확률 및 통계학

- 6장 연속형 확률분포(1) -

명 세인(sein@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

목 차

- 연속형 균일분포
- 정규분포
- 표준정규분포
- 정규분포의 적용

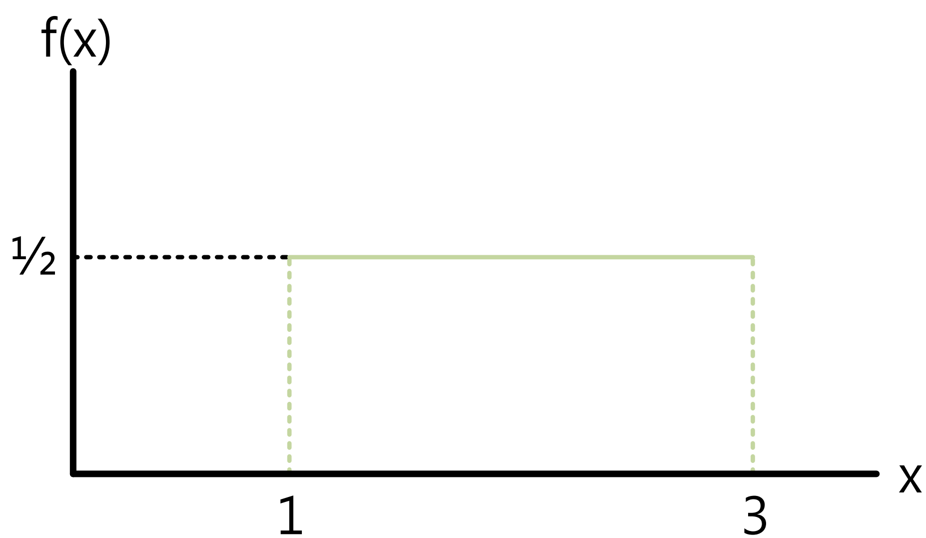
연속형 균일분포

- 의미
  - 연속형 균일분포(Continuous Uniform Distribution)는 밀도 함수가 존재하는 구간의 확률 밀도가 균일

- 정의
  - 구간 $[A, B]$에서 정의되는 연속형 균일확률변수 X의 확률밀도함수
    - $f(x; A, B) = \begin{cases} \frac{1}{B-A}, A \le x \le B \\ 0, \text{구간 외} \end{cases}$

연속형 균일분포

- 특징
  - 그림 6.1
    - 구간 [1, 3]에서 정의되는 연속형 균일분포



    - 전체 확률(면적)은 1
    - 동일한 확률이 분포하여 일자형 그래프가 나타남

연속형 균일분포

- 정의
  - 평균 $\mu = \frac{A+B}{2}$
    - $\mu = E(X) = \int_a^b xf(x)\,dx = \left[\frac{1}{b-a} \times \frac{x^2}{2}\right]\Big|_a^b = \frac{a+b}{2}$
  - 분산 $\sigma^2 = \frac{(B-A)^2}{12}$
    - $E(X^2) = \int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b-a}\,dx = \left[\frac{1}{b-a} \times \frac{x^3}{3}\right]\Big|_a^b = \frac{a^2+ab+b^2}{3}$
    - $\sigma^2 = E(X^2) - E(X)^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$

연속형 균일분포

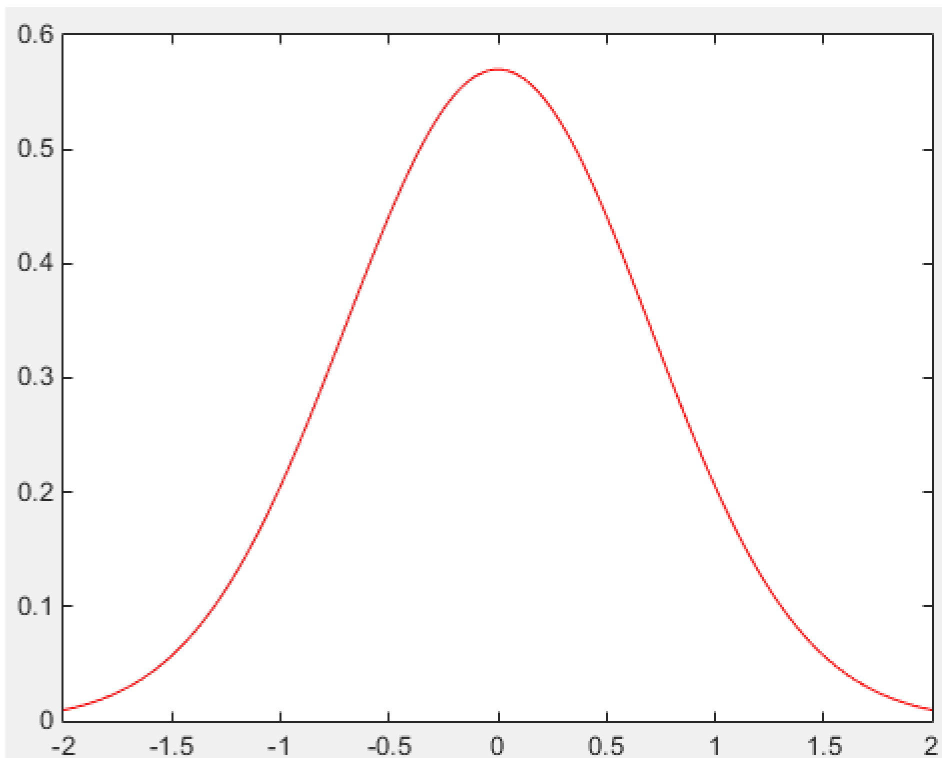
- 예제 6.1
  - 어느 회사의 회의실은 4시간을 초과하여 사용할 수 없으며, 회의실에서는 긴 회의와 짧은 회의가 자주 열릴 때, 회의시간X는 구간 [0, 4]에서 정의되는 균일분포를 따름
    - 밀도함수는?
      - $f(x) = \begin{cases} \frac{1}{4}, 0 \le x \le 4 \\ 0, \text{구간 외} \end{cases}$
    - 어떤 회의가 3시간 이상 계속될 확률은?
      - $P[X \ge 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}$

정규분포

- 의미
  - 1733년 드 무아브르가 정규곡선의 모형을 제안
    - 이항분포
  - 1820년 라플라스에 의해 분포의 함수식이 도축
    - 중심극한 정리(이항분포가 아닌 확률분포)
  - 동일한 양의 반복측정에 대한 오차에 관한 연구에서 정규분포와 관련이 있는 방정식을 유도해낸 가우스를 기리기 위해 가우시안분포로 불리기도 함
    - 최소제곱법
  - 여러가지 현상을 정규분포로 표현할 수 있음

정규분포

- 정의
  - 평균 μ와 분산 σ^2을 가지는 정규확률변수 X의 확률분포
    - $n(x; \pi,\ \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty$
    - π = 원주율, e = 오일러상수



정규분포

- 특징
  - $x = \mu$에서 곡선이 최대 값
    - 중앙값(Median), 평균(Mean), 최빈값(Mode)이 나타남
  - 곡선은 평균μ를 지나는 수직 축에 대하여 대칭
  - 곡선은 $x = \mu \pm \sigma$에서 변곡점이 나타남
  - 평균에서 멀어질수록, 정규곡선은 수평축에 접근
  - 곡선과 수평축 사이의 총 면적은 1

정규분포

- 정의
  - 평균 $\mathrm{E(X)} = \mu$
    - $E(X-\mu) = \int_{-\infty}^{\infty} \frac{x-\mu}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$
    - $z = \frac{x-\mu}{\sigma}$ 로 두고, $dx = \sigma dz$ 라 하면
    - $E(X-\mu) = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z\, e^{-\frac{1}{2}z^2} dz = 0$(기함수)
    - $\mathrm{E(X)} = \mu$

정규분포

- 정의
  - 분산 σ^2
    - $E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (X-\mu)^2 e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$
    - $z = \frac{x-\mu}{\sigma}$ 로 두고, $dx = \sigma dz$ 라 하면
    - $E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2\, e^{-\frac{z^2}{2}} dz$
    - $\mu = z, dv = ze^{\frac{-z^2}{2}}$라 하고 부분적분 하면, $dv = dz, v = -e^{\frac{-z^2}{2}}$ 이 되어 다음과 같이 변형
      - $E[(X-\mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ze^{-\frac{z^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \right) = \sigma^2(0+1) = \sigma^2$

표준 정규분포

- 의미
  - 모든 정규확률분포 X는 표준정규분포 Z로 변환가능
  - 표준정규분포 Z에 대한 확률 값을 정리한 표를 이용해 모든 정규확률분포를 쉽게 해결 가능

- 정의
  - 평균이 0이고, 분산이 1인 정규확률변수의 분포를 표준정규분포(Standard Normal Distribution) Z로 분류
  - 공식을 이용하여 x를 결정
    - $z = \frac{x-\mu}{\sigma}$
    - $x = \sigma z + \mu$

표준 정규분포

- 표준 정규분포표

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511966	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621720	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802337	0.805105	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.879000	0.881000	0.882977
1.2	0.884930	0.886861	0.888768	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903200	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935745	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959070	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965620	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999534	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999651
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999822	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

표준 정규분포

- 예제 6.2
  - 표준정규분포가 주어졌을 때, 다음의 면적
    - $z = 1.84$의 우측 면적
      - $1 - 0.9671 = 0.0329$
    - $z = -1.97, z = 0.86$ 사이의 면적
      - $0.8051 - 0.0244 = 0.7807$

표준 정규분포

- 예제 6.3
  - 표준정규분포가 주어졌을 때, 다음의 각 경우에 대한 k값
    - $P(z > k) = 0.3015$
      - $1 - 0.3015 = 0.6985$를 만족하는 값 $k = 0.52$
    - $P(k < Z < -0.18) = 0.4197$
      - k와 -0.18사이의 면적이 0.4197을 만족하기 위해,
        k 까지의 면적은 $0.4286 - 0.4197 = 0.0089$를 만족하는 $k = -2.37$

표준 정규분포

- 예제 6.4
  - $\mu = 50$이고 $\sigma = 10$인 정규분포가 주어졌을 때, X가 45와 62사이의 값을 취할 확률
    - $x_1 = 45$와 $x_2 = 62$에 대응하는 z값
      - $z_1 = \frac{45-50}{10} = -0.5, z_2 = \frac{62-50}{10} = 1.2$
      - $P(45 < X < 62) = P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5)$
        $= 0.8849 - 0.3085 = 0.5764$

표준 정규분포

- 예제 6.5
  - $\mu = 300$이고 $\sigma = 50$인 정규분포가 주어졌을 때, X가 362보다 큰 값을 취할 확률
    - $x = 362$ 에 대응하는 값
      - $z = \frac{362-300}{50} = 1.24$
    - $P(X > 362) = P(Z > 1.24) = 1 - P(Z < 1.24) = 1 - 0.8925 = 0.1075$

표준 정규분포

- 예제 6.6
  - $\mu = 40$이고 $\sigma = 6$인 정규분포가 주어졌을 때
    - 왼쪽 면적이 전체의 45%가 되는 x
      - 0.45가 되는 z값 $P(\mathrm{Z} < -0.13) = 0.45$
      - 따라서 $x = (6)(-0.13) + 40 = 39.22$
    - 오른쪽 면적이 전체 면적의 14%가 되는 x
      - 왼쪽 면적이 $1 - 0.14 = 0.86$이 되는 z값 $P(Z < 1.08) = 0.86$
      - 따라서 $x = (6)(1.08) + 40 = 46.48$

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.7
    - 어느 축전지의 평균 수명이 3년이고, 표준편차가 0.5년이고 축전지의 수명이 정규분포를 따른다고 가정할 때, 임의로 주어진 전지의 수명이 2.3년 보다 짧을 확률
      - $z = \frac{2.3-3}{0.5} = -1.4$
      - $P(X < 2.3) = P(Z < -1.4) = 0.0808$

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.8
    - 전기회사에서 평균수명이 800시간이고 표준편차가 40시간인 정규분포의 수명분포를 가지는 전구를 생산하고 있다. 임의로 선정된 전구의 수명이 778시간과 834시간 사이에 있을 확률은?
      - $z_1 = \frac{778-800}{40} = -0.55, z_2 = \frac{834-800}{40} = 0.85$

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.9
    - 어느 공정에서 볼베어링의 직경이 중요한 품질특성이 된다. 구매자 측은 직경의 규격한계를 $3.0 \pm 0.01cm$로 정해 놓고 있다. 따라서 이 규격한계를 벗어나는 부품은 불합격 처리할 때, 볼베어링의 직경은 평균이 3.0, 표준편차 0.005인 정규분포를 따른다면 생산품 중 불량으로 처리되는 것은?
      - 규격한계 $x_1 = 2.99$와 $x_2 = 3.01$에 대응되는 z값은 각각
        - $z_1 = \frac{2.99-3.0}{0.005} = -2.0,\ z_2 = \frac{2.01-3.0}{0.005} = +2.0$이 되어
        - $P(2.99 < X < 3.01) = P(-2.0 < Z < +2.0)$
        - $P(Z < -2.0) = 0.0228$가 구해지고, 정규분포의 대칭성을 이용하면
          - $P(Z < -2.0) + P(Z > 2.0) = 2(0.0228) = 0.0456$
        - 볼베어링의 4.56%가 불합격으로 처리됨

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.10
    - 어느 치수가 규격한계인 $1.50 \pm d$ 내에 들어오지 않으면 모든부품을 불합격시키는 평가기준이 사용된다고 할 때, 측정값은 평균이 1.50이고 표준편차가 0.2인 정규분포를 따르면, 측정값의 95%가 규격한계 내에 들도록 d 값을 결정
      - $P(-1.96 < Z < 1.96) = 0.95$
      - $1.96 = \frac{(1.50+d)-1.50}{0.2} \rightarrow d = (0.2)(1.96) = 0.392$

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.11
    - 평균 저항이 40Ω 이고 표준편차가 2Ω 인 저항기를 만드는 기계가 있다. 저항이 정규분포를 따르면, 43Ω 을 넘는 저항을 갖게 되는 저항기는 몇 퍼센트인가?
      - $x = 43$에 대한 오른쪽 면적을 구하기 위해 대응하는 z값을 구하면
      - $z = \frac{43-40}{2} = 1.5$
      - $P(X > 43) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - 9.332 = 0.0668$
      - 따라서 저항기의 6.68%가 $43\ \Omega$ 이 넘는 저항을 가짐

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.12
    - 예제 6.11에서 저항의 측정값을 소수 첫째 자리에서 반올림할 때, 43Ω 이 넘는 저항기의 비율을 구하라
      - 42.5Ω 보다 크고 43.5Ω 보다 작은 저항기를 모두 43의 저항으로 볼 때, 43.5Ω보다 큰 영역
      - $z = \frac{43.5-40}{2} = 1.75$ 이므로
      - $P(X > 43.5) = P(Z > 1.75) = 1 - P(Z < 1.75) = 1 - 0.9599$ $= 0.0401$
      - 반올림을 적용할 때 4.01%가 43Ω을 넘는 저항이 됨
      - 예제 6.11과의 차이는 $6.68\% - 4.01\% = 2.67\%$

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.13
    - 어느 시험 성적이 평균 74점이고, 표준편차가 7인 정규분포를 따르고, 12%의 학생에게 A학점이 주어진다. A학점을 받은 학생 중 가장 낮은 점수를 갖는 학생과, B학점을 받은 학생 중 가장 높은 점수를 받은 학생은 각각 몇 점인가?
      - 학점 A는 면적 0.12를 나타내고, 외쪽 면적이 0.88이 되는 z값을 구하면
      - $P(Z < 1.18) = 0.88$
      - $X = (7)(1.18) + 74 = 82.26$
      - A학점 중 가장 낮은 점수는 83점, B학점 중 가장 높은 학점은 82

정규분포의 적용

- 정규분포를 적용할 수 있는 예제
  - 예제 6.14
    - 예제 6.13에서 제 6십분위수를 구하라
      - 6십분위수를 D_6으로 하면, 왼쪽 면적이 총 면적의 60%가되는 x값을 구해야 함
      - $P(Z < 0.25) \approx 0.6$
      - $x = (7)(0.25) + 74 = 75.75$
      - $D_6 = 75.75$이므로 성적의 60%는 75점 이하이다.

감사합니다!

보충

- 기하분포
- 포아송분포
  - 제한된 공간, 시간에서 발생확률을 논할 때
    - 큐잉 모델(패킷 큐, 프로세스 큐)
      - 큐에 대한 입력은 포아송분포, 큐의 크기가 무한대일 때 출력또한 포아송분포
    - 트래픽 부하
      - 서비스는 일정량의 트래픽을 요구, 그 이상의 트래픽처리를 생각
- 정규분포