② 베르누이 분포 Likelihood 모형

- x_d가 0 or 1 값만 가질 수 있다. 즉, 독립변수는 D개의 독립적인 베르누이 확률변수,
즉, 동전으로 구성된 동전 세트로 표현가능. 이 동전들의 모수 θ_d는 동전 d 마다 다르다.
또한, 클래스 $y = C_k (k=1, \dots, K)$ 마다도 x_d가 1이될 확률이 다르다.
- 즉, 동전의 모수 $\theta_{d,k}$는 동전 d마다 다르고, 클래스 k 마다도 다르다.
전체 $D \times K$의 동전이 존재하고, 같은 클래스에 속하는 D개의 동전이 하나의 동전 세트를 구성하고, 이러한 동전 세트가 K개 있다고 생각할 수 있다.

$$P(x_d \mid y = C_k) = \theta_{d,k}^{x_d} (1-\theta_{d,k})^{(1-x_d)}$$

⇒ 이러한, 동전 세트마다 확률 특성이 다르므로, 베르누이 분포 likelihood 모형을 기반으로 하는 나이브 베이즈 모형은 동전 세트를 N번 던진 결과로 부터, $1, \dots, K$ 중에 어느 동전 세트를 던졌는지 찾아내는 모형.

③ 다항 분포 Likelihood 모형

- x 벡터가 다항 확률분포의 sample이라고 가정한다.
D개의 면을 가지는 주사위를 $\sum_{d=1}^{D} x_d$번 던져서 나온 결과로 본다.
ex) $x = (1, 4, 0, 5)$ ⇒ 4면체 주사위를 10번던져서 1:1번, 2:4번, 4:5번 나옴.

- $P(x_1, \dots, x_D \mid y = C_k) \propto \prod_{d=1}^{D} \theta_{d,k}^{x_{d,k}}$

$$N_k = \sum_{d=1}^{D} x_{d,k}$$

$$\sum_{d=1}^{D} \theta_{d,k} = 1$$

⇒ 주사위를 던진 결과로 부터 $1, \dots, K$ 중 어느 주사위를 던졌는지 찾아내는 모형

스무딩 : 표본 Data 수가 적은 경우에 베르누이 모수가 0 or 1 이라는 극단적인 모수 추정 값이 나올 수 있음. 하지만 현실적으로 이럴 가능성이 낮다.

i) 베르누이 : 모수가 0.5인 가장 일반적인 경우를 가정하여 0이 나오는 경우, 1이 나오는 경우, 두 개의 가상 표본 Data 추가. (라플라스 스무딩 or Add one)

$$\hat{\theta}_{d,k} = \frac{N_{d,k} + \alpha}{N_k + 2\alpha}$$

ii) 독립변수가 다항분포이고 카테고리수가 D개 : $\hat{\theta}_{d,k} = \frac{N_{d,k} + \alpha}{N_k + D\alpha}$, α가 정수가 아니어도 됨.