KAIST 전산학과 최광무

7.1 이산 확률(Discrete Probability)의 소개

세 가지 기본 용어들

실험(experiment): 행동(outcome)의 집합이 일어나는 행위(procedure)

실험의 전체 공간(sample space): 행동의 전체집합

사건(event): 전체 공간의 부분집합

개별사건(actual outcome): 임의의 개별사건 e가 사건 E의 원소($e \in E$)일 때, 사건 E가 일어났다고(occur)한다.

(정의 1-1) 실험의 전체 공간 S의 각 원소 들이 일어날 확률이 모두 같을 때,

사건 $E(E \subseteq S)$가 일어날 확률 $p(E) = \frac{|E|}{|S|}$로 정의한다.

(정리 1-1) 실험의 사건 E의 반대사건 $\overline{E}$가 일어날 확률 $p(\overline{E}) = 1 - \frac{|E|}{|S|}$이다.

(정리 1-2) 사건 E_1 또는 사건 E_2가 일어날 확률은

$p(E_1 \cup E_2) = p(E_1) + p(E_2) + p(E_1 \cap E_2)$이다.

7.2 확률 이론

확률의 배당(Assigning Probabilities)

실험의 전체 공간 S가 셀 수 있게 많을 때(countable)[1] 개별사건 $s \in S$의 확률은 아래 두 조건에 맞추어 배당한다(assigning).

(1) $\forall\, s \in S$: $0 \leqq p(s) \leqq 1$. $\qquad p : S \rightarrow [0,1]$

(2) $\bigcup_{s \in S} p(s) = 1$.

(정의 2-1) 전체 공간 S가 n개의 개별사건(원소)를 가지고 있을 때, $\forall\, s \in S$: $p(s) = 1/n$일 때 확률분포 p는 균일분포(uniform distribution)한다고 한다.

(정의 2-1') 전체 공간 S에서 $\forall\, s,t \in S$: $p(s) = p(t)$이면 확률분포 p는 균일분포(uniform distribution)한다고 한다.

(정의 2-2) 사건 E가 일어날 확률은 개별사건들의 합이다.

$P(E) = \sum_{e \in E} p(e)$.

(정의 2-3) 종속사건(conditional probability) 두 사건 E와 F에서, 사건 E가 **반드시** 사건 F가 **일어난 다음에**($p(F) > 0$) 일어날 확률 $p(E|F) = \frac{P(E \cap F)}{P(F)}$이다.

(사실 3) 사건 E가 일어난 다음에 사건 F가 일어날 확률 $p(E \cap F) = p(E|F)P(F)$이다.

(정의 2-4) 독립사건(independent events) 두 사건 E와 F에서, 사건 E가 일어난 다음에 사건 F가 일어날 확률 $p(E \cap F) = P(E)P(F)$이면 두 사건은 서로 독립(mutually independent)[2]인 사건 이라 한다.

---

1) countable은 자연수의 부분집합과 동등할 때로 정의하므로, 유한한 경우도 포함된다.

2) $p(E|F) = P(E)$이므로 사건 E와 F는 서로 독립이다.

(정의 2-5) 생략.
(정리 2-2) 베르누이(Bernoulli)의 시도(trial)와 이항분포(Binomial distribution)
성공할 확률이 p인 독립사건을 n번 시도하여 k번 성공할 확률은 $C(n,k)p^k q^{n-k}$[3]이다.
(사실 2) $\sum_{k \in N_n} C(n,k)p^k q^{n-k} = (p+q)^n = 1$.

(정의 2-6) 무작위변수(random variable) X를 실험 공간 S에서 적절한 수의 집합 A(치역)로 가는 함수로 정의한다. $X: S \to A$.
(정의 2-7) 실험 공간 S와 무작위변수 X의 치역 A에서 무작위변수 X의 분포는
$\forall r \in X(S)(\subseteq A)$: $(r,\ p(X=r))$ 순서쌍의 집합이다.

(예 12) 주사위 2개를 던진 수의 합을 치역으로 하는 무작위변수 X의 분포(교과서 447p).

(예 13) 몇 명 이상의 사람이 모여야, 생일이 같은 사람이 있을 확률이 $1/2$을 넘는가?
i번째 사람이 그 앞의 $(i-1)$명의 서로 다른 생일을 가진 사람들과 생일이 다를 확률은 $\frac{366-(i-1)}{366}$이므로, i명의 사람이 모두 다른 생일을 가질 확률 $p_i = \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367-i}{366}$이다.
$1 - \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367-i}{366} > \frac{1}{2}$ 부등식을 풀면, $i > 22.XXX$이다.
따라서 답은 23명이다.

7.3 종속사건과 독립사건(바이에(Bayes') 정리)
(정리 3-1) 바이에(Bayes')의 정리: 확률이 0이 아닌 두 사건 E와 F에서,

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F)) + p(E|\overline{F})p(\overline{F})} = \frac{p(E|F)p(F)}{p(E)} \text{ 또는}$$
$$p(F|E)P(E) = p(E|F)p(F).$$

(증명) $p(F|E) = \frac{P(E \cap F)}{P(E)}$ $p(E|F) = \frac{P(E \cap F)}{P(F)}$
$\therefore\ p(F|E)P(E) = p(E|F)P(F) = P(E \cap F)$
한편 $P(E) = p(E|F)p(F)) + p(E|\overline{F})p(\overline{F})$.

(정리 3-2) 바이에(Bayes') 정리의 확장: 확률이 0이 아닌 사건 E와 서로 독립인 n개의 사건 F_i(단 $i \in N_n$)에서,

$$p(F_i|E) = \frac{p(E|F_i)p(F_i)}{\Sigma_{j \in N_n} p(E|F_j)p(F_j)} = \frac{p(E|F_i)p(F_i)}{p(E)} \text{ 또는}$$
$$p(F_i|E)P(E) = p(E|F_i)p(F_i).$$

3) $q = \overline{p} = 1-p$이다.

7.4 기대치(Expected Value)과 분산(Variance)

(정의 4-1) $E(X) = \sum_{x \in S} p(s)X(s)$ X의 기대치(expected value; expectation, mean)

$s \in S$: $X(s) - E(X)$ 개별사건 s의 편차(deviation)는 이다.

(정리 4-1) 실험 공간 S에서 무작위변수 X의 기대치는 $E(X) = \sum_{r \in X(S)} p(X=r)r$이다.

(정리 4-2) 성공할 확률이 p인 서로 독립인 n번의 베르누이 시도도의 기대치는 np이다.

(증명) $p(X=k) = C(n,k)p^k q^{n-k}$ 정리 2-2

$$E(X) = \sum_{r \in X(S)} p(X=k)k = \sum_{k=1}^{n} k\,C(n,k)p^k q^{n-k} = \sum_{k=1}^{n} n\,C(n-1,k-1)p^k q^{n-k}$$

$$= np\sum_{k=1}^{n} n\,C(n-1,k-1)p^{k-1}q^{n-k} = np\sum_{j=0}^{n-1} n\,C(n-1,j)p^j q^{n-1-j}$$

$$= np(p+q)^{n-1} = np.$$

(정리 4-3) $i \in N_n$: X_i가 실험 공간 S에서 무작위변수이고, a와 b가 실수 상수일 때,

(i) $E\left(\sum_{i \in N_n} X_i\right) = \sum_{i \in N_n} E(X_i)$.

(ii) $E(aX+b) = aE(X) + b$.

실제 계산복잡도(Average-Case Computational Complexity)

알고리즘의 복잡도는 입력자료(input data)에 따라 다르다. 입력자료의 분포를 예상할 수 있다면 실제 계산복잡도를 예측할 수 있다.

입력자료의 집합을 실험공간 S라 보면,

$$E(X) = \sum_{a \in S} p(a)\,X(a).$$

(예 8) 선형찾기(Linear Search)의 실제 복잡도

입력자료: $i \in N_n$: $a_i = S_i$, 찾고자하는 값 a. x가 입력자료의 실험공간 S_i에 있을 확률: p/n.

(1) $x \in S$: $(2i+1)$ 비교. $E(X_i) = \frac{p}{n}\sum_{k=1}^{i}(2i+1) = \frac{p}{n}(i(i+1)+i) = \frac{p}{n}i(i+2)$.

(2) $x \notin S$: $(2n+2)$ 비교 $E(X_{\neg a}) = (1-p)(2n+2) = 2(1-p)(n+1)$.

(i) $p=1$: $E(X) = (n+2)$.

(iii) $p=0$: $E(X) = 2(n+1)$

(예 9) 삽입찾기(Linear Search)의 실제 복잡도

<u>기하급수 분포(Geometric Distribution)</u>

(예 10) 성공확률 p인 베르누이 시도에서 $k \in \boldsymbol{N_1}$에서 최초에 성공까지의 기대치는?

(정의 4-2) $p(X=k) = p(F^{k-1}S) = (1-p)^{k-1}$ 기하급수 분포(Geometric Distribution)

(정리 4-4) $E(X) = \sum_{k\in \boldsymbol{N_1}} (1-p)^{k-1}p \cdot k = p \cdot \sum_{k\in \boldsymbol{N_1}} (1-p)^{k-1}k = p\frac{1}{p^2} = \frac{1}{p}$.

독립 분포(Independent Random Variable)

(정의 4-3) $p(X=x \wedge Y=y) = p(X=x) \cdot p(Y=y)$ 서로 독립이라 한다.

(정리 4-5) $E(XY) = E(X)E(Y)$ 독립변수의 기대치

(정의 4-1) $s \in S$: $dev(X) \stackrel{\text{def}}{=} X(s)-E(X)$ s의 편차(derivation)

(정의 4-4) $V(X) \stackrel{\text{def}}{=} \sum_{x\in S} (X(s)-E(X))^2 p(s)$ X의 분산(variance)

(정의 4-4") $a(X) \stackrel{\text{def}}{=} \sqrt{V(X)}$ X의 표준편차(standard deviation)

(정리 4-6) $V(X) = E(X^2) - E(X)^2 = E(X^2) - \mu^2$. (단 $E(X)=\mu$)

(증명) $V(X) \stackrel{\text{def}}{=} \sum_{x\in S} (X(s)-\mu)^2 p(s) = \sum_{x\in S} X(s)^2 p(s) - 2\mu \sum_{x\in S} X(s)p(s) + \mu^2 \sum_{x\in S} p(s)$

$= E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 = E(X^2) - E(X)^2$.

무작위변수의 편차의 제곱의 기대치인 분산은 무작위변수 제곱의 기대치에서 무작위변수의 기대치(μ)의 제곱을 빼는 것과 같다.

(Col. 1) $V(X) = E\big((X-\mu)^2\big)$

(증명) $E\big((X-\mu)^2\big) = E(X^2-2\mu X+\mu^2) = E(X^2)-2\mu E(X)+\mu^2 = E(X^2)-\mu^2 = V(X)$.

(정리 4-7) $V(X_1+X_2+\cdots+X_n) = V(X_1)+V(X_2)+\cdots+V(X_n)$

독립변수 합에 분산(variance)는 개별분산의 합이다.