}{5}} \\
&= \frac{1292}{3393} \approx 0.3808
\end{aligned}
$$

4.3 이항분포

3. Society of Actuaries(SOA)의 확률 시험문제는 5지 선다형으로 제시된다. 15문항의 SOA 시험 문제 중에서 지문을 임의로 선택할 때, 다음을 구하라.

(1) 정답을 선택한 평균 문항 수
(2) 정답을 정확히 5개 선택할 확률
(3) 정답을 4개 이상 선택할 확률

Solution. 정답을 선택한 문제의 개수를 확률변수 X라 하면 $X \sim \mathrm{B}\left(15, \frac{1}{5}\right)$이다.

(1)

$$\mathrm{E}(X) = 15 \times \frac{1}{5} = 3$$

(2)

$$\begin{aligned} P(X = 5) &= \binom{15}{5} \left(\frac{1}{5}\right)^5 \left(1 - \frac{1}{5}\right)^{15-5} \\ &\approx 0.1032 \end{aligned}$$

(3)

$$\begin{aligned} P(X \geq 4) &= 1 - P(X \leq 3) \\ &= 1 - \sum_{i=0}^{3} \binom{15}{i} \left(\frac{1}{5}\right)^i \left(1 - \frac{1}{5}\right)^{15-i} \\ &\approx 0.3518 \end{aligned}$$

6. 지방의 어느 중소도시에서 5%의 사람이 특이한 질병에 걸렸다고 한다. 이들 중에서 임의로 5명을 선정하였을 때, 이 질병에 걸린 사람이 2명 이하일 확률을 구하라.

Solution. 질병에 걸린 사람의 수를 확률변수 X라 하면 $X \sim \mathrm{B}\left(5, \frac{1}{20}\right)$이다. 따라서

$$\begin{aligned} P(X \leq 2) &= \sum_{i=0}^{2} \binom{5}{i} \left(\frac{1}{20}\right)^i \left(1 - \frac{1}{20}\right)^{5-i} \\ &\approx 0.9988 \end{aligned}$$

7. 수도권 지역에서 B^+ 혈액형을 가진 사람의 비율이 10%라고 한다. 이 때 헌혈센터에서 20명이 헌혈을 했을 때, 그들 중 정확히 4명이 B^+ 혈액형일 확률과 적어도 3명이 B^+일 확률을 구하라.

Solution. B^+ 혈액형을 가진 사람의 수를 확률변수 X라 하면 $X \sim \mathrm{B}\left(20, \frac{1}{10}\right)$이다. 따라서

(1)

$$\begin{aligned} P(X=4) &= \binom{20}{4}\left(\frac{1}{10}\right)^4\left(1-\frac{1}{10}\right)^{16} \\ &\approx 0.0898 \end{aligned}$$

(2)

$$\begin{aligned} P(X \geq 3) &= \sum_{i=3}^{20}\binom{20}{i}\left(\frac{1}{10}\right)^i\left(1-\frac{1}{10}\right)^{20-i} \\ &\approx 0.3231 \end{aligned}$$

14. $X \sim \mathrm{H}(N, r, n)$에 대하여 $r/N = p$로 일정하고 $N \to \infty$이면, X의 확률분포는 모수 n과 p를 갖는 이항분포에 가까워짐을 보여라.

Solution. $\mathrm{H}(N, r, n)$의 확률질량함수는

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \qquad 0 \leq x < n, x \in \mathbb{Z}$$

이다. 이 때

$$\begin{aligned} &\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{r}} \\ =&\frac{\frac{r!}{x!(r-x)!} \times \frac{(N-r)!}{(n-x)!(N-r-n+x)!}}{\frac{N!}{n!(N-n)!}} \\ =&\frac{n!}{x!(n-x)!} \times \frac{r!(N-r)!}{N!} \times \frac{(N-n)!}{(r-x)!(N-n-r+x)!} \\ =&\binom{n}{x} \times \prod_{i=1}^{x}\frac{(r-x+i)}{(N-x+i)} \times \prod_{i=1}^{n-x}\frac{(N-r-(n-x)+i)}{(N-n+i)} \end{aligned}$$

이다. 한편

$$\begin{aligned} \lim_{N\to\infty}\frac{(r-x+i)}{(N-x+i)} &= \lim_{N\to\infty}\frac{r}{N} = p, \\ \lim_{N\to\infty}\frac{(N-r-(n-x)+i)}{(N-n+i)} &= \lim_{N\to\infty}\frac{N-r}{N} = 1-p \end{aligned}$$

이므로

$$
\begin{aligned}
\lim_{N\to\infty} f(x) &= \lim_{N\to\infty} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{r}} \\
&= \lim_{N\to\infty} \binom{n}{x} \times \prod_{i=1}^{x} \frac{(r-x+i)}{(N-x+i)} \times \prod_{i=1}^{n-x} \frac{(N-r-(n-x)+i)}{(N-n+i)} \\
&= \binom{n}{x} p^x (1-p)^{n-x}
\end{aligned}
$$

인데, 이는 $\mathrm{B}(n, p)$의 확률질량함수이다. 따라서 $N \to \infty$이면 $\mathrm{H}(N, r, n) \to \mathrm{B}(n, p)$이다. □

4.4 기하분포와 음이항분포

3. 매 회 성공률이 p인 베르누이 실험을 처음 성공할 때까지 독립적으로 반복 시행한 횟수를 확률변수 X라 한다. 이 때 처음 성공이 있기 전까지 실패한 횟수 Y의 확률질량함수와 평균 그리고 분산을 구하라.

Solution. 성공 확률이 p인 베르누이 시행에 대해 x번 시행 후 첫 번째 성공을 얻을 확률은

$$P(X=x)=p(1-p)^{x-1}$$

이며, 이는 기하분포를 따른다. 따라서 $X\sim \mathrm{G}(p)$에 대해

$$\begin{aligned}\mathrm{E}(X)&=\frac{1}{p}\\ \mathrm{Var}(X)&=\frac{1-p}{p^2}\end{aligned}$$

이다. 또한 $Y=X-1$이므로

$$\begin{aligned}P(Y=y)&=p(1-p)^{y}\\ \mathrm{E}(Y)&=\frac{1-p}{p}\\ \mathrm{Var}(Y)&=\frac{1-p}{p^2}\end{aligned}$$

이다.

5. 매 회 성공률이 p인 베르누이 실험을 r번째 성공이 있기까지 독립적으로 반복 시행한 횟수를 확률변수 X라 한다. 이 때 r번째 성공이 있기까지 실패한 횟수 Y의 확률질량함수와 평균 그리고 분산을 구하라.

Solution. 성공 확률이 p인 베르누이 시행에 대해 x번 시행 후 r번째 성공을 얻을 확률은

$$P(X=x)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$$

이며, 이는 음이항분포를 따른다. 따라서 $X\sim \mathrm{NB}(r,p)$에 대해

$$\begin{aligned}\mathrm{E}(X)&=\frac{r}{p}\\ \mathrm{Var}(X)&=\frac{r(1-p)}{p^2}\end{aligned}$$

이다. 또한 $Y = X - r$이므로 **3**과 마찬가지로

$$\begin{aligned} P(Y=y) &= \binom{y+r-1}{r-1} p^r (1-p)^y \\ \mathrm{E}(Y) &= \frac{r(1-p)}{p} \\ \mathrm{Var}(Y) &= \frac{r(1-p)}{p^2} \end{aligned}$$

이다.

6. 매 회 성공률이 0.4인 베르누이 실험을 3번째 성공이 있기까지 실패한 횟수 Y의 확률질량함수와 평균 그리고 분산을 구하라.

Solution. **5**에서 보인 것과 같이

$$\begin{aligned} P(Y=y) &= \binom{y+3-1}{3-1} 0.4^3 (1-0.4)^y \\ &= \binom{y+2}{2} 0.4^3 \cdot 0.6^y \end{aligned}$$

이다. 이 때

$$\begin{aligned} \mathrm{E}(Y) &= \frac{3 \times 0.6}{0.4} = 4.5 \\ \mathrm{Var}(Y) &= \frac{3(1-0.4)}{0.4^2} = 11.25 \end{aligned}$$

이다.

11. 컴퓨터 시뮬레이션을 통하여 0에서 9까지의 숫자를 무작위로 선정하며, 각 숫자가 선정될 가능성은 동일하다고 한다. 이 때 다음을 구하라.

(1) 처음으로 숫자 0이 나올 때까지 시뮬레이션을 반복한 횟수에 대한 확률분포
(2) (1)의 확률분포에 대한 평균 μ와 분산 σ^2
(3) 시뮬레이션을 10회 실시해서야 비로소 4번째 0이 나올 확률
(4) 4번째 0을 얻기 위하여 시뮬레이션을 반복한 평균 횟수

Solution.

(1) 숫자 0이 나올 확률은 0.1이므로, 주어진 확률변수는 기하분포 $\mathrm{G}(0.1)$을 따른다.

(2) (1)의 분포에 대해

$$\mu = \frac{1}{0.1} = 10$$
$$\sigma^2 = \frac{1-0.1}{0.1^2} = 90$$

(3) 주어진 확률변수를 X라 하면, X는 음이항분포 $\mathrm{NB}(4, 0.1)$을 따른다. 이 때

$$P(X=10) = \binom{10-1}{4-1} 0.1^4 (1-0.1)^6 \approx 0.0045$$

이다.

(4) (3)의 분포에 대해

$$\mu = \frac{4}{0.1} = 40$$

이다.

13. $X \sim \mathrm{G}(p)$에 대하여 $\mathrm{E}[X(X-1)] = \dfrac{2q}{p^2}$임을 보여라.

Solution. $X \sim \mathrm{G}(p)$이면 $P(X=k) = (1-p)^{k-1} p$이다. 따라서

$$\begin{aligned}
&\mathrm{E}[X(X-1)] \\
=&\sum_{x=1}^{\infty} x(x-1) P(X=x) \\
=&\sum_{x=1}^{\infty} x(x-1)(1-p)^{x-1} p \\
=&\sum_{x=1}^{\infty} x(x-1)(1-p)^{x-2} \times p(1-p) \\
=&p(1-p) \sum_{x=1}^{\infty} \frac{d^2}{dp^2} (1-p)^x \\
=&p(1-p) \left[\frac{d^2}{dp^2} \sum_{x=1}^{\infty} (1-p)^x \right] \\
=&p(1-p) \left(\frac{d^2}{dp^2} \frac{1-p}{p} \right) \\
=&p(1-p) \frac{2}{p^3} \\
=&\frac{2(1-p)}{p^2} = \frac{2q}{p^2}
\end{aligned}$$

이다.

14. $X \sim \mathrm{G}(p)$에 대한 비기억성 성질을 증명하라.

Solution. 어떤 분포가 비기억성 또는 무기억성이라 함은 정의역의 원소 s, t에 대해

$$P(X > s+t \mid X > s) = P(X > t)$$

가 성립함을 말한다. $X \sim \mathrm{G}(p)$이면 $P(X = k) = (1-p)^{k-1}p$이므로,

$$\begin{aligned} P(X > k) &= \sum_{x=k+1}^{\infty} (1-p)^{x-1} p \\ &= p(1-p)^{k-1} \sum_{x=1}^{\infty} (1-p)^{x} \\ &= p(1-p)^{k-1} \frac{1-p}{p} \\ &= (1-p)^{k} \end{aligned}$$

이고, 따라서

$$\begin{aligned} P(X > s+t \mid X > s) &= \frac{P(X > s+t)}{P(X > s)} \\ &= \frac{(1-p)^{s+t}}{(1-p)^{s}} \\ &= (1-p)^{t} = P(X > t) \end{aligned}$$

이므로 X는 비기억성이다.

4.5 푸아송분포

2. 보험회사는 10년 기간의 보험증권을 소지한 가입자 5,000명을 가지고 있다. 이 기간 동안에 12,200개의 보험금 지급 요구가 있었고, 지급 요구 건수는 푸아송 분포를 이룬다고 한다.

(1) 연간 보험증권당 요구 건수의 평균을 구하라.
(2) 1년에 한 건 이하의 요구가 있을 확률을 구하라.
(3) 보험금 지급 요구 건당 1,000만 원의 보험금이 지급된다면, 1년에 한 보험가입자에게 지급될 평균 보험금을 구하라.

Solution. 단위기간을 1년으로 두면, 지급 요구 건수는 평균적으로 1,220건일 것이다.

(1)
$$\mathrm{E}(X) = \frac{1220}{5000} = 0.244$$
이고, 단위 기간이 1년일 때 개인당 지급 요구 건수 $X \sim \mathrm{Poi}(0.244)$이다.

(2)
$$\begin{aligned} P(X \le 1) &= P(X=0) + P(X=1) \\ &= \frac{0.244^0 \times e^{-0.244}}{0!} + \frac{0.244^1 \times e^{-0.244}}{1!} \\ &= \frac{1.244}{e^{0.244}} \approx 0.9747 \end{aligned}$$

(3) $10^7\mathrm{E}(X) = 244 \times 10^4$이므로 1년에 한 보험가입자에게 지급될 평균 보험금은 244만원이다.

3. 보험회사는 보험 가입자가 보험금 지급 요구를 네 번 신청할 확률이 두 번 신청할 확률의 3배가 되는 것을 발견했다. 보험금 지급 요구 건수가 푸아송 분포를 이룬다고 할 때, 요구 건수의 분산을 구하라.

Solution. 지급 요구 건수 $X \sim \mathrm{Poi}(\lambda)$에 대하여 $P(X=4) = 3P(X=2)$로 두면
$$\begin{aligned} &\frac{\lambda^4 e^{-\lambda}}{4!} = 3 \times \frac{\lambda^2 e^{-\lambda}}{2!} \\ \Rightarrow &\lambda^2 = \frac{3 \times 4!}{2!} = 36 \\ \Rightarrow &\lambda = 6 \end{aligned}$$
이므로, 분산 $\mathrm{Var}(X) = 6$이다.

8. 우리나라 동남부 지역에서 단위시간으로써 1년에 3번 지진이 일어난다고 한다.

(1) 앞으로 2년간 적어도 3번의 지진이 일어날 확률을 구하라.
(2) 지금부터 다음 지진이 일어날 때까지 걸리는 시간 T의 확률분포를 구하라.

Solution. 1년간 지진의 발생 횟수 $X \sim \text{Poi}(3)$이다.

(1) 2년간 지진의 발생 횟수 $Y \sim \text{Poi}(6)$이므로

$$\begin{aligned} P(Y \geq 3) &= 1 - P(Y \leq 2) \\ &= 1 - \sum_{k=0}^{2} P(Y = k) \\ &= 1 - \sum_{k=0}^{2} \frac{6^k e^{-6}}{k!} \\ &= 1 - 25e^{-6} \approx 0.9380 \end{aligned}$$

(2) $T \sim \text{Exp}(3)$이다.

9. 해저 케이블에 생긴 결함의 수는 1km당 발생 비율 $\lambda = 0.15$인 푸아송 과정에 따른다고 한다.

(1) 처음 3km에서 결함이 발견되지 않을 확률을 구하라.
(2) 처음 3km에서 결함이 발견되지 않았다고 할 때, 처음부터 3km 지점과 4km 지점에서 결함이 발견되지 않을 확률을 구하라.
(3) 처음 3km 지점 이전에 1개, 3km 지점과 4km 지점에서 1개의 결함이 발견되지만, 4km 지점과 5km 지점에서 결함이 발견되지 않을 확률을 구하라.

Solution. 1km당 결함의 수 $X \sim \text{Poi}(0.15)$이다. 또한 $x$km당 결함의 수 $X(x) \sim \text{Poi}(0.15x)$이다.

(1) 처음 3km에서 결함이 발견되지 않을 확률은 $P(X(3) = 0)$이다. $X(3) \sim \text{Poi}(0.45)$이므로

$$\begin{aligned} P(X(3) = 0) &= \frac{0.45^0 e^{-0.45}}{0!} \\ &= e^{-0.45} \approx 0.6376 \end{aligned}$$

이다.
(2) 처음부터 3km 지점과 4km 지점에서 결함이 발견되지 않는 사건은 처음 3km에서 결함이 발견되지 않는 사건과 독립이다. 따라서 3km 지점과 4km 지점의 사이 1km 구간에서 결함이 발견되지 않을 확률은 $P(X(1) = 0)$이고, (1)과 같은 방법으로 계산하면 $e^{-0.15} \approx 0.8607$이다.
(3) 처음 3km 지점 이전에 1개가 발견될 확률은 $P(X(3) = 1)$, 3km 지점과 4km 지점에서 1개의 결함이 발견될 확률은 $P(X(1) = 1)$, 4km 지점과 5km 지점에서 결함이 발견되지 않을 확률은 $P(X(1) = 0)$이므로 이 사건들이 동시에 일어날 확률은

$$\begin{aligned} &P(X(3) = 1) P(X(1) = 1) P(X(1) = 0) \\ =&\frac{0.45^1 e^{-0.45}}{1!} \times \frac{0.15^1 e^{-0.15}}{1!} \times \frac{0.15^0 e^{-0.15}}{0!} \\ =&0.45 \times 0.15 \times e^{-0.75} \approx 0.03188 \end{aligned}$$

이다.

11. 어느 상점에 찾아오는 손님은 시간당 $\lambda = 4$인 푸아송 과정에 따른다고 한다. 아침 8시에 문을 열었을 때, 8시 30분까지 꼭 한 사람이 찾아오고 11시까지 찾아온 사람이 모두 5명일 확률을 구하라.

Solution. 문제 **9**의 (3)과 같은 방법으로 생각하면, x시간당 방문하는 손님의 수 $X(x) \sim \mathrm{Poi}(4x)$에 대해 8시 30분까지 정확히 한 명이 찾아오는 확률은 $P(X(0.5)=1)$, 8시 30분부터 11시까지 4명이 찾아오는 확률은 $P(X(2.5)=4)$이므로 이 사건들이 동시에 일어날 확률은

$$\begin{aligned}
&P(X(0.5)=1)\,P(X(2.5)=4)\\
=&\frac{2^1e^{-2}}{1!}\times\frac{10^4e^{-10}}{4!}\\
=&\frac{1256}{3}e^{-12}\approx 0.002572
\end{aligned}$$

이다.

12. 이항확률변수 $X \sim \mathrm{B}(n,p)$에 대하여 $np=\mu$로 일정하고 $n\to\infty$이면, X의 확률질량함수는 모수 μ인 푸아송 확률변수의 확률질량함수에 근사함을 보여라.

Solution. $X\sim\mathrm{B}(n,p)$에서 $P(X=x)=\binom{n}{x}p^x(1-p)^{n-x}$이다. 이 때

$$\begin{aligned}
\lim_{n\to\infty}P(X=x) &= \lim_{n\to\infty}\left[\binom{n}{x}p^x(1-p)^{n-x}\right]\\
&= \lim_{n\to\infty}\left[\frac{n!}{x!\,(n-x)!}p^x(1-p)^{n-x}\right]\\
&= \frac{1}{x!}\lim_{n\to\infty}\left[\frac{n!}{(n-x)!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}\right]\\
&= \frac{\mu^x}{x!}\lim_{n\to\infty}\left[\left(\prod_{k=n-x+1}^{n}k\right)\frac{1}{n^x}\left(1-\frac{\mu}{n}\right)^{n-x}\right]\\
&= \frac{\mu^x}{x!}\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}\\
&= \frac{\mu^x}{x!}\lim_{n\to\infty}\left[\left(1+\frac{1}{n(-\mu^{-1})}\right)^{n(-\mu^{-1})}\right]^{-\mu}\\
&= \frac{\mu^x e^{-\mu}}{x!}
\end{aligned}$$

인데, 이는 $\mathrm{Poi}(\mu)$의 확률질량함수이다. 따라서 $n\to\infty$이면 $\mathrm{B}(n,p)\sim\mathrm{Poi}(\mu)$이다. □

13. $X\sim\mathrm{P}(\mu)$에 대하여 $\mathrm{E}[X(X-1)]=\mu^2$임을 보여라.

Solution. $X \sim \mathrm{Poi}(\mu)$이면 $P(X=x) = \dfrac{\mu^x e^{-\mu}}{x!}$이다. 따라서

$$\begin{aligned}
\mathrm{E}[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)\,P(X=x) \\
&= \sum_{x=0}^{\infty} x(x-1)\,\frac{\mu^x e^{-\mu}}{x!} \\
&= \sum_{x=2}^{\infty} \frac{\mu^x e^{-\mu}}{(x-2)!} \\
&= \mu^2 \sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!}
\end{aligned}$$

인데, $\dfrac{\mu^x e^{-\mu}}{x!}$은 $X \sim \mathrm{Poi}(\mu)$의 확률질량함수이다. 따라서

$$\sum_{x=0}^{\infty} \frac{\mu^x e^{-\mu}}{x!} = 1$$

이므로 $\mathrm{E}[X(X-1)] = \mu^2$이다.

5 연속확률분포

5.1 균등분포

1. $X \sim \mathrm{U}(-1, 1)$에 대하여 다음을 구하라.

(1) 확률밀도함수와 분포함수
(2) 평균과 분산
(3) $P(\mu - \sigma < X < \mu + \sigma)$
(4) 사분위수

Solution.

(1) 확률밀도함수 f는 다음과 같다.

$$f(x) = \frac{1}{2} \qquad -1 \le x \le 1$$

분포함수 F는 다음과 같다.

$$F(x) = \frac{x+1}{2} \qquad -1 \le x \le 1$$

(2)

$$\mu = \frac{1+(-1)}{2} = 0$$
$$\sigma^2 = \frac{(1-(-1))^2}{12} = \frac{1}{3}$$

(3)

$$\begin{aligned} &P(\mu - \sigma < X < \mu + \sigma) \\ &= P\left(-\sqrt{\frac{1}{3}} < X < \sqrt{\frac{1}{3}}\right) \\ &= \frac{1}{2} \times 2\sqrt{\frac{1}{3}} \\ &\approx 0.5773 \end{aligned}$$

(4) $Q_1 = -0.5, Q_2 = 0, Q_3 = 0.5$

3. 특수직에 근무하는 사람들 중에서 임의로 선정한 사람의 출생에서 사망에 이르기까지 걸리는 시간을 X라 하자. 그러면 이 사람의 생존 시간 X는 사고나 질병에 의하지 않는다면 $[0, 65.5]$에서 균등분포를 이룬다고 한다.

(1) X의 밀도함수와 분포함수를 구하라.
(2) X의 평균을 구하라.
(3) 임의로 선정된 사람이 60세 이상 생존할 확률을 구하라.
(4) 임의로 선정된 사람이 45세 이상 살았다는 조건 아래서, 60세 이상 생존할 확률을 구하라.
(5) (4)의 조건 아래서, x $(45 < x \leq 65.5)$세 이상 생존할 확률을 구하라.

Solution.

(1) 확률밀도함수 f는 다음과 같다.

$$f(x) = \frac{1}{65.5} \qquad 0 \leq x \leq 65.5$$

분포함수 F는 다음과 같다.

$$F(x) = \frac{x}{65.5} \qquad 0 \leq x \leq 65.5$$

(2)

$$E(X) = \frac{0 + 65.5}{2} = 32.75$$

(3)

$$\begin{aligned} P(X \geq 60) &= 1 - F(60) \\ &= 1 - \frac{60}{65.5} \approx 0.0840 \end{aligned}$$

(4)

$$\begin{aligned} P(X \geq 60 \,|\, X \geq 45) &= \frac{1 - F(60)}{1 - F(45)} \\ &= \frac{1 - \frac{60}{65.5}}{1 - \frac{45}{65.5}} \\ &\approx 0.2683 \end{aligned}$$

(5)

$$\begin{aligned} P(X \geq x \,|\, X \geq 45) &= \frac{1 - F(x)}{1 - F(45)} \\ &= \frac{1 - \frac{x}{65.5}}{1 - \frac{45}{65.5}} \\ &= \frac{65.5 - x}{20.5} = \frac{131 - 2x}{41} \qquad 45 < x \leq 65.5 \end{aligned}$$

5.2 지수분포와 감마분포

3. 약속장소에서 친구를 만나기로 하고, 정시에 도착하였으나 친구가 아직 나오지 않았다. 그리고 친구를 만나기 위하여 기다리는 시간은 $\lambda = 0.2$인 지수분포에 따른다고 한다.

(1) 친구를 만나기 위한 평균 시간을 구하라.
(2) 3분이 경과하기 이전에 친구를 만날 확률을 구하라.
(3) 10분 이상 기다려야 할 확률을 구하라.
(4) 6분이 경과했다고 할 때, 추가적으로 더 기다려야 할 시간에 대한 확률분포를 구하고, 모두 합쳐서 10분 이상 걸릴 확률을 구하라.

Solution. 기다리는 시간을 X라고 하자. $X \sim \mathrm{Exp}(0.2)$이다. 따라서 분포함수 $F(x) = 1 - e^{-0.2x}$이다.

(1) $\mu = \dfrac{1}{0.2} = 5$
(2)

$$\begin{aligned} P(X \leq 3) &= F(3) \\ &= 1 - e^{-0.2 \cdot 3} \approx 0.4512 \end{aligned}$$

(3)

$$\begin{aligned} P(X \geq 10) &= 1 - F(10) \\ &= e^{-0.2 \cdot 10} \approx 0.1353 \end{aligned}$$

(4) 지수분포는 무기억성을 가지므로, 추가적으로 기다려야 할 시간은 여전히 $\mathrm{Exp}(0.2)$를 따른다. 따라서 모두 합쳐 10분 이상 기다려야 할 확률은 4분 이상 더 기다리는 확률과 같으므로,

$$\begin{aligned} P(X \geq 4) &= 1 - F(4) \\ &= e^{-0.2 \cdot 4} \approx 0.4493 \end{aligned}$$

5. 어떤 기계의 고장나는 날의 간격이 $\lambda = 0.3$인 지수분포에 따른다고 한다.

(1) 이 기계가 고장나는 날의 평균 간격을 구하라.
(2) 고장나는 날의 간격에 대한 표준편차를 구하라.
(3) 고장나는 날의 간격에 대한 중앙값을 구하라.
(4) 이 기계가 수리된 후 다시 고장나기까지 적어도 일주일 이상 사용할 수 있을 확률을 구하라.
(5) 이 기계를 5일 동안 정상적으로 사용했을 때, 고장나기까지 적어도 이틀 이상 사용할 수 있을 확률을 구하라.

Solution. 고장나는 날의 간격을 $X \sim \mathrm{Exp}(0.3)$라고 하자.

(1)

$$\mu = \frac{1}{0.3} = \frac{10}{3}$$

(2)

$$\begin{aligned}\sigma^2 &= \left(\frac{10}{3}\right)^2 \\ \Rightarrow \sigma &= \frac{10}{3}\end{aligned}$$

(3) X의 분포함수는 $F(x) = 1 - e^{-0.3x}$이다. 중앙값은 $F^{-1}(0.5)$이므로, $F(m) = 0.5$가 되는 m을 구하면

$$\begin{aligned}&F(m) = 1 - e^{-0.3m} = 0.5 \\ &\Rightarrow e^{-0.3m} = 0.5 \\ &\Rightarrow -0.3m = -\ln 2 \\ &\Rightarrow m = \frac{\ln 2}{0.3} \approx 2.310\end{aligned}$$

이다.

(4)

$$\begin{aligned}P(X > 7) &= 1 - F(7) \\ &= 1 - \left(1 - e^{-0.3\cdot 7}\right) \\ &= e^{-2.1} \approx 0.1225\end{aligned}$$

(5) 지수분포는 무기억성을 가지므로, 구하는 확률은

$$\begin{aligned}P(X > 2) &= 1 - F(2) \\ &= 1 - \left(1 - e^{-0.3\cdot 2}\right) \\ &= e^{-0.6} \approx 0.5488\end{aligned}$$

이다.

8. 어느 상점에 매 시간당 평균 30명의 손님이 푸아송 과정을 따라서 찾아온다고 한다.

(1) 상점 주인이 처음 두 손님을 맞이하기 위하여 5분 이상 기다릴 확률을 구하라.
(2) 처음 두 손님을 맞이하기 위하여 3분에서 5분 정도 기다릴 확률을 구하라.

Solution. 처음 두 손님을 맞이하기까지 걸리는 시간을 X라 하면, $X \sim \Gamma(2,2)$이다. 이 때 확률밀도함수 $f(x) = \frac{x}{4}e^{-\frac{x}{2}}$이다.

(1)

$$\begin{aligned} P(X \geq 5) &= \int_5^\infty \frac{x}{4}e^{-\frac{x}{2}}\,dx \\ &= -\frac{1}{2}\left[(x+2)\,e^{-\frac{x}{2}}\right]_5^\infty \\ &= \frac{7}{2}e^{-\frac{5}{2}} \approx 0.2873 \end{aligned}$$

(2)

$$\begin{aligned} P(3 < X < 5) &= \int_3^5 \frac{x}{4}e^{-\frac{x}{2}}\,dx \\ &= -\frac{1}{2}\left[(x+2)\,e^{-\frac{x}{2}}\right]_3^5 \\ &= \frac{5}{2}e^{-\frac{3}{2}} - \frac{7}{2}e^{-\frac{5}{2}} \approx 0.2705 \end{aligned}$$

9. 전화 교환대에 1분당 평균 2번의 비율로 신호가 들어오고 있으며, 교환대에 도착한 신호는 푸아송 과정을 따른다고 한다.

(1) 푸아송 과정의 비율 λ를 구하라.
(2) 교환대에 들어오는 두 신호 사이의 평균 시간을 구하라.
(3) 2분과 3분 사이에 신호가 없을 확률을 구하라.
(4) 교환원이 교환대에 앉아서 3분 이상 기다려야 첫 번째 신호가 들어올 확률을 구하라.
(5) 처음 2분 동안 신호가 없으나 2분과 4분 사이에 4건의 신호가 있을 확률을 구하라.
(6) 처음 신호가 15초 이내 들어오고, 그 이후 두 번째 신호가 들어오기까지 3분 이상 걸릴 확률을 구하라.

Solution. 1분간 들어오는 신호의 수를 X라고 하자. $X \sim \mathrm{Poi}(\lambda)$이다.

(1) 1분당 평균 2번의 비율로 신호가 들어오므로, $\lambda = 2$이다.
(2) 두 신호 사이의 간격 T는 $\mathrm{Exp}(2)$를 따르므로, 이 때 $\mu = 0.5$이다. 또한 분포함수 $(x) = 1 - e^{-2x}$이다.
(3) 2분과 3분 사이에 신호가 없을 확률은 1분간 신호가 없었을 확률과 같으므로,

$$P(T > 1) = 1 - F(1) = e^{-2} \approx 0.1353$$

이다.

(4)

$$P(T > 3) = 1 - F(3) = e^{-6} \approx 0.002479$$

(5) $Y \sim \mathrm{Poi}(4)$에 대해 처음 2분 동안 신호가 없을 확률은 $P(Y=0)$, 2분과 4분 사이에 4건의 신호가 있을 확률은 $P(Y=4)$이므로 동시에 일어날 확률은

$$\begin{aligned} P(Y=0)P(Y=4) &= \frac{4^0 e^{-4}}{0!} \times \frac{4^4 e^{-4}}{4!} \\ &= \frac{4^4 e^{-8}}{4!} \approx 0.003528 \end{aligned}$$

이다.

(6)

$$\begin{aligned} P(T < 0.25)P(T > 3) &= \left(1 - e^{-0.5}\right)\left[1 - \left(1 - e^{-6}\right)\right] \\ &= e^{-6}\left(1 - e^{-0.5}\right) \\ &\approx 0.0009753 \end{aligned}$$

16. 가격이 200(천원)인 프린터의 수명은 평균 2년인 지수분포를 이룬다고 한다. 구매한 날로 1년 안에 프린터가 고장이 나면 제조업자는 구매자에게 전액을 환불하고, 2년 안에 고장이 나면 반액을 환불할 것을 약속하였다. 만일 제조업자가 100대를 판매하였다면, 환불로 인하여 지불해야 할 평균 금액을 구하라.

Solution. 프린터의 수명을 $X \sim \mathrm{Exp}\left(\frac{1}{2}\right)$라고 하면, 누적분포함수 $F(x) = 1 - e^{-\frac{1}{2}x}$이다. 이 때 프린터가 1년 이내에 고장날 확률 $P(X \le 1) = 1 - e^{-\frac{1}{2}}$이고, 1년과 2년 사이에 고장날 확률 $P(1 < X \le 2) = P(X \le 2) - P(X \le 1) = e^{-\frac{1}{2}} - e^{-1}$이므로 프린터 한 대에 대한 평균 환불 금액은

$$\begin{aligned} &200P(X \le 1) + 100P(1 < X \le 2) \\ =&200\left(1 - e^{-\frac{1}{2}}\right) + 100\left(e^{-\frac{1}{2}} - e^{-1}\right) \\ =&200 - 100e^{-\frac{1}{2}} - 100e^{-1} \approx 102.6 \text{ (천원)} \end{aligned}$$

이고, 100대에 대해서는

$$100\left(200 - 100e^{-\frac{1}{2}} - 100e^{-1}\right) \approx 10256 \text{ (천원)}$$

이 된다. 따라서 평균적으로 10,255,899원을 환불해야 한다.

19. 임의의 양수 a, b에 대하여 모수 λ인 지수분포를 갖는 확률변수 X는 다음 성질을 가짐을 보여라.

$$P(X > a + b | X > a) = P(X > b)$$

Solution. 지수분포의 누적분포함수 $F(x) = 1 - e^{-\lambda x}$이다. 따라서

$$\begin{aligned} P(X > a+b|X > a) &= \frac{P(X > a+b)}{P(X > a+b|X > a)} \\ &= \frac{1-F(a+b)}{1-F(a)} \\ &= \frac{1-\left(1-e^{-\lambda(a+b)}\right)}{1-\left(1-e^{-\lambda a}\right)} \\ &= \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} \\ &= e^{-b} = P(X > b) \end{aligned}$$

이다. □

19. 임의의 양수 α에 대하여 $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$이 성립한다. 이것을 이용하여, $X \sim \Gamma(\alpha, \beta)$에 대한 평균과 분산은 각각 $\mu = \alpha\beta$, $\sigma^2 = \alpha\beta^2$임을 보여라.

Solution. $X \sim \Gamma(\alpha, \beta)$의 확률밀도함수 $f(x) = \dfrac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$이다.

$$\begin{aligned} \mu &= \int_0^\infty x \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^\alpha e^{-\frac{x}{\beta}}\, dx \\ &= \frac{\Gamma(\alpha+1)\beta}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^\alpha e^{-\frac{x}{\beta}}\, dx \end{aligned}$$

에서 $\dfrac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^\alpha e^{-\frac{x}{\beta}}$은 $\Gamma(\alpha+1, \beta)$의 확률밀도함수이므로 $\displaystyle\int_0^\infty \frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}} x^\alpha e^{-\frac{x}{\beta}}\, dx = 1$이고, 감마함수의 정의에 의해 $\dfrac{\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha$이므로 $\mu = \alpha\beta$이다.

분산을 구하기 위해 $\mathrm{E}\left(X^2\right)$를 계산하면

$$\begin{aligned} \mathrm{E}\left(X^2\right) &= \int_0^\infty x^2 \cdot \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}\, dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha+1} e^{-\frac{x}{\beta}}\, dx \\ &= \frac{\Gamma(\alpha+2)\beta^2}{\Gamma(\alpha)} \int_0^\infty \frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}} x^{\alpha+1} e^{-\frac{x}{\beta}}\, dx \end{aligned}$$

에서, $\frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}}x^{\alpha+1}e^{-\frac{x}{\beta}}$은 $\Gamma(\alpha+2,\beta)$의 확률밀도함수이므로 $\int_0^\infty \frac{1}{\Gamma(\alpha+2)\beta^{\alpha+2}}x^{\alpha+1}e^{-\frac{x}{\beta}}\,dx = 1$이다. 따라서

$$\begin{aligned}
\mathrm{E}\left(X^2\right) &= \frac{\Gamma(\alpha+2)\beta^2}{\Gamma(\alpha)} \times 1 \\
&= \frac{(\alpha+1)\Gamma(\alpha+1)\beta^2}{\Gamma(\alpha)} \\
&= \frac{\alpha(\alpha+1)\Gamma(\alpha)\beta^2}{\Gamma(\alpha)} = \alpha(\alpha+1)\beta^2
\end{aligned}$$

이다. 분산 $\sigma^2 = \mathrm{E}\left(X^2\right) - [\mathrm{E}(X)]^2$이므로

$$\begin{aligned}
\sigma^2 &= \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 \\
&= \alpha\beta^2
\end{aligned}$$

이다. □

5.3 정규분포

3. $X \sim \mathrm{N}(77, 16)$에 대하여 다음을 구하라.

(1) X가 평균을 중심으로 10% 안에 있지 않은 확률
(2) X의 분포에 대한 25-백분위수
(3) X의 분포에 대한 75-백분위수
(4) $P(\mu - x_0 < X < \mu + x_0) = 0.95$인 x_0

Solution. X를 표준화시키면 $Z = \dfrac{X-77}{\sqrt{16}} = \dfrac{X-77}{4}$이다.

(1) X가 평균을 중심으로 10% 안에 있을 확률을 구하면

$$\begin{aligned} P(|X-\mu| \le 0.1\mu) &= P(-0.1\mu \le X-\mu \le 0.1\mu) \\ &= P(69.3 \le X-\mu \le 84.7) \\ &= P\left(\frac{69.3-77}{4} \le \frac{X-77}{4} \le \frac{84.7-77}{4}\right) \\ &= P(-1.925 \le Z \le 1.925) \\ &\approx 0.9729 \end{aligned}$$

이므로, 구하고자 하는 확률은 약 $1-0.9729 = 0.02712$이다.

(2) 25-백분위수 p_{25}에 대해 $P(X \le p_{25}) = 0.25$를 만족하므로

$$\begin{aligned} P(X \le p_{25}) &= P\left(\frac{X-77}{4} \le \frac{p_{25}-77}{4}\right) \\ &= P\left(Z \le \frac{p_{25}-77}{4}\right) = 0.25 \end{aligned}$$

이다. 표준정규분포표를 이용해 $\dfrac{p_{25}-77}{4} \approx -0.675$를 얻을 수 있으며, 이 때 p_{25}는 약 74.3이다.

(3) (2)와 같은 방법으로 구하면 p_{75}는 약 79.7이다.

(4)

$$\begin{aligned} P(\mu - x_0 < X < \mu + x_0) &= P\left(\frac{-x_0}{4} < \frac{X-77}{4} < \frac{x_0}{4}\right) \\ &= P\left(-\frac{x_0}{4} < Z < \frac{x_0}{4}\right) \end{aligned}$$

이므로, $\dfrac{x_0}{4} \approx 1.96$을 얻는다. 따라서 $x_0 \approx 7.84$이다.

4. $X \sim \mathrm{N}(4, 9)$에 대하여 다음을 구하라.

(1) $P(X < 7)$
(2) $P(X \le x_0) = 0.9750$인 x_0
(3) $P(1 < X < x_0) = 0.756$인 x_0

Solution. X를 표준화시키면 $Z = \dfrac{X-4}{\sqrt{9}} = \dfrac{X-4}{3}$이다.

(1)

$$\begin{aligned} P(X<7) &= P\left(Z < \frac{7-4}{3}\right) \\ &= P(Z<1) &&\approx 0.8413 \end{aligned}$$

(2) 표준정규분포표에서 $P(Z<1.96) \approx 0.9750$이다. 따라서

$$\begin{aligned} \frac{x_0-4}{3} &\approx 1.96 \\ x_0 &\approx 9.88 \end{aligned}$$

이다.

(3) $P(1<X<x_0) = 0.95$이면 $P(X<x_0) - P(X<1) = 0.95$인데, $P(X<1)$은 $P(Z<-1) \approx 0.1587$이므로 $P(X<x_0) \approx 0.9147$이다. $P(Z<1.37) \approx 0.9147$이므로 $\dfrac{x_0-4}{3} \approx 1.37$이고, $x_0 \approx 8.11$이다.

10. 대구에서 서울까지 기차로 걸리는 시간은 $X \sim \mathrm{N}(3.5, 0.4)$인 정규분포를 이루고, 고속버스로 걸리는 시간은 $Y \sim \mathrm{N}(3.8, 0.9)$인 정규분포를 이룬다고 한다.

(1) $P(X \le 3.2)$일 확률을 구하라.
(2) $P(Y \le 3.2)$일 확률을 구하라.
(3) $X - Y$의 확률분포를 구하라.
(4) $P(|X-Y| \le 0.1)$일 확률을 구하라.

Solution. $\dfrac{X-3.5}{\sqrt{0.4}} = Z$이고, $\dfrac{Y-3.8}{\sqrt{0.9}} = Z$이다.

(1)

$$\begin{aligned} P(X \le 3.2) &= P\left(Z \le \frac{3.2-3.5}{\sqrt{0.4}}\right) \\ &\approx P(Z \le -0.4743) \\ &\approx 0.3176 \end{aligned}$$

(2)

$$\begin{aligned} P(Y \le 3.2) &= P\left(Z \le \frac{3.2-3.8}{\sqrt{0.9}}\right) \\ &\approx P(Z \le -0.6325) \\ &\approx 0.2635 \end{aligned}$$

(3) $X-Y \sim \mathrm{N}(-0.3,\, 1.3)$
(4) $\dfrac{X-Y+0.3}{\sqrt{1.3}} = Z$이다.

$$\begin{aligned} P(|X-Y| \le 0.1) &= P\left(\frac{0.2}{\sqrt{1.3}} \le Z \le \frac{0.4}{\sqrt{1.3}}\right) \\ &\approx P(0.1754 \le Z \le 0.3508) \\ &\approx 0.06752 \end{aligned}$$

16. 컴퓨터를 이용하여 얻은 [0, 1]에서 균등분포를 이루는 100개의 독립인 확률변수들 $X_1, X_2, \cdots, X_{100}$에 대하여 다음을 구하여라.

(1) $S = X_1, X_2, \cdots, X_{100}$이 45와 55 사이일 근사확률
(2) 표본평균 $\overline{X}$가 0.55 이상일 확률

Solution. $X_i \sim \mathrm{U}(0,1)$이다. 따라서 $\mu = \dfrac{1}{2}$, $\sigma^2 = \dfrac{1}{12}$이다.

(1) 중심극한정리에 의하면

$$X = \sum_{i=1}^{100} X_i \approx \mathrm{N}\left(100\mu,\, 100\sigma^2\right)$$

이므로, X의 분포는 $\mathrm{N}\left(50,\, \dfrac{25}{3}\right)$로 근사되며 $\dfrac{X-50}{\sqrt{\dfrac{25}{3}}} = \dfrac{\sqrt{3}(X-50)}{5} \approx Z$이다. 이 때 45와 55 사이일 근사확률은

$$\begin{aligned} P(45 \le X \le 55) &\approx P\left(-\sqrt{3} \le Z \le \sqrt{3}\right) \\ &\approx P(-1.732 \le Z \le 1.732) \\ &\approx 0.9167 \end{aligned}$$

이다.

(2) $\overline{X} \approx \mathrm{N}\left(\mu, \dfrac{\sigma^2}{n}\right) = \mathrm{N}\left(\dfrac{1}{2}, \dfrac{1}{1200}\right)$이다. 이 때 $\dfrac{\overline{X}-\dfrac{1}{2}}{\sqrt{\dfrac{1}{1200}}} = 10\sqrt{3}\left(2\overline{X}-1\right) \approx Z$이다. 따라서

$$\begin{aligned} P\left(\overline{X} \ge 0.55\right) &\approx P\left(Z \ge 10\sqrt{3}(0.10)\right) \\ &\approx P(Z \ge 1.732) \\ &\approx 0.04163 \end{aligned}$$

이다.

19. $X \sim \mathrm{B}(16, 0.5)$일 때, 연속성을 수정한 근사확률 $P(8 \le X \le 10)$과 $P(X \ge 10)$을 구하라.

Solution. $\mu = np = 8$, $\sigma^2 = np(1-p) = 4$이다. 따라서 $\mathrm{B}(16, 0.5) \approx \mathrm{N}(8, 4)$이고, $\dfrac{X-8}{2} \approx Z$이다. 연속성을 수정하면

$$P(8 \le X \le 10) \approx P(-0.25 \le Z \le 1.25) \approx 0.4931$$
$$P(X \ge 10) \approx P(Z \ge 0.75) \approx 0.2266$$

이다.

21. Society of Actuaries(SOA)의 확률 시험문제는 5지 선다형으로 제시된다. 100문항의 SOA 시험문제 중에서 지문을 임의로 선택할 때, 다음을 구하라.

(1) 평균적으로 정답을 선택한 문항 수
(2) 정답을 정확히 8개 선택할 근사확률

Solution. 정답을 선택할 확률은 $\dfrac{1}{5}$이므로 정답의 개수 $X \sim \mathrm{B}\left(100, \dfrac{1}{5}\right)$이다.

(1) $\mu = 20$
(2) $\sigma^2 = 20 \times \dfrac{1}{5} \times \dfrac{4}{5} = 16$이므로 X를 정규분포에 근사하면 $\mathrm{N}(20, 16)$을 얻을 수 있다. 이 때 $\dfrac{X-20}{4} \approx Z$이므로, 따라서

$$\begin{aligned} P(X = 8) &= P(7.5 \le X \le 8.5) \\ &\approx P(-3.125 \le Z \le -2.875) \\ &\approx 0.001131 \end{aligned}$$

이다.

5.4 정규분포에 관련된 연속분포들

1. $X \sim \chi^2(12)$에 대하여, 다음을 구하라.

(1) $\mu = \mathrm{E}(X)$
(2) $\sigma^2 = \mathrm{Var}(X)$
(3) $P(X > 5.23)$
(4) $P(X < 21.03)$
(5) $\chi^2_{0.995}(12)$
(6) $\chi^2_{0.005}(12)$

Solution. χ^2-분포표로부터 값들을 얻을 수 있다.

(1) $\mu = 12$
(2) $\sigma^2 = 2 \times 12 = 24$
(3) $P(X > 5.23) \approx 0.95$
(4) $P(X > 21.05) \approx 0.05$이므로 $P(X < 21.05) \approx 0.95$
(5) 3.07
(6) 28.30

Solution.

2. $X \sim \chi^2(10)$에 대하여, $P(X < a) = 0.05$, $P(a < X < b) = 0.90$을 만족하는 상수 a와 b를 구하라.

Solution. $P(X < a) = 0.05$라면 $P(X > a) = 0.95$이고, $P(a < X < b) = 0.90$이라면 $P(X > b) = 0.05$이다. 따라서 χ^2-분포표로부터 $a \approx 3.94$, $b \approx 18.31$을 얻을 수 있다.

3. t-분포표를 이용하여 $T \sim \mathrm{t}(12)$일 때, 다음을 구하라.

(1) $t_{0.1}(12)$
(2) $t_{0.01}(12)$
(3) $P(T \le t_0) = 0.995$을 만족하는 t_0

Solution.

(1) $P(T > t_{0.1}(12)) = 0.1$에서 $t_{0.1}(12) \approx 1.356$
(2) $P(T > t_{0.01}(12)) = 0.01$에서 $t_{0.01}(12) \approx 2.681$
(3) $P(T \le t_0) = 0.995$이라면 $P(T > t_0) = 0.005$이므로 이 때의 $t_0 \approx 3.055$

4. F-분포표를 이용하여 $F \sim \mathrm{F}(8,6)$일 때, 다음을 구하라.

(1) $f_{0.01}(8,6)$
(2) $f_{0.05}(8,6)$
(3) $f_{0.90}(8,6)$
(4) $f_{0.99}(8,6)$

Solution.

(1) $P(F > f_{0.01}(8,6)) = 0.01$에서 $f_{0.01}(8,6) \approx 6.37$
(2) $P(F > f_{0.05}(8,6)) = 0.05$에서 $f_{0.05}(8,6) \approx 3.58$
(3) $f_{0.90}(8,6) = [f_{0.10}(8,6)]^{-1} \approx 0.34$
(4) $f_{0.99}(8,6) = [f_{0.01}(8,6)]^{-1} \approx 0.12$

7. 어느 기업의 주식을 10,000원에 구입하였고, 이 주식의 가치는 연간 10%의 연속적인 성장을 한다고 한다. 그리고 이 주식의 성장비율 Y는 $\mu_Y = 0.1$, $\sigma_Y^2 = 0.04$인 정규분포에 따른다고 한다.

(1) 6개월 후, 이 주식의 가치를 구하라.
(2) 1년 후의 주식 가격 $X = 10000e^Y$의 평균과 분산을 구하라.
(3) 1년 후 주식 가격이 11,750원 이상 12,250원 이하일 확률을 구하라.

Solution.

(1) $V(0) = 10000$이고 $r = 0.1$에서

$$V(0.5) = 10000e^{0.1\times 0.5} \approx 10510\ (\text{원})$$

을 얻는다.

(2) $\mu_Y = 0.1$, $\sigma_Y^2 = 0.04$에서 X의 평균 μ_X는

$$\begin{aligned}\mu_X &= \mathrm{E}\left(10000e^Y\right) \\ &= 10000e^{0.1+\frac{0.04}{2}} \\ &\approx 11270\ (\text{원})\end{aligned}$$

이고, 분산 σ_X^2는

$$\begin{aligned}\sigma_X^2 &= \mathrm{Var}\left(10000e^Y\right) \\ &= 10000^2\left(e^{0.04}-1\right)e^{0.2+0.04} \\ &\approx 51680000\ (\text{원})\end{aligned}$$

이다.

(3) 1년 후 주식 가격이 11,750원 이상 12,250원 이하일 확률 $P(11750 \le X \le 12250)$은

$$\begin{aligned}
P(11750 \le X \le 12250) &= P\left(11750 \le 10000e^{Y} \le 12250\right) \\
&= P\left(1.175 \le e^{Y} \le 1.225\right) \\
&= P(\ln 1.175 \le Y \le \ln 1.225) \\
&\approx P(0.1613 \le Y \le 0.2029) \\
&= P(0.3065 \le Z \le 0.5145) \\
&\approx 0.07616
\end{aligned}$$

이다.