3. 확률의 성질

[정의 1.3-1] 확률이란 표본공간 S에 있는 각 사건과 실수 $P(A)$가 대응하는 집합함수로 정의되며 다음의 성질을 만족한다.

(a) $0 \le P(A) \le 1$

(b) $P(S) = 1$

(c) $A_1, A_2, A_3, \cdots$가 $A_i \cap A_j = \varnothing$, $i \ne j$를 만족하는 사건이면, $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$ 즉, $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$.

성질 (c)를 가산 가법성 (Countable additivity)이라고 부르기도 하며, 사건 $A_1, A_2, \cdots, A_n$이 서로 배반사건일 경우 다음이 성립한다.

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n)$$

[정리 1.3-1] 각 사건 A에 대하여 $P(A) = 1 - P(A^c)$

증명> $S = A \cup A^c$이고 $A \cap A^c = \varnothing$이므로 성질 (b)와 (c)로부터 $1 = P(A) + P(A^c)$이다.

[정리 1.3-2] $P(\varnothing) = 0$

증명> 앞의 [정리 1.3-1]에서 $A = \varnothing$이라 두면 $A^c = S$이다. 따라서,

$$P(\varnothing) = 1 - P(S) = 1 - 1 = 0$$

[정리 1.3-3] $A \subset B$이면 $P(A) \le P(B)$이다.

증명> $B = A \cup (B \cap A^c)$이고 $A \cap (B \cap A^c) = \varnothing$이므로 성질 (c)로부터 $P(B) = P(A) + P(B \cap A^c) \ge P(A)$ (성질 (a)에서 $P(B \cap A^c) \ge 0$이므로)를 얻는다.

[정리 1.3-4] 각 사건 A에 대하여 $0 \le P(A) \le 1$이다.

증명> $0 \subset A \subset S$이므로 [정리 1.3-3]에 의하여 $P(\varnothing) \le P(A) \le P(S)$이다. $P(\varnothing) = 0$, $P(S) = 1$.

[정리 1.3-5] 각 사건 A에 대하여 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

증명> $A \cup B = A \cup (A^c \cap B)$이고 $A \cap (A^c \cap B) = \varnothing$이므로 성질(c)에 의하여 $P(A \cup B) = P(A) + P(A^c \cap B)$--(*)이다. 또한, $B = (A \cap B) \cup (A^c \cap B)$이고 $(A \cap B) \cap (A^c \cap B) = \varnothing$이므로 $P(B) = P(A \cap B) + P(A^c \cap B)$이다. $P(A^c \cap B) = P(B) - P(A \cap B)$을 (*)에 대입하면 위의 식이 성립한다.

[정리 1.3-6] 사건 A, B, C에 대하여 다음이 성립한다.

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

증명> $A \cup B \cup C = A \cup (B \cup C)$로 두고 [정리 1.3-5]를 적용시키면 된다.

[참고: 정리 1.3-5, 1.3-6 일반화]

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} \cap E_{i_2}) + \cdots + (-1)^{r+1} \sum_{i_1 < \cdots < i_r} P(E_{i_1} \cap \cdots \cap E_{i_r})$$
$$+ \cdots + (-1)^{n+1} \sum_{i_1 < \cdots < i} P(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_n)$$

위 식에서 $\sum_{i_1 < \cdots < i_r} P(E_{i_1} \cap \cdots \cap E_{i_r})$은 집합 $\{1, 2, \cdots, n\}$에서 $\binom{n}{r}$개의 부분집합에 대하여 연산을 수행한다는 뜻이다.

[예 1.3-1] 동전을 두 번 던지는 실험을 하여 표본공간 $S = \{(H,H), (H,T), (T,H), (T,T)\}$를 얻었다. S의 각 원소가 얻어질 확률은 1/4이다. 이제 $A_1 = \{(H,H), (H,T)\}$, $A_2 = \{(H,H), (T,H)\}$라 두면 $P(A_1) = P(A_2) = 1/2$, $P(A_1 \cap A_2) = 1/4$, $P(A_1 \cup A_2) = 1/2 + 1/2 - 1/4 = 3/4$.

[예 1.3-2] 표본공간을 $S = A \cup B$라 하고 $P(A) = 0.6$, $P(B) = 0.7$이라고 하자. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$에서 $P(A \cup B) = 1$이므로 $1 = 0.6 + 0.7 - P(A \cap B)$이다. 따라서 $P(A \cap B) = 0.3$

4. 확률의 계산

[곱셈원리] 실험 E_1에서 n_1개의 결과를 얻고, 실험 E_2에서 n_2개의 결과를 얻는다고 할 때, 먼저 E_1를 시행하고 다음에 E_2를 시행한 실험 E_1E_2는 n_1n_2개의 결과를 갖는다.

[예 1.4-1] E_1은 주사위를 던지는 실험이고, E_2는 동전을 던지는 실험이라고 하면 $n_1=6$, $n_2=2$이다. 따라서 실험 E_1E_2는 12가지의 결과를 갖는다. 구체적인 결과는 다음과 같다.

$$(1,H),\ (2,H),\ (3,H),\ (4,H),\ (5,H),\ (6,H),$$
$$(1,T),\ (2,T),\ (3,T),\ (4,T),\ (5,T),\ (6,T).$$

[참고] (곱셈원리의 일반화)
실험 E_i을 통하여 n_i, $i=1,2,\cdots,m$개의 결과를 얻는다고 하자. 이때 실험 $E_1E_2\cdots E_m$은 $n_1n_2\cdots n_m$개의 결과를 갖는다.

[정의 1.4-1] n개의 서로 다른 결과(혹은 물건)를 배열하는 방법을 순열(Permutation)이라 한다.

[정의 1.4-2] $P(n,r)$은 n개에서 r개를 뽑는 순열을 뜻한다.

$$P(n,r)=n(n-1)\cdots(n-r+1)=\frac{n!}{(n-r)!},\ 0\neq 1$$

[정의 1.4-3] $\binom{n}{r}$은 순서를 고려하지 않고 n개에서 r개를 선택하는 방법(조합)의 개수이다.

$$\binom{n}{r}=\frac{n!}{r!(n-r)!}=\binom{n}{n-r},\ \binom{n}{n_1,n_2,\cdots,n_s}=\frac{n!}{n_1!n_2!\cdots n_s!},\ \sum_{j=1}^{s}n_j=n$$

5. 복원추출 비복원추출

[정의 1.5-1] n개의 표본 중에서 r개를 지정된 순서대로 뽑는 경우 표본의 크기가 r인 순서표본(Ordered sample)이라고 한다.

[정의 1.5-2] 복원추출(Sample with replacement)이란 다음 표본이 추출되기 전에 추출된 표본이 복원되는 추출방법을 말한다.

[정의 1.5-3] 비복원추출(Sample with replacement)이란 추출된 표본이 복원되지 않는 추출방법을 말한다.

[예 1.5-1] 상자에 100개의 퓨즈가 들어있는데 그 중에서 20개는 불량품이다. 퓨즈 5개를 꺼내어 모두 정상이면 합격시킨다고 할 때 합격할 확률을 구해보자. X를 불량 퓨즈의 개수라고 할 때,

$$P(X=0)=\frac{\binom{20}{0}\binom{80}{5}}{\binom{100}{5}}=0.32$$

일반적인 경우 $P(X=x)=\dfrac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}$, $x=0,1,2,3,4,5.$

- 초기하 분포(Hypergeometric distribution)

(i) $n=n_1+n_2$개의 표본에서 r개를 추출하는 방법

$$P(X=x)=\frac{\binom{n_1}{x}\binom{n_2}{r-x}}{\binom{n}{r}},\ x\le r,\ x\le n_1,\ r-x\le n_2$$

(ii) $n=n_1+n_2+\cdots+n_s$개의 표본에서 r개를 추출하는 방법($x_i\le n_i$, $x_1+x_2+\cdots+x_s=r$)

$$P(X_1=x_1,X_2=x_2,\cdots,X_s=x_s)=\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}\cdots\binom{n_s}{x_s}}{\binom{n}{r}}.$$

[예 1.5-2] 13장의 카드를 뽑을 경우 클로버 2장, 다이아몬드 4장, 하트 3장, 스페이드 4장을 뽑을 확률은?

$$\frac{\binom{13}{2}\binom{13}{4}\binom{13}{3}\binom{13}{4}}{\binom{52}{13}}=0.018.$$

[예 1.5-3] (생일문제) k명이 있을 때 같은 생일을 가진 사람이 최소 2인 이상일 확률은?

풀이> k명 모두가 생일이 다를 확률 $\frac{365\cdot 364\cdot\cdots(365-k+1)}{365^k}$

구하고자 하는 확률 $1-\frac{365\cdot 364\cdot\cdots(365-k+1)}{365^k}$

k	10	22	23	40	50	70
확률	0.092	0.475	0.507	0.891	0.970	0.999

- **이항정리**

[이항정리] $(a+b)^n=\sum_{r=0}^{n}\binom{n}{r}b^r a^{n-r}$, $2^n=\sum_{r=0}^{n}\binom{n}{r}$

$b^r a^{n-r}$의 계수는 $\binom{n}{r}$이며 이를 이항계수(Binomial coefficient)라고 한다.

[파스칼의 정리] $\binom{n}{r}=\binom{n-1}{r}+\binom{n-1}{r-1}$

[이항정리의 일반화] $(a_1+a_2+\cdots+a_s)^n$을 전개할 때 $a_1^{n_1}a_2^{n_2}\cdots a_s^{n_s}$항의 계수는 $\binom{n}{n_1,n_2,\cdots,n_s}=\frac{n!}{n_1!n_2!\cdots n_s!}$이며 다항계수 (Multinomial coefficient)라 한다.

6. 조건부확률

[정의 1.6-1] 사건 B가 주어졌을 때, 사건 A가 일어나는 조건부 확률은

$$P(A|B)=\frac{P(A\cap B)}{P(B)},\ P(B)>0$$

[예 1.6-1] $P(A)=0.4$, $P(B)=0.5$, $P(A\cap B)=0.3$이면
$P(A|B)=0.3/0.5=0.6$이고, $P(B|A)=0.3/0.4=0.75$

[예 1.6-2] 5장의 카드를 비복원 추출하여 게임을 한다. A는 스페이드 2장이 나오고 B는 하트가 3장이 나올 사건이라면,

$$P(A)=\frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}},\ P(B)=\frac{\binom{13}{3}\binom{39}{2}}{\binom{52}{5}},\ P(A\cap B)=\frac{\binom{13}{2}\binom{13}{3}}{\binom{52}{3}}$$

[참고] 조건부 확률은 $P(B)>0$일 때, 확률함수의 공리를 만족한다.
① $P(A|B)\geq 0$ ② $P(S|B)=1$
③ 사건 $A_1, A_2, \cdots$가 서로 배반사건이라면,

$$P(A_1\cup A_2\cup\cdots|B)=P(A_1|B)+P(A_2|B)+\cdots.$$

①은 $P(A|B)=\frac{P(A\cap B)}{P(B)}\geq 0$이므로 분명하다.

②는 $P(B)>0$이면 $P(S|B)=\frac{P(S\cap B)}{P(B)}=\frac{P(B)}{P(B)}=1$이므로 성립한다.

③ 또한

$$\begin{aligned}P(A_1\cup A_2\cup\cdots|B)&=\frac{P[(A_1\cup A_2\cup\cdots)\cap B]}{P(B)}=\frac{P[(A_1\cap B)\cup(A_2\cap B)\cup\cdots]}{P(B)}\\&=\frac{P(A_1\cap B)+P(A_2\cap B)+\cdots}{P(B)}\\&=\frac{P(A_1\cap B)}{P(B)}+\frac{P(A_2\cap B)}{P(B)}+\cdots\\&=P(A_1|B)+P(A_2|B)+\cdots\end{aligned}$$

이므로 성립함을 알 수 있다. 위 식의 세 번째 등호에서는 $(A_1\cap B)$, $(A_2\cap B)$, $\cdots$가 서로 배반사건임을 이용하였다.

[확률의 곱셈정리] $P(A\cap B)=P(B)\cdot P(A|B)$.
[확률의 곱셈정리의 일반화] $P(A_1\cap A_2\cap\cdots\cap A_{n-1})>0$이면,
$P(A_1\cap A_2\cap\cdots\cap A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots P(A_n|A_1\cap\cdots\cap A_{n-1})$.

[예 1.6-3] 항아리에 푸른 공 7개, 붉은 공 3개가 들어있다. 공 2개를 차례로 뽑을 경우 첫 번째는 붉은 공 (A), 두 번째는 푸른 공 (B)이 나올 확률은?
풀이> $P(A)=3/10$이고 $P(B|A)=7/9$이므로 구하고자 하는 확률

$$P(A\cap B)=\left(\frac{3}{10}\right)\left(\frac{7}{9}\right)=\frac{7}{30}.$$

[예 1.6-3] 항아리에 검은 공 (B)이 b개, 붉은 공 (R)이 r개 들어있다. 공을 한 개 꺼내어 검은 공이면 꺼낸 공과 함께 c개의 검은 공을 항아리에 넣고, 붉은 공이면 꺼낸 공과 함께 c개의 붉은 공을 항아리에 다시 넣은 후 공을 한 개 꺼낼 때 검은 공일 확률은?

풀이> 첫 번째 시행에서 검은 공 (B_1)이 나올 확률은 $P(B_1)=\frac{b}{b+r}$이며 두 번째 시행에서 검은 공 (B_2)이 나올 확률은 다음과 같다.

$$\begin{aligned}P(B_2)&=P(B_1\cap B_2)+P(R\cap B_2)\\&=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}+\frac{r}{b+r}\cdot\frac{b}{b+r+c}=\frac{b}{b+r}.\end{aligned}$$

[정리 1.6-1] (전확률의 정리(Theorem of total probability)

$S=\bigcup_{i=1}^{n}B_i$, $B_i\cap B_j=\varnothing$, $i\neq j$ (이 경우 B_i를 S의 분할 **(Partition)**이라고 함)이면, $P(B_i)>0$일 때, $P(A)=\sum_{i=1}^{n}P(A|B_i)P(B_i)$이다.

증명> $A=A\cap S=\bigcup_{i=1}^{n}(A\cap B_i)$, $(A\cap B_i)\cap(A\cap B_j)=\varnothing$ 이므로

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

[따름정리] $0<P(B)<1$일 때,
$P(A)=P(A|B)P(B)+P(A|B^c)P(B^c)$이다.

7. 베이즈정리

[예 1.7-1] 항아리 B_1에는 붉은 구슬 2개와 흰 구슬 4개가 들어있고, 항아리 B_2에는 붉은 구슬 1개와 흰 구슬 2개가 들어있으며, 항아리 B_3에는 붉은 구슬 5개와 흰 구슬 4개가 들어있다. 각 항아리를 택하는 확률이 $P(B_1)=1/3$, $P(B_2)=1/6$, $P(B_3)=1/2$일 때, 항아리를 랜덤하게 선택하여 구슬을 한 개 뽑을 경우 붉은 구슬일 확률 $P(R)$을 구하라.

풀이>

$$\begin{aligned} P(R) &= P(B_1 \cap R) + P(B_2 \cap R) + P(B_3 \cap R) \\ &= P(B_1)P(R \mid B_1) + P(B_2)P(R \mid B_2) + P(B_3)P(R \mid B_3) \\ &= \left(\frac{1}{3}\right)\left(\frac{2}{6}\right) + \left(\frac{1}{6}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{2}\right)\left(\frac{5}{9}\right) = \frac{4}{9}. \end{aligned}$$

[예 1.7-2] [예 1.7-1]에서 꺼낸 구슬이 붉은 구슬일 때, 항아리 B_1에서 나왔을 확률 $P(B_1|R)$, 항아리 B_2에서 나왔을 확률 $P(B_2|R)$, 항아리 B_3에서 나왔을 확률 $P(B_3|R)$를 구하라.

풀이>

$$\begin{aligned} P(B_1 \mid R) &= \frac{P(B_1 \cap R)}{P(R)} \\ &= \frac{P(B_1)P(R \mid B_1)}{P(B_1)P(R \mid B_1) + P(B_2)P(R \mid B_2) + P(B_3)P(R \mid B_3)} \\ &= \frac{(1/3)(2/6)}{(1/3)(2/6) + (1/6)(1/3) + (1/2)(5/9)} = 2/8, \end{aligned}$$

$$P(B_2 \mid R) = \frac{P(B_2 \cap R)}{P(R)} = \frac{(1/6)(1/3)}{4/9} = 1/8,$$

$$P(B_3 \mid R) = \frac{P(B_3 \cap R)}{P(R)} = \frac{(1/2)(5/9)}{4/9} = 5/8.$$

조건부확률(사후확률: Posterior probability) $P(B_1|R)$, $P(B_2|R)$, $P(B_3|R)$은 각 항아리를 선택하는 확률(사전확률: Prior probability) $P(B_1)$, $P(B_2)$, $P(B_3)$에 따라 변한다.

[정리 1.7-1] (베이즈 정리)

$S = B_1 \cup B_2 \cup \cdots \cup B_m$, $B_i \cap B_j = \emptyset$, $i \neq j$ 즉, B_1, B_2, $\cdots$, B_m이 S의 분할이라고 하면,

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{m} P(B_i)P(A \mid B_i)}, \quad k = 1, 2, \cdots, m$$

증명> $A = (B_1 \cap A) \cup (B_2 \cap A) \cup \cdots \cup (B_m \cap A)$이므로

$P(A) = \sum_{i=1}^{m} P(B_i)P(A|B_i)$이며, $P(A) > 0$이면,

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{m} P(B_i)P(A|B_i)}, \quad k = 1, 2, \cdots, m$$

[예 1.7-3] 1부터 5까지의 번호가 적혀있는 항아리에 각각 10개의 공이 들어있다. i번째 항아리에는 i개의 검은 공과 $10-i$개의 흰 공이 들어있다. 한 개의 항아리를 랜덤하게 택하여 한 개의 공을 꺼내는 실험을 수행한다.

(1) 꺼낸 공이 검은 공일 확률.

(2) 꺼낸 공이 검은 공이었다면 이 공이 5번째 항아리에서 나왔을 확률.

풀이> A: 검은 공이 꺼내어지는 사건, B_i: i번째 항아리에서 공이 꺼내지는 사건, $P(B_i)=1/5$, $P(A|B_i)=i/10$

(1) $P(A) = \sum_{i=1}^{5} P(A \mid B_i)P(B_i) = \sum_{i=1}^{5}\left(\frac{i}{10}\right)\left(\frac{1}{5}\right) = \frac{3}{10}$.

(2) $P(B_5 \mid A) = \dfrac{P(B_5 \cap A)}{P(A)} = \dfrac{P(A \mid B_5)P(B_5)}{\sum_{i=1}^{5} P(A \mid B_i)P(B_i)} = \dfrac{\frac{1}{2} \cdot \frac{1}{5}}{\frac{3}{10}} = \dfrac{1}{3}$.

8. 독립사건

[정의 1.8-1] 사건 A와 B가 서로 **독립**일 필요충분조건은

$$P(A\cap B)=P(A)P(B)$$

이다. 그렇지 않으면 A와 B는 **종속**사건이라고 한다.

[예 1.8-1] 동전 한 개를 두 번 던져서 앞면과 뒷면이 나오는 경우의 표본공간은 $S=\{HH, HT, TH, TT\}$이다. 사건 A와 B를 다음과 같이 정의하면
$A=$\{처음 던진 동전이 앞면이 나오는 경우\}$=\{HH, HT\}$
$B=$\{두 번째 던진 동전이 뒷면이 나오는 경우\}$=\{HT, TT\}$,
$P(A|B)=\dfrac{P(A\cap B)}{P(B)}=\dfrac{1/4}{1/2}=1/2=P(A)$이므로 사건 A와 B는 서로 독립이다.

[정리 1.8-1] A와 B가 독립사건이면 다음도 역시 독립이다.
(1) A와 B^c (2) A^c와 B (3) A^c와 B^c

증명 (1) $P(B\mid A)+P(B^c\mid A)=P(A\cap B)/P(A)+P(A\cap B^c)/P(A)$

$$=\frac{P[(B\cap A)\cup(B^c\cap A)]}{P(A)}=P(A)/P(A)=1$$

이므로 $P(B^c\mid A)=1-P(B\mid A)$이다. 따라서,

$$P(A\cap B^c)=P(A)P(B^c\mid A)=P(A)(1-P(B\mid A))$$
$$=P(A)(1-P(B))=P(A)P(B^c).$$

(2)는 (1)과 증명방법이 동일함.

(3) $P(A^c\cap B^c)=P((A\cup B)^c)=1-P(A\cup B)$

$$=1-(P(A)+P(B)-P(A\cap B))$$
$$=(1-P(A))(1-P(B))=P(A^c)P(B^c).$$

[정의 1.8-1] 사건 A, B와 C가 서로 독립일 필요충분조건
$P(A\cap B)=P(A)P(B)$, $P(A\cap C)=P(A)P(C)$, $P(B\cap C)=P(B)P(C)$
그리고 $P(A\cap B\cap C)=P(A)P(B)P(C)$

[예 1.8-2] 동전을 세 번 던지는 실험을 한다. i번째 던져서 앞면이 나오는 사건을 A_i라고 하면 표본공간은

$$S=\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

이고, $P(A_i)=4/8$, $P(A_i\cap A_j)=2/8$, $i\neq j=1,2,3$, $P(A_1\cap A_2\cap A_3)=1/8$
$P(A_i\cap A_j)=1/4=P(A_i)P(A_j)$, $i\neq j$
$P(A_1\cap A_2\cap A_3)=1/8\,P(A_1)P(A_2)P(A_3)$
이므로 사건 A_1, A_2, A_3는 서로 독립이다.

[정의 1.8-2] 사건 A, B, C가 서로 독립이면, 다음이 성립한다.
(1) A와 $(B\cap C)$는 서로 독립이다.
(2) A와 $(B\cup C)$는 서로 독립이다.
(3) A^c와 $(B\cap C^c)$는 서로 독립이다.
(4) A^c, B^c, C^c는 서로 독립이다.
[증명] 생략

[참고] 일반적으로 사건 $A_1, A_2, \cdots, A_n$이 서로 독립일 필요충분조건은 다음의 식이 성립해야 한다.

$$P(A_i\cap A_j)=P(A_i)P(A_j),\quad i\neq j$$
$$P(A_i\cap A_j\cap A_k)=P(A_i)P(A_j)P(A_k),\quad i\neq j,\quad j\neq k,\quad i\neq k$$
$$\vdots$$
$$P\left(\bigcap_{i=1}^{n}A_i\right)=\prod_{i=1}^{n}P(A_i).$$