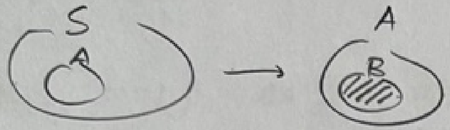
기초통계학 중간고사

학번 : 2011724570 이름 : 정석규

1. 조건부 확률 : 사건A가 발생했다는 전제하에 사건 B가 발생할 확률은

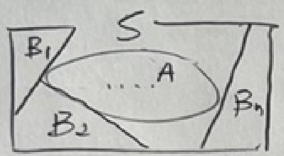
$P(B|A) = \dfrac{P(A \cap B)}{P(A)}$, $P(A) > 0$ 으로 정의할 수 있으며,



전체 표본공간을 A로 축소시켰을 때 $A \cap B$의 사건이 발생과 의미가 같다.

전확률의 정리 : $S = \bigcup_{i=1}^{n} B_i$. $B_i \cap B_j = \emptyset$. $i \neq j$ 의 경우 $P(B_i) > 0$ 일때.

$P(A) = \sum_{i=1}^{n} P(A|B_i) P(B_i)$ 이다.



다음의 그림에서 집합A는 $(A \cap B_1) \cup \cdots \cup (A \cap B_n)$ 으로 정리가 가능하며 $P(A) = \sum_{i=1}^{n} P(A \cap B_n)$ 으로 표현이 가능하다. 조건부 확률의 정의에 따라 $P(A \cap B_n)$은 $P(A|B_i) P(B_i)$ 로 표현이 가능하다.

베이즈 정리 : $S = \bigcup_{i=1}^{n} B_i$. $B_i \cap B_j = \emptyset$, $i \neq j$의 경우

$$P(B_k|A) = \frac{P(B_k) P(A|B_k)}{\sum_{i=1}^{n} P(B_i) P(A|B_i)} \quad (k = 1, \cdots, m)$$

$P(B_k|A)$는 $P(B_k \cap A) / P(A)$ 이며 전확률의 정의에 따라

$\dfrac{P(B_k \cap A)}{\sum_{i=1}^{n} P(A|B_i) P(B_i)}$. 이며 . 분자는 조건부 확률의 정의에 따라 $\dfrac{P(B_k) P(A|B_k)}{\sum_{i=1}^{n} P(A|B_i) P(B_i)}$

로 표현가능하다

2. 수학적 기댓값 : 확률변수 X가 $f(x)$에 대하여 $E(u(x)) = \sum_{x \in R} u(x) f(x)$의 값이 존재할때의 $E(u(x))$. 각 사건에 대한 변수와 사건에 부여된 확률을 곱한것을 전체사건에 대하여 합한값

평균 : $u(x) = x$ 일 때의 수학적 기댓값 $E(x) = \sum_{x \in R} x f(x) = \int_{-\infty}^{\infty} x f(x) dx$

분산 : $u(x) = (X - \mu)^2$ 일때의 수학적 기댓값으로, 편차제곱의 기댓값과 동일한 의미이다.

$Var(x) = E((X - \mu)^2) = \sum_{R} (X - \mu)^2 f(x) = E(X^2) - E(X)^2$

3. 적률생성함수 : X의 거듭제곱의 기댓값을 X의 1차 적률로 정의 할때,

$M^{(r)}(0) = E(X^r)$ 이 되는 함수 $M(t)$를 적률생성 함수라고 하며,

$M(t) = E(e^{tX}) = \sum_R e^{tx} f(x) = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$ 로 계산이 가능하다.

적률 생성함수를 통해 평균 $E(X) = M'(0)$ 으로,

분산 $Var(X) = M''(0) - M'(0)^2$ 으로 쉽게 계산이 가능하다.

누가적률 : 적률생성함수 $M(t)$에 ln 함수를 적용한 값은 누가적률이라 하며,

$R(t) = \ln M(t)$ 로 정의한다.

$R'(t) = \frac{M'(t)}{M(t)}$ 이므로 $R'(0) = M'(0) = \mu$

$R''(t) = \frac{M(t)M''(t) - M'(t)^2}{(M(t))^2}$ 으로 $R''(0) = Var(X)$ 가 되어

평균과 분산을 쉽게 계산이 가능하다.

4. 베르누이 분포 : 결과가 두가지 중 하나로만 나오는 실험이나 시행을 베르누이 시행이라고 할때, 확률변수 $X = 0, 1$ 즉, 단한번의 실행에 의하여 결정되는 베르누이 시행의 확률 분포는 $f(x) = p^x (1-p)^x$ $(x = 0, 1,\ 0 \le p \le 1)$ 이다.

이항분포 : 성공할 확률을 p, 실패할 확률이 $1-p$인 베르누이 시행을 독립적으로 n번 시행하였을 때, X번 성공하는 확률변수를 이항분포라고 하며,

$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $\quad x = 0, 1, \cdots, n$ 으로 표현한다.

초기하 분포 : 비복원추출에서 N개중에 n개를 뽑았을 때, 원하는 집합 k에서 x개가 뽑힐 확률의 분포이다.

$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ 의 확률 분포 함수를 가진다.

기하분포 : 성공할 확률이 p인 베르누이 실험에 대하여 처음 성공이 일어 날때까지의 실패횟수를 나타내는 분포로 $f(x) = (1-p)^x p$ 의 분포를 가진다.

음이항 분포 : 성공확률이 p인 베르누이 실험에 대하여 r회 성공이 날때까지 실패 횟수를 X 라고 하면, $r-1$ 번 성공이 나타날 경우의 수는 $\binom{x+r-1}{r-1}$ 로 $\binom{x+r-1}{x}$ 와 같은 값을 가진다.

$$f(x) = \binom{x+r-1}{x} p^r q^x \quad (x = 0, 1, \cdots) \sim Nb(r, p)$$ 의

확률 밀도 함수를 가진다.

포아송 분포 : 단위시간 안에 어떤 사건이 몇번 발생할 것인지를 표현하는 분포로. 어떤 단위구간 의 사건은 다른 단위구간에서의 발생으로부터 독립적이고, 둘 이상의 사건이 동시에 발생할 확률이 0이며, 겹치지 않는 구간에서 일어나는 사건의 수가 독립적이고, 어느 부분에서나 사건이 동일하게 발생할때 정의 된다. $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$ $(\lambda > 0)$ 의 확률 밀도 함수를 가진다.

다항분포 : 여러개의 값을 가지는 독립확률변수에 대하여 여러번의 독립 시행 에 각각의 값이 특정 횟수가 나타날 확률

$$f(x_1, \cdots, x_n) = \frac{n!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_k^{x_k} \quad \Sigma x_i = n, \ \Sigma p_i = 1$$

의 확률 밀도 함수를 가진다.

5. 균일분포 : 특정구간에서 ([a,b]) 균등한 값을 가지는 연속분포.

$$f(x) = \frac{1}{b-a} \quad a \le x < b$$ 의 함수를 가진다.

지수분포 : 사건이 독립적일때, 일정시간 발생하는 사건의 횟수가 푸아송 분포를 따른다면 다음 사건이 발생하기 까지 걸리는 시간의 분포.

$$f(x) = \lambda e^{-\lambda t} \quad x > 0,\ \lambda > 0$$ 의 pdf 를 가진다.

정규분포 : X의 pdf가 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ $-\infty < x, \mu < \infty$, $0 < \sigma < \infty$

를 가질때 모수가 $N(\mu, \sigma^2)$인 정규분포를 따른다고 한다

표준정규분포 : 모수가 $N(0, 1)$인 정규분포.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$ 의 pdf를 가진다.

6. 결합 확률 밀도 함수 : 확률 변수가 여러개 일때, 이를 함께 고려하는 확률 분포로.

$f(x, y)$로 표시하며, $0 \le f(x, y) \le 1$. $\sum\sum_{(x, y \in R)} f(x, y) = 1$

$P((X, Y) \in A) = \sum\sum_{(x, y \in A)} f(x, y)$ 의 성질을 가진다.

주변 밀도함수 : 결합 밀도함수 $f(x, y)$에 대해 $f(x), f(y)$의 개별적인

확률 밀도함수를 말한다.

$f(x) = \sum_y f(x, y) = \int_{-\infty}^{\infty} f(x, y)\,dy$

$f(y) = \sum_x f(x, y) = \int_{-\infty}^{\infty} f(x, y)\,dx$ 로 계산한다.

조건부 확률밀도 함수 : 조건부 확률 정의 $P(B|A) = \frac{P(A \cap B)}{P(A)}$ 에 대하여

$A = \{X = x\}$, $B = \{Y = y\}$ 라고 하면

$g(y|x) = \frac{f(x, y)}{f(x)}$ 로 표현이 가능하다.

7. 독립의 필요 충분조건 : $f(x, y) = f(x) f(y)$

척도 1) 공분산 : $Cov(X, Y)$로 표시하며 $E((X - \mu_x)(Y - \mu_Y))$로 계산한다.

확률변수 2개의 선형관계를 나타낸다.

2) 상관계수 : $\rho_{X,Y} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$ 로 나타내며 -1에서 +1까지의 값을 가진다.