2017/07/19, 2017 확률 세미나

확률 및 통계학

- 3장 확률변수와 확률분포 -

송 영 준(youngjun@pel.smuc.ac.kr)

상명대학교 프로토콜공학연구실

목 차

- 확률변수 개념
- 이산형 확률분포
- 연속형 확률분포
- 결합확률분포

확률변수 개념

- 확률변수(Random Variable)
  - 표본공간 내에 있는 각 원소에 하나의 실수 값을 대응시키는 함수
  - 변수가 취하는 값에 확률이 대응하고 있는 경우
  - 일정한 확률을 갖고 발생하는 사건(Event)에 수치가 부여되는 변수
  - 주로 대문자 X로 표현, 그에 대응하는 하나의 값은 소문자 x로 표현
  - 종류
    - 이산형 확률변수(Discrete Random Variable)
      - 확률변수들의 집합이 셀 수 있는 집합인 경우
    - 연속형 확률변수(Continuous Random Variable)
      - 확률변수가 연속적인 구간 내의 값을 취하는 경우

확률변수 개념

- 이산형 확률변수 예제
  - 4개의 붉은 공(R)과 3개의 검은 공(B)이 들어 있는 항아리에서 연속적으로 2개의 공을 비복원추출하는 실험에서 X를 붉은 공의 개수라 할 때, 출현 가능한 결과와 확률변수 X의 값 x

표본공간	x
RR	2
RB	1
BR	1
BB	0

- 연속형 확률변수 예제
  - 과속탐지 카메라에 적발되는 과속 차량들 사이의 시간간격을 확률변수 X라고 하면 X의 값 x의 범위는 $x \geq 0$

확률변수 개념

- 표본공간(Sample Space)
  - 통계적 실험에서 발생 가능한 모든 결과들의 집합
  - 기호 S로 표시
  - 종류
    - 이산표본공간(Discrete Sample Space)
      - 표본공간이 유한 개 혹은 셀 수 있는 무한 개의 원소로 이루어졌을 경우
      - 계수자료(Count Data)
        - 사고횟수, 불량품횟수, 발생건수
    - 연속표본공간(Continuous Sample Space)
      - 표본공간의 실선의 어떤 구간 내의 모든 수를 포함하는 경우
      - 측정자료(Measured Data)
        - 높이, 무게, 온도, 거리, 수명

이산형 확률분포

- 확률분포
  - 확률변수가 특정한 값을 가질 확률을 나타내는 함수를 의미
  - 확률변수가 어떤 종류의 값을 가지는가에 따라서 이산 확률분포와 연속 확률분포로 나뉨

- 이산형 확률분포(Discrete Probability Destribution)
  - 이산형 확률변수 X의 모든 값과 이에 대응하는 확률을 표나 그래프로 나타낸 것을 X의 이산형 확률분포라 함
  - 확률 질량 함수를 통하여 표현

이산형 확률분포

- 이산형 확률분포 예제
  - 어느 대리점에서 판매된 외제차의 50%에 디젤엔진이 장착되었다고 할 때, 이 대리점에서 다음에 판매될 4대의 외제차 가운데 디젤엔진이 장착된 차의 수의 확률분포에 대한 식을 구하라

x	0	1	2	3	4
$f(x)$	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$

$DDDD$					
$DDDG$	$DDGD$	$DGDD$	$GDDD$		
$DDGG$	$DGGD$	$GDDG$	$DGDG$	$GGDD$	$GDGD$
$DGGG$	$GDGG$	$GGDG$	$GGGD$		
$GGGG$					

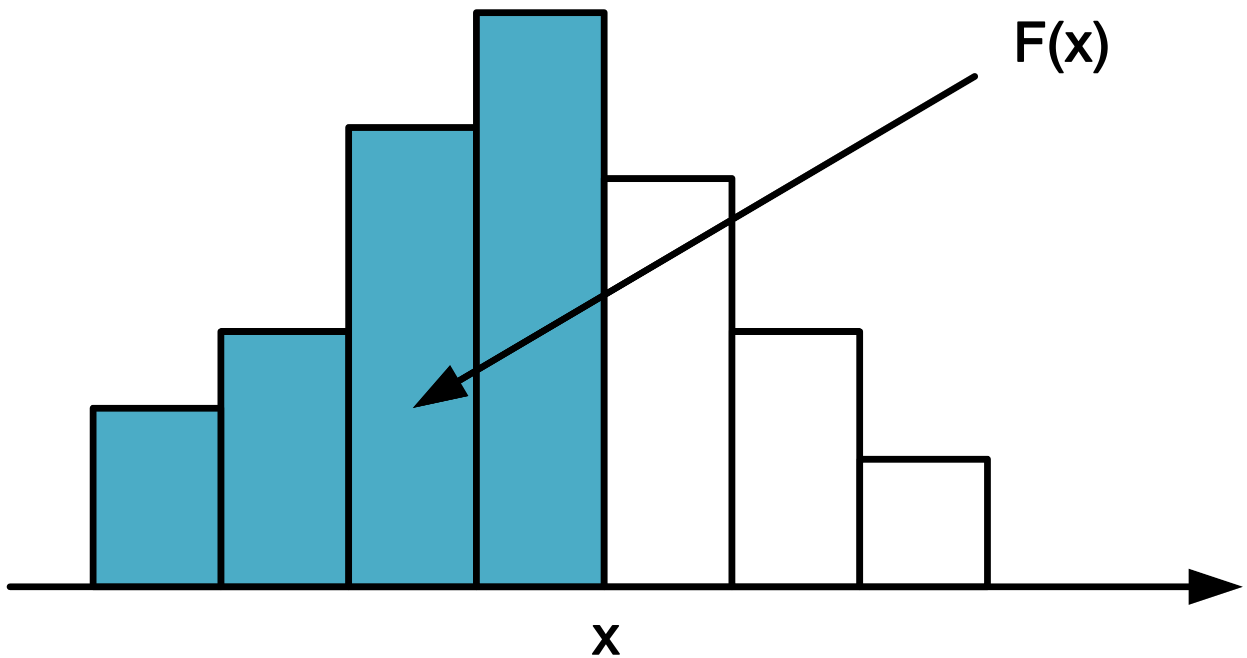
- 발생확률이 동일한 표본점 : $2^{14} = 16$
- $f(x) = \frac{\binom{4}{x}}{16}, x = 0,1,2,3,4$

이산형 확률분포

- 확률질량함수(Probability Mass Function)
  - 모든 x에 대해 순서쌍 $(x, f(x))$의 집합이 다음 조건을 만족하면, 이를 이산형 확률변수 X의 확률함수, 확률질량함수, 확률분포라 함
    - $f(x) \geq 0$
    - $\sum_x f(x) = 1$
    - $P(X = x) = f(x)$
- 누적분포함수(Cumulative Distribution Function)
  - 확률변수 X가 특정한 값 x를 넘지 않을 확률을 나타내는 함수 $F(x)$를 누적분포함수라 함
  - 확률분포 $f(x)$를 가지는 이산형 확률변수 X의 누적분포함수 $F(x)$
    - $F(x) = P(X \leq x) = \sum_{t \leq x} f(t), -\infty < x < \infty$

이산형 확률분포

- 누적 분포함수(Cumulative Distribution Function)



이산형 확률분포

- 누적분포 함수 예제
  - 어느 대리점에서 판매된 외제차의 50%에 디젤엔진이 장착되었다고 할 때, 이 대리점에서 다음에 판매될 4대의 외제차 가운데 디젤엔진이 장착된 차의 수의 확률분포에 대한 식을 구하라. 그리고 F(x)를 사용하여 f(2)=3/8이 됨을 증명하라

x	0	1	2	3	4
$f(x)$	1/16	4/16	6/16	4/16	1/16

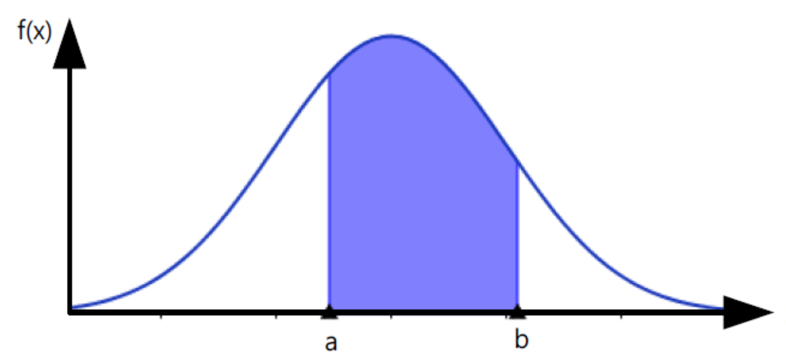
- $F(0) = f(0) = \frac{1}{16}$
- $F(1) = f(0) + f(1) = \frac{1}{16} + \frac{4}{16} = \frac{5}{16}$
- $F(2) = f(0) + f(1) + f(2) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} = \frac{11}{16}$
- $F(3) = f(0) + f(1) + f(2) + f(3) = \frac{1}{16} + \frac{4}{16} + \frac{6}{16} + \frac{4}{16} = \frac{15}{16}$
- $F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1$

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{16}, & 0 \le x < 1 \\ \frac{5}{16}, & 1 \le x < 2 \\ \frac{11}{16}, & 2 \le x < 3 \\ \frac{15}{16}, & 3 \le x < 4 \\ 1, & x \ge 4 \end{cases}$$

- 따라서, $f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{6}{16} = \frac{3}{8}$

연속형 확률분포

- 연속형 확률분포(Continuous Probability Distribution)
  - 연속확률변수가 가지는 확률분포
  - 확률변수 값의 어느 한 점보다는 어떤 구간에 더 관심을 가지는 경우
  - 확률밀도함수를 이용해 분포를 표현
- 확률밀도함수(Probability Density Function)
  - 다음 조건이 만족하면 $f(x)$를 실수의 집합 R상에서 정의된 연속형 확률변수에 대한 확률밀도함수라 함
    - 모든 $x \in R$에 대하여 $f(x) \geq 0$
    - $\int_{-\infty}^{\infty} f(x)dx = 1$
    - $P(a < X < b) = \int_{a}^{b} f(x)dx$

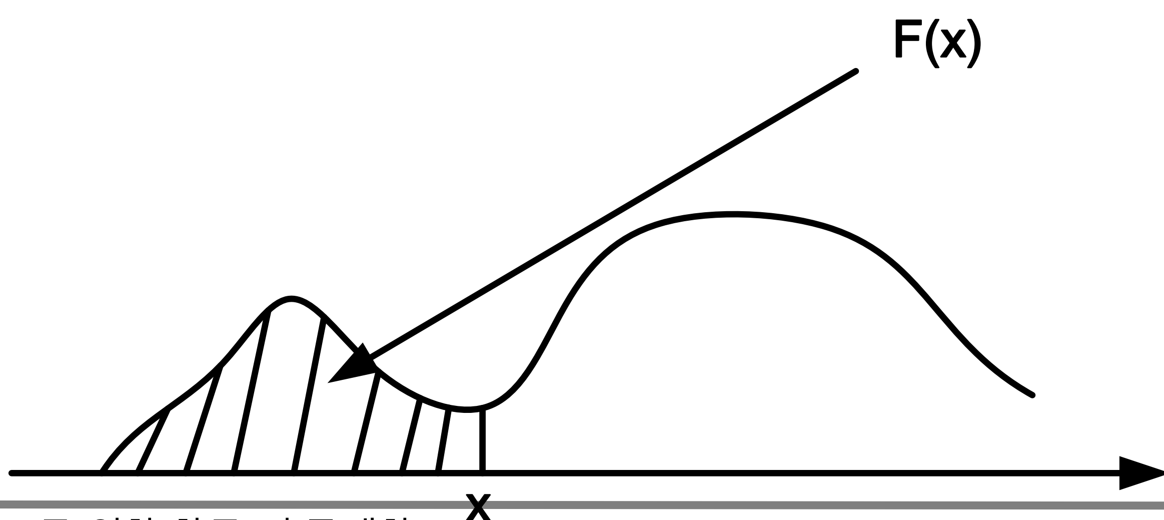


연속형 확률분포

- 연속형 확률분포 예제
  - 제어실험에서 반응온도(˚C)의 변화에 따른 오차는 다음과 같은 확률분포를 가지는 연속확률변수 X라고 가정하자
    - $f(x) = \begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, \text{다른 곳에서} \end{cases}$
    - (a) $f(x)$가 확률밀도함수임을 증명하라
      - $f(x) \geq 0$임은 명확하고, $f(x)$가 확률밀도함수가 됨은 다음과 같이 확인 할 수 있다.
      - $\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{2} \frac{x^2}{3} dx = \frac{x^3}{9} \Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$
    - (b) P(0<X<=1)을 구하라
      - $P(0 < X \leq 1) = \int_{0}^{1} \frac{x^2}{3} dx = \frac{x^3}{9} \Big|_{0}^{1} = \frac{1}{9}$

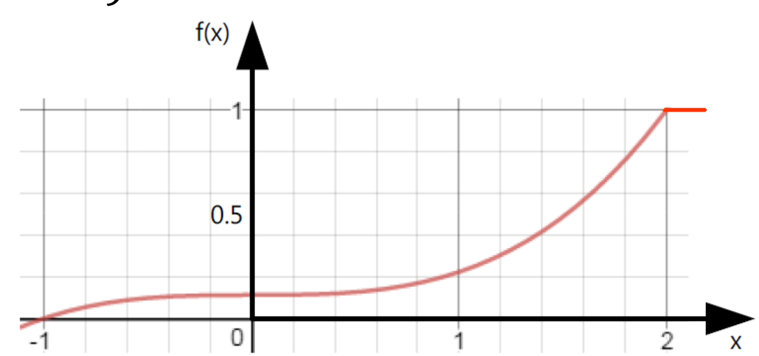
연속형 확률분포

- 연속형 누적분포함수
  - 확률밀도함수가 $f(x)$인 연속형 확률변수 X의 누적분포함수 $F(x)$
    - $F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(t)dt, -\infty < x < \infty$
    - $P(a < X < b) = F(b) - F(a)$
    - 미분 가능하면, $f(x) = \frac{dF(x)}{dx}$



연속형 확률분포

- 연속형 누적분포함수 예제
  - 제어실험에서 반응온도(° C)의 변화에 따른 오차는 다음과 같은 확률분포를 가지는 연속확률변수 X라고 가정하자
    - $f(x) = \begin{cases} \frac{x^2}{3}, -1 < x < 2 \\ 0, \text{다른 곳에서} \end{cases}$
  - 위 확률밀도함수에 대하여 $F(x)$를 구하고, 그것을 이용하여 $P(0 < X \leq 1)$을 구하라
    - $F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} \frac{t^2}{3} dt = \frac{t^3}{9} \Big|_{-1}^{x} = \frac{x^3+1}{9}$
    - $\therefore F(x) = \begin{cases} 0, \ x < -1 \\ \frac{x^3+1}{9}, \ -1 \leq x < 2 \\ 1, \ x \geq 2 \end{cases}$



결합확률분포

- 결합확률분포(Joint Probability Distribution)
  - 여러 개의 확률변수들의 결과를 동시에 취급하는 경우
  - 두 확률변수 X, Y의 모든 값과 이에 대응하는 확률을 표나 그림으로 나타낸 것
    - e.g., 화학실험 중 침전물의 양 P와 방출되는 가스 부피 V를 측정하면 (p, v)로 구성되는 2차원 표본공간을 얻게 됨
  - 이산형 결합확률분포
    - X와 Y가 두 이산형 확률변수라 할 때, 두 변수의 확률분포는 확률변수 X와 Y의 범위 내에서 어떤(x, y)에대한 $f(x, y)$를 값으로 가지는 함수로 표시 됨
    - $f(x, y)$는 x와 y가 동시에 일어날 확률

결합확률분포

- 결합확률분포(Joint Probability Distribution)
  - 이산형 결합확률분포
    - 결합확률질량함수(Joint Probability Mass Function)
      - 다음 조건이 만족될 때 함수 $f(x, y)$를 이산형 확률변수 X와 Y의 결합 확률질량함수라 함
        - 모든 (x, y)에 대하여 $f(x, y) \geq 0$
        - $\sum_x f(x, y) \sum_y f(x, y) = 1$
        - $P(X = x, Y = y) = f(x, y)$
        - x, y평면상의 어떤 영역 A에 대하여, $P[(x, y) \in A] = \sum \sum_A f(x, y)$

결합확률분포

- 결합확률분포(Joint Probability Distribution)
  - 연속형 결합확률분포
    - 결합밀도함수(Joint Density Function)
      - 다음 조건이 만족될 때 함수 $f(x, y)$를 연속확률변수 X와 Y의 결합밀도함수라 함
        - 모든 (x, y)에 대하여 $f(x, y) \geq 0$
        - $\int_{-\infty}^{\infty} f(x, y) \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$
        - $P[(X, Y) \in A] = \iint_A f(x, y) dx dy$, 이 때 A는 xy평면상의 임의의 영역

- 결합확률분포 예제
  - 3개의 청색, 2개의 적색, 3개의 녹색볼펜이 들어 있는 상자에서 임의로 2개를 추출하고자 할 때, X를 청색볼펜의 수, Y를 적색볼펜의 수라고 하자
    - 결합확률분포 $f(x,y)$를 구하라

$f(x,y)$		x			행의 합
		0	1	2	
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14		3/7
	2	1/28			1/28
열의 합		5/14	15/28	3/28	1

    - 발생확률이 동일한 표본점 : $\binom{8}{2} = 28$
    - $f(x,y) = \dfrac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$

결합확률분포

- 결합확률분포 예제
  - (b) $A = \{(x, y) | x + y \leq 1\}$이라고 할 때, $P[(X, Y) \in A]$를 구하라
    - $\{(x, y) | x + y \leq 1\}$를 만족하는 경우 : $f(0,0), f(0,1), f(1,0)$
    - $f(0,0) + f(0,1) + f(1,0) = \frac{3}{28} + \frac{6}{28} + \frac{9}{28} = \frac{18}{28}$

결합확률분포

- 주변분포(Marginal Distribution)
  - 결합확률분포로부터 각 각의 이산확률분포 또는 연속확률분포를 구하는 경우
  - 결합확률변수$P(X, Y)$의 결합되기 전 함수인 $P(X), P(Y)$를 의미
  - 이산형 확률변수 X와 Y의 결합확률분포 $f(x, y)$가 주어진 경우
    - X만의 확률분포$g(x)$는 $f(x, y)$를 Y의 모든 값에 대해 합하면 얻을 수 있음
      - 연속인 경우는 합 대신 적분을 적용
    - Y만의 확률분포 $h(y)$는 $f(x, y)$를 X의 모든 값에 대해 합하면 얻을 수 있음
      - 연속인 경우 합 대신 적분을 적용
    - $g(x)$와$h(y)$를 각 각 X와 Y의 주변분포

결합확률분포

- 주변분포(Marginal Distribution)
  - X, Y의 주변분포
    - 이산
      - $g(x) = \sum_y f(x, y)$, $h(y) = \sum_x f(x, y)$
    - 연속
      - $g(x) = \int_{-\infty}^{\infty} f(x, y)dy$, $h(y) = \int_{-\infty}^{\infty} f(x, y)dx$

결합확률분포

- 주변분포 예제

$f(x,y)$		x			행의 합
		0	1	2	
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14		3/7
	2	1/28			1/28
열의 합		5/14	15/28	3/28	1

- 결합확률분포 $f(x,y)$의 $g(x), h(y)$구하는 방법
  - $g(x) = \sum_y f(x,y)$
    - $g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{5}{14}$
    - $g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{15}{28}$
    - $g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28}$
    - 따라서 $g(x) = g(0) + g(1) + \ g(2) = \frac{5}{14} + \frac{15}{28} + \frac{3}{28} = 1$

결합확률분포

- 주변분포 예제

$f(x,y)$		x			행의 합
		0	1	2	
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14		3/7
	2	1/28			1/28
열의 합		5/14	15/28	3/28	1

- 결합확률분포 $f(x,y)$의 $g(x), h(y)$구하는 방법
  - $h(y) = \sum_x f(x,y)$
    - $h(0) = f(0,0) + f(1,0) + f(2,0) = \frac{15}{28}$
    - $h(1) = f(0,1) + f(1,1) + f(2,1) = \frac{3}{7}$
    - $h(2) = f(0,2) + f(1,2) + f(2,2) = \frac{1}{28}$
    - 따라서 $h(x) = h(0) + h(1) + h(2) = \frac{15}{28} + \frac{3}{7} + \frac{1}{28} = 1$

결합확률분포

- 조건부 분포(Conditional Distribution)
  - 조건부 확률(Conditional Probability)
    - 어떤 사건이 전제되어 있을 때 다른 사건의 확률을 구하는 것
    - 사건 A가 일어났다는 전제 조건에서 사건 B가 일어날 확률
      - 즉, 어떤 조건하에서 추가적으로 일어날 사건의 확률을 의미
      - $P(B|A)$식으로 표기
        - 사건 A가 일어났을 때 B가 일어날 확률
        - $P(B|A) = \frac{P(A \cap B)}{P(A)}, P(A) > 0$

결합확률분포

- 조건부 분포(Conditional Distribution)
  - X와 Y를 이산형 또는 연속형인 두 확률변수라고 할 때, $X = x$로 주어졌을 때 확률변수 Y의 조건부 분포
    - $f(y|x) = \frac{f(x,y)}{g(x)}, g(x) > 0$
  - 같은 방법으로 $Y = y$로 주어졌을 때 확률변수 X의 조건부 분포
    - $f(x|y) = \frac{f(x,y)}{h(y)}, h(y) > 0$

결합확률분포

- 조건부 분포(Conditional Distribution)
  - 이산형 확률변수 $Y = y$라고 주어졌을 때 이산형 확률변수 X가 a와 b 사이의 값을 가질 확률
    - $P(a < X < b|Y = y) = \sum_{a<x<b} f(x|y)$
  - X와 Y가 연속형인 경우
    - $P(a < X < b|Y = y) = \int_a^b f(x|y)dx$

결합확률분포

- 조건부 분포 예제
  - $Y = 1$로 주어졌을 때 X의 조건부 분포를 구하고 그것을 이용하여 $P(X = 0|Y = 1)$을 구하라
    - $f(x|y) = \frac{f(x,y)}{h(y)}$
    - $y = 1$일 때, $f(x|y)$가 필요하므로 $h(1)$를 구해야 함
      - $h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$
      - $f(x|1) = \frac{f(x,1)}{h(1)} = \frac{7}{3} f(x,1), x = 0,1,2$
        - $f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}$
        - $f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}$
        - $f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right)(0) = 0$
      - $\therefore P(X = 0|Y = 1) = f(0|1) = \frac{1}{2}$

x	0	1	2
$f(x\|1)$	1/2	1/2	0

$f(x,y)$		x			행의 합
		0	1	2	
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14		3/7
	2	1/28			1/28
열의 합		5/14	15/28	3/28	1

결합확률분포

- 통계적 독립(Statistically Independent)
  - X와 Y를 결합확률분포 $f(x,y)$와 주변분포 $g(x), h(y)$를 가지는 이산형 혹은 연속형 확률변수라 할 때, 모든(x,y)에 대하여 $f(x,y) = g(x)h(y)$가 성립하면 확률변수 X와 Y는 통계적으로 독립이라 함
  - 결합밀도함수를 통한 증명
    - $f(x,y)$가 y에 종속되어 있지 않으면, $f(x|y) = g(x)$이고 $f(x,y) = g(x)h(y)$
      - $f(x,y) = f(x|y)h(y)$
      - $g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} f(x|y)h(y)dy$
      - $g(x) = f(x|y)\int_{-\infty}^{\infty} h(y)dy$
      - $\int_{-\infty}^{\infty} h(y)dy = 1$, 따라서 $g(x) = f(x|y)$
      - $f(x,y) = g(x)h(y)$

결합확률분포

- 통계적 독립 예제

$f(x,y)$		x			행의 합
		0	1	2	
y	0	3/28	9/28	3/28	15/28
	1	3/14	3/14		3/7
	2	1/28			1/28
열의 합		5/14	15/28	3/28	1

- 위 예제가 통계적 독립이 아님을 증명하라
  - 점(0,1)
  - $f(0,1) = \frac{3}{14}$
  - $g(0) = \sum_{y=0}^{2} f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$
  - $\mathrm{h}(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$
  - $f(0,1) \neq g(0)h(1)$
  - ∴따라서 X와 Y는 통계적으로 독립이 아님

감사합니다!