ABSTRACT CONTRIBUTED PAPER POSTER PLUS MODE

59 ISI WORLD STATISTICS CONGRESS 25-30 AUGUST HONG KONG CHINA

UNITED STATISTICAL ALGORITHMS, LP CO-MOMENTS, COPULA DENSITY, NONPARAMETRIC MODELING

By EMANUEL PARZEN, Texas A&M University

Teaching and practice of statistics: “Las Vegas syndrome”, what happens in this chapter stays in this chapter. We propose alternative: unified algorithms that yield traditional methods (chi square, F tests, goodness of fit, orthogonal series density estimation) for (X,Y) data, continuous, discrete, binary variables X and Y. To integrate methods for continuous and discrete data, first for X construct orthonormal functions T_j(x;X) which are functions of mid-distribution Fmid(x;X)=(rank(x;X)-.5)/n. For Y construct T_k(y;Y). Second compute LP co-moments LP(0,j;X)= E[X T_j(X;X)], LP(j,0;X,Y)=E[T_j(X,X)Y], LP(j,k;X,Y)=E[T_j(X,X)T_k(Y;Y)], which include Spearman correlation, Gini correlation, Wilcoxon rank two sample. Third, LP represent VAR[X], COV[X,Y], E[Y|X], VAR[E[Y|X]], dep(x,y;X,Y)=Pr[Y=y|X=x]/Pr[Y=y], denoted copula density d(u,v) where x=Q(u;X), y=Q(v;Y). Four, compute raw L2 divergence equals SUM(|LP(j,k)|^2). Information divergence estimator LPINFOR sums over largest |LP(j,k;X,Y)|^2 selected by AIC criterion, which identifies influential scores T(x;X),T(y;Y) and thus sources of dependence and classification. Five, Smooth estimator of d(u,v) obtained by retaining in sum only largest LP co-moments. Plot copula density (conditional comparison density) d(u,v) as function of v for u=.05 (.1) .95. Real data examples are given to show how LP algorithm compares with traditional statistical methods. Many methods of density estimation (maximum entropy, logistic regression, autoregressive, kernel) can be applied to non-parametrically estimate copula density d(u,v), a hot problem.