ELEG 5633 Detection and Estimation
Kullback-Leibler (KL) Divergence

Jingxian Wu

Department of Electrical Engineering
University of Arkansas

Outline

- Motivation: LLR of i.i.d. Samples
- Kullback-Leibler Divergence
- KL Divergence of Gaussian Distributions
- Bounding Error Probabilities

Motivation

Example

Consider a binary Hypothesis:

$$H_1 : X \sim p_1(X)$$
$$H_0 : X \sim p_0(X)$$

Based on a sequence of independent observations $x_1, x_2, \cdots, x_n$, design the ML test.

Log-likelihood Ratio (LLR) of Independent Samples

Consider a sequence of idependent and identically distributed (i.i.d.) random variables (RVs) $X_1, X_2, \cdots, X_n \sim p_m(x)$, for $m \in \{0, 1\}$. For a given observation $x = [x_1, x_2, \cdots, x_n]$

- Likelihood ratio

$$L(x) = \frac{p_1(x_1, x_2, \cdots, x_n)}{p_0(x_1, x_2, \cdots, x_n)} = \prod_{i=1}^{n} \frac{f_1(x_i)}{f_0(x_i)}$$

- Log-likelihood ratio (LLR)

$$\log L(x) = \sum_{i=1}^{n} \log \frac{p_1(x_i)}{p_0(x_i)}$$

- Normalized LLR

$$\Lambda(x) = \frac{1}{n} \sum_{i=1}^{n} \log \frac{p_1(x_i)}{p_0(x_i)}$$

Since x is a random vector, $\Lambda(x)$ is a random variable.

Convergence of Normalized LLR

- Strong Law of Large Numbers
  Consider a sequence of i.i.d. RV $X_1, X_2, \cdots$ with $\mathbb{E}(X_i) = \mu$. Then the sample mean, $\bar{x}_n := \frac{1}{n}\sum_{i=1}^{n} x_i$, converges to the expected value μ almost surely (a.s.) as $n \to \infty$,

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \xrightarrow{a.s.} \quad \mu, \text{ when } n \to \infty$$

  which means $\Pr\left(\lim_{n\to\infty} \bar{x}_n = \mu\right) = 1$.
- Convergence of $\Lambda(x)$: Assume $X \sim q(x)$. Then as $n \to \infty$,

$$\Lambda(x) = \frac{1}{n}\sum_{i=1}^{n} \log\frac{p_1(x_i)}{p_0(x_i)} \quad \xrightarrow{a.s.} \quad \mathbb{E}_q\left[\log\frac{p_1(X)}{p_0(X)}\right], \text{ when } n \to \infty$$

  where $q(x)$ could be the same or different as $p_m(x)$.

Kullback-Leibler (KL) Divergence

The KL divergence between from distribution $p_1(x)$ to $p_0(x)$ is

$$D(p_1 \| p_0) = \mathbb{E}_{p_1}\left[\log \frac{p_1(X)}{p_0(X)}\right] \tag{1}$$

- Continuous RV

$$D(p_1 \| p_0) = \int \log \frac{p_1(x)}{p_0(x)} p_1(x) dx$$

- Discrete RV

$$D(p_1 \| p_0) = \sum_k \log \frac{P_1(X = x_k)}{P_0(X = x_k)} P_1(X = x_k)$$

- KL divergence measures the non-symmetric difference between $p_1(x)$ and $p_0(x)$
  - Note that $D(p_1 \| p_0) \neq D(p_0 \| p_1)$ in general

Example

Consider Bernoulli distributions $P_m \sim \text{Bernoulli}(\rho_m)$, for $m = 0, 1$. Find $D(P_1 \| P_0)$ and $D(P_0 \| P_1)$.

KL Divergence Properties

- Gibb's inequality

$$D(p_1 \| p_0) \geq 0$$

Proof: Jensen's inequality: $\mathbb{E}\left[\log(X)\right] \leq \log\left(\mathbb{E}(X)\right)$

$$\begin{aligned}
D(p_1 \| p_0) &= \mathbb{E}_{p_1}\left[\log \frac{p_1(x)}{p_0(x)}\right] = -\mathbb{E}_{p_1}\left[\log \frac{p_0(x)}{p_1(x)}\right] \\
&\geq -\log\left(\mathbb{E}_{p_1}\left[\frac{p_0(x)}{p_1(x)}\right]\right) \\
&= -\log\left(\int \frac{p_0(x)}{p_1(x)} p_1(x) dx\right) = -\log(1) = 0
\end{aligned}$$

- $D(p_1 \| p_0) = 0$ if and only if (iff) $p_1(x) = p_0(x)$

Example

Consider a binary Hypothesis: $H_1 : X \sim p_1(X)$ and $H_0 : X \sim p_0(X)$. Based on a sequence of i.i.d. observations $x_1, x_2, \cdots, x_n$, design the LRT.

- The ML test is

$$\Lambda(x) = \frac{1}{n}\sum_{i=1}^{n} \log \frac{p_1(x_i)}{p_0(x_i)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{n}\log\lambda$$

- When $n \to \infty$, $\frac{1}{n}\log\lambda \to 0$

  - $H_1 : \Lambda(x|H_1) \xrightarrow{a.s.} E_{p_1}\left[\log \frac{p_1(X)}{p_0(X)}\right] = D(p_1 \| p_0) \geq 0.$
    Thus $P_{\mathsf{MD}} \to 0$ as $n \to \infty$

  - $H_0 : \Lambda(x|H_0) \xrightarrow{a.s.} E_{p_0}\left[\log \frac{p_1(X)}{p_0(X)}\right] = -D(p_0 \| p_1) \leq 0$
    Thus $P_{\mathsf{FA}} \to 0$ as $n \to \infty$

KL Divergence of Gaussian Distributions

Example

Find the KL divergence from $p_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ to $p_0 \sim \mathcal{N}(\mu_0, \sigma^2)$

Answer: $D(p_1 \| p_0) = \frac{(\mu_1 - \mu_0)^2}{2\sigma^2}$

KL divergence of multivariate Gaussian RV

The KL divergence from $p_1 \sim \mathcal{N}(\mu_1, \Sigma_1)$ to $p_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$ is

$$D(p_1 \| p_0) = \frac{1}{2}\left[\mathrm{tr}(\Sigma_0^{-1}\Sigma_1) + (\mu_0 - \mu_1)^T \Sigma_0^{-1}(\mu_0 - \mu_1) - n + \log\left(\frac{\det \Sigma_0}{\det \Sigma_1}\right)\right]$$

where n is the dimension of the RV.

- Special Cases:
  - If $\Sigma_0 = \Sigma_1$, then

$$D(p_1 \| p_0) = \frac{1}{2}(\mu_0 - \mu_1)^T \Sigma_0^{-1}(\mu_0 - \mu_1) = D(p_0 \| p_1)$$

  - If $\mu_0 = \mu_1$, then

$$D(p_1 \| p_0) = \frac{1}{2}\left[\mathrm{tr}(\Sigma_0^{-1}\Sigma_1) - n + \log\left(\frac{\det \Sigma_0}{\det \Sigma_1}\right)\right]$$

Example

Consider a binary Hypothesis: $H_1 : X \sim \mathcal{N}(\mu_1, \sigma^2 \mathbf{I}_n)$ and $H_0 : X \sim \mathcal{N}(\mu_0, \sigma^2 \mathbf{I}_n)$. where X is an n dimensional vector. Find the ML detector and P_e.

Solution:

- ML test:

$$-\|x-\mu_1\|^2 \underset{H_0}{\overset{H_1}{\gtrless}} -\|x-\mu_0\|^2$$

$$y=(\mu_1-\mu_0)^T x \underset{H_0}{\overset{H_1}{\gtrless}} \frac{1}{2}(\mu_1^T\mu_1-\mu_0^T\mu_0)=\lambda$$

- $y|H_i \sim \mathcal{N}(\mu_{y|H_i}, \sigma^2_{y|H_1})$

$$\mu_{y|H_0}=(\mu_1-\mu_0)^T\mu_0, \quad \sigma^2_{y|H_0}=\sigma^2\|\mu_1-\mu_0\|^2$$

- $P_{\mathsf{FA}}=\Pr(y>\lambda|H_0)=Q\left(\frac{\lambda-\mu_{y|H_0}}{\sigma\|\mu_1-\mu_0\|}\right)=Q\left(\frac{\|\mu_1-\mu_0\|}{2\sigma}\right)=$
  $Q\left(\sqrt{\frac{\|\mu_1-\mu_0\|^2}{4\sigma^2}}\right)=Q\left(\sqrt{\frac{1}{2}D(p_1\|p_0)}\right)$

$$P_e=\pi_0 P_{\mathsf{FA}}+\pi_1 P_{\mathsf{MD}}=Q\left(\sqrt{\frac{1}{2}D(p_1\|p_0)}\right)$$

Example

Consider a binary Hypothesis:

$$H_0 : X \sim \mathcal{N}(a_0, \sigma^2)$$

$$H_1 : X \sim \mathcal{N}(a_1, \sigma^2)$$

Where X is a scalar. If we have n i.i.d. observations $x_1, x_2, \cdots, x_n$, find the ML detector and P_e.

- Solution: Let $\mu_i = [a_i, a_i, \cdots, a_i]^T$

$$P_e = Q\left(\sqrt{\frac{\|\mu_1 - \mu_0\|^2}{4\sigma^2}}\right) = Q\left(\sqrt{\frac{n(a_1 - a_0)^2}{4\sigma^2}}\right) = Q\left(\sqrt{\frac{n}{2}D(p_1\|p_0)}\right)$$

where $D(p_1\|p_0) = \frac{(a_1-a_0)^2}{2\sigma^2}$

Can we get P_e if X is non-Gaussian distributed?

Bounding Error Probabilities

Hoedings Inequality

If $Z_1, Z_2, \cdots, Z_n$ are i.i.d. and $a \leq Z_i \leq b$, then

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n} Z_i - \mathbb{E}[Z] > \epsilon\right) \leq e^{-\frac{2n\epsilon^2}{c^2}}$$
$$\Pr\left(\mathbb{E}[Z] - \frac{1}{n}\sum_{i=1}^{n} Z_i > \epsilon\right) \leq e^{-\frac{2n\epsilon^2}{c^2}}$$

where $c^2 = (b-a)^2$.

Bounding Error Probabilities

Theorem

Consider a binary Hypothesis: $H_1 : X \sim p_0(X)$ and $H_0 : X \sim p_1(X)$. Assume $0 < \alpha \leq p_i(x) \leq \beta < \infty$, for $i = 0, 1$. Based on a sequence of i.i.d. observations $X_1, X_2, \cdots, X_n$, then for ML test

$$P_{\mathsf{FA}} \leq \exp\left(-\frac{2}{c^2} n D^2(p_0 \| p_1)\right)$$

$$P_{\mathsf{MD}} \leq \exp\left(-\frac{2}{c^2} n D^2(p_1 \| p_0)\right)$$

where $c^2 = 4 \log^2 \frac{\beta}{\alpha}$.

Sketch of Proof

- Let $Z_i = \log \frac{p_1(X_i)}{p_0(X_i)}$, $\Lambda_n = \frac{1}{n}\sum_{i=1}^{n} Z_i$. The ML test is $\Lambda_n \underset{H_0}{\overset{H_1}{\gtrless}} 0$
- $\mathbb{E}[Z_i|H_0] = -D(p_0\|p_1)$, $\mathbb{E}[Z_i|H_1] = D(p_1\|p_0)$
- $$\begin{aligned} P_{\mathsf{FA}} &= \Pr(\Lambda_n > 0|H_0) = \Pr(\Lambda_n - \mathbb{E}[Z_i|H_0] > -\mathbb{E}[Z_i|H_0]|H_0) \\ &\leq \exp\left(-\frac{2}{c^2} n D^2(p_0\|p_1)\right) \end{aligned}$$

  where $c^2 = (b-a)^2 = \left(\log \frac{\beta}{\alpha} - \log \frac{\alpha}{\beta}\right)^2 = 4\log^2 \frac{\beta}{\alpha}$
- $$\begin{aligned} P_{\mathsf{MD}} &= \Pr(\Lambda_n < 0|H_1) = \Pr(\Lambda_n - \mathbb{E}[Z_i|H_1] < -\mathbb{E}[Z_i|H_1]|H_1) \\ &\leq \exp\left(-\frac{2}{c^2} n D^2(p_1\|p_0)\right) \end{aligned}$$

Example

Consider a binary hypothesis:

$$H_0 : X \sim \text{Bernoulli}(0.3)$$
$$H_1 : X \sim \text{Bernoulli}(0.5)$$

- Find $D(P_1 \| P_0)$ and $D(P_0 \| P_1)$
- Based on a sequence of i.i.d. observations $x_1, \cdots, x_n$, design the ML test
- What is the upper bound of P_{FA}?
- How many i.i.d. samples do we need to achieve $P_{\text{FA}} < 0.1$?

Solution:

- $$D(P_1 \| P_0) = \mathbb{E}_1\left[\log \frac{P_1(X)}{P_0(X)}\right] = 0.5 \log \frac{0.5}{0.3} + 0.5 \log \frac{0.5}{0.7} = 0.0872$$
$$D(P_0 \| P_1) = \mathbb{E}_0\left[\log \frac{P_1(X)}{P_0(X)}\right] = 0.3 \log \frac{0.3}{0.5} + 0.7 \log \frac{0.7}{0.5} = 0.0823$$
- Under H_0: $0.3 \le \Pr(X = x_i | H_0) \le 0.7$
Under H_1: $\Pr(X = x_i | H_1) = 0.5$
Thus $0.3 \le \Pr(X = x_i | H_m) \le 0.7$, $\alpha = 0.3, \beta = 0.7$.
- $c^2 = 4 \log^2 \frac{\beta}{\alpha} = 2.8717$
- $P_{\mathsf{FA}} \le \exp\left(-\frac{2}{c^2} n D^2(P_0 \| P_1)\right) = \exp(-0.0047n)$
- $$\exp(-0.0047n) \le 0.1 \Rightarrow n \ge \frac{1}{0.0047} |\log 0.1| \approx 490$$