# RICH TRANSFORMATION MODELS AND MORE

KJELL DOKSUM

Department of Statistics
University of Wisconsin-Madison

May, 2008

# SUMMARY

POWER TRANSFORMATIONS TO NORMALITY IN REGRESSION.

TRANSFORMATION PARAMETERS KNOWN.

TRANSFORMATION PARAMETERS UNKNOWN.

CONTROVERSY. WHAT CONTROVERSY?

BRILLINGER-STOKER AVERAGE DERIVATIVE COEFFICIENTS.

TOPICS:

1. REGRESSION TRANSFORMATION MODELS.
2. YEO-JOHNSON (2000 BIOMETRIKA).
3. HERNANDEZ-JOHNSON (1980 JASA).
4. WONG TRANSFORMATIONS(1983).
5. KLAASSEN-WELLNER COPULA (1997).

$(\mathbf{X}_1, Y_1), \cdots, (\mathbf{X}_n, Y_n)$ i.i.d $\sim (\mathbf{X}, Y)$,
$\mathbf{X} \in R^d, Y \in R$.
$p(\mathbf{x}, y) = f(\mathbf{x})p(y|\mathbf{x})$.
RANDOM $\mathbf{X}$ OR FIXED $\mathbf{x}$, "*SAME*" RESULT.

TRANSFORMATIONS:
$y^{(\lambda_0)}, (x_1^{(\lambda_1)}, \cdots, x_d^{(\lambda_d)}) = \mathbf{x}^{(\boldsymbol{\lambda})}$.
e.g.

$$
\begin{aligned}
y^{(\lambda_0)} = y_{BC}^{(\lambda_0)}(BoxCox) \; &= (y^{\lambda_0} - 1)/\lambda_0, y > 0, \lambda_0 \; \neq 0. \\
&= \log y, \qquad\qquad\qquad \lambda_0 \; = 0.
\end{aligned}
$$

TRANSFORMATION MODEL:
$Y^{(\lambda_0)} = \alpha + \beta^T \mathbf{X}^{(\boldsymbol{\lambda})} + \epsilon,\ \epsilon \sim normal(0, \sigma^2)$.

BOX-COX: MUST HAVE $y > 0$.
TROUBLE: WHEN $\epsilon \sim N(0, \sigma^2)$, $y \in R$.
SIGNED POWER: $\{sgn(y)|y|^\lambda - 1\}/\lambda$.
TROUBLE:CHANGES FROM CONVEX TO CONCAVE IN y, $y \in R$. FUNNY LOOKING.

YEO-JOHNSON (2000 BIOMETRIKA)

$$y_{YJ}^{(\lambda)} = \begin{cases} \{(y+1)^\lambda - 1\}/\lambda, & y \geq 0, \lambda \neq 0. \\ log(1+y), & y \geq 0, \lambda = 0. \\ -\{(-y+1)^{2-\lambda} - 1\}/(2-\lambda), & y < 0, \lambda \neq 2. \\ -log(-y+1), & y < 0, \lambda = 2. \end{cases}$$

PROPERTIES OF YJ:

(1) CONVEX IN y FOR $\lambda > 1$.

(2) CONCAVE IN y FOR $\lambda < 1$.

(3) $y_{YJ}^{(\lambda)}$ AND ITS DERIVATIVES $(\partial^k/\partial\lambda^k)y_{YJ}^{(\lambda)}$ ARE CONTINUOUS IN $(\lambda, y)$.

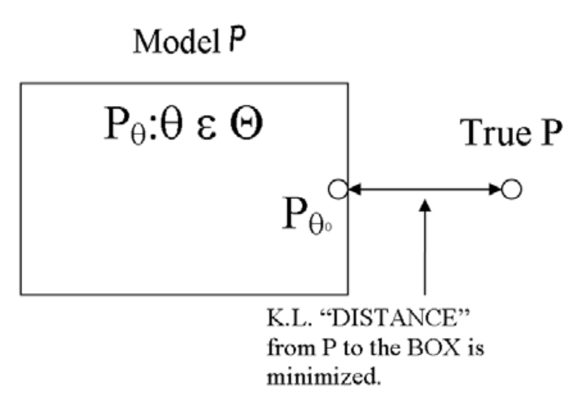
(4) GOOD LOOKING.

HERNANDEZ JOHNSON: TRANSFORMATIONS
JASA (1980).
X=OBSERVATIONS.
P=TRUE DISTRIBUTION OF **X**.
$\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ WORKING MODEL.

BOX AXIOMS:

(I) $\mathcal{P}$ IS NOT TRUE.
(II) $\mathcal{P}$ IS USEFUL.

# THINKING INSIDE/OUTSIDE THE BOX:




DOKSUM, OZEKI, KIM, NETO. RICH VOLUME 2007.
HUBER (1967) SANDWICH FORMULA: $VAR_P(\hat{\theta})$, WHERE $\hat{\theta} = MLE$ FOR $\mathcal{P}$.

DISCUSSION: WHAT IF $\mathcal{P}=$ CUBIC SPLINES WITH 5 KNOTS AND P=CUBIC SPLINE WITH 10 KNOTS?
THE MLE IS INCONSISTENT? NOT?

AWARD WINNING GENOMIC (GWAS) RESEARCHERS PHILOSOPHY:

WE PREFER SIMPLE, PROVEN, CLASSICAL MODELS. IF WE FIND A SIGNIFICANT GENETIC MARKER EFFECT USING A SIMPLE MODEL, THE SIGNIFICANCE ALSO HOLDS FOR THE TRUE MODEL PROVIDED THAT WE HAVE INCLUDED ALL CONFOUNDING VARIABLES. LOGISTIC REGRESSION. VALIDATION.

HERNANDEZ JOHNSON (1980) MULTIVARIATE TRANSFORMATIONS: $\theta = (\boldsymbol{\lambda}, \boldsymbol{\mu}, \Sigma)$

$(X_1^{(\lambda_1)}, \cdots, X_p^{(\lambda_p)}) \sim \textit{normal}(\boldsymbol{\mu}, \Sigma)$. FIX $\boldsymbol{\lambda}$,

$(\boldsymbol{\mu}(\boldsymbol{\lambda}), \Sigma(\boldsymbol{\lambda})) = \arg\min\{d_{KL}(P, P_\theta) : \theta \in \Theta\}$.

THEN SET $\theta(\boldsymbol{\lambda}) = (\boldsymbol{\lambda}, \boldsymbol{\mu}(\boldsymbol{\lambda}), \Sigma(\boldsymbol{\lambda}))$.

$\boldsymbol{\lambda_0} = \arg\min\{d_{KL}(P, P_{\theta(\boldsymbol{\lambda})}) : \boldsymbol{\lambda} \in R^p\}$.

$\boldsymbol{\mu}_0, \Sigma_0 = (\boldsymbol{\mu}(\boldsymbol{\lambda_0}), \Sigma(\boldsymbol{\lambda_0}))$.

$P_{\theta_0}$ IS A WORKING MODEL WHERE $\theta_0 = (\boldsymbol{\lambda}_0, \boldsymbol{\mu}_0, \Sigma_0)$ .

$P_{\theta_0}$ IS NOT TRUE.

$P_{\theta_0}$ IS USEFUL.

MLE $\hat{\theta}$ of $\theta$: REPLACE $d_{KL}$ WITH ($-$*LOG LIKELIHOOD*).

REGRESSION $(Y^{(\lambda_0)}, \mathbf{X}^{(\boldsymbol{\lambda})}) \sim \textit{normal}(\boldsymbol{\mu}, \Sigma)$.
REGRESS $Y^{(\lambda_0)}$ ON $\mathbf{X}^{(\boldsymbol{\lambda})}$.
ALL THE USUAL REGRESSION FORMULAS
HOLD WITH $\mathbf{X}$, $Y$ REPLACED BY $\mathbf{X}^{(\hat{\boldsymbol{\lambda}})}$, $Y^{(\hat{\lambda}_0)}$.
$VAR_P(\hat{\theta})=$ SANDWICH FORMULA FOR $P \notin \mathcal{P}$.

CONTROVERSY: IF $\lambda$ IS UNKNOWN, $\beta$ IS A COEFFICIENT VECTOR ON AN UNKNOWN "*SCALE*" $= Y^{(\lambda)}$:

$$Y^{(\lambda)} = \alpha + \Sigma\beta_j X_j + \epsilon,$$

BRILLINGER-STOKER SOLUTION:
ADC= AVERAGE DERIVATIVE COEFFICIENT: NOTE THAT

$$E(Y|\mathbf{X}) = g(\alpha + \Sigma\beta_j X_j),$$

g()=UNKNOWN, e.g. g() COULD DEPEND ON $\lambda$.

$$E(\frac{\partial}{\partial X_j} E(Y|\mathbf{X})) = c\beta_j,\ \ WHERE\ c = E(g'(\alpha + \Sigma\beta_j X_j)).$$

DEF: $\beta_j^{ADC} = (c\beta_j)/\sqrt{\Sigma(c\beta_j)^2} = \beta_j/|\boldsymbol{\beta}|$.
$\beta_j^{ADC}$ IS AN IDENTIFIABLE PARAMETER WHEN $g()$ ($OR$ $\lambda$) IS UNKNOWN.

WHEN ARE THE $\lambda$ KNOWN AND $\lambda$ UNKNOWN CASES THE SAME?

DOKSUM AND CHI-WING WONG (1983) JASA.
CONSIDER CONTIGUOUS MODELS WHERE
$\beta_j/\sigma = \beta_{nj}/\sigma_n \rightarrow 0$ as $n \rightarrow \infty$.

SUPPOSE $Y_i^{(\lambda)} = \beta_0 + \sum_{j=1}^{d} \beta_j x_{ij} + \epsilon_i$,

MORE PRECISELY, SET $\theta_{ni} = \sum_{j=1}^{d} \beta_{nj} x_{ij}/\sigma_n$, $Var(\epsilon_i) = \sigma_n^2$,

$\Omega_n = \{\boldsymbol{\beta}_n : |\boldsymbol{\theta}_n| \leq K, max|\theta_{ni}| \rightarrow 0\}(BICKEL-DOKSUM 1981)^c$

THEN THE ASYMPTOTICS FOR THE $\lambda$ KNOWN AND UNKNOWN CASE COINCIDE FOR ESTIMATION AND TESTING OF $\beta_1, \cdots, \beta_d$. NOT FOR $\beta_0$.

CHI-WING WONG (1983) THESIS.

$$\begin{aligned} Y_i &= \beta_0 + \sum_{j=1}^{d} \beta_j x_{ij}^{(\lambda_j)} + \epsilon_i \\ &= \alpha_0 + \sum_{j=1}^{d} \alpha_j z_{ij}^{(\lambda_j)} + \epsilon_i \end{aligned}$$

WHERE $z_{ij}^{(\lambda_j)} = [x_{ij}^{(\lambda_j)} - \bar{x}_j^{(\lambda_j)}]/SE(x_j^{(\lambda_j)})$. THE MODEL HAS BEEN REPARAMETERIZED BY STANDARDIZING THE TRANSFORMED PREDICTORS.
RESULT: IF d=1, THE $\lambda$, KNOWN AND UNKNOWN CASES ARE ASYMPTOTICALLY EQUIVALENT.

HERE $\alpha_0, \alpha_1$ AND $\lambda_1$ ARE ORTHOGONAL PARAMETERS IN THE SENSE OF COX AND REID, 1987 JRSS B.

IF $d > 1$, TRANSFORM $x_{ij}^{(\lambda_j)}$ TO HAVE IDENTITY SAMPLE COVARIANCE MATRIX. THE ASYMPTOTIC EQUIVALENCE OF THE $\lambda$ KNOWN, UNKNOWN CASES STILL HOLDS.

HERE THE ESTIMATES OF $\boldsymbol{\lambda}$ ARE THE SAME IN THE STANDARDIZED AND UNSTANDARDIZED CASE. FOR ESTIMATING THE $\beta$'s, THE $\boldsymbol{\lambda}$ KNOWN AND UNKNOWN CASES ARE ASYMPTOTICALLY DIFFERENT EVEN THOUGH THE $\beta$'s ARE SIMPLE TRANSFORMS OF THE $\alpha$'s.

THE BIVARIATE NORMAL COPULA.
$(X, Y) \sim H(x, y)$, CONTINUOUS, $X \sim F$, $Y \sim G$,
$Z \equiv \Phi^{-1}(F(X))$, $W \equiv \Phi^{-1}(G(Y))$.
DEF:

$$\mathcal{P}_C \equiv \{H \ : (\Phi^{-1}(F(X)), \Phi^{-1}(G(Y))) \sim N(0, 0, 1, 1, \rho)$$

NOTE: $h_0(y) = \Phi^{-1}(G(y))$ AND
$h_1(x) = (\Phi^{-1}(F(x))$ ARE NP INCREASING
TRANSFORMATIONS.
IF $F(), G()$ KNOWN, $\hat{\rho} = n^{-1}\Sigma Z_i W_i$ IS THE MOM
ESTIMATE.

IF $F(), G()$ UNKNOWN, REPLACE $Z_{(i)}$ WITH $E(Z_{(i)})$, $W_{(i)}$ WITH $E(W_{(i)})$. CALLED NORMAL SCORES.
FISHER-YATES 1938. GET $\hat{\rho}_{NS}$

KLAASSEN AND WELLNER (1997).
THEOREM: FOR $H \in \mathcal{P}_C$,

(a) $\sqrt{n}(\hat{\rho}_{NS} - \rho) \to N(0, (1 - \rho^2)^2)$.
(b) $\hat{\rho}_{NS}$ IS SEMIPARAMETRICALLY EFFICIENT.

NOTE: IF THE MODEL IS BIVARIATE NORMAL,

$$\sqrt{n}(\hat{\rho}_{MLE} - \rho) \to N(0, (1 - \rho^2)^2)$$

SPECIAL CASE: $(X, Y) \sim H(x, y)$ WITH

$$Y^{(\lambda_0)} = \beta_0 + \beta_1 X^{(\lambda_1)} + \epsilon, \ \ \textit{WHERE}$$

$(Y^{(\lambda_0)}, X^{(\lambda_1)}, \epsilon) \sim$ *NORMAL*.
ZOU AND HALL (2002):
THE MLE OF $\rho = \textit{CORR}$ WITH $\lambda_0, \lambda_1$ UNKNOWN HAS ASYMPTOTICALLY THE SAME DISTRIBUTION AS THE MLE WITH $\lambda_0, \lambda_1$ KNOWN.
PROOF: $H \in \mathcal{P}_C$. MLE NO WORSE THAN NORMAL SCORES.

NEED: SEMIPARAMETRIC EFFICIENCY.
$\Rightarrow$ PARAMETRIC EFFICIENCY.

CONSISTENT WITH WONG RESULT.
IN COX-REID LANGUAGE:
$\rho, \lambda_0$, and $\lambda_1$ ARE ORTHOGONAL PARAMETERS.

SUMMARY:

1. REGRESSION TRANSFORMATION MODELS.
2. YEO-JOHNSON (2000 BIOMETRIKA).
3. HERNANDEZ-JOHNSON (1980 JASA).
4. WONG TRANSFORMATIONS(1983).
5. KLAASSEN-WELLNER COPULA (1997).