경영통계

원광대학교 경영학부

담당교수: 정호일

제7장 확률분포 II

연속확률분포

-균등분포

-정규분포

-이항분포의 정규근사법

-지수분포

I. 균등분포

균등분포의 정의

확률변수 X가 두 점 a와 b사이에 있는 어떤 값을 취할 수 있고 이러한 값들의 발생확률이 모두 일정할 때 확률변수 X는 균등분포를 따른다고 한다

균등확률밀도함수

$$f(X) = \frac{1}{b-a} \quad (a \le X \le b)$$

$$= 0 \qquad (\text{다른 값을 가지는 경우})$$

I. 균등분포

균등분포의 기댓값과 분산

기대값 : $E(x) = \mu = \dfrac{a+b}{2}$

분산 : $Var(x) = \sigma^2 = \dfrac{(b-a)^2}{12}$

균등분포 예제

예1)제주도의 1월중 낮 최고기온은 0도~6도 사이에서 균등분포를 한다고 하자.

1) 3.5도 이상의 온도를 갖는 날의 비율은 얼마인가?

<풀이> $P(0 \le X \le 3.5) = \dfrac{6-3.5}{6} = 0.417$

2) 1월 중 어느 날의 온도가 1도를 넘지 않을 확률은?

<풀이> 1/6

I. 균등분포

균등분포 예제

예2) 한국제철에서는 여러가지 두께의 강판을 생산하고 있다. 강판의 두께를 150mm와 200mm 사이의 값을 갖는 균등확률변수라고 할 때

1) 균등확률변수의 기대값과 표준편차를 구하라.

<풀이>

$$E(X) = \frac{200+150}{2} = 175$$

$$\sigma = \frac{200-150}{\sqrt{12}} = 14.43$$

2) 두께가 160mm 미만인 철판을 폐기한다고 할 때 폐기물의 비율을 구하라,

<풀이>

$$\frac{160-150}{200-150} = \frac{1}{5}$$

3) 균등확률분포를 구하라.

<풀이>

$$f(X) = \frac{1}{b-a} = \frac{1}{50}$$

Ⅱ. 정규분포

정규분포의 정의

정규분포는 확률변수 X가 음의 무한대에서 양의 무한대사이의 어떠한 값도 취할 수 있는 연속확률분포로서 μ(평균)와 σ(표준편차) 2개의 모수에 의해 특징지어지는 분포 .
한편 확률변수 X가 평균 μ, 분산 σ^2정규분포를 따른다면 간단히 $X \sim N(\mu, \sigma^2)$라고 일반적으로 표기한다.

정규확률밀도함수

•

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$$

$\pi = 3.1416$

$e = 2.71828\ ..$

μ : 분포의 평균 $(-\infty < \mu < +\infty)$

σ : 분포의 표준편차 $(\sigma > 0)$

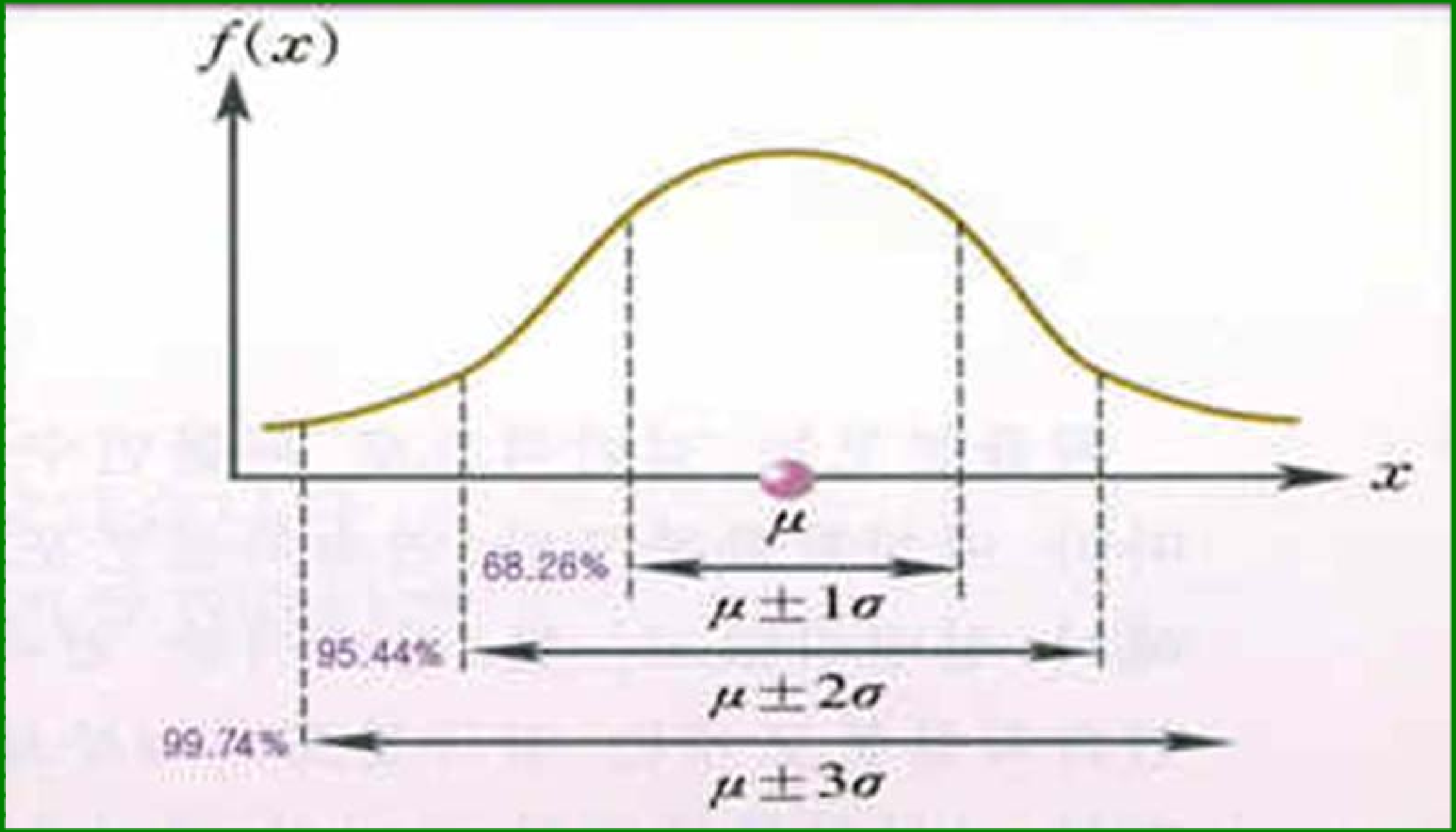
Ⅱ. 정규분포

정규분포의 특징

- 정규곡선은 종모양을 나타낸다.
- 정규분포는 평균을 중심으로 좌우대칭을 이룬다. 따라서 평균=중앙치=최빈치이다.
- 정규분포의 형태와 위치는 평균과 표준편차가 결정한다.
- 분포의 최고 높이는 $\frac{0.39894}{\sigma}$이며 표준편차에 따라 값이 달라진다.
- 정규곡선은 x축에 닿지 않으므로 확률변수 X의 범위는 $-\infty < x < +\infty$이다.
- 정규곡선 밑의 면적은 1이다.
- 정규곡선 밑의 두 점 사이의 면적은 정규확률변수가 이들 두 점 사이를 취할 확률이다.
- $-\infty$와 $+\infty$의 범위 사이에는 무수한 값이 있기 때문에 정규확률변수가 어떤 특정한 값을 취할 확률은 0이다.

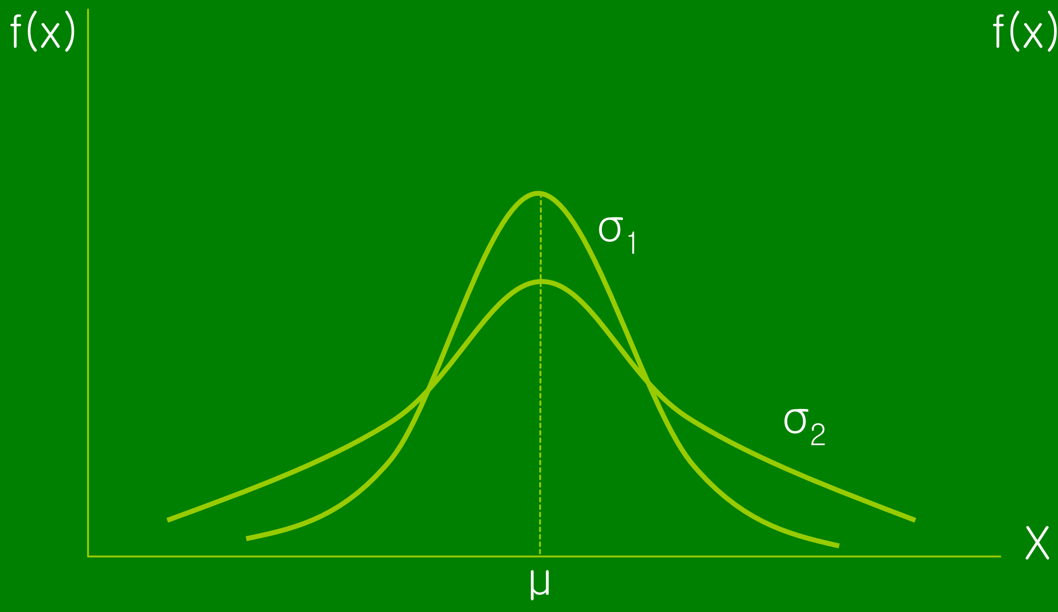
II. 정규분포

정규분포의 형태

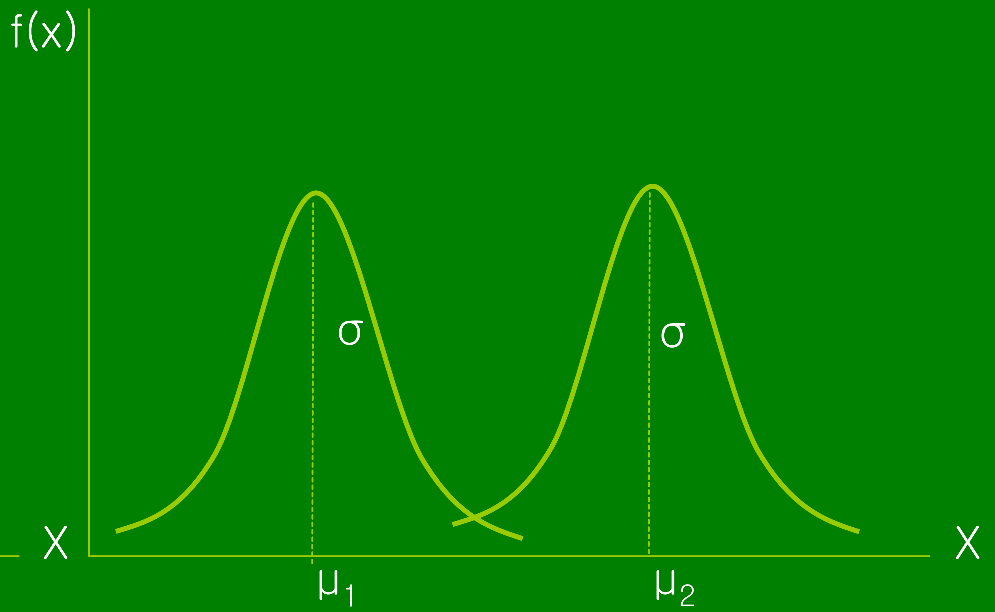


Ⅱ. 정규분포

μ와 σ에 따른 정규분포의 형태



평균이 동일하고 표준편차가 다른 경우



평균이 다르고 표준편차가 동일한 경우

Ⅱ. 정규분포

정규분포의 기댓값과 분산

$$E(X) = \int_{-\infty}^{\infty} X \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{X-\mu}{\sigma}\right]^2} dX = \mu$$

$$Var(X) = \int_{-\infty}^{\infty} X^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left[\frac{X-\mu}{\sigma}\right]^2} dX - \mu^2 = \sigma^2$$

Ⅱ. 정규분포

표준정규분포

정규분포를 표준화한 특수한 형태의 정규분포를 말하며 이는 확률변수 X를 Z값으로 표준화한 분포로서 평균(μ)=0이고 표준편차(σ)가 1인 정규분포를 따를 때 이를 Z분포라고 하며 Z~N(0,1)로 표기한다.

$$Z = \frac{X - \mu}{\sigma}$$

표준정규확률밀도함수

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[Z]^2} \quad (-\infty \le Z \le +\infty)$$

Ⅱ. 정규분포

표준정규분포의 기댓값과 분산

$$E(Z) = E[\frac{X-\mu}{\sigma}] = E[\frac{1}{\sigma}X - \frac{\mu}{\sigma}] = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma}$$

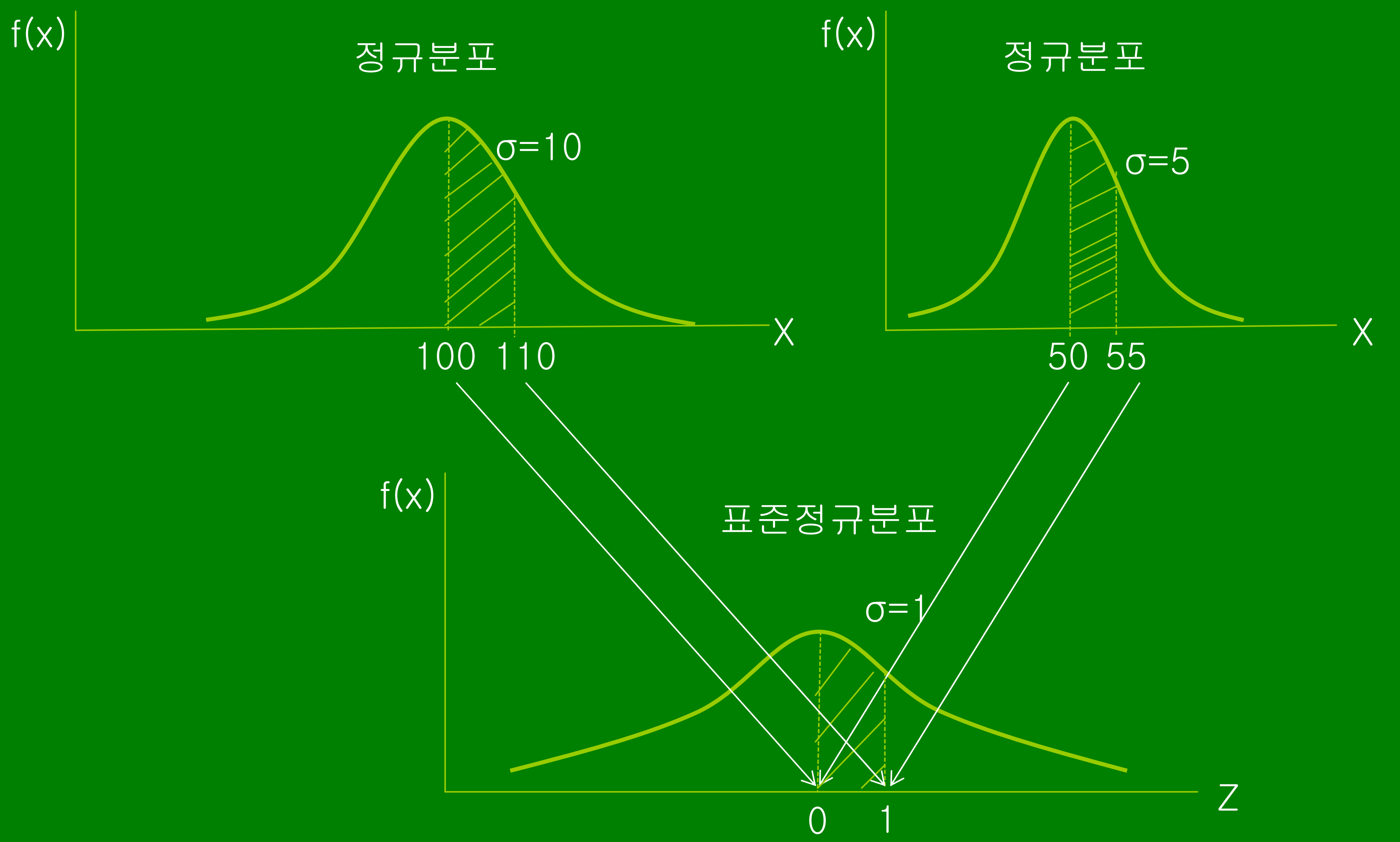
$$= \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$Var(Z) = Var(\frac{X-\mu}{\sigma}) = Var[\frac{1}{\sigma}(X-\mu)] = \frac{1}{\sigma^2}Var(X)$$

$$= \frac{\sigma^2}{\sigma^2} = 1$$

II. 정규분포

정규분포와 표준정규분포의 관계



Ⅱ. 정규분포

표준정규분포의 특징

- 표준정규분포의 확률변수 Z는 평균이 0, 표준편차가 1인 특수한 경우의 정규분포이다.(이때 분포의 최고 높이는 0.39894로 일정하다.)

확률변수 Z가 평균 0으로부터 ±1표준편차 거리에 분포할 확률이 0.6826이다.

확률변수 Z가 평균 0으로부터 ±2표준편차 거리에 분포할 확률이 0.9544이다.

확률변수 Z가 평균 0으로부터 ±3표준편차 거리에 분포할 확률이 0.9974이다.

$$P(-1 \le Z \le +1) = P(\mu - 1\sigma \le X \le \mu + 1\sigma) = 0.6826$$

$$P(-2 \le Z \le +2) = P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.9544$$

$$P(-3 \le Z \le +3) = P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.9974$$

Ⅱ. 정규분포

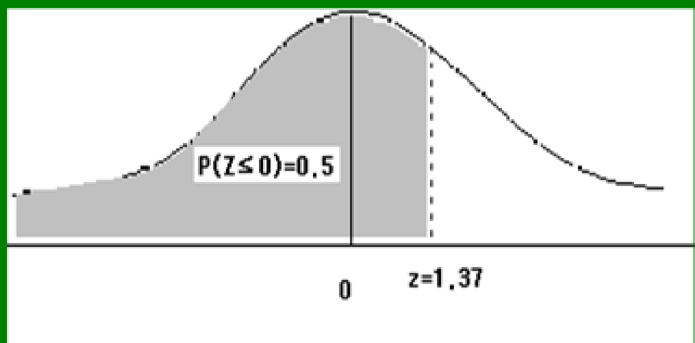
표준정규분포 이용연습

예1) Z값이 1.37이하일 확률은?

(풀이) $P(Z \le 1.37)$

$= P(Z \le 0) + P(0 \le Z \le 1.37)$

$= 0.5 + 0.4147 = 0.9147$



예2) $P(Z \ge 1.6) = ?$

(풀이)

$P(Z \ge 1.6) = 0.5 - P(0 \le Z \le 1.6)$

$= 0.5 - 0.4452$

$= 0.0584$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

II. 정규분포

정규분포 문제연습

예3) 스마트폰의 밧데리 충전시간이 평균 50시간, 표준편차 15시간의 정규분포를 따른다고 가정하자. 밧데리 충전시간을 나타내는 확률변수 X에 대하여 다음을 Z값으로 전환시켜라.

1) X=20일 때 Z=? (풀이) $Z = (X-\mu)/\sigma = (20-50)/15 = -2$
2) X=75일 때 Z=? (풀이) $Z = (X-\mu)/\sigma = (75-50)/15 = 1.67$

예4) 확률변수 X가 $X \sim N(50, 15^2)$일 때 $P(20 \le X \le 75)$를 구하라.

(풀이) $P(20 \le X \le 75)$를 Z값으로 표준화하면 $P(-2 \le Z \le 1.67)$이므로

$P(-2 \le Z \le 0)+P(0 \le Z \le 1.67)$은 $P(0 \le Z \le 2)+P(0 \le Z \le 1.67)$이다.

따라서 표준정규분포표를 활용하면 $P(0 \le Z \le 2)=0.4772$, $P(0 \le Z \le 1.67)=0.4526$

예5) 평균 $\mu=100$, 표준편차 $\sigma=20$인 정규분포의 확률변수를 X라고 할 때 다음을 구하라.

1) $P(X \le 65)$ (풀이) $P(Z \le (65-100)/20)=P(Z \le -1.75)=0.5-P(0 \le Z \le 1.75)=0.5-0.4599=0.0401$

2) $P(80 \le X \le 130)$ (풀이) $P(-1 \le Z \le 1.5)=P(0 \le Z \le 1)+P(0 \le Z \le 1.5)=0.3413+0.4332=0.7745$

3) $P(X \ge 140)$ (풀이) $P(Z \ge 2)=0.5-P(0 \le Z \le 2)=0.5-0.4772=0.0228$

II. 정규분포

정규분포 문제연습

예6) 통계학 과목의 학년말 성적을 확률변수 X라고 할 때 X는 평균 70점, 표준편차 10점인 정규분포를 따른다.

1) 상위 20.9%의 학생에게 A학점을 주려고 한다. 몇 점 이상이어야 하는가?

(풀이) $P=0.291$에 해당하는 $Z=0.81$ $Z=(X-\mu)/\sigma$에서 $X=\mu+Z\sigma$ ∴ $X=70+0.81(10)=78.1$

2) 하위 5%의 학생에게 F학점을 주려고 한다. 몇 점 이하이어야 하는가?

(풀이) $P=0.45$에 해당하는 $Z=-1.645$ ∴ $X=70-1.645(10)=53.55$

예7) 통계학 기말고사를 완료하는데 소요되는 시간은 평균 80분, 표준편차 10분인 정규분포를 이룬다.

1) 한 학생이 시험을 한 시간 이내에 끝낼 확률은?

2) 시험을 보는 학생의 90%가 충분한 시간을 갖도록 하기 위해서 시험시간을 몇분으로 제한해야 하는가?

3) 시험을 보는 학생은 60명이다. 시험시간을 90분으로 제한 할 때 이 시간 내에 시험을 완료하지 못할 학생의 수는 몇 명인가?

Ⅱ. 정규분포

이항분포의 정규근사법

- 이항분포는 시행횟수가 커지면 포아송분포 또는 연속확률분포인 정규분포에 근사시킬 수 있다.[$np>5$, $n(1-P)>5$ 즉 성공확률과 실패확률이 비슷]
- 이항분포를 정규분포로 근사 시키기 위해서는 연속성 조정이 필요.
  이것은 연속확률분포의 분포곡선 이하의 면적이 1이라는 속성을 충족시켜 주기 위해 이항분포는 정수 값만을 변수 값으로 취할 수 있으므로 이항분포에서는 $P(X \le K)+P(X \ge K+1)=1$이지만 확률변수 x가 정규분포를 따른다고 할 때는 변수값의 이산성 때문에 $P(X \le K)+P(X \ge K+1) \ne 1$가 된다.
- K와 K+1 사이의 값을 고려해주기 위해 이 구간을 양쪽에 귀속시켜 확률을 계산한다.
- 연속성 조정 방법
  - If 이항확률 $P(X \le K)$이면 $P(X \le K+0.5)$로 연속성 조정
  - If 이항확률 $P(X \ge K)$이면 $P(X \ge K-0.5)$로 연속성 조정
  - If 이항확률 $P(X=K)$이면 $P(K-0.5 \le X \le K+0.5)$로 연속성 조정

II. 정규분포

이항분포의 정규근사 예제

예제1) 어느 회사에서 신용장을 작성할 때 신용장의 10% 정도가 결점을 가진다고 한다. 100개의 신용장을 표본으로 추출할 때 13개의 신용장에서 실수를 발견할 확률, 즉 P(x=13)을 정규분포에 근사시켜 구하라.

(풀이) X~B(100, 0.1) E(x)=μ=100 x 0.1=10, Var(x)=100 x 0.1 x 0.9=9

X~B(100, 0.1)≒X~N(10,9) 이다.

따라서 P(x=13)에 대하여 연속성을 조정하면 P(13−0.5≤X≤13+0.5)를 정규분포를 이용하여 확률을 구하는 것과 같다.

따라서 P(12.5≤X≤13.5)를 Z값으로 표준화 시키면

$$P\left(\frac{12.5-10}{\sqrt{9}} \le Z \le \frac{13.5-10}{\sqrt{9}}\right) = P(0.83 \le Z \le 1.17)$$

$$P(0 \le Z \le 1.17) - P(0 \le Z \le 0.83) = 0.3790 - 0.2967$$

$$= 0.0823$$

II. 정규분포

이항분포의 정규근사 예제

예제2) 어느 회사에서 신용장을 작성할 때 신용장의 10% 정도가 결점을 가진다고 한다. 100개의 신용장을 표본으로 추출할 때 13개의 신용장에서 실수를 발견할 확률, 즉 $P(x \le 13)$을 정규분포에 근사시켜 구하라.

(풀이) $X \sim B(100, 0.1)$ $E(x)=\mu=100 \times 0.1=10$, $Var(x)=100 \times 0.1 \times 0.9=9$

$X \sim B(100, 0.1) \fallingdotseq X \sim N(10,9)$ 이다.

따라서 $P(x \le 13)$에 대하여 연속성을 조정하면 $P(X \le 13+0.5)$를 정규분포를 이용하여 확률을 구하는 것과 같다.

따라서 $P(X \le 13.5)$를 Z값으로 표준화 시키면

$$
\begin{aligned}
& P(Z \le \frac{13.5 - 10}{\sqrt{9}}) = 0.5 + P(0 \le Z \le 1.17) \\
& = 0.5 + P(0 \le Z \le 1.17) \\
& = 0.5 + 0.2967 \\
& = 0.7967
\end{aligned}
$$

Ⅲ. 지수분포

지수분포의 개념

지수분포는 포아송분포와 역의 관계로 연속하여 발생하는 두 사상사이의 시간간격, 공간간격, 거리간격 등을 확률변수로 하는 확률분포이다.

지수확률밀도함수

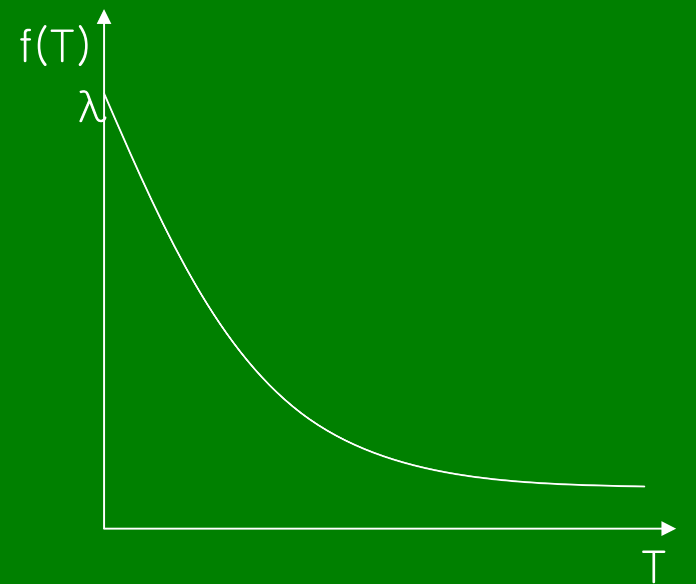
두 사상사이의 시간간격, 공간간격, 거리간격 등을 나타내는 확률변수를 T라고 할 때 지수확률밀도함수는 다음과 같다.

- T가 증가함에 따라 확률함수의 크기가 지수적으로 감소한다.

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

λ : 단위당평균발생횟수

t : 두사건사이의경과시간 또는거리



포아송분포

정 의

- 일정한 시간이나 구간에서 발생하는 특정한 사건의 수를 변수값으로 하는 확률변수의 분포

확률변수의 예

- 새로 구입한 자동차에서 발견되는 흠집 수
- 하루에 고장나는 기계의 대수
- 1달 동안 아파트 엘리베이터의 고장 건수
- 1주일 간에 발생하는 정전 건수
- 112, 114, 119 전화교환대에 걸려오는 시간당 전화 건수

지수분포

정 의

- 한 사건이 발생한 후에 다음 사건이 발생할 때까지의 시간이나 면적을 변수값으로 하는 확률변수의 분포

확률변수의 예

- 새로 구입한 자동차에서 1개의 흠집이 발견되고 또 다른 1개의 흠집이 발견되기까지의 면적
- 공장에서 기계 1대가 고장난 후에 또 다른 기계가 고장날 때까지 걸리는 시간
- 아파트 엘리베이터가 1번 고장난 후에 또 다시 고장날 때까지 걸리는 시간
- 112, 114, 119 전화교환대에 전화가 걸려온 후 다음 전화가 걸려올 때까지의 시간

III. 지수분포

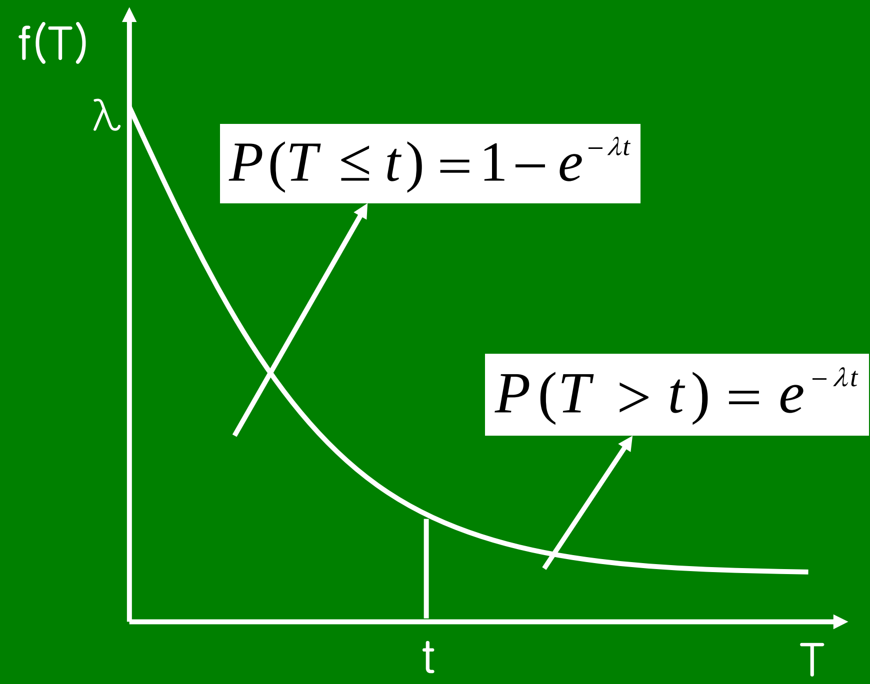
지수분포의 확률

두 사상의 발생사이의 시간간격이 특정시간간격 t보다 클 확률

$$P(T > t) = \int_t^\infty \lambda e^{-\lambda t} dt = e^{-\lambda t}$$

두 사상의 발생사이의 시간간격이 특정시간간격 t보다 작을 확률

$$P(T \le t) = \int_0^t \lambda e^{-\lambda t} dt = 1 - e^{-\lambda t}$$



Ⅲ. 지수분포

지수분포의 기댓값과 분산

$$E(T) = \mu = \frac{1}{\lambda}$$

$$Var(T) = \sigma^2 = \frac{1}{\lambda^2}$$

$$\sigma = \frac{1}{\lambda}$$

II. 정규분포

지수분포 문제연습

예1) 어느 은행에 도착하는 고객의 수는 분당 평균 0.75명인 포아송 분포를 따른다고 한다. 만일 도착 사이의 시간이 3분 이하라고 한다면 고객이 지루하지 않게 서비스를 받을 수 있다고 할 때

1) 지루하게 느끼지 않을 고객의 비율을 구하라. (풀이)

$$P(T \le 3) = \int_0^3 0.75\, e^{-0.75 \times 3}\, dt$$
$$= 1 - e^{-0.75 \times 3}$$
$$= 0.8946$$

2) 도착 사이의 시간이 4분 이상일 확률을 구하라. (풀이)

$$P(T \ge 4) = \int_4^\infty 0.75\, e^{-0.75 \times 4}\, dt$$
$$= e^{-0.75 \times 4}$$
$$= 0.0498$$

3) 고객 도착 사이의 평균시간과 표준편차를 구하라. (풀이)

$$E(T) = \frac{1}{\lambda} = \frac{1}{0.75} = 1.33$$

$$\sigma(T) = \frac{1}{\lambda} = \frac{1}{0.75} = 1.33$$