Department of Statistics and Actuarial Science
The University of Hong Kong

Friday 22 February 2013, 2:30-3:30 pm, Room: MW524

Seminar:

The (Unspoken) Truth of Spurious Correlations

Speaker: Kai W. Ng

Patrick S C Poon Professor in Statistics and Actuarial Science

The University of Hong Kong

(URL: http://www.hku.hk/statistics/staff/kaing/)

1. **J. Fan and J. Lv (2008),** *JRSS(B)*, **commented by 41 Discussants in Royal Stat Society Meeting; p.852, p.884**
2. **J.Fan, R. Samsworth, Y. Wu (2009)** *Journal of Machine Learning Research*, **p.2014**
3. **J. Fan and J. Lv (2010), Invited Review Article,** *Statistica Sinica*, **p.102-3**
4. **J. Fan, J. Lv and L. Qi (2011)** *Annu. Rev. Econ.*, **p.293**
5. **J. Fan, S. Guo and N. Hao (2012),** *JRSS(B)*, **p.39**
6. **G. Li, H. Peng, J Zhang and L. Zhu (2012)** *Ann. Stat.*, **p.5, p.17-19**
7. **D.A.S. Fraser and K.W. Ng (1980) Multivariate regression: Analysis with Spherical Error, in** *Multivariate Analysis V (Editor: P.R. Krishnaiah)*, **369-386. New York: Elsevier North-Holland.**

Bivariate normal: μ_x, μ_y, σ_x^2, σ_y^2, ρ. i.i.d. with size n: $(\boldsymbol{x}, \boldsymbol{y})$
Sample correlation coefficient R:

$$R =: \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} = \frac{\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{y}}{\sqrt{\boldsymbol{x}'\boldsymbol{Q}\boldsymbol{x}}\ \sqrt{\boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}}},$$

where $\boldsymbol{Q} = \boldsymbol{I}_n - n^{-1}\mathbf{1}\mathbf{1}'$, $\mathbf{1} = [1 \cdots 1]'$ and $\boldsymbol{I}_n$ is identity matrix.
If $\rho = 0$,

$$T_{corr} = \sqrt{n-2}\frac{R}{\sqrt{1-R^2}} \sim t(n-2), \quad n > 2 \tag{1}$$

$$f_R(r) = \frac{\Gamma((n-1)/2)}{\sqrt{\pi}\Gamma((n-2)/2)}(1-r^2)^{(n-4)/2}, \quad -1 < r < 1 \tag{2}$$

$$R^2 \sim \text{Beta}(1/2, (n-2)/2)$$

C&D approach: Conditioning

$\mathbf{E}(X|Y) = (\mu_x - \rho(\sigma_x/\sigma_y)\mu_y) + \rho(\sigma_x/\sigma_y)Y$, $\mathbf{Var}(X|Y)(1-\rho^2)\sigma_x^2$
$\because \mathbf{E}(X|Y) = \beta_0 + \beta_1 Y \therefore \beta_0 = \mu_x - \rho(\sigma_x/\sigma_y)\mu_y$ **and** $\beta_1 = \rho\sigma_x/\sigma_y$
$\because$ **Two** $\mathbf{E}(X|Y)$ **are equal** $\therefore \rho = 0$ **iff** $\beta_1 = 0$ **iff** $\beta_0 = \mu_x$
$(X \; ind \; Y) \Rightarrow H_0 : \beta_1 = 0 \therefore T_{regr} = \hat{\beta}_1/S.E.(\hat{\beta}_1) \sim t(n-2)$, **but**

$$T_{regr} = \frac{\hat{\beta}_1}{S.E.(\hat{\beta}_1)} = \sqrt{n-2}\frac{R}{\sqrt{1-R^2}} = T_{corr} \tag{3}$$

C&D approach: De-conditioning

1. **The distribution of** T_{regr} **and, hence of** R**, does not depend on** Y**. So** R **is independent of** Y**,** $R^2 \sim \text{Beta}(1/2, (n-2)/2)$
2. Y **can have any dist with** $P(\boldsymbol{y} = a\mathbf{1}) = 0 \; \forall a$**; e.g. continuous pdf of** n**-dimension**

Theorem 1

Assume: (a) $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p$ independent, $\boldsymbol{x}_j \sim N_n(\mu_j \mathbf{1}, \sigma_j^2 \boldsymbol{I}_n)$, $n \geq 3$, whether $p < n$ or $p \geq n$ (b) $\boldsymbol{y}$ independent with any dist. with $P(\boldsymbol{y} = a\mathbf{1}) = 0 \; \forall a$, e.g. a continuous pdf of dimension n. Let R_j be the sample corr. coef. of $(X_j, Y), j = 1, 2, \ldots, p$.

(i) $\{R_1^2, \ldots, R_p^2\}$ are i.i.d. $\text{Beta}(1/2, (n-2)/2)$, so that the kth largest has the following c.d.f. evaluated at x

$$F_{[k]}(x|p,n) = \sum_{i=0}^{k-1} \binom{p}{i} [1 - B(x|1/2,(n-2)/2)]^i \; B(x|1/2,(n-2)/2)^{p-i}$$

(ii) $(R_1, \ldots, R_p)$ is independent of $\boldsymbol{y}$, and of the sample means and corrected sums of squares of all $\boldsymbol{x}_j$, $j = 1, \ldots, p$. □

Exact distribution of "γ_n=$\max_{j\leq p}|\text{corr}_n(X_j, Y)|$"

$$G_{\gamma_n}(u|p) = Pr(\gamma_n \leq u) = Pr(\gamma_n^2 \leq u^2) = [B(u^2|1/2, (n-2)/2)]^p$$

$$g_{\gamma_n}(u|p) = p[B(u^2|1/2, (n-2)/2)]^{p-1} \frac{2(1-u^2)^{(n-4)/2}}{B(1/2, (n-2)/2)},\ 0 < u < 1$$

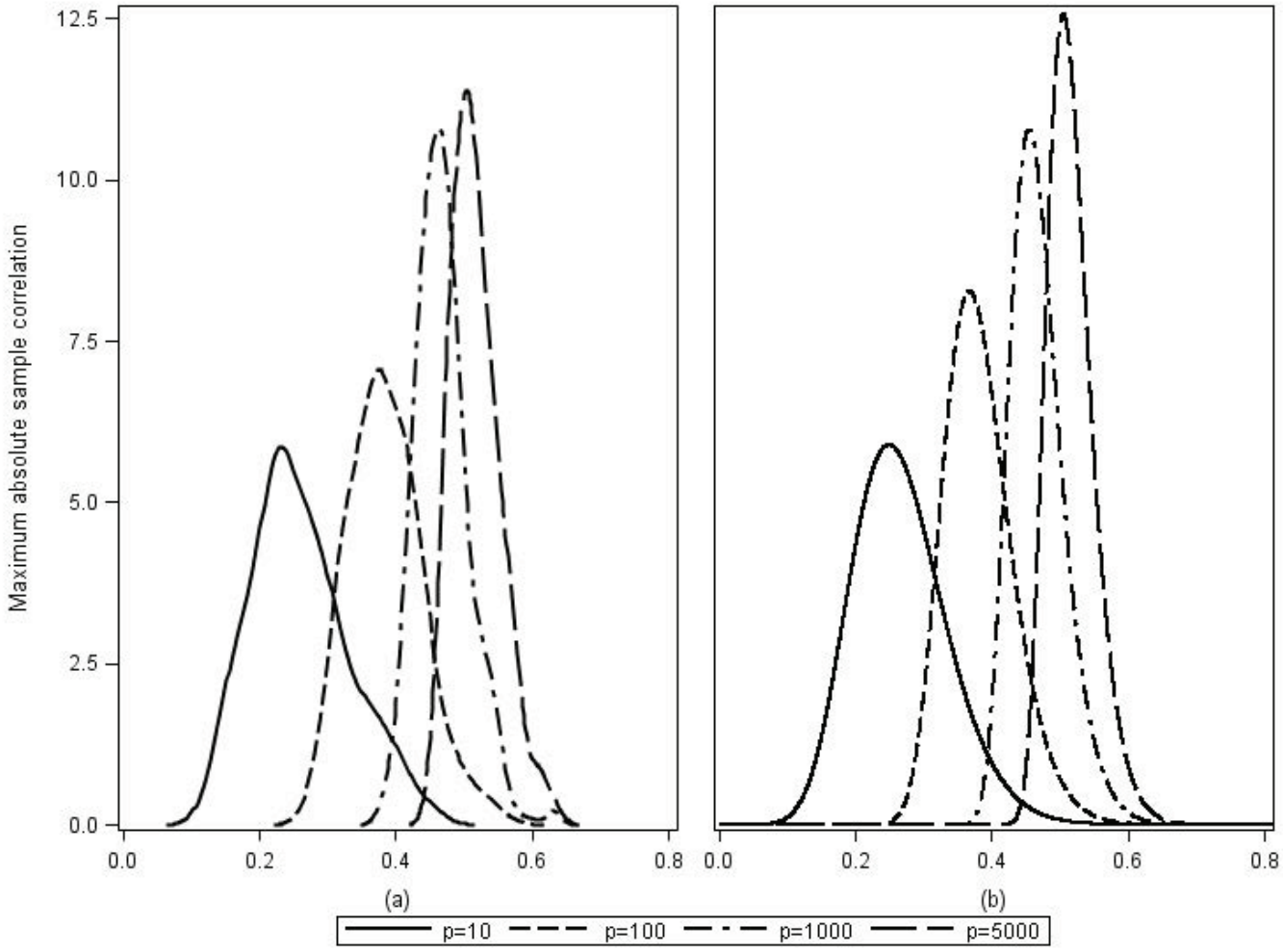
Critical Sample Size $N(\gamma, r^2, p)$

The smallest integer such that the max squared correlation of p false predictors is at most r^2 with probability $\gamma = 1 - \alpha$.

Critical Sample Size Algorithm **(given (γ, r^2, p), start n=1)**

1. **RETURN** n **if** $[B(r^2|1/2, (n-2)/2)]^p \geq \gamma$
2. **Else let** $n = n + 1$ **and go to Step 1**

"γ_n=$\max_{j\leq p}|\text{corr}_n(X_j, Y)|$" in Fig.1(a) of Fan, Guo and Hao (2012), $n = 50$ **and** $p = 10, 100, 1000, 5000,$ **where** $(Y, X_1, \ldots, X_p)$ **i.i.d** $N(0, 1)$. **(a) 500 simulated data sets vs. (b) Exact pdf**



Critical Sample Size $N(\gamma, r^2, p)$

$r^2(\%)$	Different p with fixed $\gamma = 0.99$									
	1000	**2000**	**3000**	**4000**	**5000**	**6000**	**7000**	**8000**	**9000**	**10000**
1	**1943**	**2075**	**2153**	**2208**	**2250**	**2285**	**2315**	**2340**	**2363**	**2383**
2	**968**	**1034**	**1072**	**1100**	**1121**	**1138**	**1153**	**1166**	**1177**	**1187**
5	**383**	**409**	**424**	**435**	**443**	**450**	**456**	**461**	**465**	**469**
10	**188**	**201**	**208**	**213**	**217**	**221**	**223**	**226**	**228**	**230**
15	**123**	**131**	**136**	**139**	**142**	**144**	**146**	**148**	**149**	**150**
20	**90**	**96**	**100**	**102**	**104**	**106**	**107**	**108**	**109**	**110**
25	**71**	**75**	**78**	**80**	**81**	**83**	**84**	**85**	**85**	**86**
30	**58**	**61**	**64**	**65**	**66**	**67**	**68**	**69**	**69**	**70**

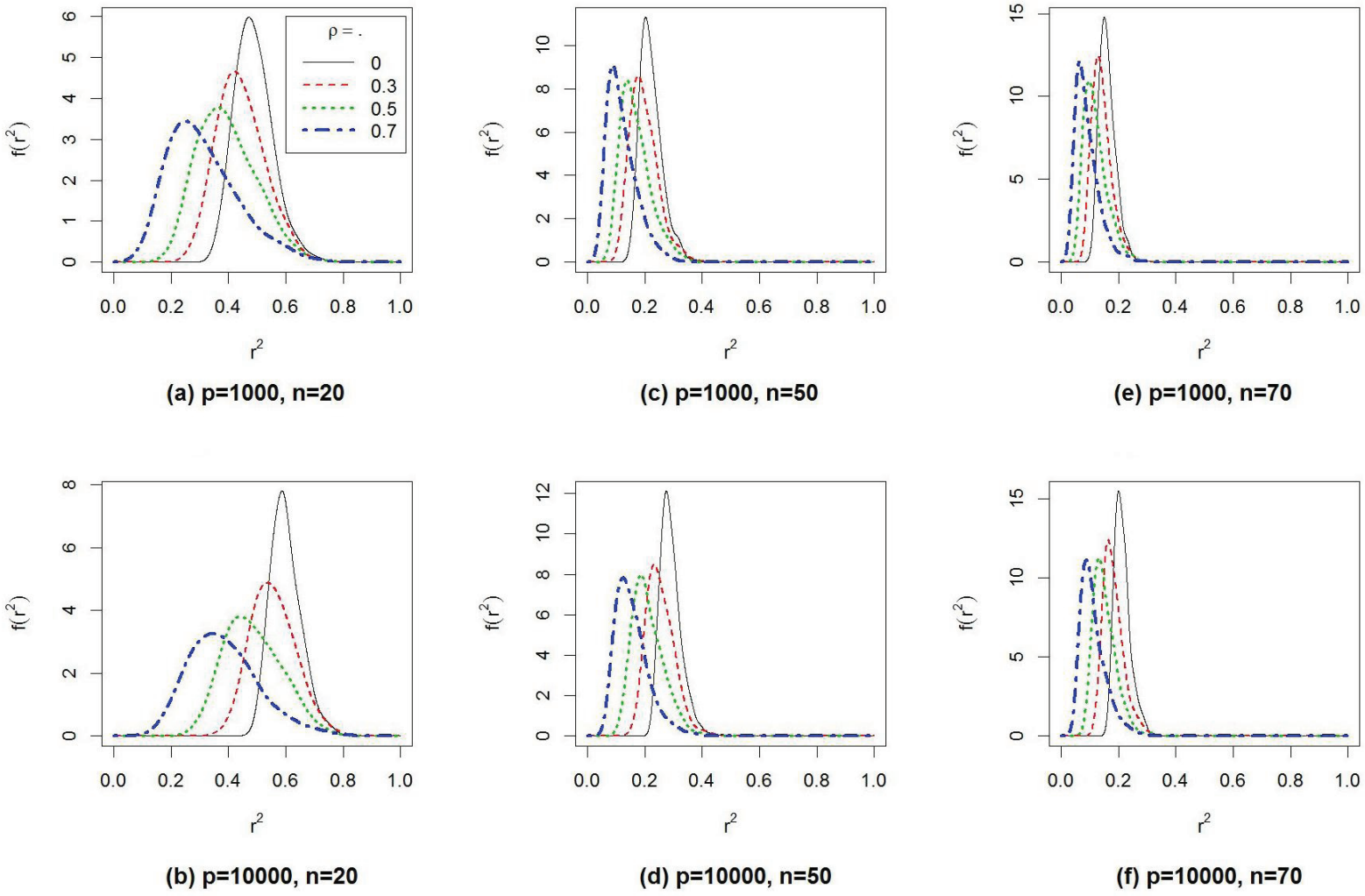
$r^2(\%)$	Different p with fixed $\gamma = 0.99$								
	2E4	**3E4**	**4E4**	**5E4**	**6E4**	**7E4**	**8E4**	**9E4**	**1E5**
1	**2516**	**2594**	**2649**	**2692**	**2727**	**2757**	**2782**	**2805**	**2825**
2	**1253**	**1292**	**1319**	**1341**	**1358**	**1373**	**1386**	**1397**	**1407**
5	**495**	**511**	**521**	**530**	**537**	**543**	**548**	**552**	**556**
10	**243**	**250**	**255**	**259**	**263**	**266**	**268**	**270**	**272**
15	**158**	**163**	**167**	**169**	**171**	**173**	**175**	**176**	**178**
20	**116**	**120**	**122**	**124**	**126**	**127**	**128**	**129**	**130**
25	**91**	**93**	**95**	**97**	**98**	**99**	**100**	**101**	**102**
30	**74**	**76**	**78**	**79**	**80**	**81**	**81**	**82**	**83**

Critical Sample Size $N(\gamma, r^2, p)$

$r^2(\%)$	Different p with fixed $\gamma = 0.999$									
	1000	**2000**	**3000**	**4000**	**5000**	**6000**	**7000**	**8000**	**9000**	**10000**
1	**2384**	**2517**	**2594**	**2650**	**2693**	**2728**	**2757**	**2783**	**2806**	**2826**
2	**1187**	**1253**	**1292**	**1320**	**1341**	**1358**	**1373**	**1386**	**1397**	**1407**
5	**469**	**496**	**511**	**522**	**530**	**537**	**543**	**548**	**552**	**556**
10	**230**	**243**	**250**	**255**	**260**	**263**	**266**	**268**	**270**	**272**
15	**150**	**158**	**163**	**167**	**169**	**171**	**173**	**175**	**176**	**178**
20	**110**	**116**	**120**	**122**	**124**	**126**	**127**	**128**	**129**	**130**
25	**86**	**91**	**94**	**95**	**97**	**98**	**99**	**100**	**101**	**102**
30	**70**	**74**	**76**	**78**	**79**	**80**	**81**	**81**	**82**	**83**

$r^2(\%)$	Different p with fixed $\gamma = 0.999$								
	2E4	**3E4**	**4E4**	**5E4**	**6E4**	**7E4**	**8E4**	**9E4**	**1E5**
1	**2960**	**3038**	**3093**	**3137**	**3172**	**3202**	**3227**	**3250**	**3271**
2	**1474**	**1513**	**1540**	**1562**	**1579**	**1594**	**1607**	**1618**	**1629**
5	**582**	**598**	**609**	**617**	**624**	**630**	**635**	**639**	**643**
10	**285**	**292**	**298**	**302**	**305**	**308**	**311**	**313**	**315**
15	**186**	**191**	**194**	**197**	**199**	**201**	**202**	**204**	**205**
20	**136**	**140**	**142**	**144**	**146**	**147**	**148**	**149**	**150**
25	**106**	**109**	**111**	**112**	**114**	**115**	**116**	**116**	**117**
30	**86**	**89**	**90**	**91**	**92**	**93**	**94**	**94**	**95**

$N(\gamma, r^2, p)$ gives higher γ for equally correlated spurious predictors



$\because |\boldsymbol{R}| > 0 \Rightarrow \rho > -1/(p-1)$ **in equal correlations** $\therefore$ **Not consider** $\rho < 0$

Corollary 1

Under condition (a) of Theorem 1. The following are valid for the sample correlation matrix $\boldsymbol{R} = (R_{jk})$.

(i) Each R_{jk}^2 $(j \neq k) \sim \text{Beta}(1/2, (n-2)/2)$, indep. of $\boldsymbol{x}_j$ & $\boldsymbol{x}_k$

(ii) All R_{jk} $(j \neq k)$ are pairwise independent

(iii) For any fixed j, all R_{jk} $(j \neq k)$ are mutually independent among themselves and jointly independent of $\boldsymbol{x}_j$

(iv) Any subset of R_{jk} $(j \neq k)$ cannot be mutually independent if it contains a triangular block of three elements totally below (or totally above) the diagonal

(v) $\boldsymbol{R}$ is independent of the sample means and corrected sums of squares of all $\boldsymbol{x}_j$, $j = 1, \ldots, p$ □

Theorem 2

Given $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_q\}$, $q<min\{p,n-2\}$, whether $p<n$ or $p\geq n$, assume cond. ind. of $\{\boldsymbol{x}_{q+1},\cdots,\boldsymbol{x}_p,\boldsymbol{y}\}$, $\boldsymbol{x}_j=\boldsymbol{X}\boldsymbol{\beta}_j+\sigma_j\boldsymbol{z}_j$ with $\boldsymbol{X}=[\mathbf{1}\ \boldsymbol{x}_1\cdots\boldsymbol{x}_q]$ and $\boldsymbol{z}_j\sim N_n(\mathbf{0},\boldsymbol{I}_n)$, and $\boldsymbol{y}$ has any dist. with $P(\boldsymbol{y}=\boldsymbol{\beta}\boldsymbol{X})=0\ \forall\boldsymbol{\beta}$ (e.g. a continuous pdf of dimension n).
(i) $R^2_{yj|1\cdots q}$ are i.i.d. $\text{Beta}(1/2,(n-q-2)/2)$ with beta c.d.f. $B(x|1/2,(n-q-2)/2)$, and the kth largest of $R^2_{yj|1\cdots q}$ $(j=q+1,\ldots,p)$ has the following c.d.f. at x

$$F_{[k]}(x|p-q,n-q)=\sum_{i=0}^{k-1}\binom{p-q}{i}\left[1-B(x|1/2,(n-q-2)/2)\right]^i\ B(x|1/2,(n-q-2)/2)^{p-q-i}$$

(ii) $(R_{y,q+1|1\cdots q},\ldots,R_{yp|1\cdots q})$ is independent of $\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_q,\boldsymbol{y}\}$; in particular, $(R_{y1},\ldots,R_{yq})$ and $(R_{y,q+1|1\cdots q},\ldots,R_{yp|1\cdots q})$ are independently distributed. □

▶ Mathematically, one may view Theorem 1 as a corollary of Theorem 2 when $q = 0, \boldsymbol{X} = \mathbf{1}, \boldsymbol{Q} = \boldsymbol{I}_n - n^{-1}\mathbf{1}\mathbf{1}'$; i.e. the ordinary correlation coefficient is a special case of partial correlation coefficient. ∴ Use $\boldsymbol{X}$ for both cases thereafter.

▶ A pdf of $\boldsymbol{z}$ is "spherically symmetric" if it is a function of the sum of squares, like $f(\boldsymbol{z}) = h(\boldsymbol{z}'\boldsymbol{z})$ where $h(\cdot)$ is a non-negative function of a non-negative variable.

Examples for spherically symmetric $\boldsymbol{z}_j$ in $\boldsymbol{x}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \sigma_j \boldsymbol{z}_j$:

(a) $\boldsymbol{z}_j \sim N_n(\mathbf{0}, \boldsymbol{I}_n)$

(b) $\boldsymbol{z}_j \sim \left(\sum_{k=1}^m \theta_k N_n(\mathbf{0}, \boldsymbol{I}_n)\right)$, **where** $0 < \theta_k < \sum_{k=1}^m \theta_k = 1$

(c) Multivariate $t(\nu)$ **[Cauchy if** $\nu = 1$**;** $N_n(\mathbf{0}, \boldsymbol{I}_n)$ **if** $\nu \to \infty$**]:**

$$f(\boldsymbol{z}_j) = \frac{\Gamma((n+\nu)/2)}{\nu^{n/2}\pi^{n/2}\Gamma(\nu/2)}(1 + \nu^{-1}\boldsymbol{z}_j'\boldsymbol{z}_j)^{-(n+\nu)/2}, \quad \nu > 0.$$

Fraser-Ng results on Spherically symmetric error

In $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \sigma\boldsymbol{z}$, where $\boldsymbol{X}$ is $n \times k$ with rank $k < n$, if $\boldsymbol{z}$ has a spherically symmetric pdf, then the following are valid, where $\boldsymbol{Q} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $RSS = \boldsymbol{y}'\boldsymbol{Q}\boldsymbol{y}$.

(i) $(RSS)^{-1/2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ is ind. of $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ and RSS.

(ii) $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})/\sqrt{RSS}$ and $(RSS)^{-1/2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ have the same distributions as if the elements of $\boldsymbol{z}$ are i.i.d. standard normal; for the jth component of $\boldsymbol{\beta}$, $(\hat{\beta}_j - \beta_j)/S.E.(\hat{\beta}_j)$ has the same $t(n-k)$ distribution as in the normal case, and $(RSS)^{-1/2}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$ is uniformly distributed on the surface of the unit sphere in the $n-k$ dimension subspace orthogonal to the subspace spanned by $\boldsymbol{X}$.

(iii) The dist of RSS depends on the spherically symmetric distribution; RSS/σ^2 has $\chi^2(n-k)$ dist iff $\boldsymbol{z} \sim N_n(\boldsymbol{0}, \boldsymbol{I}_n)$.

∴ **The proof of Thms. 1 and 2 depend only on parts (i) and (ii) of the above Fraser-Ng results which happen to be the same as if normality is assumed. So we have:**

Theorem 3

(a) The conclusions in (i) and (ii) of Theorem 1 are still valid if we change only the normality assumption about $(\boldsymbol{x}_1 - \mu_1\mathbf{1}), \ldots, (\boldsymbol{x}_p - \mu_p\mathbf{1})$ **so that each** $(\boldsymbol{x}_1 - \mu_1\mathbf{1})$ **independently has a spherically symmetric density function of dimension** n**, possibly from a different family and possibly with a different scaling parameter.**

(b) The conclusions in (i) and (ii) of Theorem 2 are still valid if we change only the normality assumption $\boldsymbol{z}_j \sim N_n(\mathbf{0}, \boldsymbol{I}_n)$ **so that each** $\boldsymbol{z}_j$ **now has a possibly different spherically symmetric density function with a possibly different scaling parameter.**