UNITED STATISTICAL ALGORITHMS, LP COMOMENTS, COPULA DENSITY, NONPARAMETRIC MODELING

Emanuel Parzen[1] and Subhadeep Mukhopadhyay[2]
[1]Texas A&M University, College Station, TX, USA
[2]Temple University, Philadelphia, PA and Texas A&M University, College Station, TX, USA

ABSTRACT

In addition to solving problems "retail" (one at a time for one client/collaborator), academic statistics should aim to solve problems "wholesale"(algorithms and computer code that can be applied for many clients). We call this approach to teaching and practice SMART COMPUTATIONAL STATISTICS = united data science algorithms providing methods for Small Data and Big Data. It practices that the goal of computing is insight (from graphical presentations) rather than numbers (especially accurate numbers which can be suboptimal (right answer to wrong question)). Our extensive theory simultaneously extends and integrates traditional (classical) statistical methods, treats multivariate discrete and continuous variables, parametric and nonparametric modeling, classification and measuring dependence relationships of (X, Y) where X can be a high dimensional vector of features and Y a scalar variable to predict or classify. The goal of this paper is to gently introduce many basic concepts: quantile; mid-distribution; mid-quantile; comparison density; comparison distribution; skew-G distribution; LP coefficients, moments, co-moments; orthonormal score functions series density estimation; custom construction of mid-distribution based score functions; comparison probability version of Bayes theorem; copula density function series nonparametric estimation. The practice of these algorithms is illustrated with examples of real data.

Keywords: Bayes theorem proof, copula densitiy, comparison density, information, LPINFOR, LP moment, LP Comoment, LP orthogonormal score function, nonparametric data modeling, quantile.

1 Qunatiles, Mean, Variance, Median, Quartiles

Probability law of random variable X is described by distribution function $F(x)$ or quantile function $Q(u), 0 < u < 1$. Denote by U Uniform$(0, 1)$. *Theorems*: In distribution $X = Q(U)$. With probability 1, $Q(F(X)) = X, \mathbb{E}[Y|X] = \mathbb{E}[Y|F(X)]$ When X continuous, $F(X) = U$ in distribution; $F(Q(u)) = u, 0 < u < 1$; $f[Q(u)]Q'(u) = 1$.
Mean $\mathbb{E}[X] = \int_0^1 Q(u)du$; sort before adding for insight about mean of sample. Normalize X by transforming to $\mathcal{Z}(X) = (X - \mathbb{E}[X])/\sigma(X)$, $\mathrm{QI}(X) = (X - \mathrm{MQ})/\mathrm{DQ}$, where measures of location mid-quartile MQ and scale DQ are defined in terms of mid-quantile $Q^{\mathrm{mid}}(u)$, defined below, by $\mathrm{MQ} = .5(\mathrm{Q1} + \mathrm{Q3})$, $\mathrm{DQ} = 2(\mathrm{Q3} - \mathrm{Q1})$, quartiles $Q1 = Q^{\mathrm{mid}}(.25), Q3 = Q^{\mathrm{mid}}(.75)$, median $Q2 = Q^{\mathrm{mid}}(.5)$.
Definitions: When F is continuous, define $Q^{\mathrm{mid}}(u) = Q(u)$. When F is discrete (true for sample distribution) define mid-distribution $F^{\mathrm{mid}}(x) = F(x) - .5p(x)$, probability mass function $p(x) = Pr[X = x]$. *Theorem*: $\mathbb{E}[F^{\mathrm{mid}}(X)] = .5, \mathrm{Var}[F^{\mathrm{mid}}(X)] = (1/12)(1 - \sum_x p^3(x))$. Compute F^{mid} from sample of size n : $F^{\mathrm{mid}}(x) = (\mathrm{Midrank(x)} - .5)/\mathrm{n}$. Call x probable value if $p(x) > 0$. Let $x_1 < \cdots < x_r$ denote probable values of X. Define and plot $Q^{\mathrm{mid}}(u)$: connect linearly $(0, x_1), (F^{\mathrm{mid}}(x_j), x_j), (1, x_r)$. Probability laws to fit (model) data are identified from comparing plots of $QI(Q^{\mathrm{mid}}(u))$ and $\mathrm{QI}(Q(u))$ for famous distributions $Q(u)$. A value X is an outlier if $|\,\mathrm{QI}(X)| > 1$. Symmetry and tails are identified from a five number summary of a data set: MQ, DQ, QI(Q2), QI(MIN), QI(MAX). Novel asymptotic distribution theory of sample mid-quantiles

is given by Ma et al. (2011). United statistical analysis began with Parzen (1979, 1983), Eubank et al. (1987), Kallenberg and Ledwina (1999), Rayner et al. (2009). .

2 Comparison Distribution, Comparison Density, X Continuous

A parametric continous model G for the true unknown continuous distribution F of X has parameters estimated so that G is as close as possible to F by the criterion that $W = G(X)$ is close in distribution to uniform U. The distrubution function of W is called a comparison distribution, denoted $D(u) = D(u; G, F) = FG^{-1}(u)$, with derivative called *comparison density*

$$d(u) = d(u; G, F) = \frac{fG^{-1}(u)}{gG^{-1}(u)} \tag{2.1}$$

Measure distance between G and F by information distances of $d(u)$ from 1 (Kullback-Liebler, Renyi). Related concepts are relative density and grade density. A constraint on choice of parametric G is that comparison density $d(u)$ is bounded. An estimator $\widehat{d}(u)$ provides estimator $\widehat{f}(x)$ by

$$\widehat{f}(x) \;=\; g(x)\,\widehat{d}\big(G(x)\big). \tag{2.2}$$

Definition: Define $f(x) = g(x)d(G(x))$ to be skew-G distribution; it can be simulated by accept-reject sampling from G.
Modeling Principle for Practical Research: Model F by skew-G whose parametric G and comparison density $d(u)$ minimize estimate of $\int [d(u) - 1]^2 \, \mathrm{d}u$. If integral not significantly different from zero accept G as true distribution of F.
Density estimation has many methods and an enormous literature. We emphasize Neyman estimators pioneered by Neyman (1937); they are orthonormal series representations in terms of custom built score functions. When F is continuous: score functions are orthonormal Legendre polynomials $\mathrm{Leg}_j(u), 0 < u < 1$. Initial (not guaranteed non-negative) estimators, called $L2$, have form

$$\widehat{d}(u) - 1 \;=\; \sum_j C_j \,\mathrm{Leg}_j(u). \tag{2.3}$$

MaxEnt (or exponential model) estimators have form

$$\log \widehat{d}(u) = \sum_j \theta_j \,\mathrm{Leg}_j(u) \tag{2.4}$$

Estimating equations for both methods can be expressed in terms of LP coefficients, defined as mean with respect to sample distribution $\tilde{F}$, of score function $T_j(X; G) = \mathrm{Leg}_j(G(X))$. Definition: For X continuous, define

$$\mathrm{LP}(j; X, G) \;=\; \mathbb{E}\big[\,\mathrm{Leg}_j(G(X) \mid \tilde{F}\big]. \tag{2.5}$$

Estimating equations for parameters of orthonormal series density estimators are

$$\mathbb{E}\big[\,\mathrm{Leg}(G(X)) \mid \text{estimated}\, \widehat{\mathrm{d}}(\mathrm{u})\big] \;=\; \mathrm{LP}(\mathrm{j}; \mathrm{X}, \mathrm{G}) \tag{2.6}$$

Which Comes First, Parameters or Score Functions (Sufficient Statistics)? We solve the model selection problem by identifying most significant score functions $\mathrm{Leg}_j(u)$ whose LP coefficients $\mathrm{LP}(j; X, G)$ are significantly different from zero. *Theorem*: Under the null hypothesis that $G = F$, the true distribution, $\mathrm{LP}(j; X, G)$ are zero mean, unit variance, uncorrelated asymptotically normal.

3 Comparison Distribution, Density, X Discrete, Chi-Squared

When observe sample of size n of discrete X with probable values x_j, one compares sample probabilities $\tilde{p}(x)$ with population model $p_0(x)$ by Chi-Squared statistic CHISQ which we represent CHISQ $= n$ CHIDIV,

$$\text{CHIDIV} = \sum_x (\tilde{p}(x) - p_0(x))^2/p_0(x) \;=\; \sum_x p_0(x)\left[\tilde{p}(x)/p_0(x) - 1\right]^2 \;=\; \int_0^1 \left[d(u;p_0;\tilde{p}) - 1\right]^2 \,\mathrm{d}u. \quad (3.1)$$

defining *comparison density*

$$d(u) \;=\; d(u;p_0,\tilde{p}) \;=\; \tilde{p}(Q_0(u))/p_0(Q_0(u)), \quad (3.2)$$

comparison distribution $D(u) = D(u;p_0,\tilde{p}) = \int_0^u d(t)\,\mathrm{d}t$. *Theorem*: $D(u) = \tilde{F}(Q_0(u))$ at "exact" $u = F_0(x_j)$ for some probable x_j. Plot of $D(u)$, called PP plot, connects linearly $(0,0), (F_0(x_j), \tilde{F}(x_j))$. A diagnostic of null hypothesis H_0 that p_0 is true probability density is area of region between $D(u)$ and u; one can show

$$\text{AREA} \;=\; \sum_{\mathrm{x}} \tilde{\mathrm{p}}(\mathrm{x})(\mathrm{F}_0^{\mathrm{mid}}(\mathrm{x}) - .5)). \quad (3.3)$$

Using formula $\mathrm{Var}_0(F^{\mathrm{mid}}(X)) = (1/12)(1 - \sum_x p_0^3(x))$, normalize statistic AREA to have mean 0, variance 1; we express $\mathcal{Z}(\text{AREA})$ in terms of $T_1(X;p_0) = \mathcal{Z}(F_0^{\mathrm{mid}}(X))$

$$\text{LP}(1;p_0,\tilde{p}) \;=\; \mathbb{E}\big[\mathcal{Z}(F_0^{\mathrm{mid}}(X) \mid \tilde{F}\big] \;=\; \mathbb{E}[T_1(X;p_0) \mid \tilde{F}]. \quad (3.4)$$

Definition: For X discrete define higher order LP coefficients

$$\text{LP}(j;p_0,X) \;=\; \mathbb{E}[T_j(X;p_0) \mid \tilde{p}]. \quad (3.5)$$

In terms of higher order orthonormal score functions $T_j(X;p_0)$, custom constructed by Gram Schmidt orthonormalization of powers of $T_1(X;p_0) = \mathcal{Z}(F_0^{\mathrm{mid}}(X))$. When $p_0(x_j) = 1/r$, the higher order score functions are discrete Legendre polynomials. A comparison density estimator by an orthonormal series representation of $d(u)$ is obtained by representing LP coefficients as Fourier-type coefficients of comparison density:

$$\text{LP}(j;X,p_0) \;=\; \int_0^1 S_j(u;p_0)d(u)\,\mathrm{d}u, \quad (3.6)$$

defining $S_j(u;p_0) = T_j(Q_0(u);p_0)$, which are orthonormal functions on $0 < u < 1$. Further we have orthogonal components decomposition of Chi Square pre-statistic CHIDIV. *Theorem*: CHIDIV $= \sum_j |\,\text{LP}(j;X,p_0)|^2$ also called raw-LPINFOR. Define $\widehat{\text{CHIDIV}}$ or $\widehat{\text{LPINFOR}}$ to be the sum of squares of LP coefficients only over LP coefficients not significantly different from 0. This equals norm squared of $\widehat{d}(u) - 1$. As a test statistic for H_0 (goodness of fit of model p_0 to data) $\widehat{\text{CHIDIV}}$ approximately numerically equals raw-CHIDIV but has smaller degrees of freedom equal to number of non-zero LP coefficients. Therefore it is less likely for the null model p_0 to be accepted using the Smooth-Chi Square test.

4 LP Moments of Univariate X

For X continuous quantile $Q(u)$ has orthonormal series representation

$$Q(u) \;=\; \sum_j \mathrm{LP}(j;X)\,\mathrm{Leg}_j(u), \tag{4.1}$$

defining LP moments: $\mathrm{LP}(j;X) \;=\; \mathbb{E}\big[Q(U)\,\mathrm{Leg}_j(U)\big]$, U Uniform$(0,1)$.

$$\mathrm{Theorem:}\;\; \mathrm{LP(j;X)} \;=\; \mathbb{E}\big[\mathrm{X\,Leg_j(F(X)}\big],\;\; \mathrm{LP(j;}\mathcal{Z}\mathrm{(X))} \;=\; \mathbb{E}[\mathcal{Z}\mathrm{(X)\,Leg_j(F(X))}]. \tag{4.2}$$

To define LP moments for X discrete define $T_0(X;X) = 1, T_1(X;X) = \mathcal{Z}(F^{\mathrm{mid}}(X;X))]$, $\mathrm{LP}(j;X) = \mathbb{E}[XT_j(X;X)]$, also denoted $\mathrm{LP}(j,0;X,X)$.
Construction: $T_j(X;X)$ are orthonormal constructed by Gram Schmidt orthonormalization of powers of $T_1(X;X)$. *Definition*: Score functions $S_j(u;X) = T_j(Q(u;X);X), 0 < u < 1$; their plots are similar to piecewise constant versions of orthonormal Legendre polynomials on $0 < u < 1$. Our definition of LP moments when X is continuous is an orthonormal version of L moments as defined by Hosking (1990). As estimators of L moments we recommend the LP moments of the discrete sample distribution of X. The diagnostic powers of LP moments are illustrated (see Table on web) by their values for standard distributions, and for Tukey-λ distributions:

$$Q(u;\mathrm{sym}) \;=\; (\mathrm{u}^{\lambda} - (1-\mathrm{u})^{\lambda})/\lambda,\;\; \mathrm{Q(u;asym)} = -((1-\mathrm{u})^{\lambda} - 1)/\lambda \tag{4.3}$$

Variance of X (finite second moment) has representation: *Theorem*: $1 \;=\; \sum_j |\,\mathrm{LP}(j;\mathcal{Z}(X))|^2$. We identify X short or medium tail, possibly normal, if $|\,\mathrm{LP}(1;\mathcal{Z}(X))|^2 > .95$.
Two Sample Equality: Research "quick" test for equality of distributions: compute LP moments of pooled sample. A complete analysis computes comparison density(pooled sample, sample 1) denoted d(u; X pooled sample, X sample 1). Wilicoxon linear rank statistic is equivalent to (with asymptotic $\mathcal{N}(0,1/n)$ under null hypothesis)

$$\mathbb{E}\big[\mathcal{Z}(F^{\mathrm{mid}}(X;X\,\mathrm{pooled}))\mid\;\mathrm{sample\,1\,X}]\sqrt{\mathrm{odds(Pr[X\,in\,sample\,1])}} \tag{4.4}$$

5 (X,Y) Copula Density, Discrete or Continuous Variables

For X,Y both discrete joint probability is described by joint probability mass function $p(x,y;X,Y)$. For X,Y both continuous joint probability law described by joint probability density $f(x,y:X,Y)$. For Y discrete and X continuous joint probability is described by product of marginal of one and conditional of other. That the product is the same whether we condition on X or condition on Y follows by proving integral up to x' of both sides of equation equals $\Pr[X \le x' \mid Y = y]$)

$$\mathit{Theorem}\,(\mathrm{PRE-BAYES}):\;\; \mathrm{Pr[Y=y|X=x]f(x;X)} \;=\; \mathrm{f(x;X|Y=y)\,Pr[Y=y]} \tag{5.1}$$

Comparison Probability: for Y discrete

$$\mathrm{ComPr}[Y=y|X=x] \;=\; \Pr[Y=y|X=x]/\Pr[Y=y] \tag{5.2}$$

For Y continuous

$$\mathrm{ComPr}[Y=y|X=x] \;=\; f(y;Y|X=x)/f(y;Y) \tag{5.3}$$

THEOREM (Bayes Theorem): $\mathrm{ComPr}[Y = y|X = x] = \mathrm{ComPr}[X = x|Y = y]$. *Definitions*: EQUAL concepts conditional comparison density, copula density of (X, Y)

$$
\begin{aligned}
d(u; X, X|Y = Q(v; Y)) &= \mathrm{ComPr}[X = Q(u; X)|Y = Q(v; Y)] \\
d(v; Y, Y|X = Q(u; X)) &= \mathrm{ComPr}[Y = Q(v; Y)|X = Q(u; X)]
\end{aligned}
$$

Relation between conditional comparison density and copula density,

$$
\mathrm{cop}(u, v; X, Y) = d(u; X, X|Y = Q(v; Y)) = d(v; Y, Y|X = Q(u; X))) \tag{5.4}
$$

Theorem: For (X continuous, Y continuous) $\mathrm{cop}(u, v; X, Y) = \mathrm{dep}(Q(u; X), Q(v; Y); X, Y)$, where Dependence function $\mathrm{dep}(x, y; X, Y) = f(x, y; X, Y)/f(x; X)f(y; Y)$. Similarly define $\mathrm{dep}(x, y; X, Y)$ for X,Y both discrete. Note fundamental formulas

$$
\begin{aligned}
\mathrm{Corr}[I(X = x), I(Y = y)] &= \big[\,\mathrm{dep}(x, y; X, Y) - 1\big]\sqrt{\mathrm{odds}(\Pr[\mathrm{X = x}])\mathrm{odds}(\Pr[\mathrm{Y = y}])} \\
\mathrm{Corr}[T(X), I(Y = y)] &= \mathbb{E}\big[\mathcal{Z}(T(X))|Y = y\big]\sqrt{\mathrm{odds}(\Pr[\mathrm{Y = y}])}.
\end{aligned} \tag{5.5}
$$

6 LP-Comoment, LPINFOR

A contingency table test of independence is traditionally $\mathrm{CHISQ} = n\,\mathrm{CHIDIV}$ which we express in terms of dependence function, and also in terms of LP co-moments (compare Serfling and Xiao (2007)) and fundamental concept LPINFOR.
Definition: LP COMOMENTS are defined as $\mathrm{LP}(j, k; X, Y) = \mathbb{E}\big[T_j(X; X)T_k(Y; Y)\big]$.

$$
\begin{aligned}
\mathrm{CHIDIV} &= \sum_x \tilde{p}(x; X)\tilde{p}(y; Y)(\tilde{\mathrm{dep}}(x, y; X, Y) - 1)^2 \\
&= \iint_{[0,1]^2} \mathrm{d}u\ \mathrm{d}v\big[\,\mathrm{cop}(u, v; X, Y) - 1\big]^2 = \mathrm{LPINFOR}(X, Y) \\
&= \sum \big|\,\mathrm{LP}(j, k; X, Y)\big|^2
\end{aligned} \tag{6.1}
$$

Define smooth information $\widehat{\mathrm{LPINFOR}}(X, Y)$ to be sum over LP comoments that are significantly different from zero. *Theorem*: Under null hypoithesis of independence (zero population LP comoment) sample LP comoments have mean 0, variance 1, uncorrelated, asymptotically normal. *Nonparametric estimator* of population copula density:

$$
\widehat{\mathrm{cop}}(u, v; X, Y) = \sum \mathrm{LP}(j, k; X, Y)S_j(u; X)S_k(v; Y) \tag{6.2}
$$

We find interpretation easier if we plot slices $\mathrm{cop}(u, v; X, Y), 0 < v < 1$, for selected values of u.

7 LP Comoments Zero Order, Nonparametric Regression

Pearson correlation (Linear) regression of Y on X plots line $\mathcal{Z}(Y) = \mathrm{Corr}(X, Y)\mathcal{Z}(X)$. Spearman correlation regression of $F^{\mathrm{mid}}(Y; Y)$ on $F^{\mathrm{mid}}(X; X)$ plots line $\mathcal{Z}(F^{\mathrm{mid}}(Y; Y)) = \mathrm{LP}(1, 1; X, Y)\mathcal{Z}(F^{\mathrm{mid}}(X; X))$. IMPORTANT FOR PRACTICE! Our definition of $\mathrm{LP}(1, 1; X, Y)$ successfully defines Spearman correlation for tied data which is asymptotically $\mathcal{N}(0, 1/n)$ under null hypothesis. Nonparametric regression conditional expectation $\mathbb{E}[Y|X]$ is a function of X satisfying for all suitable $h(X) : \mathbb{E}\big[\mathbb{E}[Y|X]h(X)\big] = \mathbb{E}[Y h(X)]$. We can represent $\mathbb{E}[Y|X] - \mathbb{E}[Y] = \sum_j C_jT_j(X; X)$, a linear combination of score functions of X with coefficients $C_j = \mathbb{E}[\mathbb{E}[Y|X]T_j(X;X)] = \mathbb{E}[YT_j(X;X)] = \mathrm{LP}(j,0;X,Y)$. *Definition* LP comoment zero order:

$$\mathrm{LP}(j,0;X,Y) = \mathbb{E}[T_j(X;X)Y], \text{ Similarly } \mathrm{LP(0,k;X,Y)} = \mathbb{E}[\mathrm{XT_k(Y;Y)}]. \tag{7.1}$$

Practice uses $\mathrm{LP}(j,0;X,\mathcal{Z}(Y)), \mathrm{LP}(0,k;\mathcal{Z}(X),Y)$. GINI correlation have two versions equivalent to $\mathrm{LP}(1,0;X,Y)$ and $\mathrm{LP}(0,1;X,Y)$. Spearman Correlation is $\mathrm{LP}(1,1;X,Y)$ Define second order Spearman correlation of Y on X to be $\mathrm{LP}(2,1;X,Y)$. A second order (nonlinear) regression of Y on X on the scatter diagram $(T_1(X;X), T_1(Y;Y))$ is defined to be

$$T_1(Y;Y) \;=\; \mathrm{LP}(1,1;X,Y)T_1(X;X) + \mathrm{LP}(2,1;X,Y)T_2(X;X) \tag{7.2}$$

Diagnostics of dependence of (X,Y) are plots of this curve, and calculate multiple correlation coefficients: $\mathrm{R2second(Y|X)} = \left|\,\mathrm{LP(1,1;X,Y)}\right|^2 + \left|\,\mathrm{LP(2,1;X,Y)}\right|^2$, $\mathrm{R2spearman(X,Y)} = \left|\,\mathrm{LP(1,1;X,Y)}\right|^2$. $\mathrm{R2pearson(X,Y)} = \left|\,\mathrm{Corr[X,Y]}\right|^2$. Examples are the best way to interpret this high dimension data analysis strategy (algorithm).

References

Eubank, R. L., LaRiccia, V. N. and Rosenstein, R. B. (1987). Test statistics derived as components of pearson's phi-squared distance measure. *Journal of the American Statistical Association*, **82** 816–825.

Hosking, J. R. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 105–124.

Kallenberg, W. C. (2009). Estimating copula density using model selection techniques. *Insurance: Mathematics and Economics*, **45** 209–223.

Kallenberg, W. C. and Ledwina, T. (1999). Data-driven rank tests for independence. *Journal of the American Statistical Association*, **94** 285–301.

Ma, Y., Genton, M. G. and Parzen, E. (2011). Asymptotic properties of sample quantiles of discrete distributions. *Annals of the Institute of Statistical Mathematics*, **63** 227–243.

Neyman, J. (1937). Smooth tests for goodness of fit. *Skand. Aktuar.*, **20** 150–199.

Parzen, E. (1979). Nonparametric statistical data modeling (with discussion). *Journal of the American Statistical Association*, **74** 105–131.

Parzen, E. (1983). Fun.stat quantile approach to two sample statistical data analysis. *Technical Report, Texas A&M University.*

Parzen, E. (1991). Goodness of fit tests and entropy. *Journal of Combinatorics, Infor- mation, and System Science*, **16** 129–136.

Parzen, E. (1992). Time series, statistics, and information. *New Directions in Time Series Analysis (ed. E. Parzen et al). Springer Verlag: New York.* 65–286.

Parzen, E. and Mukhopadhyay, S. (2012). Modeling, dependence, classification, united statistical science, many cultures. `arXiv:1204.4699`.

Provost, S. and Jiang, M. (2012). Orthogonal polynomial density estimates: alternative representations and degree selection. *International Journal of Computational and Mathematical Sciences*, **6** 12–29.

Rayner, J., Thas, O. and Best, D. (2009). *Smooth tests of goodness of fit: using R.* Wiley: Singapore.

Serfling, R. and Xiao, P. (2007). A contribution to multivariate L-moments: L-comoment matrices. *Journal of Multivariate Analysis*, **98** 1765–1781.