표본 Data가 주어진 경우

↳ 확률변수 X의, 즉 이론적인 확률밀도(질량) 함수가 아닌 실제 Data가 주어진 경우, 확률질량 함수를 추정하여 엔트로피를 계산한다.

ex) 1) $p(y=0)=\frac{40}{80}=\frac{1}{2}$, $p(y=1)=\frac{40}{80}=\frac{1}{2}$

$H[Y]=-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}=1$

2) $p(y=0)=\frac{20}{60}=\frac{1}{3}$, $p(y=1)=\frac{40}{60}=\frac{2}{3}$

$H[Y]=-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}=0.92$

3) $p(y=0)=\frac{30}{40}=\frac{3}{4}$, $p(y=1)=\frac{10}{40}=\frac{1}{4}$

$H[Y]=-\frac{3}{4}\log_2\frac{3}{4}-\frac{1}{4}\log_2\frac{1}{4}=0.81$

4) $p(y=0)=\frac{20}{20}=1$, $p(y=1)=\frac{0}{20}=0$

$H[Y]=-1\log_2 1-0\log_2(0)=0$

↓ 수치가 원래는 같은 엔트로피를 0으로 향해 줄어든 것.

조건부 엔트로피 - 조건부 엔트로피는 상관관계가 있는 두 확률변수 X, Y가 있고 X의 값을 안다면 Y확률변수가 가질 수 있는 정보의 양을 뜻함

$$H[Y|X]=-\sum_i\sum_j p(x_i,y_j)\log_2 p(y_j|x_i)$$

$$H[Y|X]=-\int\int p(x,y)\log_2 p(y|x)\,dx\,dy$$

$$\Downarrow$$

$$H[Y|X]=\sum_i p(x_i)H[Y|x_i]$$

$$H[Y|X]=\int p(x)H[Y|x]\,dx$$

증명: $H[Y|X]=-\sum_i\sum_j p(x_i,y_j)\log_2 p(y_j|x_i)$

$$=-\sum_i\sum_j p(y_j|x_i)p(x_i)\log_2 p(y_j|x_i)$$

$$=-\sum_i p(x_i)\sum_j p(y_j|x_i)\log_2 p(y_j|x_i)$$

$$=\sum_i p(x_i)H[Y|x_i]$$