Thesis Notes

Ethan Ashby

9/10/2020

Contents

1 Statistical Inference: The Minimum Distance Approach **1**

1.0.1 Distances Based on Distribution Functions 1

1.1 Density-Based Distances 1

1.1.1 The Distances in Discrete Models 2

1.1.2 The Hellinger Distance 2

References **3**

1 Statistical Inference: The Minimum Distance Approach

Statistical modeling relies on the quantification of how much the data disagrees with the model. This is assessed through divergence. For example, one might want to measure the distance between a nonparametric density estimates. A good example of this is the chi-square distance of Pearson. (Ayanendranath (2011))

Most density-based divergences are not mathematical distances because they are not metrics: most are not symmetric. Sometimes the asymmetry is a desireable property. These divergences are non-negative and should equal 0 if the data match the model precisely.

1.0.1 Distances Based on Distribution Functions

Suppose $G_n(x)$ is an empirical distributino function and let $F_\theta : \theta \in \Theta \subset \mathbb{R}^p$ be a parametric family of distributions used to describe the true distribution. A general way to measure distacne between G_n and F_θ is $\rho(G_n, F_\theta)$. The weighted *Kolmorgorov-Smirnov* distance is usually given by the below formula with $\psi(u) = 1$:

$$\rho_{KS}(G_n, F_\theta) = \sup_{-\infty<z<\infty} |G_n(z) - F_\theta(z)| \sqrt{\psi(F_\theta(z))}$$

The ordinary Kolmorgorov-Smirnov distance can be used to test the null that the known distribution G_n represents the true data generating distribution. The Kolmogorov-Smirnov distance has been used in pattern recognition, image comparison and segementation, signature verification, credit scoring, and library design. The ordinary distance related to continous distributions, so modifications are needed to apply to discrete and discontinuous distributions.

The *Cramér-von Mises distance* bewteen the empirical dist and distribution function is given by:

$$\rho_{CM}(G_n, F_\theta) = \int_{-\infty}^{\infty} (G_n(z) - F_\theta(z))^2 \psi(F_\theta(z)) dF_\theta(z)$$

Where $\psi(u) = 1$ gives the usual distance.

1.1 Density-Based Distances

Focus of this book is on chi-square type distances, ϕ-divegrences, f-divergences, g-divergences, or disparities.

1.1.1 The Distances in Discrete Models

Let's say your goal is to estimate the parameter θ efficiently and robustly, by determining the element of the model family which provides the closest match to the data in terms of the distance. This is akin to quantifying the sepeartion between two vectors that sum to 1.

Definition 1 *Let C be a thrice differentiable, strictly convex function on [-1, ∞] satisfying $C(0) = 0$. Let the Pearson residual at the value X be defined by*

$$\delta(x) = \frac{d_n(x)}{f_\theta(x)} - 1$$

*. Then the **disparity** between vector $\vec{d}$ and $\vec{f_\theta}$ is:*

$$\rho_C(d_n, f_\theta) = \sum_{x=0}^{\infty} C(\delta(x)) f_\theta(x)$$

Jensen's inequality shows that the disparity is nonzero and that it only equals 0 when the two vectors are euqal.

Specific forms of the function C generate many well known disparities. For example:

1. $C(\delta) = (\delta + 1)log(\delta + 1) - \delta$ genereates the **likelihood disparity (LD)**
2. The symmetric opoosite of the likelihood disparity (swapping the vectors) is the **Kullback-Leibler divergence (KLD)**
3. The distance that corresponds to $C(\delta) = \delta - log(\delta + 1)$ is the **Hellinger distance (HD)**
4. $C(\delta) = \delta - log(\delta + 1)$ yields the **Pearson's chi-square (PCS)**
5. $C(\delta) = \frac{\delta^2}{2}$ yields the **Neyman's chi-square (NCS)**.

There are several important families of disparities. The *Cressie-Read* family of power divergences is indexed by a real tuning parameter λ where different values of lambda retyurn PCS, LD, HD, KLD, NCS respectively.

Other subfamilies include the blended weight Hellinger distance, blended weight chi-square divergence, negative exponential disparities, generalized KL divergence

Lemma 1 *Suppose that C(-1) and C'(∞) are finite. Then the disparity $\rho_C(g, f)$ is bounded above by $C(-1)+ C'(\infty)$*

1.1.2 The Hellinger Distance

Note that the actual Hellinger distance is one half square root of the hellinger disparity measure:

$$\left\{\sum (d_n^{1/2} - f_\theta^{1/2})^2\right\}^{1/2} = \left\{\frac{1}{2} HD(d_n, f_\theta)\right\}^{1/2}$$

This distance satifies the triangle inequality and the disparity measure $HD(d_n, f_\theta)$ is very popular in robust minimum distance literature. Notice that the term inside the left hand side is related to:

$$B(d_n, f_\theta) = -log\left(\sum_x d_n^{1/2} f_\theta^{1/2}\right)$$

This is the Bhattacharyya coefficient, which can be thought of as the approximate overlap between two probability densities.

Bhattacharyya's distacne may be looked at as a special case of the **Rényi divergence**:

$$RD_r(d_n, f_\theta) = \frac{1}{r(r-1)} log\left(\sum_x d_n^r(x) f_\theta^{1-r}(x)\right), r \neq 0, 1$$

When $r = 0$ we get the LD (Likelihood disparity). When $r = 1$ we get the KLD.

Stopped at page 14. Will resume tomorrow.

References

Ayanendranath, Basu. 2011. *Statistical Inference: The Minimum Distance Approach.* Boca Raton, Florida: Chapman; Hall/CRC Press.