확률통계 및 공통 알고리즘

2. 확률변수 및 확률분포

임채영

서울대학교

이번강의에서 다룰 내용

- 확률변수, 이산, 연속 확률변수
- 확률분포, 기대값, 분산, 표준편차
- 이산, 연속확률분포 예시
- 결합확률분포, 주변확률분포, 독립성, 공분산, 상관계수, 조건부 확률분포
- 다변량 분포의 예시

(실) 확률변수 (random variable)

표본공간의 각 원소를 하나의 실수로 대응하는 함수

$$c \in \mathcal{S}, \ X(c) = x \in \mathbb{R}$$

예시

- 동전 1개를 던졌을 때 표본공간은 $\mathcal{S} = \{H, T\}$ (앞면 : H, 뒷면 : T)
- $P(\{H\}) = P(\{T\}) = \frac{1}{2}$
- $X(H) = 1, \ X(T) = 0$
- 함수 X의 정의역은 $\mathcal{S} = \{H, T\}$, 치역은 $\{0, 1\}$

확률과 확률변수

확률변수가 가지는 값에 대한 확률의 의미?

- 앞의 예시에서 $X = 1$일 확률은?
  - 동전 1개를 던졌을 때, 앞면이 나올 확률과 같다.
  - 따라서, 확률변수 X가 1의 값을 가질 확률을 다음과 같이 표현한다.
  - $P(X = 1) = P(\{c \in \mathcal{S} \mid X(c) = 1\}) = P(\{H\}) = 1/2$

확률분포 (probability distribution)

확률변수 X의 확률분포 (probability distribution)란: 확률변수 X가 가질 수 있는 값과 해당하는 확률에 대해 나타낸 것으로, 확률을 계산 할 수 있는 정보를 제공.

- 이산확률변수: X가 취할 수 있는 값이 $x_1, x_2, x_3, \ldots$와 같이 이산 일 때
  - 해당 값과 대응하는 확률을 제공
- 연속확률변수: X의 취할 수 있는 값이 셀 수 없이 많을 때
  - 특정 구간에 속하는 확률을 계산할 수 있는 정보를 제공.

이산확률변수 (discrete random variable)

확률분포는 다음과 같은 확률질량함수 (probability mass function) $p(x)$로 표현 가능

$$p(x) = P(X = x) = \begin{cases} P(X = x_i) & , x = x_i\text{일 때}(i = 1, 2, \ldots) \\ 0 & \textit{otherwise.} \end{cases}$$

- $0 \leq p(x) \leq 1$
- $\sum_{all\ x} p(x) = 1$
- $P(a < X \leq b) = \sum_{a<X\leq b} p(x)$

예제

15개의 상품 중 5개가 불량품이다. 3개를 단순랜덤추출하였을 때, 그 중 불량품의 개수를 X라 하자. 확률변수 X의 확률분포를 구하여라.

연속확률변수 (continuous random variable)

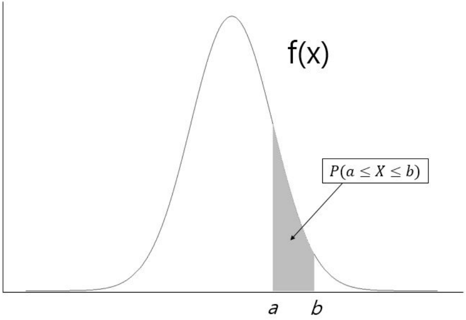
확률분포는 확률밀도함수 (probability density function) $f(x)$를 도입하여 X의 값이 $a < X \le b$일 확률로 표현

$$P(a < X \le b) = \int_a^b f(x)dx$$

- $f(x) \ge 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \le X \le b) = \int_a^b f(x)dx$

연속확률변수의 성질

- 연속확률변수의 한 점에서의 확률은 0이다. $P(X = a) = 0$
- $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b)$
  $= P(a < X < b)$



예제

연속확률변수 X의 확률밀도함수가

$$f(x) = \begin{cases} bx(1-x) & (0 \le x \le 1\text{일 때}) \\ 0 & (x < 0 \text{ 또는 } x > 1\text{일 때}) \end{cases}$$

로 주어질 때, 상수 b와 확률 $P(0 \le X \le \frac{3}{4})$를 구하여라.

누적확률분포 함수

- pmf, pdf 외에 확률분포를 나타내는 또 다른 함수
- Cumulative Distribution Function, CDF
- $F_X(x) = P(X \le x)$
- 이산확률변수, 연속확률변수에 상관없음
- Non-decreasing 함수
- 연속인 경우: $\frac{d}{dx}F(x) = f(x)$

기대값(expected value)

확률변수 X의 중심을 나타내는 값, 평균

$$\mu = E(X) = \begin{cases} \sum_x xp(x) & \text{(이산확률변수)} \\ \int_{-\infty}^{\infty} xf(x)dx & \text{(연속확률변수)} \end{cases}$$

예제: 동전을 2회던지는 실험에서 앞면의 개수를 X라고 할 때, X의 기대값을 구하여라.

- 확률변수 X의 함수 $g(X)$의 기대값

$$E(g(X)) = \begin{cases} \sum_x g(x)p(x) & \text{(이산확률변수)} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{(연속확률변수)} \end{cases}$$

- 기대값의 성질 : 선형성
  - $E(aX + b) = aE(X) + b$ (a, b는 상수)
  - $E[ag(X) + bh(X)] = aE(g(X)) + bE(h(X))$ (a, b는 상수)

분산(variance)과 표준편차(standard deviation)

X의 평균을 μ라고 하자.

(1) 분산

$$Var(X) = E[(X-\mu)^2] = \begin{cases} \sum_x (x-\mu)^2 p(x) & \text{(이산확률변수)} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx & \text{(연속확률변수)} \end{cases}$$

(2) 표준편차

$$sd(X) = \sqrt{Var(X)}$$

- $Var(X) = E(X^2) - [E(X)]^2$
- $Var(aX + b) = a^2 Var(X)$

예제

X의 확률밀도함수가 $f(x) = \begin{cases} 1 & , 0 \leq x \leq 1 \\ 0 & , otherwise \end{cases}$ 일 때,
X의 평균, 분산, 표준편차를 구하여라.

베르누이 분포(Bernoulli distribution)

베르누이 시행 (Bernoulli trial)

- 실험의 결과 두 가지 중의 하나로 나오는 시행
- 표본 공간 $\mathcal{S} = \{성공(s), 실패(f)\}$
- 성공 확률 $p = P(\{s\})$

베르누이 확률변수 (Bernoulli random variable)

- 베르누이 시행의 결과를 0 또는 1의 값으로 대응시키는 확률변수
- $X(s) = 1, X(f) = 0$인 확률변수

- 베르누이 확률변수의 확률분포를 베르누이 분포라 한다
- $X \sim Ber(p)$
- $p(x) = p^x(1-p)^{1-x}$, $x = 0, 1$.
- $E(X) = p$
- $Var(X) = E(X^2) - [E(X)]^2 = p(1-p)$

이항분포 (Binomial distribution)

베르누이 시행을 n번 시행할 때 성공횟수의 분포

- $X \sim B(n, p)$ 또는 $Bin(n, p)$.
- $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, x=0, $\cdots$, n.
- $n = 1$이면 베르누이 분포

이항분포의 성질

$X \sim B(n, p)$

- $E(X) = np$
- $Var(X) = np(1-p)$

예시

5개 중 하나를 택하는 선다형 문제가 20문항 있는 시험에서 랜덤하게 답을 써 넣는 경우
X : 20문항 중 정답의 수

- $X \sim B(20, 0.2)$
- 정답이 하나도 없을 확률 :
- 8개 이상의 정답을 맞힐 확률 :
- $E(X) = 20 \times 0.2 = 4$
- $Var(X) = 20 \times 0.2 \times 0.8 = 3.2$

균일분포 (Uniform distribution)

- 확률변수 X가 a와 b 사이에서 같은 정도로 값을 가질 때 균일분포를 따른다고 한다.
- $X \sim Uniform(a, b)$
- $f(x) = \frac{1}{b-a}$, $a < x < b$
- $E(X) = \frac{a+b}{2}$, $Var(X) = \frac{(b-a)^2}{12}$

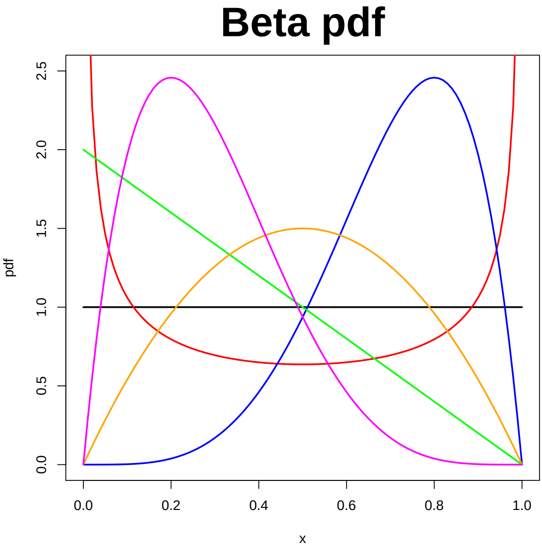
예제

전구의 수명은 10년과 20년 사이의 균일분포를 따른다. 전구의 평균 수명과 분산을 구하고 전구가 17년 이상의 수명을 가질 확률을 구하여라.

베타 분포 (Beta Distribution)

- 연속확률분포중의 하나로 $0 \le X \le 1$인 확률변수가 다음의 확률밀도함수를 가지는 경우이다.
- $f(x) = f(x|\alpha, \beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $x \in [0, 1]$, $\alpha > 0, \beta > 0$.
- $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ 는 정규화 상수 (normalizing constant)
- $X \sim Beta(\alpha, \beta)$
- $E(X) = \frac{\alpha}{\alpha+\beta}$
- $\alpha = \beta = 1$이면 베타분포는 균일분포와 같다.

두 모수 α, β의 값에 따라 다양한 형태의 확률밀도함수가 나온다.



지수분포 (Exponential distribution)

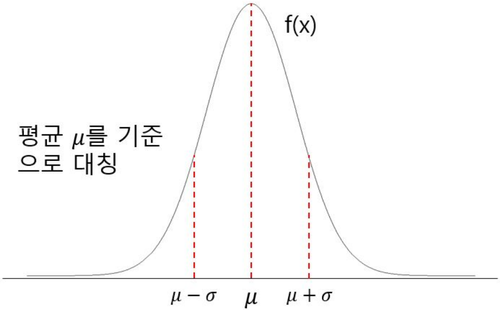
- 하나의 사건이 일어난 후 독립인 그 다음 사건이 일어날대까지 기다리는 시간 (waiting time)을 모형화 한 분포
- $X \sim Exp(\lambda)$
- $f(x) = \lambda \exp(-\lambda x)$, $x > 0$. λ: rate parameter
- 또는 $f(x) = \frac{1}{\rho} \exp(-x/\rho)$. ρ: scale parameter
- $E(X) = \frac{1}{\lambda} = \rho$
- $Var(X) = \frac{1}{\lambda^2} = \rho^2$
- Memoryless property: P(X > s+t | X > s)=P(X > t), $s, t > 0$.

정규분포 (Normal distribution)

- 가우스(Gauss, 1777-1855)에 의해 제시된 분포로서 가우스분포(Gaussian distribution)라고도 불린다.
- 물리학 실험 등에서 오차에 대한 확률분포를 연구하는 과정에서 발견된 연속확률분포.
- 통계학 초기 발전 단계에서 모든 자료의 히스토그램이 가우스분포의 형태와 유사하지 않으면 비정상적인 자료라고 믿어서 "정규(normal)"라는 이름이 붙게 되었다.

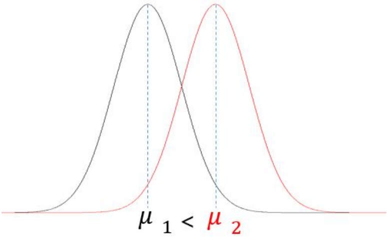
- $X \sim N(\mu, \sigma^2)$, μ: 평균(mean), σ^2: 분산(variance), $\tau^2 = 1/\sigma^2$: *precision*

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\Big(- \frac{(x-\mu)^2}{2\sigma^2} \Big), -\infty < x < \infty, \ \sigma > 0$$

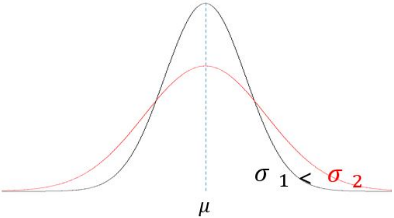


정규분포의 성질

표준편차는 같고 평균이 다른 두 정규분포



평균은 같고 표준편차가 다른 두 정규분포



정규분포의 성질

- $X \sim N(\mu, \sigma^2)$

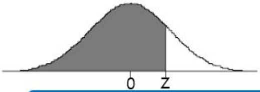
$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$

표준정규분포(standard normal distribution)

- 평균이 0이고 표준편차가 1인 정규분포를 표준정규분포 (standard normal distribution)라고 한다.
- 보통 Z로 표기.

표준정규분포표

- 표준정규분포표($P(Z \leq z)$)



소수점 둘째 자리

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830

소수점 첫째 자리

- $P(Z \leq 0.93) = 0.8238$

표준정규분포에서 확률 구하기

- $P(a \leq Z \leq b) = P(Z \leq b) - P(Z < a)$
- $P(Z \geq z) = 1 - P(Z < z)$

예시

$P(Z < -1.9 \text{ 또는 } Z > 2.1) = ?$

$$
\begin{aligned}
P(Z < -1.9 \text{ 또는 } Z > 2.1) &= P(Z < -1.9) + P(Z > 2.1) \\
&= P(Z < -1.9) + 1 - P(Z \leq 2.1) \\
&= 0.0287 + 1 - 0.9821 \\
&= 0.0179
\end{aligned}
$$

αth percentile (백분위수)

- αth 백분위수는 $P(Z \leq z) = \alpha$를 만족하는 z를 말하며 z_α로 표기하자.
- $Z_{0.975} = 1.96$. 즉, $P(Z < 1.96) = 0.975$.
- $Z_{0.995} = 2.58$
- 정규분포가 아니더라도, 같은 방식으로 정의할 수 있다.
- 교재에 따라 $P(Z > z) = \alpha$인 z를 z_α로 표기하기도 한다. 따라서, 문제에서 주어지는 z_α의 정의에 유의한다.

정규분포에서 확률 구하기

- 일반적인 정규분포 $X \sim N(\mu, \sigma^2)$의 확률 계산시에 표준정규분포를 이용한다
- 표준화(standardization) :
  $X \sim N(\mu, \sigma^2)$일 때 $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$

  $X = \mu + \sigma Z$

$$
\begin{aligned}
P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)
\end{aligned}
$$

예제

학생들의 통계학 성적의 분포가 근사적으로 $N(60, 10^2)$을 따른다고 한다. 45점 이하인 학생에게 F학점을 준다고 할 때, F학점을 받게 될 학생의 비율을 근사적으로 구하여라.

결합분포(joint probability distribution)

두 개의 확률변수가 취할 수 있는 값들의 모든 쌍의 확률을 나타낸 것

- 이산형 결합확률질량함수

$$p(x, y) = P(X = x, Y = y)$$

  - $0 \leq p(x, y) \leq 1$
  - $\sum_x \sum_y p(x, y) = 1$
  - $P(a < X \leq b, c < Y \leq d) = \sum_{a<x\leq b} \sum_{c<y\leq d} p(x, y)$

- 연속형 결합확률밀도함수

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d f(x, y) dydx$$

  - $f(x, y) \geq 0$
  - $\int f(x, y) dxdy = 1$
  - $P(a < x \leq b, c < y \leq d) = \int_c^d \int_a^b f(x, y) dxdy$
  - 결합 누적 확률분포 함수: $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$.

예제

서로 다른 동전 A,B,C를 동시에 던지는 실험에서 확률변수

$$X = \begin{cases} 1 & ,\text{동전 A가 } H \\ 0 & , otherwise \end{cases}$$

$$Y = \begin{cases} 1 & ,\text{동전 A,B가 } H \\ 0 & , otherwise \end{cases}$$

$$Z = \begin{cases} 1 & ,\text{동전 B,C가 } H \\ 0 & , otherwise \end{cases}$$

Table: 표본공간과 확률변수

S	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
X	1	1	1	1	0	0	0	0
Y	1	1	0	0	0	0	0	0
Z	1	0	0	0	1	0	0	0

Table: X와 Y의 결합확률분포

y \ x	0	1	행의 합
0	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{3}{4}$
1	0	$\frac{1}{4}$	$\frac{1}{4}$
열의 합	$\frac{2}{4}$	$\frac{2}{4}$	1

주변확률밀도함수(Marginal PDF)

- 이산형 : $p_X(x) = \sum_y p(x, y)$
- 연속형 : $f_X(x) = \int f(x, y)dy$

예제: 앞의 예제에서

Table: X의 주변확률분포

x	0	1	계
$p_X(x)$	$\frac{1}{2}$	$\frac{1}{2}$	1

두 확률변수의 함수의 기대값

두 확률변수 X, Y의 함수의 기대값

- $E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y)p(x,y) & \text{(이산형)} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y) & \text{(연속형)} \end{cases}$
- $E[ag(X,Y) + bh(X,Y)] = aE[g(X,Y)] + bE[h(X,Y)]$ (a, b 는 상수)

두 확률변수의 독립성

두 확률변수 X, Y 가 다음을 만족할때:
모든 x, y에 대해

$$\begin{aligned} p(x,y) &= p_1(x)p_2(y) \text{ (이산형)} \\ f(x,y) &= f_1(x)f_2(y) \text{ (연속형)} \end{aligned}$$

- X와 Y는 서로 독립이면, $E(XY) = E(X)E(Y)$
- X, Y가 서로 독립인 경우 $X \perp Y$로 표시한다.

예제

하나의 동전을 세 번 던질 때, X는 처음 두번에서 나오는 앞면의 개수, Y는 세번째에 나오는 앞면의 개수라 하자. X와 Y는 서로 독립인가?

공분산(Covariance)과 상관계수(Correlation coefficient)

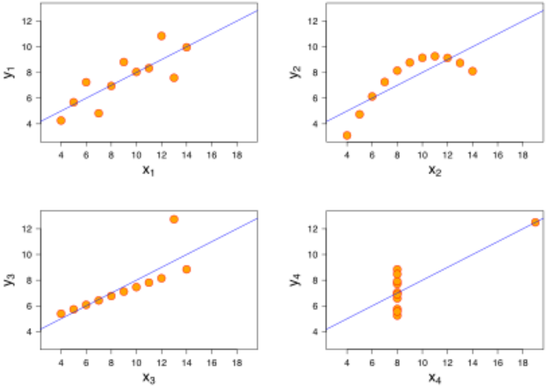
- 공분산(Covariance)

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_2)] \\ &= E(XY) - \mu_X\mu_Y = E(XY) - E(X)E(Y) \end{aligned}$$

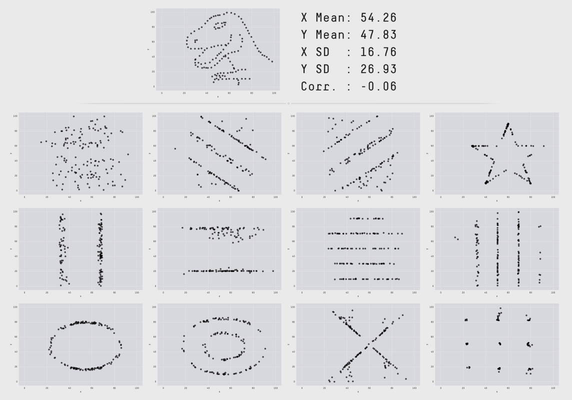
- 상관계수(Correlation coefficient) -선형의 연관성을 나타냄

$$Corr(X, Y) = \rho_{XY} = \frac{Cov(X, Y)}{sd(X)sd(Y)}$$

Anscombe's quartet



From Wikipedia



www.autodeskresearch.com/publications/samestats

공분산과 상관계수의 성질

확률변수 X, Y에 대해 다음과 같은 성질들이 있다.

- $Cov(aX + b, cY + d) = acCov(X, Y)$
- $Corr(aX + b, cY + d) = sign(ac)Corr(X, Y)$
- $Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X, Y)$
- $Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$
- $-1 \leq \rho \leq 1$
- $Y = a + bX$ 이면 $\rho = \pm 1$

확률변수 X, Y가 독립일 경우

- $E(XY) = E(X)E(Y)$
- $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$
- $Cov(X, Y) = 0$, $Corr(X, Y) = 0$
  (주의 : $Cov(X, Y) = 0$인 것이 X, Y의 독립을 의미하지 않음)
- $Var(X \pm Y) = Var(X) + Var(Y)$

조건부 확률분포

- 두개의 확률변수가 있을 때, 하나의 확률변수의 값이 주어졌을때, 나머지 하나의 확률변수의 확률분포를 말함.
- 예) $X \sim N(0,1)$, $Y|X = x \sim N(x,4)$.
- 두개의 이산 확률변수 X, Y에 대하여 $X = x$가 주어졌을때의 Y의 확률질량함수:

$$p(y|x) = P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

- $p(y|x)$는 확률질량함수이다.

- 두개의 연속 확률변수 X, Y에 대하여 $X = x$가 주어졌을때의 Y의 확률밀도함수:

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

- $f(y|x)$는 확률밀도함수이다.
- 하나가 이산 확률변수이고, 다른 하나가 연속 확률변수 여도 상관없다.

조건부 확률분포를 이용한 예제

- 불량 배터리의 평균 사용시간은 4시간이고 표준편차가 1 시간인 정규분포를 따른다고 하고, 정상 배터리의 평균 사용시간은 5시간, 표준편차가 1시간인 정규분포를 따른다고 하자. 배터리가 불량일 확률은 1%라고 알려져 있다. 하나의 배터리를 랜덤추출하여 조사하였을때, 평균사용시간이 5시간 이상일 확률은?
- X :불량 여부: 1 (불량) or 1 (정상)

  Y : 배터리 사용시간

  $Y|X=1 \sim N(4,1)$, $Y|X=0 \sim N(5,1)$
- $P(Y \geq 5) = P(Y \geq 5|X=1)P(X=1) + P(Y \geq 5|X=0)P(X=0) = P(N(4,1) \geq 5) \times 0.01 + P(N(5,1) \geq 5) \times 0.99 = P(Z \geq 1) \times 0.01 + P(Z \geq 0) \times 0.99 = 0.1587 \times 0.01 + 0.5 \times 0.99 = 0.4966.$

조건부 독립

- 두 확률변수 X, Y가 또 다른 확률변수 Z가 주어졌을때 서로 독립인 경우 X, Y는 조건부 독립이라고 부른다.
- 즉, 모든 x, y, z에 대하여, $p(x, y|z) = p(x|z)p(y|z)$ 또는 $f(x, y|z) = f(x|z)f(y|z)$.
- $X \perp Y \mid Z$

데이터가 여러개 있을때 (1)

- 데이터가 n개가 있다고 하자.
  즉, $X_1, X_2, \cdots, X_n$.
- n개의 데이터들의 결합확률분포는 $f(x_1, \cdots, x_n)$
- 일반적 가정: 데이터들은 서로 독립적으로 같은분포로부터 관측된 값이다.
- $X_1, X_2, \cdots, X_n \overset{i.i.d.}{\sim} f$ (or F)
- 이 경우 $f(x_1, \cdots, x_n) = f(x_1) \cdots f(x_n) = \prod_{i=1}^{n} f(x_i)$

확률벡터 -다변량 확률변수

- 각 원소 X_i 가 확률변수인 크기가 $p \times 1$인 벡터 $\boldsymbol{X} = (X_1, \cdots, X_p)^T$를 확률벡터(random vector)라고 부른다.
- 확률벡터의 평균은 다음과 같이 정의한다.

$$E(\boldsymbol{X}) = E\begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \boldsymbol{\mu},$$

여기서 $\mu_i = E(X_i)$.

- 확률벡터 $\boldsymbol{X}$의 공분산 행렬 (covariance matrix), Σ는 다음과 같이 정의한다.
  $cov(\boldsymbol{X}) = E((\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T)$
- $var(X_i) = \sigma_i^2$, $cov(X_i, X_j) = \sigma_{ij}$라고 하고, $\sigma_{ii} = \sigma_i^2$ 라고 하면, 공분산 행렬은 다음과 같이 표현된다.

$$\Sigma = cov(\boldsymbol{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

- Σ^{-1}: Precision matrix

다변량 확률분포의 예 - 다항 분포 (Multinomial Distribution)

- 독립시행에서 나오는 결과 (outcome)가 두가지 이상일 때를 모형화 한 것이다.
- k의 서로 다른 결과가 나오는 독립시행을 n번 시도 하였을때 각각의 결과가 나오는 횟수를 X_j라고 하자. 즉, X_j 는 n번의 독립 시행에서 범주 j가 나온 횟수. $X_1 + \cdots + X_k = n$
- 한번의 시행에서 j번째 범주가 나올 확률을 p_j라고 하자. $p_1 + \cdots + p_k = 1$.
- 이때,각 범주별로 횟수 $(X_1, \cdots, X_k)$ 는 다항분포 (multinomial distribution)을 따르고 다음과 같이 표시한다.
- $\boldsymbol{X} = (X_1, \cdots, X_k) \sim Multi(n, (p_1, \cdots, p_k))$

- 다항분포의 확률질량함수는

$$
\begin{aligned}
p(n_1, \cdots, n_k) &= p(n_1, \cdots, n_k | \, \boldsymbol{p}) \\
&= P(X_1 = n_1, \cdots, X_k = n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k},
\end{aligned}
$$

  $\boldsymbol{p} = (p_1, \cdots, p_k)$.
- 이항분포의 확장으로 볼 수 있다. $k = 2$이면 다항분포는 이항분포와 같다.
- $E(X_j) = np_j$, $var(X_j) = np_j(1 - p_j)$, $cov(X_j, X_{j'}) = -np_j p_{j'}$

디리클레분포 (Dirichlet Distribution)

- 연속 확률분포중의 하나로, $0 \le X_j \le 1$이면서 $\sum_{j=1}^{k} X_j = 1$을 만족하는 확률변수들의 벡터 $\boldsymbol{X} = (X_1, \cdots, X_k)$ $(k \ge 2)$가 다음의 확률밀도함수를 가지는 경우이다.
- $f(x_1, \cdots, x_k) = f(x_1, \cdots, x_k|\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{j=1}^{k} x_j^{\alpha_j - 1}$, $x_j \in [0, 1]$, $\sum_j x_j = 1$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k)$.

  $\alpha_j > 0$은 확률밀도함수를 정하는 모수(parameter)이고,

  $B(\boldsymbol{\alpha}) = \frac{\prod_{j=1}^{k} \Gamma(\alpha_j)}{\Gamma(\sum_j \alpha_j)}$는 정규화 상수 (normalized constant)이다.
- $\boldsymbol{X} \sim Dir(\boldsymbol{\alpha})$
- $E(X_j) = \alpha_j / \sum_i \alpha_i$
- $k = 2$이면 디리클레분포는 베타분포와 같다.

다변량 가우시안 분포 (Multivariate Gaussian Distribution)

- 각 원소가 가우시안 분포 (정규분포)를 따르는 확률벡터의 분포를 다변량 가우시안분포라고 한다.
- 가우시안 확률벡터 (크기 p)의 확률밀도함수는 다음과 같이 정의된다.

$$
\begin{aligned}
& f(x_1, \cdots, x_p) \\
&= f(x_1, \cdots, x_p | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right),
\end{aligned}
$$

  $|\Sigma|$는 Σ의 행렬식 (determinant)이다.
- $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$.

- 각 원소가 표준정규분포이고 서로 독립이면 , $\boldsymbol{Z} \sim N_p(\mathbf{0}, \boldsymbol{I})$로 표현된다. $\boldsymbol{I}$는 단위행렬 (identity matrix)이다.
- Σ는 일반적으로 양의 정 부호 행렬 (positive definite matrix) 이다.
- 양의 정부호 행렬 $\Sigma = \boldsymbol{A}\boldsymbol{A}^T$로 표현된다면 $\boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \Sigma)$.
- $\sigma_{ij} = E((X_i - \mu_i)(X_j - \mu_j)) = 0$이면, 즉 Σ의 (i, j)원소가 0 이면, X_i, X_j는 서로 독립이다.
- 따라서 서로 독립인 가우시안 확률변수로 이루어진 다변량 가우시안 확률벡터의 공분산 행렬은 대각행렬이다. 즉, $\Sigma = \text{diag}(d_1, \cdots, d_p)$.
- $a_1X_1 + \cdots + a_pX_p$ (적어도 하나의 a_i가 0이 아닌경우)는 가우시안분포(정규분포)를 따른다.

- $X_1, \cdots X_p$중에 $k(k \le p)$개의 원소를 뽑아 만든 벡터 $\boldsymbol{X}_s = (X_{i_1}, \cdots, X_{i_k})$도 가우시안분포를 따른다.
- $\boldsymbol{X}_s \sim N_s(\boldsymbol{\mu}_s, \Sigma_s)$, $\boldsymbol{\mu}_s = (\mu_{i_1}, \cdots, \mu_{i_k})^T$, Σ_s의 (l, m) 원소는 σ_{i_l, i_m}이다.
- $p = 2$인 경우, 이변량 가우시안 (bivariate Gaussian) 분포이며, 확률밀도함수는 다음과 같이 표현 할 수도 있다.

$$
\begin{aligned}
& f(x_1, x_2) \\
&= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} \left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} \right.\right. \\
&\qquad \left.\left. + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} \right] \right)
\end{aligned}
$$

분할 가우시안 분포 (Partitioned Gaussian Distribution)

- 가우시안 확률벡터의 일부로 만든 벡터의 평균벡터와 공분산행렬은 원 확률벡터의 평균벡터와 공분산행렬를 분할하여 표현할 수 있다.
- $\boldsymbol{X} = (X_1, \cdots, X_p)^T \sim N_p(\boldsymbol{\mu}, \Sigma)$일때, $\boldsymbol{X} = (\boldsymbol{X}_1^T, \boldsymbol{X}_2^T)^T$로 나누어 진다고 하자. 편의상 $\boldsymbol{X}_1 = (X_1, \cdots, X_m)^T$, $\boldsymbol{X}_2 = (X_{m+1}, \cdots, X_p)^T$라고 하자. 실제로는 순서상관없이 두개의 그룹으로 묶어도 된다.
- 이때, $\boldsymbol{X}_1 \sim N_m(\boldsymbol{\mu}_1, \Sigma_{11})$, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T)^T$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$.

조건부 분할 가우시안 분포

- $\boldsymbol{X}_2 = \boldsymbol{a}$로 주어졌을때 $\boldsymbol{X}_1$의 조건부 확률분포는

  $$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{a} \sim N_m\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{a} - \boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

- $\boldsymbol{X} = (X_1, X_2)$일때, 즉 이변량 가우시안 일때,

  $$X_1 | X_2 = a \sim N\left(\mu_1 + \frac{\sigma_1}{\sigma_2}\rho(a - \mu_2), (1 - \rho^2)\sigma_1^2\right)$$

혼합 분포 (믹스쳐 분포, Mixure Distribution)

- 여러개의 분포의 선형결합으로 이루어진 분포를 혼합분포 (믹스쳐 분포)라고 한다.
- 이산확률분포에서는 k개의 이산확률분포의 선형결합으로 이루어진 다음과 같은 확률질량함수 를 가진다.
  $p(x) = w_1 p_1(x) + \cdots + w_k p_k(x) = \sum_{i=1}^{k} w_i p_i(x)$
- 이때 $p_k(x)$는 확률질량함수이고, $w_i \geq 0$, $\sum w_i = 1$을 만족한다.
- 연속확률분포에서는 다음과 같은 확률밀도함수를 가진다.
  $f(x) = w_1 f_1(x) + \cdots + w_k f_k(x) = \sum_{i=1}^{k} w_i f_i(x)$.

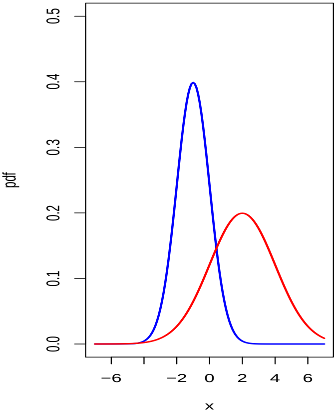
가우시안 혼합 분포 (Gaussian Mixure Distribution)

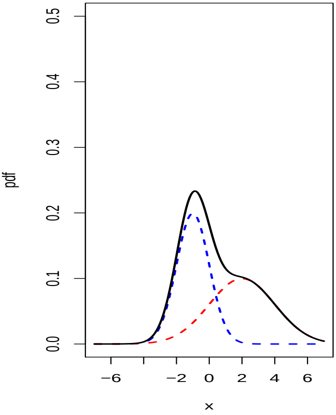
- f_i들이 가우시안 확률밀도함수인 경우 가우시안 혼합 분포라고 한다.
- $\phi(x)$를 표준정규분포의 확률밀도함수라고 하자. 즉, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.
- $X \sim N(\mu, \sigma^2)$인 경우, X의 확률밀도함수는 $\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)$로 표현할 수 있다.
- 이 경우 k개의 구성원을 가지는 가우시안 혼합 분포의 확률밀도함수는 다음과 같이 쓸 수 있다.
- $f(x) = \sum_{i=1}^{k} w_i \frac{1}{\sigma_i}\phi(\frac{x-\mu_i}{\sigma_i})$.

- $k=2$인 경우 $f(x) = w_1 \frac{1}{\sigma_1}\phi(\frac{x-\mu_1}{\sigma_1}) + (1-w_1)\frac{1}{\sigma_2}\phi(\frac{x-\mu_2}{\sigma_2})$
- $X_1, \cdots, X_n \overset{i.i.d.}{\sim} f(x) = \sum_{i=1}^{k} w_i \frac{1}{\sigma_i}\phi(\frac{x-\mu_i}{\sigma_i})$, 즉, 가우시안 혼합 분포를 따르는 랜덤 추출된 데이터가 있다고 할때, 각 X_j는 w_i의 확률로 $N(\mu_i, \sigma_i^2)$을 따른다고 해석할 수 있다.
- 군집분석의 모델로 사용할 수 있다.

가우시안 혼합 분포 예

- 왼쪽: 파란선 $N(-1, 1^2)$, 빨간선 $N(2, 2^2)$
- 오른쪽: 파란점선 $0.5 \times N(-1, 1^2)$, 빨간점선 $0.5 \times N(2, 2^2)$
  까만선: $0.5 \times N(-1, 1^2) + 0.5 \times N(2, 2^2)$





지수족 (Exponential Family)

- 지수족 혹은 지수 분포족은 확률분포들의 집합으로 확률밀도함수 또는 확률질량함수를 다음과 같이 표현할 수 있는 분포들의 집합이다.

$$\begin{aligned} f(x) &= f(x|\theta) = h(x)g(\theta)\exp(\eta(\theta)T(x)) \\ &= h(x)\exp\left(\eta(\theta)T(x) - A(\theta)\right) \end{aligned}$$

- 모수 θ가 벡터인 경우, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_s)^T$,

$$\begin{aligned} f(x|\boldsymbol{\theta}) &= h(x)g(\boldsymbol{\theta})\exp(\boldsymbol{\eta}(\boldsymbol{\theta})^T \boldsymbol{T}(x)), \\ &= h(x)\exp(\boldsymbol{\eta}(\boldsymbol{\theta})^T \boldsymbol{T}(x) - A(\boldsymbol{\theta})), \end{aligned}$$

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \cdots, \eta_s(\boldsymbol{\theta}))^T,\ \boldsymbol{T}(x) = (T_1(x), \cdots, T_s(x))^T$$

- 이미 배운 많은 분포들이 지수족에 속한다.
- 베르누이분포, 이항분포, 포아송분포, 베타분포, 지수분포, 감마분포, 정규분포 등등...
- 믹스쳐분포, t-분포, F-분포등은 지수족에 포함되지 않는다.

지수족 예

- 베르누이분포
- $p(x) = \theta^x(1-\theta)^{1-x} = \exp\left([\log(\theta/(1-\theta))]x + \log(1-\theta)\right),$

  $\eta(\theta) = \log(\theta/(1-\theta))$, $T(x) = x$, $A(\theta) = -\log(1-\theta)$,
  $h(x) = 1$.
- 포아송 분포
- $p(x) = \frac{\theta^x}{x!}e^{-\theta} = \frac{1}{x!}\exp([\log\theta]x - \theta),$

  $\eta(\theta) = \log\theta$, $T(x) = x$, $A(\theta) = \theta$, $h(x) = \frac{1}{x!}$

- 정규분포 (분산을 알때, 즉, 모수는 평균 μ하나일때)
- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} =$
  $\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) \exp\left(\frac{\mu}{\sigma}\frac{x}{\sigma} - \frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2\right)$,

  $\eta(\mu) = \frac{\mu}{\sigma}$, $T(x) = \frac{x}{\sigma}$, $A(\mu) = \frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2$,
  $h(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$

- 정규분포 (모수는 평균, 분산으로 이루어진 벡터일때 $\boldsymbol{\theta} = (\mu, \sigma^2)^T$)
- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} =$
  $\frac{1}{\sqrt{2\pi}} \exp\left(\frac{\mu}{\sigma^2}x - \frac{1}{2\sigma^2}x^2 - \frac{1}{2}\frac{\mu^2}{\sigma^2} - \frac{1}{2}\log(\sigma^2)\right),$
  $\eta(\boldsymbol{\theta}) = (\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2})^T$, $T(x) = (x, x^2)^T$,
  $A(\boldsymbol{\theta}) = \frac{1}{2}\frac{\mu^2}{\sigma^2} + \frac{1}{2}\log(\sigma^2)$, $h(x) = \frac{1}{\sqrt{2\pi}}$

데이터가 여러개 있을때 (2)

- 데이터가 서로 독립이 아닐때, 예를들어 시간에 따라 관측되는 시계열 데이터를 생각해보자.
- $X_1, \cdots, X_{n-1}$이 시간 $t = 1, \cdots, n-1$일때의 값이라고 하면, 과거의 데이터 $X_1, \cdots, X_{n-1}$가 주어졌을때 그 다음 시간 $t = n$ 일때의 값인 X_n의 분포는 $f(x_n|x_{n-1}, \cdots, x_1)$ 로 표현될수 있다.
- 또한, 조건부 확률분포의 성질을 이용하면 결합확률분포는 다음과 같이 표현할 수 있다.

$$f(x_1, \cdots, x_n) = f(x_n|x_{n-1}, \cdots, x_1)f(x_{n-1}|x_{n-2}, \cdots, x_1) \times \cdots \times f(x_2|x_1)f(x_1)$$