Stochastic Divergences and Applications to Multi-Dimensional Goodness-of-Fit Tests for Spectrum Sensing

Sanjeev G.
SPC Lab.,
Dept. of ECE,
IISc

06 October 2012

Introduction to Stochastic Divergences

- It all started with Measure theory (Hellinger's Integral) $\longrightarrow$ Statistics (ISI) $\longrightarrow$ Information Theory!
- Roughly, a Stochastic Divergence measures the similarity between two discrete/continuous probability distributions
- Earliest examples include Hellinger's distance, Mahalanobis distance, Bhattacharyya distance, Kullback-Leibler divergence etc.

Connection to Information Theory (1/4)

- Logarithm as a measure of information [Shannon1948] (restricted scope).
- $\mathcal{P} \triangleq \{P = (p_1, p_2, \cdots, p_n) : \sum_{i=1}^{n} p_i = 1, p_i > 0, i = 1, 2, \cdots, n\}$
  $\mathcal{Q} \triangleq \{Q = (q_1, q_2, \cdots, q_n) : \sum_{i=1}^{n} q_i = 1, q_i \geq 0, i = 1, 2, \cdots, n\}$
- A measure of information, $H(P) = -\sum_{i=1}^{n} p_i \log p_i$ (following four postulates [Ash]).
- "Conditional entropy", $H(P||Q) = -\sum_{i=1}^{n} p_i \log q_i$
- $H(P)$ satisfies additivity, recursivity and sum representations

Connection to Information Theory (2/4)

- In other problems, other functions serve better as measures of information! [Renyi1961]
- Renyi's entropy of order α:
  $H_\alpha(P) = (1-\alpha)^{-1} \log \left(\sum_{i=1}^{n} p_i^\alpha\right), \alpha > 0, \alpha \neq 1$
- $\lim_{\alpha \to 1} H_\alpha(P) = H(P)$
- $H_\alpha(P)$ satisfies additivity, but not recursivity and sum representations
- Renyi's entropy finds its use in numerous applications as well!

Connection to Information Theory (3/4)

- [AczelDaroczy1963], [Varma1966], [Kapur1967], [HavrdaCharvat1967], [BelisGuiasu1968], [Rathie1970], [Arimoto1971], [SharmaMittal1975], [Taneja1975], [Picard1979], [Ferreri1980], [Sant'annaTaneja1983], ...

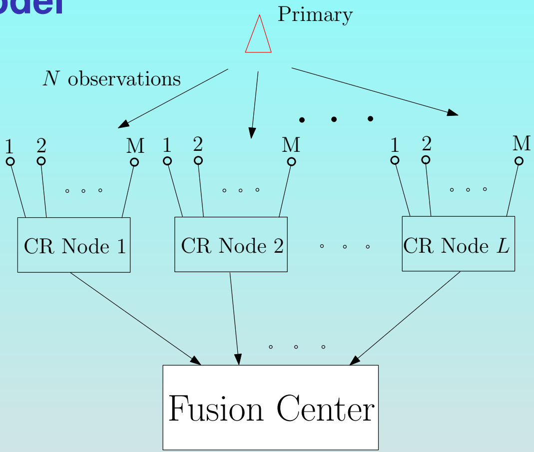
Connection to Information Theory (4/4)

- It can be shown that $H(P) \leq H(P||Q)$ (Shannon-Gibbs inequality)
- Consider $D(P||Q) \triangleq H(P||Q) - H(P)$
- Obviously, $D(P||Q) > 0$, with equality iff $P = Q$. Visualized as a distance metric. Not an actual metric, since it does not satisfy the symmetric property and the triangle inequality
- $D(P||Q)$ is called as the Kullback-Leibler divergence [KullbackLeibler1951]
- Generalizations on entropy $\longrightarrow$ Generalizations on divergences!

An Important Generalization [Csiszar1967]

- ϕ- (or f-) divergences $\phi(P||Q) = \int_{\mathcal{X}} Q(x)\phi\left(\frac{P(x)}{Q(x)}\right) d\mu(x)$, with ϕ being a real valued convex function on $(0, \infty)$.
- Special cases include the KL, variational, χ^2, Matusita, Balakrishnan and Sanghvi, Havrda and Charvat etc.
- Notation : $\delta(A, B)$ represents a stochastic divergence

System Model



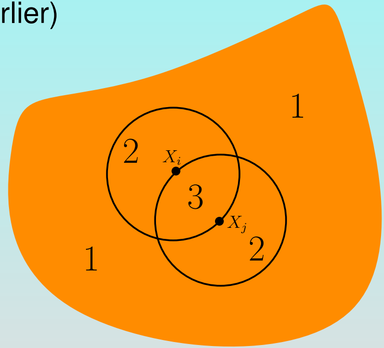
Test I - Interpoint Distances Theorem

Theorem (MaaPearlBartoszynski1996)

Let $X_1, X_2, \cdots, X_N$ be M-dimensional observations from a known distribution $\mathcal{F}$. Let Y_1 and Y_2 be the samples from the hypothesized distribution $\mathcal{G}$. Let $\delta(\cdot, \cdot)$ be an appropriately chosen stochastic distance function. Then, $\mathcal{F} = \mathcal{G}$ if and only if the distribution of $\delta(X_i, X_j)$ and $\delta(X_i, Y_1)$ (and $\delta(X_j, Y_1)$) are same and equal to $\delta(Y_1, Y_2)$, for all $1 \leq i, j \leq N$, $i \neq j$.

Test I - Intuition

- Triangle formed between any two points $X_i, X_j \sim \mathcal{F}$, and $Y \sim \mathcal{G}$. Null hypothesis is true if and only if the three sides of the triangle have the same distribution (which was specified earlier)



Test I - Construction

- Define $p_1(\cdot,\cdot)$, $p_2(\cdot,\cdot)$ and $p_3(\cdot,\cdot)$ as the probabilities associated with $Y \sim \mathcal{G}$ falling in the regions $1, 2$ and 3 respectively
- $U_k \triangleq \frac{1}{\binom{N}{2}} \sum_{i<j} p_k(X_i, X_j)$, $k = 1, 2, 3$, with $1 \leq i, j \leq N$

Lemma (BartoszynskiPearlLawrence1997)

If $Z_k \triangleq \frac{U_k - \frac{1}{3}}{\sqrt{var(U_k|H_0)}}$, *and* $\rho_{k,l} \triangleq cov(Z_k, Z_l|H_0)$ *for* $k, l \in \{1, 2, 3\}$ *and* $k \neq l$, *then*

$$Q \triangleq \frac{Z_k^2 + Z_l^2 - 2\rho_{k,l} Z_k Z_l}{1 - \rho_{k,l}^2} \sim \chi_2^2, \text{ as } N \to \infty$$

Test I - Observations and Problems

- In the present form, the probabilities $p_1(\cdot,\cdot)$, $p_2(\cdot,\cdot)$ and $p_3(\cdot,\cdot)$ have to be estimated for a particular choice of $\delta(\cdot,\cdot)$
- The asymptotic result is very tight for $N \geq 10$, for the M-antenna, N-observations SS problem
- For a range of alternate hypothesis, this test performs better than some of the sphericity tests
- Q statistic can be constructed from any Z_k, Z_l pair. "Better" combining of Z_1, Z_2 and Z_3?
- What is the optimum choice of $\delta(\cdot,\cdot)$?

Test II - Model

- Let $\mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_L$ be the matrices of observations from the L sensors
- Assumption : Physical layer fusion at the FC, without noise and fading. Idea still holds without this assumption
- If noise at each antenna is complex Gaussian with known variance, then each $\mathbf{W}_l \sim \mathcal{CW}(N, \Sigma)$.
- Statistic at the FC is $\widehat{\Sigma} \triangleq \frac{1}{L} \sum_{l=1}^{L} \mathbf{W}_l$

Test II - $< h, \phi >$ distances [Salicru et al. 1994]

- $\delta_\phi^h(X, Y) \triangleq h\left\{\int_{\mathbb{H}} \phi\left(\frac{f_X(\mathbf{Z};\theta_1)}{f_Y(\mathbf{Z};\theta_2)}\right) f_Y(\mathbf{Z};\theta_2) d\mathbf{Z}\right\}$, where
  $\mathbb{H}$ is the space of all Hermitian PD matrices,
  $h : (0, \infty) \to (0, \infty)$ is a strictly increasing function with $h(0) = 0$, and $\phi : (0, \infty) \to (0, \infty)$ is a convex function
- The differential element
  $d\mathbf{Z} = dZ_{11} dZ_{22} \cdots dZ_{MM} \prod_{i<j}^{M} dReZ_{ij} dImZ_{ij}$
- Let $d_\phi^h(X, Y) \triangleq \frac{\delta_\phi^h(X,Y) + \delta_\phi^h(Y,X)}{2}$
- Many divergence measures (including f-divergences) are special cases of δ_ϕ^h

Test II - Main Theorem

- Since the noise PDF is known, let $\widehat{\Sigma_0} \triangleq \hat{\Sigma}|\mathcal{H}_0$.

Theorem (Salicru et al. 1994)

Under some regularity conditions, if $\widehat{\Sigma_0} = \hat{\Sigma}$ *then*

$$\frac{L d_\phi^h(\hat{\Sigma}, \widehat{\Sigma_0})}{h'(0)\phi''(1)} \xrightarrow[L \to \infty]{d.} \chi^2_{M^2}$$

Test II - Observations and Problems

- Closed form to $d_\phi^h(\cdot,\cdot)$ for a suitable choice of $\delta_\phi^h(\cdot,\cdot)$
- The asymptotic result is very tight for $L \geq 10$, for the *M*-antenna, *N*-observations, *L* sensors SS problem
- For a range of alternate hypothesis, this test performs better than some of the sphericity tests
- Suitable choice of $\delta_\phi^h(\cdot,\cdot)$?

Conclusions

- Studied a family of stochastic divergences
- Test I based on interpoint distances; suitable for M-antenna CR system with N-observations
- Test II based on $< h, \phi >$ distances; suitable for CR system having L sensors with M-antenna each with N-observations
- Comparison to Sphericity tests