L03

Wednesday, August 24, 2022 5:02 PM

* $\boxed{\text{Chernoff Bound}}$

Consider the log likelihood ratio test

$$\hat{X} = \begin{cases} 0 & \\ 0 \text{ or } 1 & \text{if } \log \dfrac{P_{Y|X}(y|0)}{P_{Y|X}(y|1)} \triangleq T(y) \begin{matrix} > \\ = \\ < \end{matrix} \tau \\ 1 & \end{cases}$$

The false alarm prob is then

$$P(T(Y) < \tau \mid X=0) \overset{\text{shorthand}}{=\!=} P_0(T(Y) < \tau)$$

Misdetection prob.

$$P_1(T(Y) > \tau)$$

Many times they are hard to compute.

$\Rightarrow$ Find a bound instead.

$\boxed{\text{Markov Inequality}}$

Assuming $P(X < 0) = 0$

$$\Rightarrow P(X \geq d) \leq \frac{E(X)}{d}$$

$\boxed{\text{Chernoff Bound}}$

For general X

$$P(X \geq d)$$

$$\leq P(sX \geq sd) \quad \text{for any } s \geq 0$$

$$= P(e^{sX} \geq e^{sd}) \leq \frac{E(e^{sX})}{e^{sd}} \quad \text{for any } s > 0$$

$$\leq \min_{s \geq 0} \frac{E(e^{sX})}{e^{sd}}$$

For i.i.d. $X_1, \dots, X_n$.

Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ be the sample mean

$$P(\bar{X} \geq d) = P\left(\sum_{i=1}^{n} X_i \geq nd\right)$$

$$\leq \min_{s \geq 0} \frac{E(e^{s \sum X_i})}{e^{snd}}$$

indep.

$$\prod^{n} E(e^{sX_i})$$

$$s \geq 0 \qquad e^{snd} \qquad \Big\} \text{ indep.}$$

$$= \min_{s \geq 0} \frac{\prod_{i=1}^{n} E(e^{sX_i})}{e^{snd}}$$

$$= \min_{s \geq 0} \frac{\left(E(e^{sX})\right)^n}{e^{snd}} = \left(\min_{s \geq 0} \frac{E(e^{sX})}{e^{sd}}\right)^n$$

$\Rightarrow$ For a sample mean $\overline{X}$

$$P(\overline{X} \geq d) \leq \left(\text{The CB for each } X\right)^n$$

for any n. (exponential decay)

Moreover, the Chernoff bound is asymptotically tight (under general conditions)

I.e. If $E(X) < d$ then $\forall \varepsilon > 0, \exists n_0$ such that

$$P(\overline{X} \geq d) \geq \left(\min_{s \geq 0} \frac{E(e^{sX})}{e^{sd}} - \varepsilon\right)^n$$

$\forall n \geq n_0$

Example: X_i : i.i.d. Bernoulli R.V.

$$X_i \in_R \{0,1\}.$$

$$\overline{X}_{1000} = \frac{1}{1000} \sum_{i=1}^{1000} X_i$$

Q: $P(\overline{X}_{1000} \geq 0.75)$

Solution 1: Binomial distribution

Solution 2: Gaussian approximation (by the central limit theorem)

Solution 3: By Chernoff bound, $d = 0.75$

$$\min_{s \geq 0} \frac{E(e^{sX})}{e^{s 0.75}} = \min_{s \geq 0} \frac{\frac{1}{2} + \frac{1}{2} e^s}{e^{s \times 0.75}}$$

$$s^* = \log \frac{0.75}{0.25} \Big) = 2 \times 3^{-3/4} \simeq 0.8774$$

$$\Rightarrow P(\overline{X}_{1000} \geq 0.75) \leq (2 \times 3^{-3/4})^{1000}$$

* Application of the Chernoff Bound to

hypothesis testing with i.i.d. observations

H_0: $Y_1 \cdots Y_n$ are i.i.d with $Y_i \sim P_0$

H_1: $Y_1 \cdots Y_n$ are i.i.d with $Y_i \sim P_1$

Q: Find the expression of the misdetection prob of a <u>Log likelihood ratio</u> test with threshold γ.

Ans: $P_1\left(\log \frac{P_0(Y_1, Y_2, \cdots, Y_n)}{P_1(Y_1, Y_2, \cdots, Y_n)} \geq \gamma \right)$

Or equivalently

$$\because \log \frac{P_0(Y_1, \cdots, Y_n)}{P_1(Y_1, \cdots, Y_n)} = \log \frac{P_0(Y_1) P_0(Y_2) \cdots P_0(Y_n)}{P_1(Y_1) P_1(Y_2) \cdots P_1(Y_n)}$$

$$= \sum_{i=1}^{n} \log \frac{P_0(Y_i)}{P_1(Y_i)} \triangleq \sum_{i=1}^{n} T(Y_i) \text{ where } T(y) = \log \frac{P_0(y)}{P_1(y)}$$

$$\Rightarrow P_1\left(\frac{1}{n} \sum_{i=1}^{n} T(Y_i) \geq \tau \right)$$

$$\tau \triangleq \frac{\gamma}{n}$$

If $\tau = 0 \Rightarrow$ MD prob of the ML detector

If $\tau = \frac{\log(P(X=1)) - \log(P(X=0))}{n}$

$\Rightarrow$ MD prob of the MAP detector

What if $\tau = E_0(T(Y))$?

or more rigorously $\tau = E_0(T(Y)) - \varepsilon$ for some small $\varepsilon > 0$

What is the physical meaning of

$$① = P_1\left(\frac{1}{n}\sum_{i=1}^{n} T(Y_i) \geq E_0(T(Y)) - \varepsilon\right) ?$$

Ans: $T(Y_i)$ are i.i.d. when H_0 is true, the sample mean $\frac{1}{n}\sum_{i=1}^{n} T(Y_i)$ concentrate around $E_0(T(Y))$

$\Rightarrow$ ① is the prob when H_1 is true but the observations $Y_1, \ldots, Y_n$ look as if H_0 is true.

What is the value of

$$P_1\left(\frac{1}{n}\sum_{i=1}^{n} T(Y_i) \geq \underbrace{E_0(T(Y))}_{\tau^*}\right) ?$$

By Chernoff bound.

$$P_1\left(\frac{1}{n}\sum_{i=1}^{n} T(Y_i) \geq \tau^*\right)$$

$$\approx \left(\min_{s\geq 0} \underbrace{\frac{E_1(e^{sT(Y)})}{e^{s\tau^*}}}_{\triangleq f(s)}\right)^n$$

We need to solve

$$f'(s) = E_1\left((T(Y)-\tau^*)\, e^{s(T(Y)-\tau^*)}\right) = 0$$

Try $s=1$

$$e^{\tau^*} f'(1) = E_1\left((T(Y)-\tau^*)\, e^{T(Y)}\right)$$

$$e^{\tau^*} f'(1) = E_1\left(\left(T(Y) - \tau^*\right) e^{T(Y)} \right)$$

$$= \int_y \left(\log \frac{p_0(y)}{p_1(y)} - \tau^* \right) e^{\log \frac{p_0(y)}{p_1(y)}} \times p_1(y)\, dy$$

$$= \int \left(\log \frac{p_0(y)}{p_1(y)} - \tau^* \right) \frac{p_0(y)}{p_1(y)}\, p_1(y)\, dy$$

$$= \int \left(\log \frac{p_0(y)}{p_1(y)} - \tau^* \right) p_0(y)\, dy$$

$$= E_0\left(T(Y) - \tau^* \right)$$

$$= E_0(T(Y)) - \tau^* = 0$$

recall that $\tau^* \triangleq E_0(T(Y))$

One can actually show that $f(s)$ is a convex function of s. So $s^* = 1$ is the minimum. H.V. Poor pp. 86 - 91

$$\Rightarrow \min_{s \geq 0} f(s) = f(1)$$

$$= \frac{E_1\left(e^{T(Y)}\right)}{e^{\tau^*}}$$

$$= \frac{\int_y e^{\log \frac{P_0(y)}{P_1(y)}} \cdot P_1(y)\,dy}{e^{\tau^*}}$$

$$= \frac{\int \frac{P_0(y)}{P_1(y)} P_1(y)\,dy}{e^{\tau^*}} = e^{-\tau^*}$$

This special $\tau^* \triangleq E_0\left(\log \frac{P_0(Y)}{P_1(Y)}\right)$

is denoted as $D(P_0 \| P_1)$ the divergence between P_0 distribution & the P_1 distribution

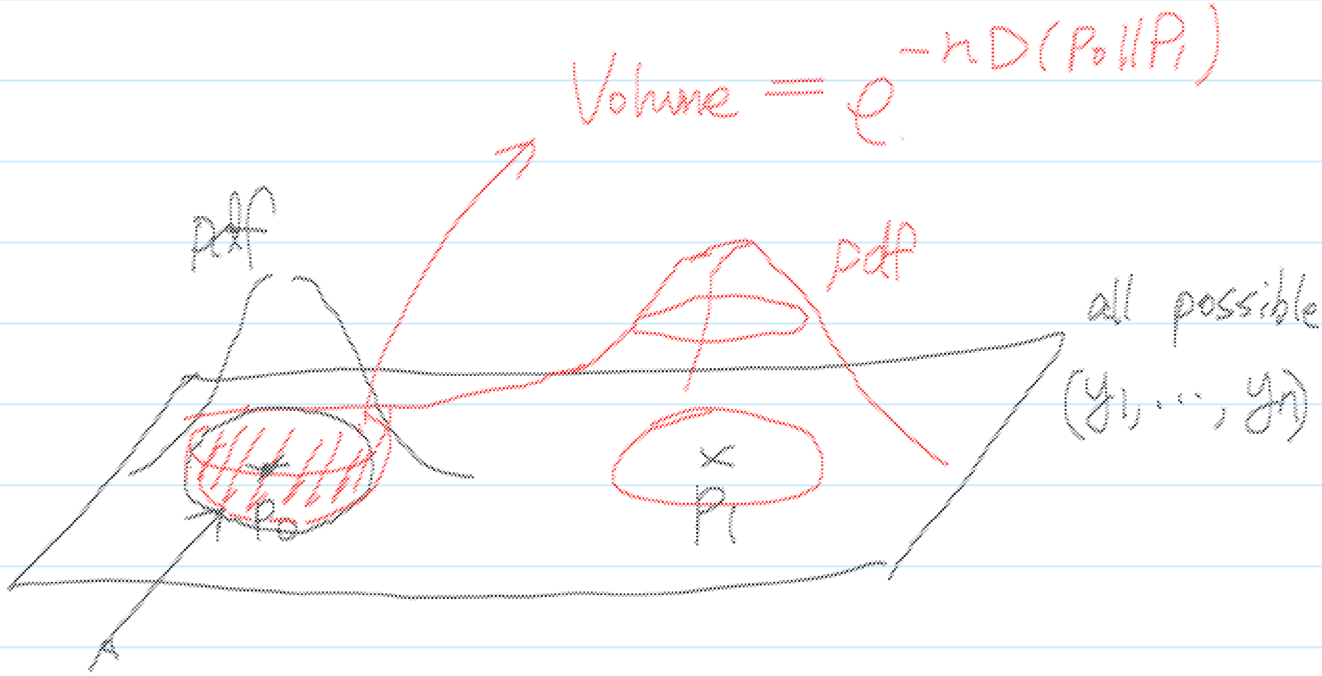
$$D(P_0\|P_1) = E_0\left(\log \frac{P_0(Y)}{P_1(Y)}\right) = \int_y P_0(y) \log \frac{P_0(y)}{P_1(y)}\,dy$$

(Also known as the Kullback-Leibler information number.)

The intuition behind $D(P_0 \| P_1)$ is:

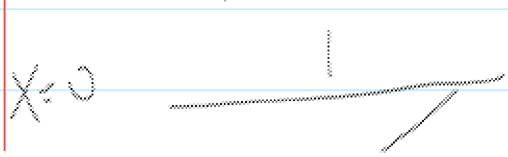
The prob that H_1 is true but the

i.i.d. $Y_1, \cdots, Y_n$ look as if H_0 is true.
is upper bounded by $e^{-nD(P_0 \| P_1)}$
(which is asymptotically tight due to the Chernoff bound.)

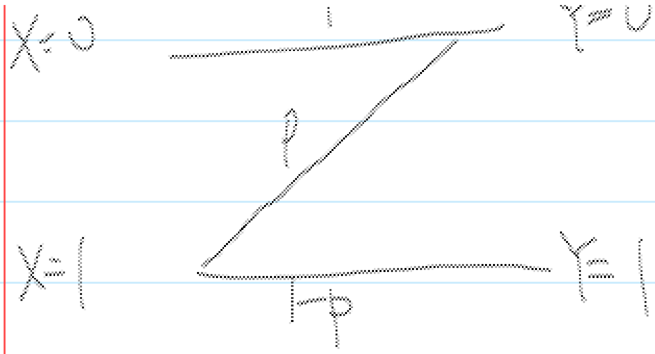


if H_0 is true, the most likely values of $(y_1, \ldots, y_n)$ is called the <u>typical set</u> of P_0

Example: The Z channel



$Y=0$ If $X=0$. $P_0(y) =$ if $y=0$ / $y=1$



If $X=0$, $P_0(y) = \begin{cases} 1 & \text{if } y=0 \\ 0 & y=1 \end{cases}$

If $X=1$, $P_1(y) = \begin{cases} p & \text{if } y=0 \\ 1-p & y=1 \end{cases}$

Q: $D(P_0 \| P_1)$ & $D(P_1 \| P_0)$

Ans: $D(P_0 \| P_1) = E_0\left(\log\left(\frac{P_0(Y)}{P_1(Y)} \right) \right)$

$$= 1 \times \log \frac{P_0(0)}{P_1(0)} + 0 = \log\left(\frac{1}{p}\right)$$

$D(P_1 \| P_0)$

$$= p \log \frac{P_1(0)}{P_0(0)} + (1-p) \log \frac{\overbrace{P_1(1)}^{=(1-p)}}{\underbrace{P_0(1)}_{=0}}$$

$$= \infty$$

Intuition:

$P(Y_1, \cdots, Y_n \text{ look like } H_0 \text{ is true} \mid H_1 \text{ is true})$

$$\underset{(\approx)}{\leq} e^{-nD(P_0 \| P_1)} = \left(\frac{1}{p}\right)^{-n} = p^n$$

I.e. receiving n consecutive zeros even though $X=1$.

$P(Y_1, \cdots, Y_n$ look like H_1 is true $| H_0$ is true$)$

$\leq e^{-nD(P_1 \| P_0)} = 0$

$(\approx)$

Why it is zero? $\because$ When $X=0$, the output (with $P_0(Y=1)=0$) can never mimic (look like) $X=1$, which has $P_1(Y=1) \neq 0$

Properties of $D(P_0 \| P_1)$

① $D(P_0 \| P_1) \geq 0$

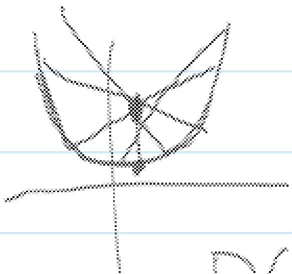
Intuition:

$e^{-nD(P_0 \| P_1)}$ is a tight prob bound $\Rightarrow e^{-nD(P_0 \| P_1)} \in [0,1]$

$\Leftrightarrow 0 \leq D(P_0 \| P_1) \leq \infty$

pf: By Jensen's inequality:

For any convex function $f(\cdot)$



$$\underbrace{E(f(X))}_{y \text{ coordinate}} \geq f(\underbrace{E(X)}_{x \text{ coordinate}})$$

$$D(P_0 \| P_1) = E_0\left(\log \frac{P_0(Y)}{P_1(Y)}\right)$$

$$= E_0\left(-\log\left(\frac{P_1(Y)}{P_0(Y)}\right)\right)$$

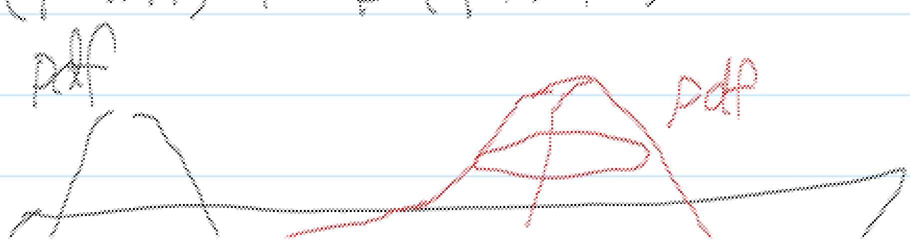
$\because -\log(\cdot)$ is convex

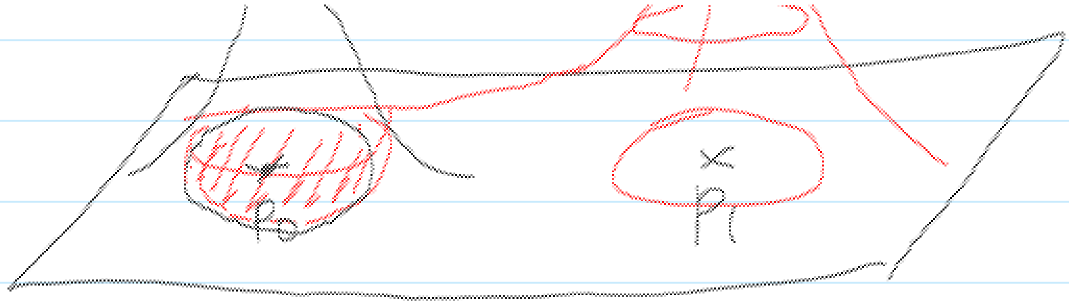
$$\geq -\log\left(E_0\left(\frac{P_1(Y)}{P_0(Y)}\right)\right)$$

$$= -\log 1 = 0$$

② In general

$$D(P_0 \| P_1) \neq D(P_1 \| P_0)$$





How to use $D(P_0 \| P_1)$?

Ex: X_i is i.i.d Bernoulli with uniform distribution

$$\bar{X}_{1000} = \frac{1}{1000} \sum_{i=1}^{1000} X_i$$

Q: $P(\bar{X}_{1000} > 0.75) = ?$

Ans: The actual distribution is $P_1 = \frac{1}{2} / \frac{1}{2}$

but the outcomes look like $P_0 = 0.75 / 0.25$

$$\Rightarrow P(\bar{X}_{1000} > 0.75) \approx e^{-1000 D(P_0 \| P_1)}$$

$$= e^{-1000\left(0.75 \times \log \frac{0.75}{0.5} + 0.25 \log \frac{0.25}{0.5}\right)}$$

$$= \left(2 \times 3^{-0.25}\right)^{1000}$$

※

Note: The actual distribution is P_1 (H1 is true) but

$D(P_0 \| P_1) = E_0\left(\log \frac{P_0(Y)}{P_1(Y)}\right)$ is evaluated by P_0

Properties of $D(P_0\|P_1)$.

③ $D(P_0\|P_1) < \infty$ if $\forall x$ we have $P_1(X=x)=0$ $\Rightarrow P_0(X=x)=0$

Mathematically Suppose $E_0\left(\log \frac{P_0(Y)}{P_1(Y)}\right) < \infty$

So if $P_1(y) = 0 \Rightarrow P_0(y) = 0$

Intuition: $D(P_0\|P_1) < \infty$

$\Leftrightarrow$ When H1 is true, the "prob of $Y_1, \dots, Y_n$ looking like H0 is true" is non-zero

$\Longleftrightarrow$ Given H_1 is true, for $Y_1, \cdots, Y_n$ to mimic H_0 is true, we must have

$$P_1(y) = 0 \Rightarrow P_0(y) = 0$$

④ $D(P_0 \| P_1)$ is a convex function with respect to P_0 & P_1

Namely for any pair of P_0 P_1, Q_0, Q_1.

We have

$$D(\alpha P_0 + (1-\alpha) Q_0 \| \alpha P_1 + (1-\alpha) Q_1)$$

$$\leq \alpha D(P_0 \| P_1) + (1-\alpha) D(Q_0 \| Q_1)$$

(Note the LHS is

$$\sum_y (\alpha P_0(y) + (1-\alpha) Q_0(y)) \log \frac{\alpha P_0(y) + (1-\alpha) Q_0(y)}{\alpha P_1(y) + (1-\alpha) Q_1(y)} \Bigg)$$

Pf: Exercise (Hint: By Jensen's inequality)

Intuition: Consider side information S

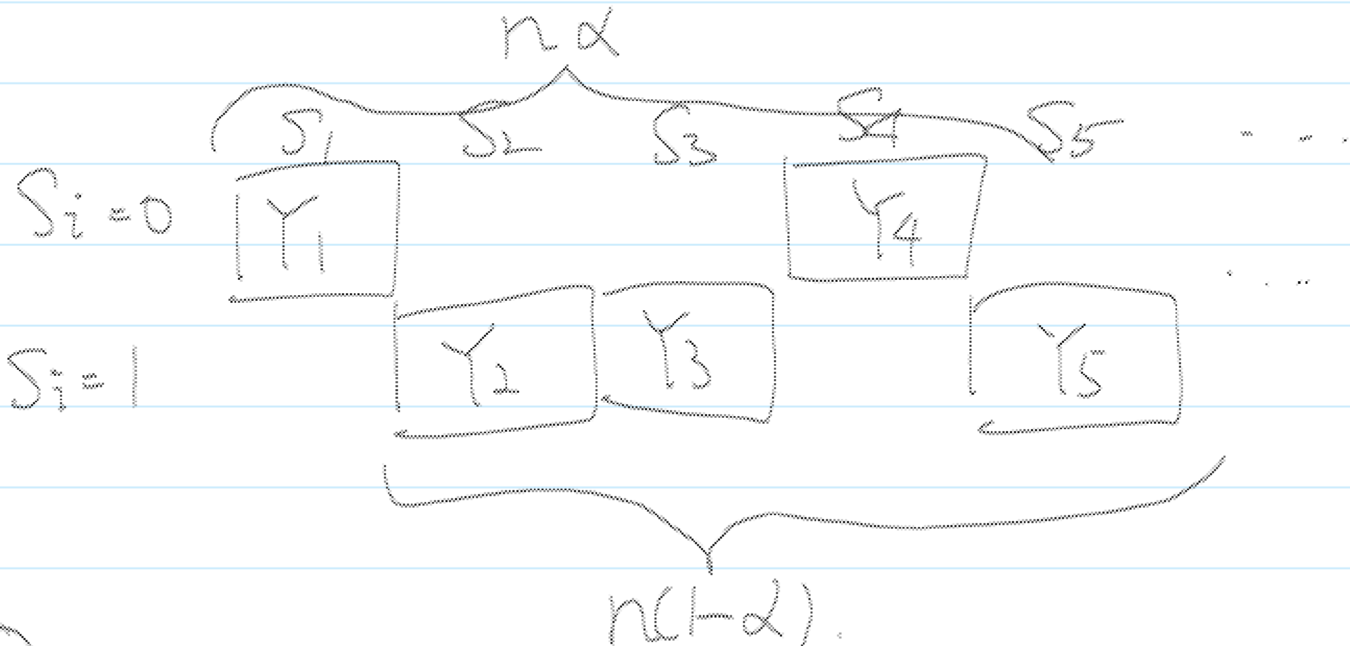
$P_0(y) = P(Y = y \mid X = 0, S = 0)$

$Q_0(y) = P(Y = y \mid X = 0, S = 1)$

$P_1(y) = P(Y = y \mid X = 1, S = 0)$

$Q_1(y) = P(Y = y \mid X = 1, S = 1)$

$P(S = 0) = \alpha \qquad P(S = 1) = 1 - \alpha$



Q: With the side info $S_1, \dots, S_n$ when will $Y_1, \dots, Y_n$ look like H_0 is true but actually H_1 is true?