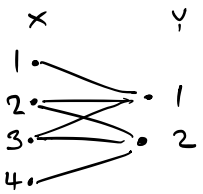
Example:

$X \in \{1,2,3,4\}$ uniform



$\longleftrightarrow$

$P_{Y|X}(\cdot|1) = (1,0)$

$P_{Y|X}(\cdot|2) = P_{Y|X}(\cdot|3) = (0.5, 0.5)$

$P_{Y|X}(\cdot|3) = (0,1)$

$P_{guess}(X) = \frac{1}{4} \Rightarrow H_{min}(X) = 2 \Rightarrow$ we can extract at most 2 bits of perfect randomness!

(achieved with trivial extractor)

$P_{X|Y}(\cdot|1) = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$

$P_{X|Y}(\cdot|2) = (0, \frac{1}{4}, \frac{1}{4}, \frac{1}{2})$

$P_{guess}(X|Y) = \frac{1}{2}$

$\Rightarrow H_{min}(X|Y) = 1 \Rightarrow$ we can extract at most 1 bit of perfect secret randomness!

extractor $f: \{1,2,3,4\} \to \{0,1\}$ given by

$f(1) = f(4) = 1$ and $f(2) = f(3) = 0$.

$Z = f(X) \Rightarrow P_{Z|Y}(1|1) = \frac{1}{2}, P_{Z|Y}(0|1) = \frac{1}{2}$

$P_{Z|Y}(1|2) = \frac{1}{2}, P_{Z|Y}(0|2) = \frac{1}{2}$

$P_{ZY}(\cdot,\cdot) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}) = P_Z(\cdot) \times P_Y(\cdot)$ ← uniform

$\Rightarrow$ perfectly random and secret

Chapter 4: Statistics / Binary Hypothesis Testing

We have a random process/source that produces random variables $X = (X_1, X_2, \dots)$ that are (discrete set) drawn independently from a distribution $Q \in \mathcal{P}(\mathcal{X})$

null hypothesis $H_0 : Q = P_0$
alternate hypothesis $H_1 : Q = P_1$

for some known distributions $P_0, P_1 \in \mathcal{P}(\mathcal{X})$.

Def. A test for the sequence $X^n = (X_1, \dots X_n)$ is a region $A_n \subseteq \mathcal{X}^n$. We declare that the alternate hypothesis is true if the observed sequence satisfies $(x_1, x_2, \dots x_n) \in A_n$.

Two kinds of error:

$\alpha_n(A_n) = P_0^n(A_n)$ error of first kind

$\beta_n(A_n) = 1 - P_1^n(A_n)$ error of second kind

$= P_1^n(A_n^c)$

4.2 Symmetric hypothesis testing

We want to minimize

$$\varepsilon^*_{sym,n}(P_0,P_1) := \min_{A_n \subset \mathcal{X}^n} \frac{1}{2}\left(\alpha_n(A_n) + \beta_n(A_n)\right)$$

(assumes that P_0 and P_1 are equally likely!)

4.2.1 Total variation distance

$$\delta_{tvd}(P_0,P_1) := \frac{1}{2}\sum_{x\in\mathcal{X}} |P_0(x) - P_1(x)|$$

Lemma: $$\delta_{tvd}(P_0,P_1) = \max_{A\subset\mathcal{X}} \sum_{x\in A} P_0(x) - P_1(x) \qquad (*)$$

$$= \max_{A\subset\mathcal{X}} \sum_{x\in A} P_1(x) - P_0(x) \qquad (**)$$

Proof:

$$\sum_{x\in\mathcal{X}} P_0(x) - P_1(x) = 0 \Rightarrow$$

$$\sum_{x\in A} P_0(x) - P_1(x) = \sum_{x\in A^c} P_1(x) - P_0(x)$$

implies that $(*) = (**)$

$A = \{x : P_0(x) \geq P_1(x)\}$ is optimal for $(*)$

A^c is optimal for $(**)$

$$\Rightarrow \quad (*) = \sum_{x\in A} P_0(x) - P_1(x) = \sum_{x\in A} |P_0(x) - P_1(x)|$$

$$(**) = \sum_{x\in A^c} P_1(x) - P_0(x) = \sum_{x\in A^c} |P_0(x) - P_1(x)|$$

$x \in A$

$$\Rightarrow (*) + (**) = 2\,\delta_{tvd}(P_0, P_1)$$

$$\Rightarrow (*) = (**) = \delta_{tvd}(P_0, P_1)$$

□

The total variation distance is closely related to the 1-norm $\delta_{tvd}(P_0, P_1) = \frac{1}{2}\|P_0 - P_1\|_1$.

Prop. Assume two hypotheses $H_0: P_0$, $H_1: P_1$ with prior probabilities p for H_0 and $1-p$ for H_1. The minimal probability of error for a hypothesis test using $n=1$ sample is given by

$$\varepsilon^*_{p,1}(P_0, P_1) = \frac{1}{2}\left(1 - \|pP_0 - (1-p)P_1\|_1\right)$$

Special case $p = \frac{1}{2}$: $\varepsilon^*_{sym,1}(P_0, P_1) = \frac{1}{2}\left(1 - \delta_{tvd}(P_0, P_1)\right)$

Proof:

$$\varepsilon^*_{p,1} = \min_{A \subset \mathcal{X}} pP_0(A) + (1-p)P_1(A^c)$$

$$= p - \max_{A \subset \mathcal{X}} \left(pP_0(A^c) - (1-p)P_1(A^c)\right) \quad (*)$$

$$= 1-p - \max_{A \subset \mathcal{X}} \left((1-p)P_1(A) - pP_0(A)\right) \quad (**)$$

optimal $A = \{x \in \mathcal{X} : (1-p)P_1(x) \geq pP_0(x)\}$ for both $(*)$ and $(**)$

$$\Rightarrow 2\varepsilon^*_{P,1} = 1 - \sum_{x \notin A} pP_0(x) - (1-p)P_1(x) - \sum_{x \in A} (1-p)P_1(x) - pP_0(x)$$

$$= 1 - \sum_{x} \left| pP_0(x) - (1-p)P_1(x) \right| - \sum_{x \in A} \left| pP_0(x) - (1-p)P_1(x) \right|$$

$$= 1 - \| pP_0 - (1-p)P_1 \|_1$$

□

4.2.2. Chernoff exponent

Prop. For any P_0, P_1, we have

$$\lim_{n \to \infty} -\frac{1}{n} \log \varepsilon^*_{\text{sym},n} \geq C(P_0, P_1),$$

in fact, equality holds (but we are not showing this)

where $C(P_0, P_1) := -\min_{0 \leq \lambda \leq 1} \log \sum_{x \in \mathcal{X}} P_0(x)^{\lambda} P_1(x)^{1-\lambda}$.

In other words: For every $\delta > 0$, we have

$$\varepsilon^*_{\text{sym},n} \leq 2^{-n[C(P_0,P_1) - \delta]}$$

for sufficiently large n

Proof: $2\varepsilon^*_{\text{sym},n} = \min_{A_n \subset \mathcal{X}^n} P_0^n(A_n) + P_1(A_n^c)$

$$= \sum_{x^n \in \mathcal{X}^n} \min \{ P_0^n(x^n), P_1(x^n) \}$$

For any $\lambda \in [0,1]$

$$\leq \sum_{x^n \in \mathcal{X}^n} P_0^n(x^n)^{\lambda} P_1^n(x^n)^{1-\lambda}$$

$$= \sum_{x_1 \in \mathcal{X}} \cdots \sum_{x_n \in \mathcal{X}} P_0(x_1)^{\lambda} \cdots P_0(x_n)^{\lambda} \cdot P_1(x_1)^{1-\lambda} \cdots P_1(x_n)^{1-\lambda}$$

$$= \left(\sum_{x \in \mathcal{X}} P_0(x)^{\lambda} P_1(x)^{1-\lambda} \right)^n$$

$$\Rightarrow \frac{1}{n} \log \varepsilon^*_{\mathrm{sym},n} + \frac{1}{n} \leq \log \sum_{x \in \mathcal{X}} P_0(x)^{\lambda} P_1(x)^{1-\lambda}$$

$$\lim_{n \to \infty} -\frac{1}{n} \log \varepsilon^*_{\mathrm{sym},n} \geq -\log \sum_{x \in \mathcal{X}} P_0(x)^{\lambda} P_1(x)^{1-\lambda}$$

Since this holds for all $\lambda \in [0,1]$, we get the desired result by maximizing the rhs. over λ. □

4.3 Asymmetric hypothesis testing

Define $\beta_n^*(\varepsilon; P_0, P_1) := \min_{\substack{A_n \\ \text{s.t. } \alpha_n(A_n) \leq \varepsilon}} \beta_n(A_n)$

In the following, we assume that $D(P_0 \| P_1) < \infty$. Otherwise, there exists $x \in \mathcal{X}$ s.t. $P_1(x) > 0$ and $P_0(x) = 0$. (Caused in the homeworks.)

We will show that:

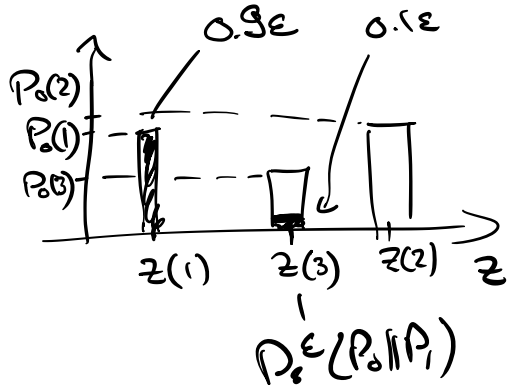
Thm: (Chernoff-Stein Lemma) For every $\varepsilon \in (0,1)$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \beta_n^*(\varepsilon) = D(P_0 \| P_1).$$

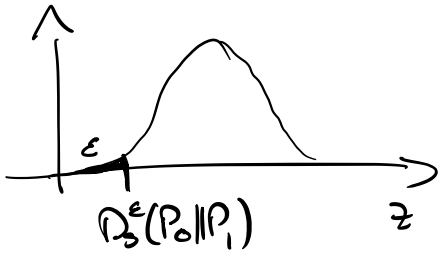
Tool: Information spectrum relative entropy

$$D_s^\varepsilon(P_0\|P_1) := \sup\left\{R \in \mathbb{R} : P_0\left[\log \frac{P_0(x)}{P_1(x)} \le R\right] \le \varepsilon\right\}$$

$Z(x)$: log-likelihood ratio



discrete case



towards central limit theorem (many copies)

Two Lemmas:

Lemma: $\lim_{n\to\infty} \frac{1}{n} D_s^\varepsilon(P_0^n \| P_1^n) = D(P_0\|P_1)$ for $\varepsilon \in (0,1)$

Lemma: Let $n \in \mathbb{N}$, $\varepsilon \in (0,1)$ and $\delta \in (0, 1-\varepsilon)$

Then $D_\varepsilon^\varepsilon(P_0^n \| P_1^n) \le -\log \beta_n^*(\varepsilon)$

$$\leq D_s^{\epsilon+\delta}(P_0^n \| P_1^n) + \log \frac{1}{\delta}$$

① ②

The theorem follows from these two lemmas.

Proof: ① To show this, we use tests of the form $A_{R,n} = \{x^n \in \mathcal{X}^n : P_0^n(x^n) \leq 2^R P_1^n(x^n)\}$

$= \{x^n \in \mathcal{X}^n : \log \frac{P_0^n(x^n)}{P_1^n(x^n)} \leq R\}$

↑ Neyman-Pearson tests (they are optimal)

Choose $R = D_s^{\epsilon}(P_0^n \| P_1^n)$, then

$$\alpha_n(A_{R,n}) = P_0^n(A_{R,n}) =$$

$$= P_0^n\left[\log \frac{P_0^n(x^n)}{P_1^n(x^n)} \leq R\right] \leq \epsilon$$

↑ by definition of R

$$\beta_n^*(\epsilon) \leq \beta_n(A_{R,n})$$

$$= P_1(A_{R,n}^c)$$

$$= \sum_{x^n \in \mathcal{X}^n} P_1^n(x^n) \mathbb{1}\{P_0^n(x^n) > 2^R P_1^n(x^n)\}$$

$$\leq 2^{-R} \sum_{x^n \in \mathcal{X}^n} P_0^n(x^n) \mathbb{1}\{ \cdots \}$$

$$\leq 2^{-R}$$

$$\Rightarrow -\log \beta_n^*(\varepsilon) \geq R$$

(2) Let A_n be the optimal test for $\beta_n^*(\varepsilon)$, thus $\alpha_n(A_n) \leq \varepsilon$, $\beta_n^*(\varepsilon) = \beta_n(A_n)$

Then,

$$P_0\Big[\underbrace{\log \frac{P_0^n(x^n)}{P_1^n(x^n)}}_{Z_n} > R\Big]$$

$$= \sum_{x^n \in \mathcal{X}^n} P_0^n(x^n) \mathbb{1}\{P_0^n(x^n) > 2^R P_1^n(x^n)\}$$

$$\geq \sum_{x^n \in \mathcal{X}^n} \left(P_0^n(x^n) - 2^R P_1(x^n)\right) \cdot \mathbb{1}\{P_0^n(x^n) > 2^R P_1^n(x^n)\}$$

$$\geq \sum_{x^n \in \mathcal{X}^n} \left(P_0^n(x^n) - 2^R P_1(x^n)\right) \mathbb{1}\{x^n \in A_n^c\}$$

$$= P_0^n(A_n^c) - 2^R P_1^n(A_n^c)$$

$$= 1 - \varepsilon - 2^R \beta_n^*(\varepsilon)$$

$$\Rightarrow P_0[Z_n \leq R] \leq \varepsilon + 2^R \beta_n^*(\varepsilon)$$

Choose $R = \log \delta - \log \beta_n^*(\varepsilon)$

$\Rightarrow P_0[Z_n \in R] \leq \varepsilon + \delta$

$\Rightarrow D_\delta^{\varepsilon+\delta}(P_0^n \| P_1^n) \geq R = -\log \beta_n^*(\varepsilon) + \log \delta$ $\square$
