Dependency/Competition in Corn Yield: A Proxy for Land Use Change

October 2009

Kobi Abayomi, Valerie Thomas, Dexin Luo

October 2009

Setup

We investigate the effect of biofuels on land use change via U.S. corn production data.

- Agricultural models are used to estimate the effect of biofuel production on crop production $\Rightarrow$ land use change
- Statistical determination of the influence of biofuel production is non-standard $\Rightarrow$ total crop use is always the sum of the different uses
- These (induced) distributions are *Compositional Distributions* $\Rightarrow$ Methodology for Dependency/Competition among Compositional Distributions

Setup

We investigate the effect of biofuels on land use change via U.S. corn production data.

- Agricultural models are used to estimate the effect of biofuel production on crop production $\Rightarrow$ land use change
- Statistical determination of the influence of biofuel production is non-standard $\Rightarrow$ total crop use is always the sum of the different uses
- These (induced) distributions are *Compositional Distributions* $\Rightarrow$ Methodology for Dependency/Competition among Compositional Distributions

Setup

We investigate the effect of biofuels on land use change via U.S. corn production data.

- Agricultural models are used to estimate the effect of biofuel production on crop production $\Rightarrow$ land use change
- Statistical determination of the influence of biofuel production is non-standard $\Rightarrow$ total crop use is always the sum of the different uses
- These (induced) distributions are *Compositional Distributions* $\Rightarrow$ Methodology for Dependency/Competition among Compositional Distributions

Outline

Context

Methodology

Illustrations

Comments

Energy Independence and Security Act

- EISA-2007 mandates an increase in ethanol production to 36 billion gallons per year by 2022.
- Ethanol production has increased more than 5000 percent since 1980
- At the same time, the total U.S. corn yield has less than doubled

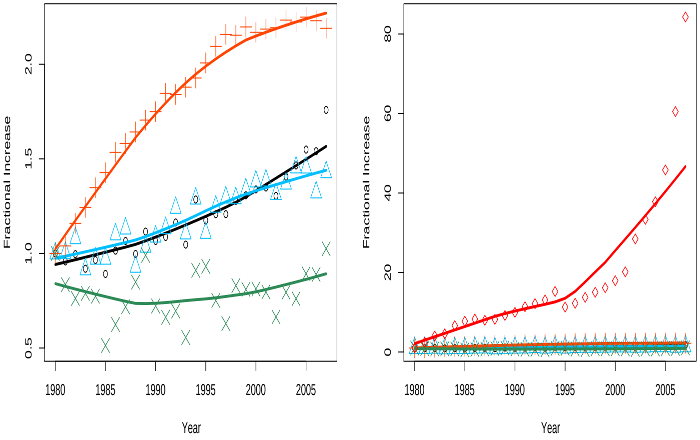
Energy Independence and Security Act

- EISA-2007 mandates an increase in ethanol production to 36 billion gallons per year by 2022.
- Ethanol production has increased more than 5000 percent since 1980
- At the same time, the total U.S. corn yield has less than doubled

Energy Independence and Security Act

- EISA-2007 mandates an increase in ethanol production to 36 billion gallons per year by 2022.
- Ethanol production has increased more than 5000 percent since 1980
- At the same time, the total U.S. corn yield has less than doubled

Fractional Increase



Environmental Impacts

- Net energy budget
- Competition with corn-based commodities
- Greenhouse emissions
- $\Rightarrow$ Competition within distribution of Corn Yield constitutents

Environmental Impacts

- Net energy budget
- Competition with corn-based commodities
- Greenhouse emissions
- $\Rightarrow$ Competition within distribution of Corn Yield constitutents

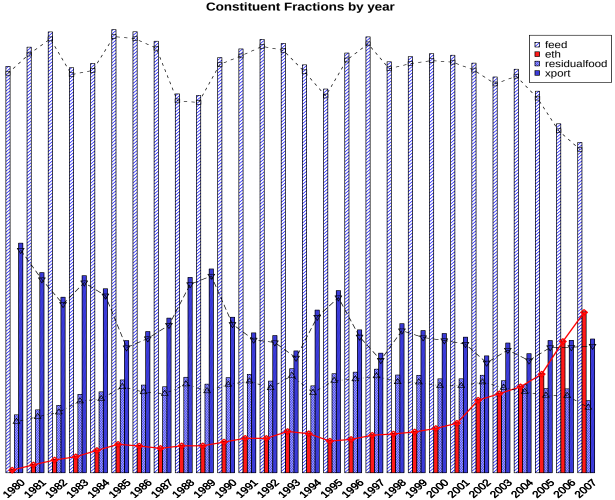
Environmental Impacts

- Net energy budget
- Competition with corn-based commodities
- Greenhouse emissions
- $\Rightarrow$ Competition within distribution of Corn Yield constitutents

Environmental Impacts

- Net energy budget
- Competition with corn-based commodities
- Greenhouse emissions
- $\Rightarrow$ Competition within distribution of Corn Yield constitutents

Constituents of Corn Yield



Outline

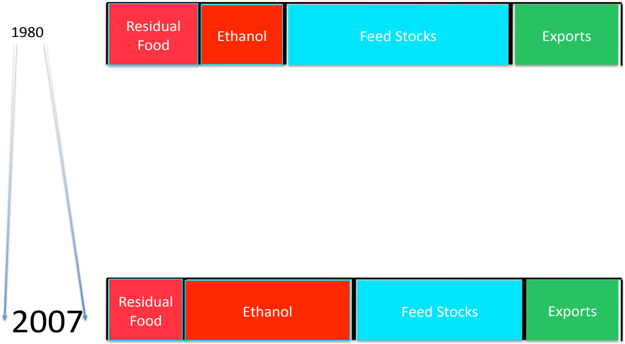
Context

Methodology

Illustrations

Comments

Compositional Distribution



Aitchison Notation

Let

$$\mathbf{x} = (x_1, ..., x_k) \tag{1}$$

be a basis or open vector of positive quantities
In this example

$$\mathbf{x} = (x_{eth}, x_{rfood}, x_{feed}, x_{xport})$$

(in bushels) corn of: ethanol production, residual food stock, feed stock, and exports.

Aitchison Notation

Let

$$y_j = x_j / \sum_j^k x_j \tag{2}$$

$\mathbf{y} = (y_1, ..., y_k)$ the vector of fractions.
Aitchison defines $y_{k+1} = 1 - \sum_j^k y_j$;
Here $\sum_j^k y_j = 1$

Aitchison Notation

A (log-ratio) transformation sets

$$v_{\substack{j\\m}} = log(\frac{y_j}{y_m}) = log\ y_j - log\ y_m \tag{3}$$

in a slight modification of Aitchison's notation (where $v_j = log(y_j/y_{k+1})$).

Modifying Aitchison Notation

- The total is fixed and known $\Rightarrow$ the residual is y_{k+1}=0 and Aitchison's v_j is undefined
- In the original notation v_j is the log of the relative fraction of constituent j to the residual component of the basis
- $v_{\cdot m}$, maps $\mathbb{S}^{m,k} = \{\mathbf{y}_{-m}, y_m\}$ to $\mathbb{R}^{k-|m|}$

Modifying Aitchison Notation

- The total is fixed and known $\Rightarrow$ the residual is y_{k+1}=0 and Aitchison's v_j is undefined
- In the original notation v_j is the log of the relative fraction of constituent j to the residual component of the basis
- $v_{\substack{\cdot\\m}}$, maps $\mathbb{S}^{m,k} = \{\mathbf{y}_{-m}, y_m\}$ to $\mathbb{R}^{k-|m|}$

Modifying Aitchison Notation

- The total is fixed and known $\Rightarrow$ the residual is y_{k+1}=0 and Aitchison's v_j is undefined
- In the original notation v_j is the log of the relative fraction of constituent j to the residual component of the basis
- $v_{\underset{m}{\cdot}}$, maps $\mathbb{S}^{m,k} = \{\mathbf{y}_{-m}, y_m\}$ to $\mathbb{R}^{k-|m|}$

Why Transform?

(Natural) Dirichlet model for $(y_1, ..., y_k)$; $\sum_j y_j = 1$; $y_j > 0 \; \forall j$ is:

$$dF(\mathbf{y}) \propto (1 - \sum_j y_j)^{\alpha_{k+1}-1} \cdot \prod_j y_j^{\alpha_j - 1} \tag{4}$$

with parameters $\alpha = (\alpha_1, ..., \alpha_{k+1})$
Insufficient for non-*neutral* proportions

Why Transform?

(Generalization) Liouville distribution is:

$$dF \propto h(\sum_j y_j) \prod y_j^{\alpha_j - 1} \tag{5}$$

with $\alpha_j > 0$ (as before) and g some function.
Note that when $h(t) = 1 - t$ the Liouville distribution is the special case Dirichlet distribution with $\alpha_{k+1} = 1$
Thus h is an additional parameter of interest for estimation

Aitchison Approach

- Aitchison fits a log-normal distribution on $\mathbf{v}$
- Σ is sufficient for dependency in the log-normal distribution

Aitchison Approach

- Aitchison fits a log-normal distribution on $\mathbf{v}$
- Σ is sufficient for dependency in the log-normal distribution

Aitchison Approach

- Under a composition

$$\Sigma_{\mathbf{v}} \propto diag(\omega_1, ..., \omega_k) + \omega_{k+1}, \tag{6}$$

- $\Sigma_{\mathbf{v}}$ is constrained to the positive orthant and proportional to the units of the residual component

Aitchison Approach

- Under a composition

$$\Sigma_{\mathbf{v}} \propto \mathit{diag}(\omega_1, ..., \omega_k) + \omega_{k+1}, \tag{6}$$

- $\Sigma_{\mathbf{v}}$ is constrained to the positive orthant and proportional to the units of the residual component

Aitchison Approach

- Aitchison uses a likelihood ratio test
- Test statistic is iteratively estimated due to the constraints on the support of the parameter space
- $\Sigma_{\mathbf{v}} \geq \mathbf{0}$ and proportional to the choice of y_{k+1}

Aitchison Approach

- Aitchison uses a likelihood ratio test
- Test statistic is iteratively estimated due to the constraints on the support of the parameter space
- $\Sigma_{\mathbf{v}} \geq \mathbf{0}$ and proportional to the choice of y_{k+1}

Aitchison Approach

- Aitchison uses a likelihood ratio test
- Test statistic is iteratively estimated due to the constraints on the support of the parameter space
- $\Sigma_{\mathbf{v}} \geq \mathbf{0}$ and proportional to the choice of y_{k+1}

Our Approach

Multivariate Version of KS distance:

$$D_{n,k} = \sup_{\mathbf{t}} |F_n(\mathbf{t}) - F(\mathbf{t})| \tag{7}$$

For $\mathbf{t} = (t_1, t_2, ...)$ the distance is a probability measure on Kendall's distributions...chi-square convergence does not hold (Nelsen 2003).

Our Approach

A Kolmogorov-Smirnov statistic (distance) for multivariate independence can be written:

$$D^{\Pi}_{n,k} = \sup_{\mathbf{t}} |F_n(\mathbf{t}) - \prod_j F_j(t_j)|. \tag{8}$$

Under independence the distance converges to zero (via Glivenko-Cantelli), but distributionally for $k \geq 2$?

Our Approach

- Let $\mathbf{u} = (u_1, ..., u_k)$, where each $u_j = F_j(v_j)$
- Let the joint distribution for $\mathbf{v}$ be $F(\mathbf{v})$
- The *copula* for $\mathbf{u}$ is

$$C(\mathbf{u}) = F(F_1(v_1), ..., F_k(v_k)) \tag{9}$$

Our Approach

- Let $\mathbf{u} = (u_1, ..., u_k)$, where each $u_j = F_j(v_j)$
- Let the joint distribution for $\mathbf{v}$ be $F(\mathbf{v})$
- The *copula* for $\mathbf{u}$ is

$$C(\mathbf{u}) = F(F_1(v_1), ..., F_k(v_k)) \tag{9}$$

Our Approach

- Let $\mathbf{u} = (u_1, ..., u_k)$, where each $u_j = F_j(v_j)$
- Let the joint distribution for $\mathbf{v}$ be $F(\mathbf{v})$
- The *copula* for $\mathbf{u}$ is

$$C(\mathbf{u}) = F(F_1(v_1), ..., F_k(v_k)) \tag{9}$$

Our Approach

With $C_n(\cdot)$ a multivariate version of the *empirical copula*:

$$C_n(\mathbf{u}) = \frac{\#\{\mathbf{t} \mid t_1 \leq F_1^{-1}(u_1), ..., t_k \leq F_k^{-1}(u_k)\}}{n} \tag{10}$$

Our Approach

- Fit a *Dirichlet* distribution (i.e. estimate $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha_k})$ for $\alpha = (\alpha_1, ..., \alpha_k)$) to the composition data $\mathbf{y}$.
- Generate T *Dirichlet* replicates, parameter $\hat{\alpha}$, each of dimension $n \times k$: $(\mathbf{y}^{\hat{\alpha},1}, ..., \mathbf{y}^{\alpha,T})$.
- Compute $m = 1...k$ versions of Aitchison's log-ratios on the replicates: $\mathbf{v}_m^{\hat{\alpha},1} ..., \mathbf{v}_m^{\hat{\alpha},T}$
- For $m = 1..k$ compute $D_{n,k}^{\Pi,1}{}_m, ..., D_{n,k}^{\Pi,T}{}_m$ of

$$D_{n,k}^{\Pi} = \sup_{\mathbf{u}} |C_n(\mathbf{u}^{\hat{\alpha}}) - \prod_j u_j^{\hat{\alpha}_j}|. \tag{11}$$

Our Approach

- Fit a *Dirichlet* distribution (i.e. estimate $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha_k})$ for $\alpha = (\alpha_1, ..., \alpha_k)$) to the composition data **y**.
- Generate T *Dirichlet* replicates, parameter $\hat{\alpha}$, each of dimension $n \times k$: $(\mathbf{y}^{\hat{\alpha},1}, ..., \mathbf{y}^{\alpha,T})$.
- Compute $m = 1...k$ versions of Aitchison's log-ratios on the replicates: $\mathbf{v}_m^{\hat{\alpha},1} ..., \mathbf{v}_m^{\hat{\alpha},T}$
- For $m = 1..k$ compute $D_{\substack{n,k\\m}}^{\Pi,1}, ..., D_{\substack{n,k\\m}}^{\Pi,T}$ of

$$D_{n,k}^{\Pi} = \sup_{\mathbf{u}} |C_n(\mathbf{u}^{\hat{\alpha}}) - \prod_j u_j^{\hat{\alpha}_j}|. \tag{11}$$

Our Approach

- Fit a *Dirichlet* distribution (i.e. estimate $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha_k})$ for $\alpha = (\alpha_1, ..., \alpha_k)$) to the composition data $\mathbf{y}$.
- Generate T *Dirichlet* replicates, parameter $\hat{\alpha}$, each of dimension $n \times k$: $(\mathbf{y}^{\hat{\alpha},1}, ..., \mathbf{y}^{\alpha,T})$.
- Compute $m = 1...k$ versions of Aitchison's log-ratios on the replicates: $\mathbf{v}_m^{\hat{\alpha},1} ..., \mathbf{v}_m^{\hat{\alpha},T}$
- For $m = 1..k$ compute $D^{\Pi,1}_{\substack{n,k\\m}}, ..., D^{\Pi,T}_{\substack{n,k\\m}}$ of

$$D^{\Pi}_{n,k} = \sup_{\mathbf{u}} |C_n(\mathbf{u}^{\hat{\alpha}}) - \prod_j u_j^{\hat{\alpha}_j}|. \tag{11}$$

Our Approach

- Fit a *Dirichlet* distribution (i.e. estimate $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha_k})$ for $\alpha = (\alpha_1, ..., \alpha_k)$) to the composition data $\mathbf{y}$.
- Generate T *Dirichlet* replicates, parameter $\hat{\alpha}$, each of dimension $n \times k$: $(\mathbf{y}^{\hat{\alpha},1}, ..., \mathbf{y}^{\alpha,T})$.
- Compute $m = 1...k$ versions of Aitchison's log-ratios on the replicates: $\mathbf{v}_m^{\hat{\alpha},1} ..., \mathbf{v}_m^{\hat{\alpha},T}$
- For $m = 1..k$ compute $D^{\Pi,1}_{\substack{n,k\\m}}, ..., D^{\Pi,T}_{\substack{n,k\\m}}$ of

$$D^{\Pi}_{n,k} = \sup_{\mathbf{u}} |C_n(\mathbf{u}^{\hat{\alpha}}) - \prod_j u_j^{\hat{\alpha}_j}|. \tag{11}$$

Our Approach

- Yields a distribution for the statistic under an independence hypothesis among the compositions...
- m versions of $D^{\Pi,1}_{\frac{n,k}{m}}, \ldots, D^{\Pi,T}_{\frac{n,k}{m}}$ are proxies for tests of *complete subcompositional independence...*
- Calculating on the log-ratios ($\mathbf{v}$) of the replicates, and not the *Dirichlet* draws picks each of m components to serve as 'residual' (via the basis $\mathbf{x}$ or composition $\mathbf{y}$) without requiring m estimates of α and m-fold random draws.

Our Approach

- Yields a distribution for the statistic under an independence hypothesis among the compositions...
- m versions of $D^{\Pi,1}_{n,k \atop m}, \ldots, D^{\Pi,T}_{n,k \atop m}$ are proxies for tests of *complete subcompositional independence*...
- Calculating on the log-ratios (**v**) of the replicates, and not the *Dirichlet* draws picks each of m components to serve as 'residual' (via the basis **x** or composition **y**) without requiring m estimates of α and m-fold random draws.

Our Approach

- Yields a distribution for the statistic under an independence hypothesis among the compositions...
- m versions of $D^{\Pi,1}_{n,k \atop m}, \ldots, D^{\Pi,T}_{n,k \atop m}$ are proxies for tests of *complete subcompositional independence...*
- Calculating on the log-ratios ($\mathbf{v}$) of the replicates, and not the *Dirichlet* draws picks each of m components to serve as 'residual' (via the basis $\mathbf{x}$ or composition $\mathbf{y}$) without requiring m estimates of α and m-fold random draws.

Outline

Context

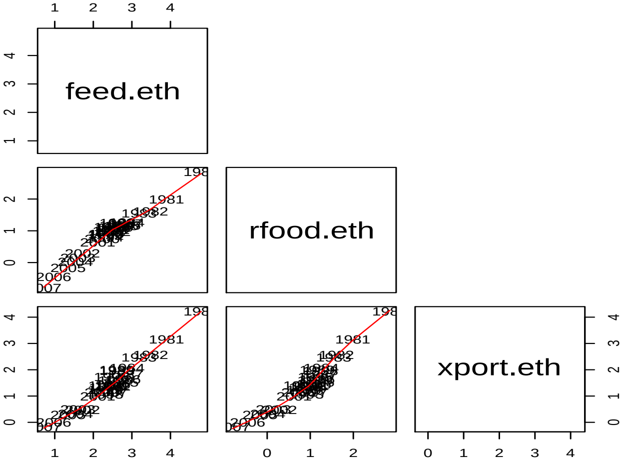
Methodology

Illustrations

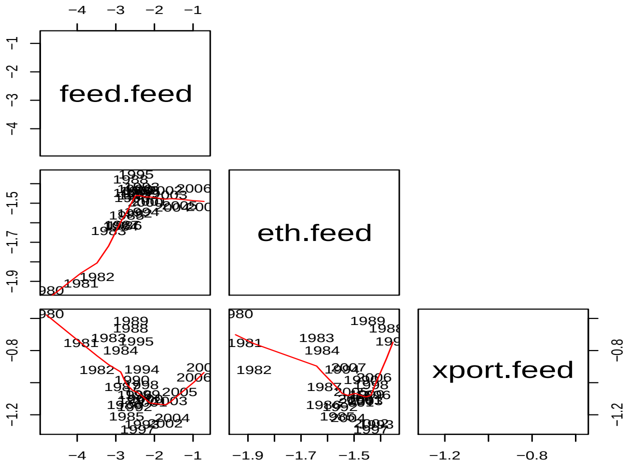
Comments

Compositional Distribution

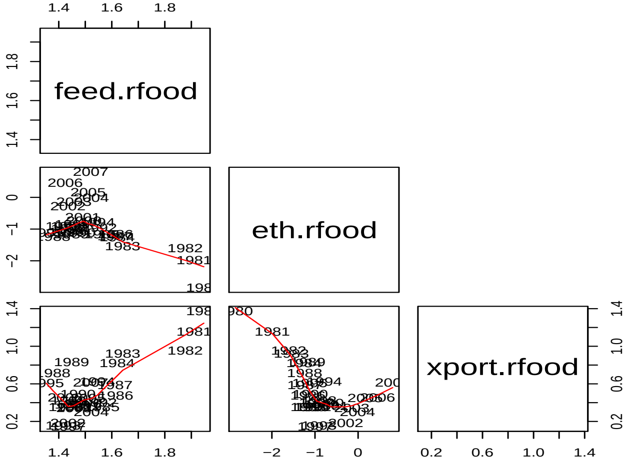
W.R.T. Ethanol



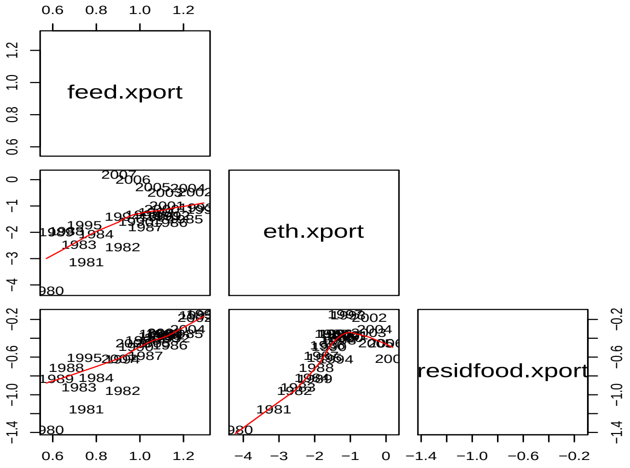
W.R.T. Feed



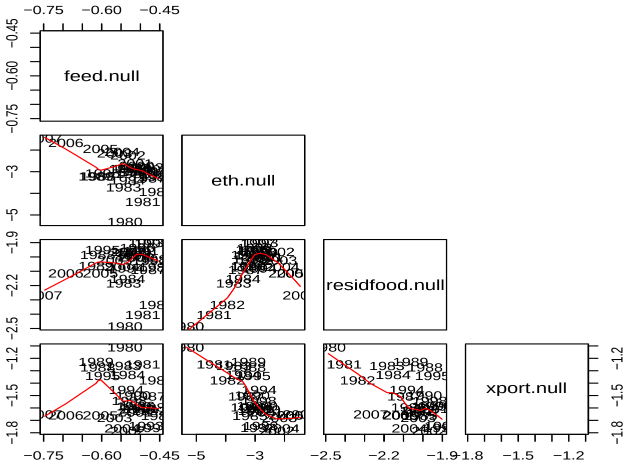
W.R.T. Food



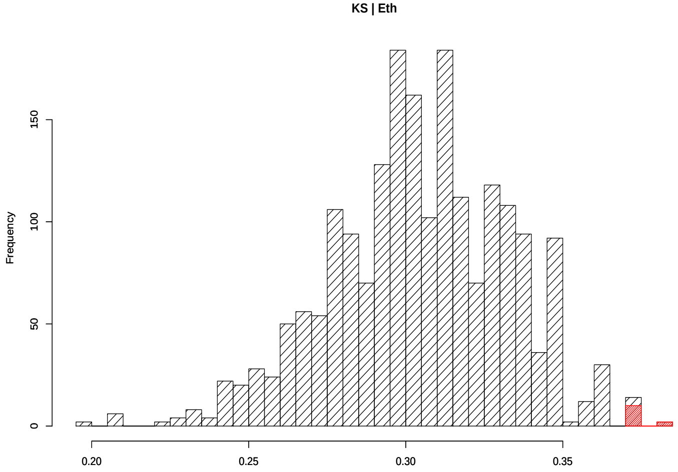
W.R.T. Export


feed.xport
eth.xport
residfood.xport
0.6
0.8
1.0
1.2
−4
−3
−2
−1
0
−1.4
−1.0
−0.6
−0.2

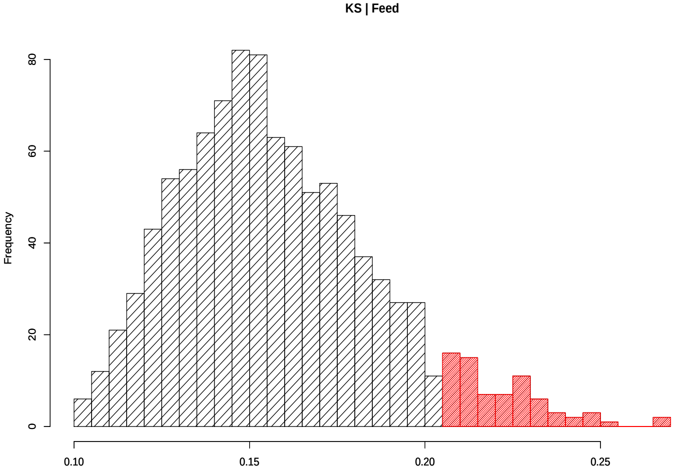
W.R.T. Null


feed.null
eth.null
residfood.null
xport.null
−0.75
−0.60
−0.45
−5
−3
−1.9
−2.2
−2.5
−1.2
−1.5
−1.8

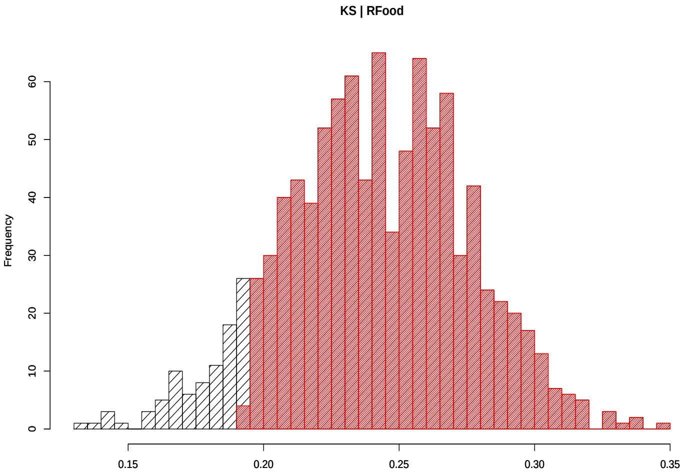
W.R.T. Ethanol



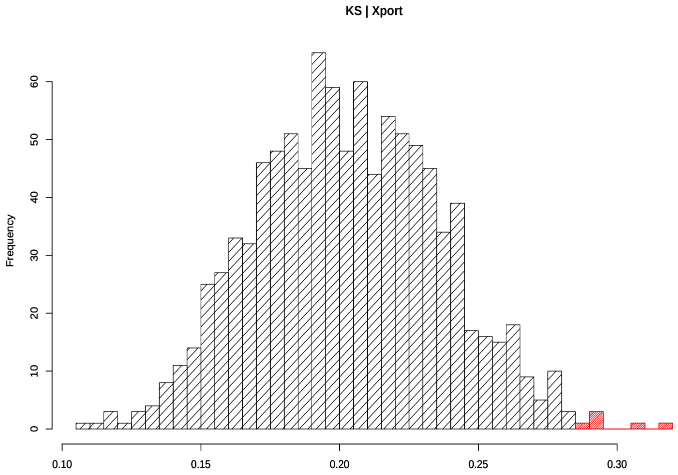
W.R.T. Feed



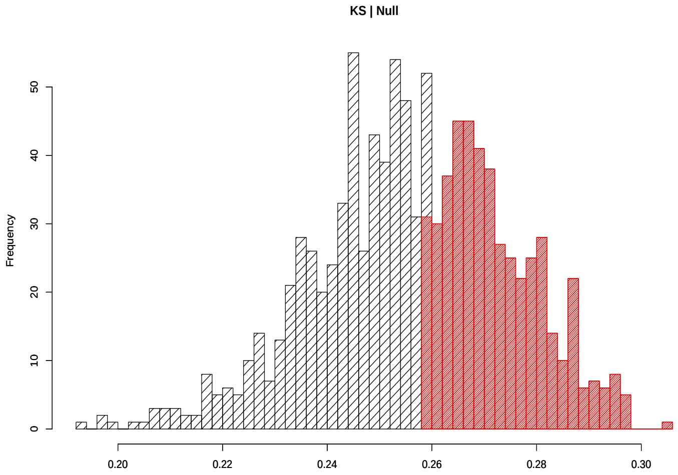
W.R.T. Residual Food



W.R.T. Export



W.R.T. Null



Outline

Context

Methodology

Illustrations

Comments

Comments

- Distance is $\mathcal{L}_\infty$ norm, dominates $\mathcal{L}_2$ - Cramer von-Mises distance
- Empirical Prob Integral Transform : : Order statistics $\Rightarrow$ Invariant to increasing (log-ratio) transform.
- In high dimensions Aitchison's method 'tail-migrates'
- Distance is Euclidean (not on Aitchison Geometry!) on prob measure space.

Comments

- Distance is $\mathcal{L}_\infty$ norm, dominates $\mathcal{L}_2$ - Cramer von-Mises distance
- Empirical Prob Integral Transform : : Order statistics $\Rightarrow$ Invariant to increasing (log-ratio) transform.
- In high dimensions Aitchison's method 'tail-migrates'
- Distance is Euclidean (not on Aitchison Geometry!) on prob measure space.

Comments

- Distance is $\mathcal{L}_\infty$ norm, dominates $\mathcal{L}_2$ - Cramer von-Mises distance
- Empirical Prob Integral Transform : : Order statistics $\Rightarrow$ Invariant to increasing (log-ratio) transform.
- In high dimensions Aitchison's method 'tail-migrates'
- Distance is Euclidean (not on Aitchison Geometry!) on prob measure space.

Comments

- Distance is $\mathcal{L}_\infty$ norm, dominates $\mathcal{L}_2$ - Cramer von-Mises distance
- Empirical Prob Integral Transform : : Order statistics $\Rightarrow$ Invariant to increasing (log-ratio) transform.
- In high dimensions Aitchison's method 'tail-migrates'
- Distance is Euclidean (not on Aitchison Geometry!) on prob measure space.

Deeper Issues

- "Richness" of *Dirichlet* replicates w.r.t generalized Liouville?
- Bayesian ‘prior’ for $\alpha \Rightarrow$ posterior distribution for D
- Time dimension?

Deeper Issues

- "Richness" of *Dirichlet* replicates w.r.t generalized Liouville?
- Bayesian 'prior' for $\alpha \Rightarrow$ posterior distribution for *D*
- Time dimension?

Deeper Issues

- "Richness" of *Dirichlet* replicates w.r.t generalized Liouville?
- Bayesian ‘prior’ for $\alpha \Rightarrow$ posterior distribution for *D*
- Time dimension?