* Information Theory

- $I(p) \geq 0$, $I(1) = 0$, $I(p_1 \cdot p_2) = I(p_1) + I(p_2)$

  $I(p) = -\log_b P$

  · Information Content $\begin{cases} b=2: \text{bits} \\ b=3: \text{trits} \\ b=e: \text{nats} \\ b=10: \text{Hartleys} \end{cases}$

- Entropy $H(x) = E(I(x)) = -\sum_{i=1}^{n} p(x_i) \cdot \log_b P(x_i)$

- Relative Entropy $D(P\|Q) = E_p\left[\log_b \frac{P(X)}{Q(X)}\right] = \sum_x p(x) \cdot \log_b \frac{p(x)}{q(x)}$

- Joint Entropy $H(X,Y) = E[-\log_b P(X,Y)]$

  $= -\sum_x \sum_y P(X=x, Y=y) \log_b P(X=x, Y=y)$

  ① $H(X,Y) \geq \max(H(X), H(Y))$

  ② $H(X,Y) \leq H(X) + H(Y)$

  ③ $H(X,Y) = H(X) + H(Y)$ if, only if X, Y are independent

- Conditional Entropy $H(X|Y) = H(X,Y) - H(Y)$

  $= -\sum_x \sum_y P(X=x, Y=y) \log_b P(X=x \mid Y=y)$

  $= -\sum_x \sum_y P(X=x, Y=y) \log_b P(X=x, Y=y) + \sum_y P(Y=y) \log_b P(Y=y)$

- Mutual Information $I(X,Y) = H(X) - H(X|Y)$

  $= H(Y) - H(Y|X)$

  $= H(X) + H(Y) - H(X,Y)$

* Central limit theorem

$\bar{X}$ follows approximately normal distribution with $(\mu, \frac{\sigma}{\sqrt{n}})$ from X,

even if X does not follow normal distribution,

if $n > 30$

* Point Estimation

- Case 1: $\bar{X}$, σ is given, $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

- Case 2: population follows Bernauli Distribution

  With given $\bar{X}$ without σ

  · Upper Bound: $\bar{X} \pm \frac{z_{\alpha/2}}{2\sqrt{n}}$

  · Estimation: $\bar{X} \pm z_{\alpha/2} \sqrt{\frac{\bar{X}(1-\bar{X})}{n}}$

- Case 3: $\bar{X}$ is given with unknown σ (distribution)

  $\bar{X} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

* Hypothesis Testing

- z-test : $\begin{cases} \text{the sample follows a normal distribution} \\ n > 30 \end{cases}$

  cf. normal dist. 이면 n<30 이어도 z-test

- t-test : $\begin{cases} 5 \leq n \leq 30 \text{ 또는} \\ s \text{가 unknown 일 때?} \end{cases}$

  $T_{n-1} = \frac{\bar{x} - \mu}{\frac{s_n}{\sqrt{n}}}$

  $S_n = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

  ① One sample T-test

  ② Two samples T-test

  $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

  $T_{n_1+n_2-2} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}}$

  ③ Paired sample T-test