Statistical Data Fusion

Benjamin Kedem

Department of Mathematics & Inst. for Systems Research
University of Maryland, College Park

"Give me a place to stand and rest my lever on, and I can move the Earth", (Archimedes, 287-212 B.C.)

Waseda University, Tokyo, Japan, March 18, 2019

ABSTRACT

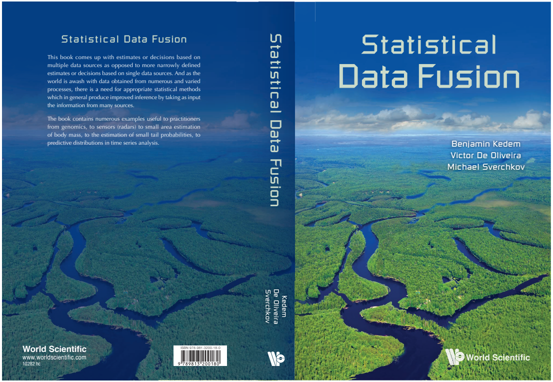
The density ratio model provides an inferential framework for semi-parametric inference vis-a-vis fused data.

a. Meteorological satellite data fused with ground truth.
b. Fused data from several sensors.
c. Fused case and control data.
d. Fused real and computer generated data.

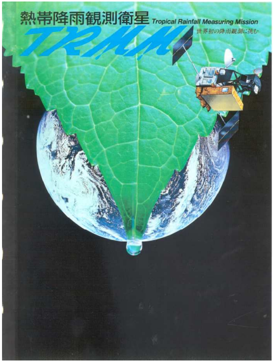
Main points:

- Review of the density ratio model and some of its basic underpinnings.
- Bayesian extension applied to radar data.
- Time series prediction by out of sample fusion.
- Augmented reality: Estimation of small tail probabilities.

Statistical Data Fusion



Motivating Example: Satellite sensors likely distortions of ground truth



Reference: Ground truth

Is there a way to relate the distribution of the satellite data to the distribution of the reference ground truth data?

Much of what we shall be dealing with has to do with this fundamental question.

A possible starting point is a **density ratio** assumption.

A: Review of the Density Ratio Model

Application to Radar Meteorology

Multiple filtering of a signal

$$\begin{aligned} f_1(\omega) &= |H_1(\omega)|^2 f(\omega) \\ &\cdot \\ &\cdot \\ &\cdot \\ f_q(\omega) &= |H_q(\omega)|^2 f(\omega) \end{aligned} \tag{1}$$

That is, q "distortions" or multiple "tilting" of the same *reference* spectral density f.

One-Way ANOVA: Testing Equi-Distribution

$$x_{11}, \quad \ldots \quad , x_{1n_1} \sim g_1(x)$$

$$\vdots$$

$$x_{q1}, \quad \ldots \quad , x_{qn_q} \sim g_q(x)$$

$$x_{m1}, \quad \ldots \quad , x_{mn_m} \sim g_m(x)$$

$$g_j(x) \sim \mathrm{N}(\mu_j, \sigma^2), \quad j = 1, \ldots, m.$$

Then, holding $g_m(x) \equiv g(x)$ as a reference:

$$
\begin{aligned}
g_1(x) &= \exp(\alpha_1 + \beta_1 x) g(x) \\
&\cdot \\
&\cdot \\
g_q(x) &= \exp(\alpha_q + \beta_q x) g(x)
\end{aligned}
$$

$$\alpha_j = \frac{\mu_m^2 - \mu_j^{\mathbf{2}}}{2\,\sigma^2}, \qquad \beta_j = \frac{\mu_j - \mu_{\mathbf{m}}}{\sigma^2}, \qquad j = 1, ..., q$$

Equidistribution testing (FKQS (2001)):

$$\mu_1 = \cdots = \mu_m \Longleftrightarrow \beta_1 = \cdots = \beta_q = 0$$

Multivariate normal

$g_j(\boldsymbol{x}) \sim \mathrm{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$, $j = 1, ..., q, m$. Reference $g_m(\boldsymbol{x}) \equiv g(\boldsymbol{x})$,

$$\frac{g_j(\boldsymbol{x})}{g(\boldsymbol{x})} = \exp[(\boldsymbol{\mu}_j - \boldsymbol{\mu}_m)'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_j'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j - \boldsymbol{\mu}_m'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_m)].$$

$$\alpha_j = -\tfrac{1}{2}(\boldsymbol{\mu}_j'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j - \boldsymbol{\mu}_m'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_m)$$

$$\boldsymbol{\beta}_j = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_m)$$

$$g_j(\boldsymbol{x}) = \exp(\alpha_j + \boldsymbol{\beta}_j'\boldsymbol{x})g(\boldsymbol{x}), \quad j = 1, ...q.$$

$$\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_m \iff \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_q = 0$$

Case-control: Multinomial logistic regression

- RV y s.t. $P(y = j) = \pi_j, \quad \sum_{j=1}^{m} \pi_j = 1.$
- Assume: For $j = 1, ..., m$, and *any* $h(x)$,

$$P(y = j|x) = \frac{\exp(\alpha_j^* + \beta_j h(x))}{1 + \sum_{k=1}^{q} \exp(\alpha_k^* + \beta_k h(x))}$$

- Define: $f(x|y = j) = g_j(x), \quad j = 1, ..., m$

Then with $\alpha_j = \alpha_j^* + \log[\pi_m/\pi_j]$, $j = 1, ..., q$, and $g_m \equiv g$,

Multinomial logistic regression

$$\begin{aligned}
g_1(x) &= \exp(\alpha_1 + \beta_1 h(x))g(x) \\
g_2(x) &= \exp(\alpha_2 + \beta_2 h(x))g(x) \\
&\vdots \\
g_q(x) &= \exp(\alpha_q + \beta_q h(x))g(x)
\end{aligned}$$

Comparison Distributions (Parzen 1977,...,2009)

CDF's: $\{F_1, ..., F_q\} \ll G$, with cont. densities $f_1, ..., f_q, g$.
Comparison Distributions defined as:

$$D_j(u; G, F_j) = F_j(G^{-1}(u)), \quad 0 < u < 1, \ j = 1, ..., q$$

Then by differentiation, with $x = G^{-1}(u)$:

$$\begin{aligned} f_1(x) &= d(G(x); G, F_1)g(x) \\ &\cdot \\ &\cdot \\ f_q(x) &= d(G(x); G, F_q)g(x) \end{aligned}$$

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$\begin{aligned} g_1 &= w_1 g \\ &\vdots \\ g_q &= w_q g \end{aligned}$$

- How can we take advantage of this?
- Assume we have data from each of $g, g_1, g_2, ..., g_q$.
- Then, the relationship between a reference distribution and its distortions or tilts opens the door to inference based on *fused* or *combined* data from many sources.

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$
\begin{aligned}
g_1 &= w_1 g \\
&\vdots \\
g_q &= w_q g
\end{aligned}
$$

- How can we take advantage of this?
- Assume we have data from each of $g, g_1, g_2, ..., g_q$.
- Then, the relationship between a reference distribution and its distortions or tilts opens the door to inference based on *fused* or *combined* data from many sources.

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$
\begin{aligned}
g_1 &= w_1 g \\
&\vdots \\
g_q &= w_q g
\end{aligned}
$$

- How can we take advantage of this?
- Assume we have data from each of $g, g_1, g_2, ..., g_q$.
- Then, the relationship between a reference distribution and its distortions or tilts opens the door to inference based on *fused* or *combined* data from many sources.

- A general structure emerges of a reference behavior (distribution) and its many distortions:

$$\begin{aligned} g_1 &= w_1 g \\ &\vdots \\ g_q &= w_q g \end{aligned}$$

- How can we take advantage of this?
- Assume we have data from each of $g, g_1, g_2, ..., g_q$.
- Then, the relationship between a reference distribution and its distortions or tilts opens the door to inference based on *fused* or *combined* data from many sources.

The previous structure suggests the following general semiparametric problem.

- Multiple data sources: $\boldsymbol{x}_1, ..., \boldsymbol{x}_q, \boldsymbol{x}_m$.
- Data fusion: $\mathbf{t} = (t_1, ..., t_n)' \equiv (\mathbf{x}_1', ..., \mathbf{x}_q', \mathbf{x}_m')'$.
- Fused data length: $n \equiv n_1 + \cdots + n_q + n_m$.
- Assume: $\boldsymbol{x}_j \sim g_j(x), \quad j = 1, ..., q, m$.
- Reference pdf: $g_m(x) = g(x)$.
- Density Ratio Assumption for a **known** $\boldsymbol{h}(x)$:

$$g_j(x) = \exp(\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(x)) g(x), \qquad j = 1, ..., q.$$

Problem

Assume DRM:

$$g_j(x) = \exp(\alpha_j + \boldsymbol{\beta}_j' \boldsymbol{h}(x)) g(x), \qquad j = 1, ..., q.$$

Use the fused data $\mathbf{t} = (t_1, ..., t_n)'$ to:

a. Estimate the reference pdf $g(x)$ and cdf $G(x)$.

b. Estimate $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_q)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', ..., \boldsymbol{\beta}_q')'$.

c. Test distribution equality,

$$\mathrm{H}_0: \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_\mathrm{q} = 0$$

Estimation

Follow Vardi (1982,1985), Qin and Zhang (1997), Owen (2001).

MLE of $G(x)$, β's, α's can be obtained by maximizing the **empirical likelihood** over the class of step cdf's with jumps at the observed values $t_1, ..., t_n$. Accordingly, if $p_i = dG(t_i)$, $i = 1, .., n$:

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\beta}, G) = \prod_{i=1}^{n} p_i \prod_{j=1}^{n_1} \exp(\ \alpha_1 +\ \boldsymbol{\beta}_1' \boldsymbol{h}(x_{1j})) \cdots$$

$$\prod_{j=1}^{n_q} \exp(\ \alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(x_{qj}))$$

1. Get p_i

Fix $\boldsymbol{\alpha}, \boldsymbol{\beta}$. Maximize $\prod_{i=1}^{n} p_i$ subject to the m constraints:

$$\sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i[w_j(t_i) - 1] = 0, \qquad j = 1, ..., q,$$

$$w_j(t_i) = \exp(\ \alpha_j + \ \boldsymbol{\beta}_j' \boldsymbol{h}(t_i)), \ j = 1, ..., q.$$

Use Lagrange multipliers $\lambda_0 = n, \quad \lambda_j = \nu_j n.$

$$(\star) \qquad p_i = \frac{1}{n_m} \cdot \frac{1}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)}$$

$$(\star) \qquad \rho_j = n_j / n_m, \quad j = 1, ..., q.$$

2. Estimate $\boldsymbol{\alpha}, \boldsymbol{\beta}$

Profile log-likelihood up to a constant as a function of $\boldsymbol{\alpha}, \boldsymbol{\beta}$ only:

$$\ell = \sum_{j=1}^{n_1} [\alpha_1 + \boldsymbol{\beta}_1' \boldsymbol{h}(x_{1j})] + \cdots + \sum_{j=1}^{n_q} [\alpha_q + \boldsymbol{\beta}_q' \boldsymbol{h}(x_{qj})]$$
$$- \sum_{i=1}^{n} \log[1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)]$$

Score equations for $j = 1, ..., q$:

$$\frac{\partial \ell}{\partial \alpha_j} = -\sum_{i=1}^{n} \frac{\rho_j w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + n_j = 0$$

$$\frac{\partial \ell}{\partial \beta_j} = -\sum_{i=1}^{n} \frac{\rho_j h(t_i) w_j(t_i)}{1 + \rho_1 w_1(t_i) + \cdots + \rho_q w_q(t_i)} + \sum_{i=1}^{n_j} h(x_{ji}) = 0$$

With

$$\nabla \equiv \left(\frac{\partial}{\partial \alpha_1}, ..., \frac{\partial}{\partial \alpha_m}, \frac{\partial}{\partial \boldsymbol{\beta}_1} ..., \frac{\partial}{\partial \boldsymbol{\beta}_m} \right)'$$

Define the matrices

$$-\frac{1}{n} \nabla \nabla' \ell(\boldsymbol{\theta}) \equiv -\frac{1}{n} \boldsymbol{S}_n \to \boldsymbol{S}, \quad n \to \infty$$

and

$$\boldsymbol{\Lambda} \equiv \mathrm{Var}\left[\frac{1}{\sqrt{n}} \nabla \ell(\boldsymbol{\theta}) \right]$$

Observe that $\boldsymbol{S}_n$ and $\boldsymbol{\Lambda}$ are $(p+1)q \times (p+1)q$ matrices.

Suppose $\boldsymbol{S}$ is positive definite. Then,

(a) The solution $\hat{\boldsymbol{\theta}}$ of the score equations is strongly consistent.

(b) As $n \to \infty$,

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 \\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \end{pmatrix} \xrightarrow{d} \mathrm{N}_{(p+1)q}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2}$$

where $\boldsymbol{\Sigma} = \boldsymbol{S}^{-1} \boldsymbol{\Lambda} \boldsymbol{S}^{-1}$.

3. Estimate $g(x), G(x)$

The solution of the score equations gives the maximum likelihood estimators $\hat{\alpha}, \hat{\beta}$, and consequently by substitution also $\hat{p}_i$. Thus,

$$\hat{p}_i = \frac{1}{n_m} \cdot \frac{1}{1 + \sum_{j=1}^{q} \rho_j \exp(\hat{\alpha}_j + \hat{\beta}_j' \boldsymbol{h}(t_i))}.$$

$$\hat{G}(t) = \sum_{i=1}^{n} \mathrm{I}(t_i \leq t)\hat{p}_i$$

Fokianos (2004):

$$\hat{g}(x) = \mathrm{Kernel}(\hat{p}_i)$$

Everything is estimated from everything

The reference $G(x)$ and all the parameters, and hence all the tilted distributions, are estimated from the entire fused data $\boldsymbol{t}$. Thus $G(x)$ is estimated from the entire fused data $\boldsymbol{t}$ and not just from the reference sample $\boldsymbol{x}_m$.

Semiparametric multivariate kernel density estimation based on many multivariate samples has been studied and applied in cancer research in Voulgaraki, Kedem, Graubard (2012).

Define the following quantities:

$$w_k(t) = \exp(\alpha_k + \beta_k^{'} h(t)),$$

$$A_j(t) = \int \frac{w_j(y) I(y \leq t)}{\sum_{k=0}^{m} \rho_k w_k(y)} dG(y), \; B_j(t) = \int \frac{w_j(y) h(y) I(y \leq t)}{\sum_{k=0}^{m} \rho_k w_k(y)} dG(y),$$

$$\bar{A}(t) = (A_1(t), \ldots, A_m(t))', \quad \bar{B}(t) = (B_1'(t), \ldots, B_m'(t))'.$$

$$\boldsymbol{\rho} = \text{diag}\{\rho_1, \ldots, \rho_m\}, \quad \mathbf{1}_p = (1, ..., 1)',$$

$\rho_j = n_j / n_m$ are sample fractions.

Theorem 1 (Y. Vardi 1982, B. Zhang 2000, G. Lu 2007)

The process $\sqrt{n}(\hat{G}(t) - G(t))$ converges weakly to a zero-mean Gaussian process in $D[-\infty, \infty]$, with covariance matrix given by

$$\begin{aligned} &\mathrm{Cov}\{\sqrt{n}(\hat{G}(t) - G(t)), \sqrt{n}(\hat{G}(s) - G(s))\} = \\ &\quad \sum_{k=0}^{m} \rho_k \left(G(t \wedge s) - G(t)G(s) - \sum_{j=1}^{m} \rho_j A_j(t \wedge s) \right) \\ &\quad + \left(\bar{A}'(s)\boldsymbol{\rho}, \bar{B}'(s)(\boldsymbol{\rho} \otimes \mathbf{1}_p) \right) S^{-1} \begin{pmatrix} \boldsymbol{\rho}\bar{A}(t) \\ (\boldsymbol{\rho} \otimes \mathbf{1}_p)\bar{B}(t) \end{pmatrix}. \end{aligned} \tag{3}$$

Estimation of threshold probabilities

- From Theorem 1, $\sqrt{n}(\hat{G}(t) - G(t))$ converges to a zero-mean Gaussian process.

- Let $\hat{V}(t)$ denote the estimated variance of $\hat{G}(t)$ obtained from the theorem by replacing parameters by their estimates.

- A $1 - \alpha$ level pointwise confidence interval for $G(t)$ is approximated by

$$\left(\hat{G}(t) - z_{\alpha/2}\sqrt{\hat{V}(t)},\ \ \hat{G}(t) + z_{\alpha/2}\sqrt{\hat{V}(t)}\right), \tag{4}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution.

- From (4) we obtain confidence intervals for $p = 1 - G(T)$ for **any** T, including relatively large T, that is, small p.

Under $\mathrm{H}_0 : \boldsymbol{\beta} = (\boldsymbol{\beta}_1^{'}, ..., \boldsymbol{\beta}_q^{'})' = \mathbf{0}$, all the moments are taken with respect to the reference g.

Define a $q \times q$ matrix $\mathbf{A}_{11}$ whose jth diagonal element is

$$\frac{\rho_j [1 + \sum_{k \neq j}^{q} \rho_k]}{[1 + \sum_{k=1}^{q} \rho_k]^2}.$$

For $j \neq j'$, the jj' element is

$$\frac{-\rho_j \rho_{j'}}{[1 + \sum_{k=1}^{q} \rho_k]^2}.$$

The elements are bounded by 1 and the matrix is nonsingular,

$$|\mathbf{A}_{11}| = \frac{\prod_{k=1}^{q} \rho_k}{[1 + \sum_{k=1}^{q} \rho_k]^m} > 0.$$

Under $\mathrm{H}_0 : \boldsymbol{\beta} = (\boldsymbol{\beta}_1^{'}, ..., \boldsymbol{\beta}_{\mathrm{q}}^{'})' = \mathbf{0}$,

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{11} \otimes E[\boldsymbol{h}'(t)] \\ \boldsymbol{A}_{11} \otimes E[\boldsymbol{h}(t)] & \boldsymbol{A}_{11} \otimes E[\boldsymbol{h}(t)\boldsymbol{h}'(t)] \end{pmatrix}$$

and

$$\boldsymbol{V} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{11} \otimes Var[\boldsymbol{h}(t)] \end{pmatrix}$$

$$(\star) \quad \mathcal{X}_1 = n\hat{\beta}'(\boldsymbol{A}_{11} \otimes Var[\boldsymbol{h}(t)])\hat{\beta} \tag{5}$$

$Var[\boldsymbol{h}(t)]$ is the covariance matrix of $\boldsymbol{h}(t)$, and all moments are evaluated with respect to the reference distribution.

$$\mathcal{X}_1 \longrightarrow \chi^2_{(qp)}$$

Application to radar meteorology (KWF 2004)

Reflectivity data obtained from two different radars (or “algorithms” or “sensors”) at two different time periods. Data are random samples of reflectivity.

Kwajalein radar: S-band (10 cm) KPOL radar, located on Kwajalein Island at the southern end of the Kwajalein Atoll, Marshall Islands.

Brown Radar: C-band radar aboard NOAA ship Ronald H. Brown (RHB) at sea near Kwajalein Island.

The data obtained during the first period are referred to suggestively as **Kwajalein1, Brown1**, and those from the second period are called **Kwajalein2, Brown2**.

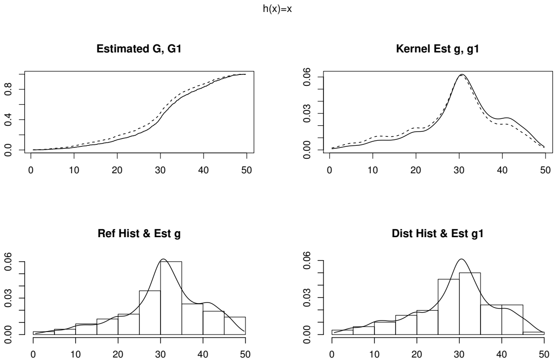
$m = 2$, $n_1 = n_2 = 500$. The hypothesis that the data come from the same radar (algorithm) is **rejected** quite conclusively.

$h(x)$	Data	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\mathcal{X}_1$	p-value
x	1	5.323	-0.164	88.332	0
	2	3.975	-0.123	52.279	4.815e-13
	3	4.695	-0.146	74.950	0
	4	5.016	-0.156	85.325	0
$\log(x)$	1	14.359	-4.142	54.526	1.534e-13
	2	18.625	-5.367	79.723	0
	3	14.880	-4.302	60.788	6.328e-15
	4	13.580	-3.921	49.771	1.727e-12

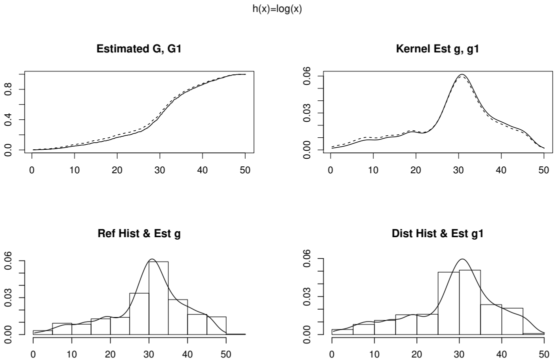
$m = 3$, $n_1 = n_2 = n_3 = 500$. The hypothesis that the data come from the same radar (algorithm) is **accepted** quite conclusively.

Data	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\mathcal{X}_1$	p-value
$h(x) = x$						
1	0.108	0.049	-0.003	-0.002	0.283	0.868
2	0.065	-0.003	-0.002	0.000	0.135	0.935
3	0.227	-0.041	-0.007	0.001	1.896	0.388
4	0.239	-0.220	-0.008	0.007	4.707	0.095
$h(x) = \log x$						
1	0.453	2.278	-0.132	-0.665	1.929	0.381
2	-0.792	-0.223	0.231	0.065	0.250	0.882
3	-0.359	0.735	0.105	-0.215	0.553	0.758
4	1.665	1.246	-0.485	-0.363	1.014	0.602

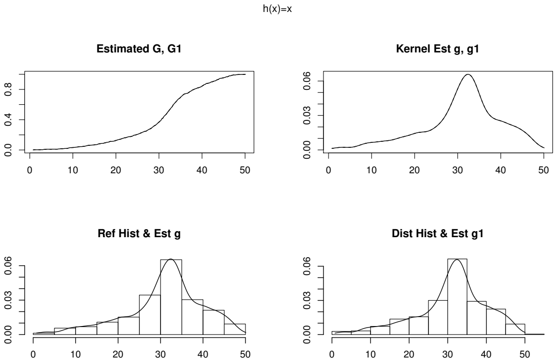
Brown1, Kwajalein1, $h = x$



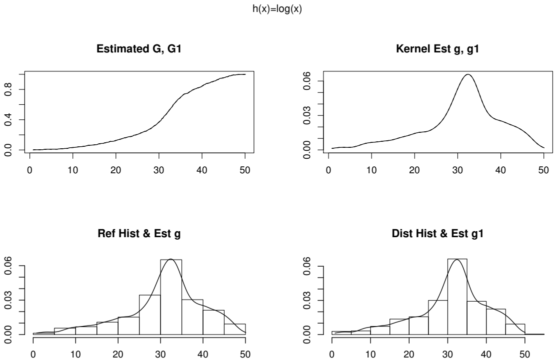
Brown1, Kwajalein1, $h = \log x$



Brown1, Brown1, $h = x$



Brown1, Brown1, $h = \log x$



B: Bayesian Extension (De Oliveira & K 2017)

Application to Radar Meteorology

We have $m = q + 1$ independent random samples following the sampling distributions

$$
\begin{aligned}
x_{11}, x_{12}, \ldots, x_{1n_1} &\overset{\text{iid}}{\sim} G_1(x) \\
x_{21}, x_{22}, \ldots, x_{2n_2} &\overset{\text{iid}}{\sim} G_2(x) \\
&\vdots \\
x_{q1}, x_{q2}, \ldots, x_{qn_q} &\overset{\text{iid}}{\sim} G_q(x) \\
x_{m1}, x_{m2}, \ldots, x_{mn_m} &\overset{\text{iid}}{\sim} G(x),
\end{aligned}
$$

$$\boldsymbol{t} = (t_1, ..., t_n)' \equiv (\boldsymbol{x}_1', ..., \boldsymbol{x}_q', \boldsymbol{x}_m')',$$

$$n = \sum_{j=1}^{q+1} n_j$$

- For Bayesian analysis we use the parametrization (β, G).
- CDF's $G_1, \ldots, G_q$ are distortions of the reference cdf G.
- Density ratio model (DRM):

$$dG_j(x) = \frac{\exp(\beta_j h(x)) dG(x)}{\int_{-\infty}^{\infty} \exp(\beta_j h(u)) dG(u)}, \qquad j = 1, \ldots, q, \tag{6}$$

- Let $A = \{c_1, c_2, \ldots, c_K\}$ be a finite but large set of points in $\mathbb{R}$, chosen to 'approximate' the support of G.

- Consider the 'nonparametric' family of distributions

$$\mathcal{G} = \Big\{ \sum_{k=1}^{K} p_k I(c_k \leq x) : p_k > 0 \text{ for all } k \text{ and } \sum_{k=1}^{K} p_k = 1 \Big\}.$$

- Assume G belongs to $\mathcal{G}$

- Then from (6) follows that

$$G_j(x) = \sum_{k=1}^{K} \Big(\frac{p_k e^{\beta_j h(c_k)}}{\sum_{l=1}^{K} p_l e^{\beta_j h(c_l)}} \Big) I(c_k \leq x), \qquad j = 1, \ldots, q. \tag{7}$$

Specialize: Use order statistics (assuming no ties)

$$A = \{t_{(1)}, t_{(2)}, \ldots, t_{(n)}\}$$

Notation:
$p_k = dG(t_{(k)})$
$\boldsymbol{p}_- = (p_1, \ldots, p_{n-1})', p_n = 1 - \sum_{k=1}^{n-1} p_k$
$\boldsymbol{p} = (\boldsymbol{p}'_-, p_n)'$

Then the DRM parametrized by $(\beta', \boldsymbol{p}'_-)' \in \mathbb{R}^q \times \mathbb{S}^{n-1}$, where

$$\mathbb{S}^{n-1} = \{\boldsymbol{p}_- \in \mathbb{R}^{n-1} : p_k > 0 \text{ for all } k \text{ and } \sum_{k=1}^{n-1} p_k < 1\},$$

is the unit simplex in $\mathbb{R}^{n-1}$.

Likelihood

Then the likelihood function of $(\boldsymbol{\beta}', \boldsymbol{p}'_{-})'$ based on the $q+1$ samples is

$$L(\boldsymbol{\beta}, \boldsymbol{p}_{-}; \boldsymbol{t}) = \prod_{k=1}^{n} p_k \cdot \prod_{i=1}^{n_1} \frac{\exp(\beta_1 h(x_{1i}))}{\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}} \cdots \prod_{i=1}^{n_q} \frac{\exp(\beta_q h(x_{qi}))}{\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}}$$

$$= \frac{\prod_{k=1}^{n} p_k \cdot \exp(\boldsymbol{\beta}' \boldsymbol{h}_{+})}{\left(\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}\right)^{n_1} \cdots \left(\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}\right)^{n_q}} \, I(\boldsymbol{p}_{-} \in \mathbb{S}^{n-1}), \quad (8)$$

where $\boldsymbol{\beta} \in \mathbb{R}^q$, $\boldsymbol{p}_{-} \in \mathbb{S}^{n-1}$, and

$$\boldsymbol{h}_{+} = \left(\sum_{i=1}^{n_1} h(x_{1i}), \ldots, \sum_{i=1}^{n_q} h(x_{qi})\right)'$$

Same as the empirical likelihood when the α_j are expressed in terms of the β_j and $\boldsymbol{p}_{-}$.

Prior

Consider transformations

$$H : \mathbb{S}^{n-1} \longrightarrow \mathbb{R}^{n-1}$$

such that

$$\sum_{k=1}^{n} c_k \log(p_k), \quad \text{with } \textstyle\sum_{k=1}^{n} c_k = 0.$$

Each such transformation is one-to-one and has Jacobian proportional to $\prod_{k=1}^{n} p_k^{-1}$ (O'Hagan, 1994).

The specific case used here is

$$H(\boldsymbol{p}_{-}) = \left(\log\left(\frac{p_1}{p_n}\right), \ldots, \log\left(\frac{p_{n-1}}{p_n}\right) \right)', \tag{9}$$

which was studied by Aitchison and Shen (1980).

Assume for hyperparameters $\boldsymbol{m}_0$ and V_0,

$$H(\boldsymbol{p}_-) \sim \mathrm{N}_{n-1}(\boldsymbol{m}_0, V_0)$$

Then the joint pdf of $\boldsymbol{p}_-$ is the *logistic-normal* distribution:

$$\pi(\boldsymbol{p}_-) \propto \Big(\prod_{k=1}^{n} p_k\Big)^{-1} \exp\big(-\frac{1}{2}(H(\boldsymbol{p}_-)-\boldsymbol{m}_0)' V_0^{-1}(H(\boldsymbol{p}_-)-\boldsymbol{m}_0)\big)\; I(\boldsymbol{p}_- \in \mathbb{S}^{n-1}). \quad (10)$$

Then for any $j, k = 1, \ldots, n-1$

$$E\Big(\frac{p_j}{p_n}\Big) = \exp\big((\boldsymbol{m}_0)_j + \frac{1}{2}(V_0)_{jj}\big), \quad (11)$$

and

$$\mathrm{cov}\Big(\frac{p_j}{p_n}, \frac{p_k}{p_n}\Big) = E\Big(\frac{p_j}{p_n}\Big)E\Big(\frac{p_k}{p_n}\Big)\Big(\exp\big((V_0)_{jk}\big) - 1\Big). \quad (12)$$

- Assume the marginal prior $\boldsymbol{\beta} \sim \mathrm{N}_q(\boldsymbol{b}_0, B_0)$.
- Assume $\boldsymbol{p}_-$ and $\boldsymbol{\beta}$ are independent.
- Then finally the prior $\pi(\boldsymbol{\beta}, \boldsymbol{p}_-)$ is proportional to

$$\Big(\prod_{k=1}^{n} p_k\Big)^{-1} \exp\Big\{ -\frac{1}{2}\Big((\boldsymbol{\beta}-\boldsymbol{b}_0)' B_0^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_0) + (H(\boldsymbol{p}_-)-\boldsymbol{m}_0)' V_0^{-1}(H(\boldsymbol{p}_-)-\boldsymbol{m}_0)\Big)\Big\} \times I(\boldsymbol{p}_- \in \mathbb{S}^{n-1}) \tag{13}$$

Posterior

Then the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{p}_{-} \mid \boldsymbol{t})$ is proportional to

$$\frac{\exp\left\{\boldsymbol{\beta}'\boldsymbol{h}_{+} - \frac{1}{2}\left((\boldsymbol{\beta}-\boldsymbol{b}_0)'B_0^{-1}(\boldsymbol{\beta}-\boldsymbol{b}_0) + (H(\boldsymbol{p}_{-})-\boldsymbol{m}_0)'V_0^{-1}(H(\boldsymbol{p}_{-})-\boldsymbol{m}_0)\right)\right\}}{\left(\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}\right)^{n_1} \cdots \left(\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}\right)^{n_q}} \times I(\boldsymbol{p}_{-} \in \mathbb{S}^{n-1}) \tag{14}$$

- The posterior distribution (14) is quite non-standard. Consequently, Bayesian inference about $(\boldsymbol{\beta}', \boldsymbol{p}'_{-})$ can benefit from the application of Markov chain Monte Carlo (MCMC).

- The underlying idea is to simulate a Markov chain that has an equilibrium distribution which agrees with the posterior distribution of interest.

- To make inference about the model parameters we will use a form of Metropolis-Hasting MCMC algorithm in which the parameters are updated separately in two blocks, $\boldsymbol{\beta}$ and $\boldsymbol{p}_{-}$.

Block 1

- By inspection of (14), the full posterior distributions of $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{p}_{-}, \boldsymbol{t}) \propto \frac{\exp\left(\boldsymbol{\beta}'\boldsymbol{h}_{+} - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{b}_0)' B_0^{-1} (\boldsymbol{\beta} - \boldsymbol{b}_0)\right)}{\left(\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}\right)^{n_1} \cdots \left(\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}\right)^{n_q}}$$

- With tuning constant $c_1 > 0$, simulate a candidate $\boldsymbol{\beta}^*$ using a random-walk with proposal

$$q_1(\boldsymbol{\beta}, \boldsymbol{\beta}^*) \sim \mathrm{N_q}(\boldsymbol{\beta}, \mathrm{c_1 I_q})$$

- Candidate $\boldsymbol{\beta}^*$ is accepted with probability

$$\alpha_1(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = \min\left\{1, \frac{\pi(\boldsymbol{\beta}^* \mid \boldsymbol{p}_{-}, \boldsymbol{t}) q_1(\boldsymbol{\beta}^*, \boldsymbol{\beta})}{\pi(\boldsymbol{\beta} \mid \boldsymbol{p}_{-}, \boldsymbol{t}) q_1(\boldsymbol{\beta}, \boldsymbol{\beta}^*)}\right\} = \min\{1, \xi_1\} \tag{15}$$

If the candidate is not accepted, the next state is set equal to the current state.

Since $q_1(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = q_1(\boldsymbol{\beta}^*, \boldsymbol{\beta})$,

$$\xi_1 = \left(\frac{\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}}{\sum_{l=1}^{n} p_l e^{\beta_1^* h(t_{(l)})}}\right)^{n_1} \cdots \left(\frac{\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}}{\sum_{l=1}^{n} p_l e^{\beta_q^* h(t_{(l)})}}\right)^{n_q}$$

$$\times \exp\left\{(\boldsymbol{\beta}^* - \boldsymbol{\beta})' \boldsymbol{h}_+ - \frac{1}{2}\Big((\boldsymbol{\beta}^* - \boldsymbol{b}_0)' B_0^{-1}(\boldsymbol{\beta}^* - \boldsymbol{b}_0) - (\boldsymbol{\beta} - \boldsymbol{b}_0)' B_0^{-1}(\boldsymbol{\beta} - \boldsymbol{b}_0)\Big)\right\}$$

Block 2

Likewise, it follows from (14) that the full posterior distributions of $\boldsymbol{p}_{-}$ is

$$\pi(\boldsymbol{p}_{-} \mid \boldsymbol{\beta}, \boldsymbol{t}) \propto \frac{\exp\left(-\frac{1}{2}(H(\boldsymbol{p}_{-}) - \boldsymbol{m}_0)' V_0^{-1}(H(\boldsymbol{p}_{-}) - \boldsymbol{m}_0)\right)}{\left(\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}\right)^{n_1} \cdots \left(\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}\right)^{n_q}} \; I(\boldsymbol{p}_{-} \in \mathbb{S}^{n-1}).$$

The candidate for $\boldsymbol{p}_{-}$ is simulated using an independence proposal $q_2(\boldsymbol{p}_{-}, \boldsymbol{p}_{-}^{*})$ being a scaled version of the logistic-normal prior distribution (10) where V_0 is replaced by $c_2 V_0$, where $c_2 > 0$ is a tuning constant. After the candidate $\boldsymbol{p}_{-}^{*}$ is simulated, it is accepted with probability

$$\alpha_2(\boldsymbol{p}_{-}, \boldsymbol{p}_{-}^{*}) = \min\left\{1, \frac{\pi(\boldsymbol{p}_{-}^{*} \mid \boldsymbol{\beta}, \boldsymbol{t}) q_2(\boldsymbol{p}_{-}^{*}, \boldsymbol{p}_{-})}{\pi(\boldsymbol{p}_{-} \mid \boldsymbol{\beta}, \boldsymbol{t}) q_2(\boldsymbol{p}_{-}, \boldsymbol{p}_{-}^{*})}\right\} = \min\{1, \xi_2\}, \qquad (16)$$

$$\xi_2 = \prod_{i=1}^{n} \frac{p_i^*}{p_i} \cdot \left(\frac{\sum_{l=1}^{n} p_l e^{\beta_1 h(t_{(l)})}}{\sum_{l=1}^{n} p_l^* e^{\beta_1 h(t_{(l)})}} \right)^{n_1} \cdots \left(\frac{\sum_{l=1}^{n} p_l e^{\beta_q h(t_{(l)})}}{\sum_{l=1}^{n} p_l^* e^{\beta_q h(t_{(l)})}} \right)^{n_q}$$

$$\times \exp\left\{ \left(\frac{c_2 - 1}{2c_2} \right) \Big((H(\boldsymbol{p}_-) - \boldsymbol{m}_0)' V_0^{-1} (H(\boldsymbol{p}_-) - \boldsymbol{m}_0) - (H(\boldsymbol{p}_-^*) - \boldsymbol{m}_0)' V_0^{-1} (H(\boldsymbol{p}_-^*) - \boldsymbol{m}_0) \Big) \right\}.$$

MCMC algorithm to simulate a Markov chain $\{(\boldsymbol{\beta}^{(m)}, \boldsymbol{p}_{-}^{(m)}) : m = 1, \ldots, M\}$ whose equilibrium distribution is $\pi(\boldsymbol{\beta}, \boldsymbol{p}_{-} \mid \boldsymbol{t})$.

Step 1: Choose the hyperparameters $\boldsymbol{b}_0$, B_0, $\boldsymbol{m}_0$, V_0, the tuning constants c_1, c_2, and the initial state $(\boldsymbol{\beta}^{(0)}, \boldsymbol{p}_{-}^{(0)})$.

For $m = 1, \ldots, M$ do the following:

Step 2: Simulate independently $\boldsymbol{\beta}^* \sim \mathrm{N}_q(\boldsymbol{\beta}^{(m-1)}, c_1 I_q)$ and $U_1 \sim \mathrm{unif}(0, 1)$, and set

$$\boldsymbol{\beta}^{(m)} = \begin{cases} \boldsymbol{\beta}^* & \text{if } U_1 < \alpha_1(\boldsymbol{\beta}^{(m-1)}, \boldsymbol{\beta}^*) \\ \boldsymbol{\beta}^{(m-1)} & \text{otherwise} \end{cases},$$

where $\alpha_1(\cdot, \cdot)$ is given by (15).

Step 3: Simulate independently $\mathbf{W} = (W_1, \ldots, W_{n-1})' \sim \mathrm{N}_{n-1}\left(\boldsymbol{m}_0, c_2 V_0\right)$ and $U_2 \sim \mathrm{unif}(0, 1)$, and compute

$$\boldsymbol{p}_{-}^{*} = (1 + \sum_{i=1}^{n-1} e^{W_i})^{-1} (e^{W_1}, \ldots, e^{W_{n-1}})'.$$

Step 4: Set

$$\boldsymbol{p}_{-}^{(m)} = \begin{cases} \boldsymbol{p}_{-}^{*} & \text{if } U_2 < \alpha_2(\boldsymbol{p}_{-}^{(m-1)}, \boldsymbol{p}_{-}^{*}) \\ \boldsymbol{p}_{-}^{(m-1)} & \text{otherwise} \end{cases},$$

where $\alpha_2(\cdot, \cdot)$ is given by (16), and $p_n^{(m)} = 1 - \mathbf{1}'\boldsymbol{p}_{-}^{(m)}$.

Bayesian Inference

- Once a large sample $\{(\boldsymbol{\beta}^{(m)}, \boldsymbol{p}_{-}^{(m)}) : m = 1, \ldots, M\}$ from the posterior distribution $\pi(\boldsymbol{\beta}, \boldsymbol{p}_{-} \mid \boldsymbol{t})$ is available, Bayesian estimates of the quantities of interest follow easily.

- Point and interval estimates of $\beta_1, \ldots, \beta_q$ are constructed from sample averages and quantiles of the corresponding chains.

- Bayesian estimate of the reference cdf G is given by its posterior expectation

$$\hat{G}^B(x) = E(G \mid \boldsymbol{t}) = \sum_{k=1}^{n} E(p_k \mid \boldsymbol{t}) I(t_{(k)} \leq x),$$

using the approximation computed from the simulated chain

$$E(p_k \mid \boldsymbol{t}) \approx \frac{1}{M} \sum_{m=1}^{M} p_k^{(m)}$$

Bayesian Inference

- Bayesian estimates of the distorted cdfs $G_1, \ldots, G_q$ are given, using (7), by

$$\hat{G}_j^B(x) = E(G_j \mid \boldsymbol{t}) = \sum_{k=1}^{n} E\Big(\frac{p_k e^{\beta_j h(t_{(k)})}}{\sum_{l=1}^{n} p_l e^{\beta_j h(t_{(l)})}} \mid \boldsymbol{t}\Big) I(t_{(k)} \leq x), \qquad j = 1, \ldots, q,$$

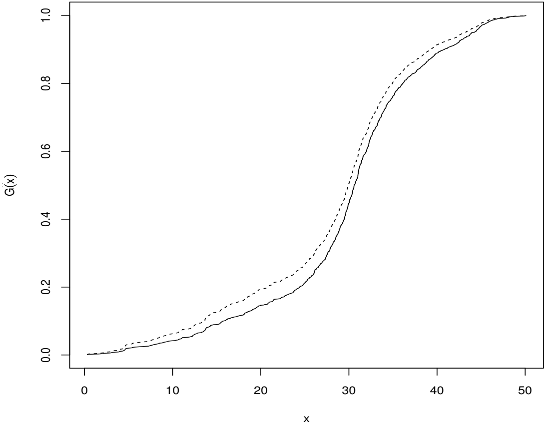
using the approximation computed from the simulated chain

$$E\Big(\frac{p_k e^{\beta_j h(t_{(k)})}}{\sum_{l=1}^{n} p_l e^{\beta_j h(t_{(l)})}} \mid \boldsymbol{t}\Big) \approx \frac{1}{M} \sum_{m=1}^{M} \frac{p_k^{(m)} e^{\beta_j^{(m)} h(t_{(k)})}}{\sum_{l=1}^{n} p_l^{(m)} e^{\beta_j^{(m)} h(t_{(l)})}}.$$

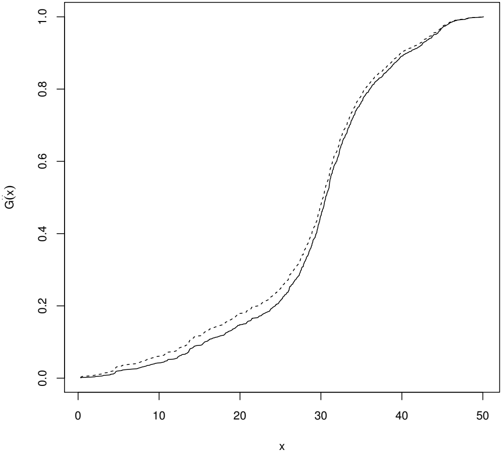
Example: Radar Meteorology

- $q = 1, m = 2, n_1 = n_2 = 500, n = 1000$.
- $\beta_1 \sim N(0, 10)$ ($b_0 = 0$, $v_0 = 10$), independent of
- $H(\boldsymbol{p}_\_) \sim \mathrm{N}_{999}\big(-0.005, (0.01 \times 0.9^{|j-k|})_{jk}\big)$, $H(\cdot)$ is given in (9).
- $h(x) = x$, and $h(x) = \log x$).
- Tuning constants $c_1 = 0.0003$ and $c_2 = 1$.
- $M = 5000$ iterations, burn-in period 500.
- β_1 and $\boldsymbol{p}_\_$ acceptance rates of 0.32 and 0.41, respectively.

$\hat{G}^B(x)$: Brown (solid), Kwajalein (dashed), $h(x) = x$



$\hat{G}^B(x)$: Brown (solid), Kwajalein (dashed), $h(x) = \log x$



Testing Distribution Equality

$$\mathrm{H}_0 : \beta_1 = \cdots = \beta_q = 0$$

- M_0 the Bayesian model under H_0. It has likelihood $L_0(\boldsymbol{p}_-; \boldsymbol{t}) = \prod_{k=1}^{n} p_k \cdot I(\boldsymbol{p}_- \in \mathbb{S}^{n-1})$ and prior $\pi_0(\boldsymbol{p}_-)$ in (10), with $\boldsymbol{p}_- \in \mathbb{S}^{n-1}$.

- M_1 Bayesian model specified by the likelihood $L_1(\boldsymbol{\beta}, \boldsymbol{p}_-; \boldsymbol{t})$ in (8) and prior $\pi_1(\boldsymbol{\beta}, \boldsymbol{p}_-)$ in (13).

- Testing H_0 versus H_1 is then equivalent to choosing between models M_0 and M_1.

- Define Marginal likelihoods $m_0(\boldsymbol{t})$ and $m_1(\boldsymbol{t})$ under M_0 and M_1:

$$m_0(\boldsymbol{t}) = \int_{\mathbb{R}^{n-1}} L_0(\boldsymbol{p}_-; \boldsymbol{t}) \pi_0(\boldsymbol{p}_-) d\boldsymbol{p}_-$$

$$m_1(\boldsymbol{t}) = \int_{\mathbb{R}^q \times \mathbb{R}^{n-1}} L_1(\boldsymbol{\beta}, \boldsymbol{p}_-; \boldsymbol{t}) \pi_1(\boldsymbol{\beta}, \boldsymbol{p}_-) d\boldsymbol{\beta} d\boldsymbol{p}_-.$$

Bayes Factor

• π_0 and $\pi_1 = 1 - \pi_0$ respective prior probabilities of models M_0 and M_1

• Bayes factor: The Bayes factor in favor of M_0 is defined as the ratio of posterior to prior odds of M_0.

$$\begin{aligned} \mathrm{BF}_{01}(\boldsymbol{t}) &= \frac{P(\mathrm{M}_0 \mid \boldsymbol{t})/(1 - P(\mathrm{M}_0 \mid \boldsymbol{t}))}{\pi_0/(1-\pi_0)} \\ &= \frac{m_0(\boldsymbol{t})}{m_1(\boldsymbol{t})} \quad \text{(From Bayes Theorem)}. \end{aligned}$$

$\mathrm{BF}_{01}(\boldsymbol{t})$ is interpreted as the relative evidence in favor of M_0 over M_1. A value of $\mathrm{BF}_{01}(\boldsymbol{t}) > 1$ points to the conclusion the data lend more support to model M_0 than to model M_1.

For the precipitation radar data the and the density ratio model with $h(x) = x$ and $\pi_0 = 1/2$.

$$\mathrm{BF}_{01}(\boldsymbol{t}) = 0.0462 \quad \text{and} \quad P(\mathrm{M}_0 \mid \boldsymbol{t}) \approx 0.044,$$

Therefore, hypothesis H_0 is rejected, and we conclude that the data produced by the Kwajalein and Brown radars come from different distributions. This agrees with the frequentist result.

C: TS Prediction by Out of Sample Fusion (OSF).

Consider the following time series regression model,

$$x_{t+1} = f(\boldsymbol{z}_t) + \epsilon_{t+1}, \quad t = 1, 2, ..., n_0 \tag{17}$$

- $\boldsymbol{z}_t$ contains past values of covariate time series possibly even past values of x_t.
- ϵ_t is an independent noise component.
- We approach time series prediction through the **distribution** of the noise component estimated by **out of sample fusion (OSF)** under a density ratio assumption (K- et al. 2005,2008, K- and Gagnon 2010).

Assume:

- $\epsilon_t \sim G$ for every t.
- $\eta_t,\ t = 1, 2, ...n_1$ is an additional source of data (real or artificial).
- Fuse the ϵ's and η's to get an estimate $\hat{G}$ under a DRM for some tilt function $\boldsymbol{h}$.
- Denote the combined "data" of size $n = n_0 + n_1$ by

$$\boldsymbol{\tau} = (\tau_1, ..., \tau_n) \equiv (\epsilon_1, ..., \epsilon_{n_0}, \eta_1, ..., \eta_{n_1})$$

We obtain the following approximation of the **predictive distribution** at $t+1$ conditional on $\boldsymbol{z}_t$,

$$\begin{aligned} P(x_{t+1} \le x \mid \boldsymbol{z}_t) &= G(x - f(\boldsymbol{z}_t)) \\ &\approx \hat{G}(x - \hat{f}(\boldsymbol{z}_t)) \\ &= \sum_{i=1}^{n} \hat{p}_i I(\tau_i \le x - \hat{f}(\boldsymbol{z}_t)), \end{aligned} \tag{18}$$

where $\hat{G}$ is obtained from the entire fused data $\boldsymbol{\tau}$.

(a) From (18) we can estimate various conditional functions of x_{t+1} given $\boldsymbol{z}_t$ as byproducts.

(b) This procedure is different from methods which use only $n_0 << n$ observations.

Tackling Dependent Residuals

(a) In practice the ϵ_t are replaced by the residuals $\hat{\epsilon}_t$.

(b) Since we are only interested in the distribution of $\hat{\epsilon}_t$, their sequential order is not important.

(c) Hence, we can use randomly shuffled or sampled residuals to induce approximate independence, while maintaining the marginal distribution.

(d) Approximate residual independence may be achieved by using subsequences $\hat{\epsilon}_{t_j}$ where the residuals are spaced sufficiently far apart in time.

(e) Using the raw residuals $\hat{\epsilon}_t$ “as is" can still lead to useful results.

Mortality Prediction

Prediction by out of sample fusion is applied here to sampled filtered total mortality data from Los Angeles County, from 01.01.1970 to 12.31.1979 (Shumway et al. 1988).

The original daily data, consisting of a response series (total mortality) and its covariate series (two weather and six pollution series), were lowpass filtered (removing frequencies above 0.10 cycles per day) and then sampled weekly to produce series of length $N = 508$ each.

Let y, T, CO denote the filtered total mortality, temperature, and carbon monoxide, respectively. A plot of y_t is shown on the next slide displaying a marked oscillation due to filtering.

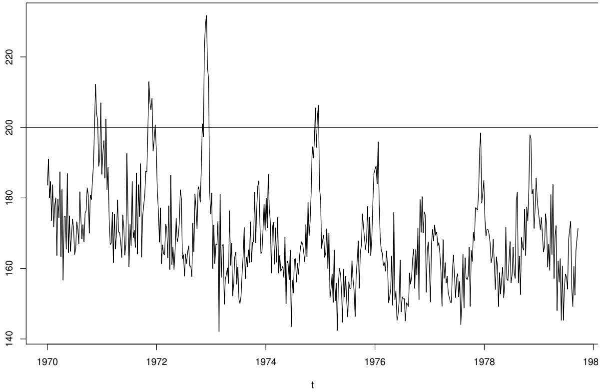


Figure: Filtered weekly mortality, Los Angeles County, 01.01.1970–12.31.1979 (Shumway et al. 1988).

Regression Model for LA Mortality

From K- and Fokianos (2002):

$$y_t = \exp\{\beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 T_{t-1} + \beta_4 \log(CO_t)\} + \hat{\epsilon}_t \tag{19}$$

Partial likelihood estimates:

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4) = (4.5051, 0.0019, 0.0018, -0.0013, 0.0468)$$

Corresponding standard errors:

$$(0.0694, 0.0004, 0.0004, 0.0004, 0.0087)$$

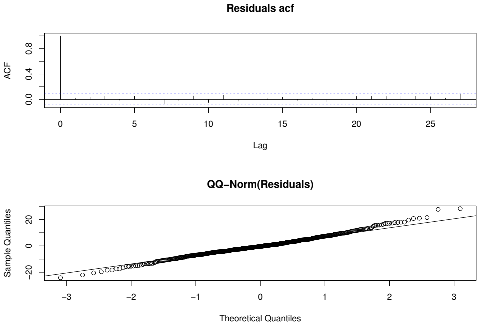


Figure: Estimated autocorrelation and qqnorm plot of $\hat{\epsilon}_t$ from model (19).

The qq-plot suggests fusion of $\hat{\epsilon} \equiv \boldsymbol{x}_0$ with i.i.d. N(mean($\hat{\epsilon}$),Var($\hat{\epsilon}$)):

$$\eta \equiv \boldsymbol{x}_1 \sim \mathrm{N}(0.0016, 59.5)$$

That is: Normal $h(x) = (x, x^2)$ in the DRM:

$$g_1(x) = \exp\{\alpha + \beta_1 x + \beta_2 x^2\} g(x).$$

Check by goodness of fit (GOF): Qin and Zhang (1997) GOF statistic

$$\Delta_n = \sup_t \sqrt{n} |\hat{G}(t) - \tilde{G}(t)|, \tag{20}$$

$\hat{G}(t)$ is the estimated reference CDF from the **fused** $(\hat{\epsilon}, \eta)$.
$\tilde{G}(t)$ is the empirical distribution from the reference sample $\boldsymbol{x}_0$ **only**.

By bootstrapping: $\Delta_n \approx 0.5409109$, p-value of 0.571.

Thus, $h(x) = (x, x^2)$ is sensible.

One-Step Predictive Distribution

With $\hat{G}(t)$ the estimated reference CDF from the **fused** $(\hat{\epsilon}, \eta)$ we have the **preictive distribution** for any level a:

$$P(y_t > a \mid y_{t-1}, y_{t-2}, T_{t-1}, \log(CO_t)) \approx 1 - \hat{G}\left(a - \exp\left\{\hat{\beta}_0 + \hat{\beta}_1 y_{t-1} + \hat{\beta}_2 y_{t-2} + \hat{\beta}_3 T_{t-1} + \hat{\beta}_4 \log(CO_t)\right\}\right). \tag{21}$$

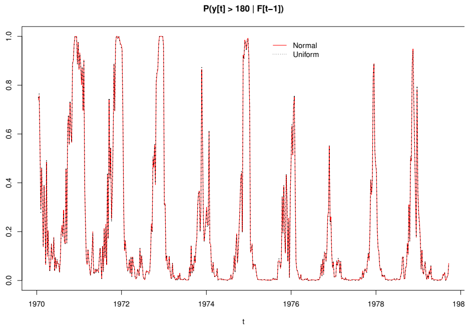


Figure: $P(y_t > 180 \mid y_{t-1}, y_{t-2}, T_{t-1}, \log(CO_t))$

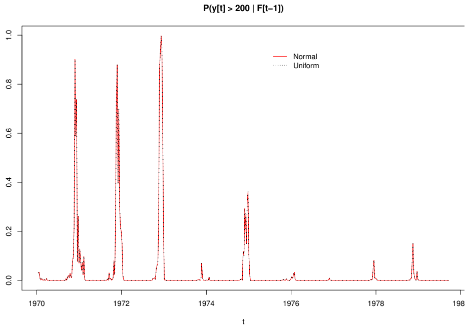


Figure: $P(y_t > 200 \mid y_{t-1}, y_{t-2}, T_{t-1}, \log(CO_t))$

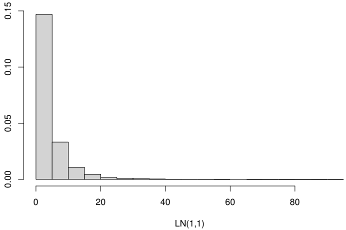
D: Estimation of Small Tail Probabilities.

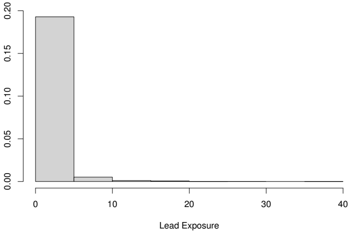
- Repeated Fusion of Real with “Fake" Data (ROSF)
- “Augmented Reality (AR), Better than Real"
  The Economist, Feb. 4, 2017, pp. 67-69.
- We’ll use Theorem 1 with misspecified h.
- B-Curve

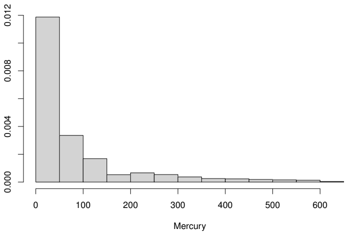
Main Points

- T is a high threshold.
- We wish to estimate a small tail probability: $p = Pr(X > T)$.
- Data (e.g. toxicity) $X_0 = (x_1, ..., x_{n_0})$ below or even far below T.
- ROSF: Fuse X_0 repeatedly with “fake" data X_1.
- With ROSF we get a curve which contains a point whose ordinate is p.
- We show how to “capture" p by Down-Up sequences.
- Comparison with an extreme value theory method POT.
- Background: Density ratio model.

Skewed Data Used







Nearly specified case: Gamma(3,1) Data, $p = 0.01$

$X_0 \sim Gamma(3, 1)$ (somewhat long tail).

$h(x) = (x, \log x)$, $T = 8.405947$, $n_0 = n_1 = 100$.
Fusion: $X_1 \sim Unif(0, 20)$.

Coverage from 100 CI's from Theorem 1: 95%.

Moderately misspecified case: f(2,12) Data, $p = 0.01$

$X_0 \sim f(2, 12)$ (long tail).

$h(x) = (x, \log x)$, $T = 6.926608$, $n_0 = n_1 = 100$.
Fusion: $X_1 \sim Unif(0, 50)$.

Coverage from 100 CI's from Theorem 1: 86%.

Misspecified case: Log-normal Data, $p = 0.01$

$X_0 \sim LN(1,1)$ (long tail).

$h(x) = (x, \log x)$, $T = 27.83649$, $n_0 = n_1 = 100$.
Fusion: $X_1 \sim Unif(0,120)$.

Coverage from 100 CI's from Theorem 1: 70%.

Misspecified case: Inverse-Gaussian Data, $p = 0.01$

$X_0 \sim IG(4, 5)$ (long tail).

$h(x) = (x, \log x)$, $T = 17.87176$, $n_0 = n_1 = 100$.
Fusion: $X_1 \sim Unif(0, 40)$.

Coverage from 100 CI's from Theorem 1: 73%.

We saw that in both the nearly specified and misspecified cases there is a **positive chance the upper bound of the semiparametric CI is above** $p = 0.01$**.**

That is all we need.

This leads to the following formulation.

Estimation of Small Tail Probabilities by ROSF

1. Suppose we wish to estimate a small tail probability $p = P(X > T)$ of some distribution and that we have a reference random sample X_0. $\max(X_0) << T$.
2. Generate a uniform sample X_1 whose support exceeds T.
3. Fuse X_0 with X_1, and get an upper bound B_1 for p from Theorem 1. Use $h = (x, \log x)$.
4. This gives a confidence interval $[0, B_1]$ for p.
5. Repeat many times to get $[0, B_1], [0, B_2], ..., [0, B_n]$.
6. Conditional on X_0, the upper bounds $B_1, B_2, ..., B_n$ are iid.
7. Assume that

$$P(B_1 \geq p) > 0. \tag{22}$$

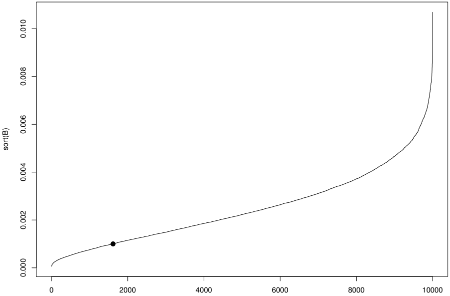
8. As $n \to \infty$, the plot of the ordered sequence $B_{(1)}, B_{(2)}, ..., B_{(n)}$ contains a point whose ordinate is p with probability approaching 1.
9. Call the plot the **B-curve**.

• Any tilt function $\boldsymbol{h}(x)$ which produces upper bounds B_i is appropriate as long as (22) holds.

• Thus, the DRM requirement of a known $\boldsymbol{h}(x)$ can be softened considerably in the present application.

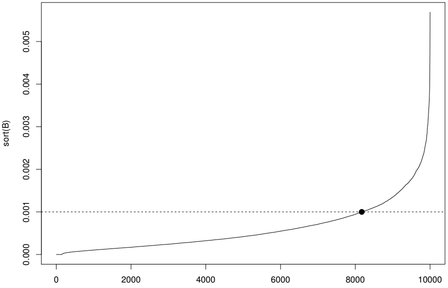
B-Curve $B_{(1)}, \ldots, B_{(10,000)}$, Fused LN(0,1)

$$p = 0.001, n0 = n1 = 100, h(x) = (x, \log x)$$
$$T = 21.98, \max X_0 = 14.46, X_1 \sim \mathrm{Unif}(0, 30)$$



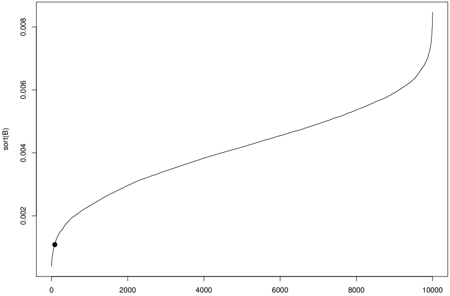
B-Curve $B_{(1)}, \ldots, B_{(10,000)}$, Fused LN(1,1)

$$p = 0.001, n0 = n1 = 100, h(x) = (x, \log x)$$
$$T = 59.75, \max X_0 = 25.17, X_1 \sim \mathrm{Unif}(0, 100)$$



B-Curve $B_{(1)}, \ldots, B_{(10{,}000)}$, Fused Mercury

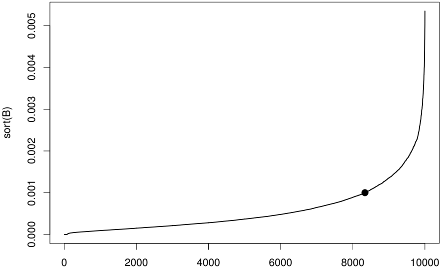
$$p = 0.001, n0 = n1 = 100, h(x) = (x, \log x)$$
$$T = 22.41, \max X_0 = 11.4, X_1 \sim \mathrm{Unif}(1, 50)$$



B-Curve $B_{(1)}, ..., B_{(10,000)}$, Fused $t_{(3)}$

$$p = 0.001, n0 = n1 = 100, h(x) = (x, \log x)$$

$$T = 12.79, \max X_0 = 4.860123, X_1 \sim \mathrm{Unif}(1, 20)$$



• The question then is how to find the point $(j, B_{(j)})$ closest to the point on the B-curve whose ordinate is p.

• That is, how to find $B_{(j)}$ closest to p.

Fact: We can get F_B from many fusions.

- Let $B_1, ..., B_n$ be a random sample of upper bounds from F_B.
- Let $\hat{F}_B$ be the corresponding empirical distribution.
- By Glivenko-Cantelli Theorem

$$\hat{F}_B \longrightarrow F_B$$

almost surely uniformly as n increases.

- Thus, since we may fuse X_0 with as many X_1 as we wish, **we know F_B for all practical purposes** (KPWC, 2016).

• Due to a large number of fusions n, with probability approaching 1

$$B_{(1)} < p < B_{(n)}. \tag{23}$$

• By the monotonicity of the B-curve as j *decreases* (e.g. from $n = 10,000$), the $B_{(j)}$ approach p from above so that there is a $B_{(j)}$ very close to p.

• The B-curve establishes a relationship between j and p approximations.

• From a basic fact about order statistics it is known that

$$P(B_{(j)} > p) = \sum_{k=0}^{j-1} \binom{n}{k} [F_B(p)]^k [1 - F_B(p)]^{n-k}. \tag{24}$$

Recall F_B is known for all practical purposes.

• Therefore, as (24) is monotone decreasing, the *smallest* p which satisfies the inequality

$$\sum_{k=0}^{j-1} \binom{n}{k} [F_B(p)]^k [1 - F_B(p)]^{n-k} \leq 0.95 \tag{25}$$

provides another relationship between j and p.

• Iterating between the two monotone relationships is an iterative method (IM).

• The iteration process starts with a sufficiently large j suggested by the B-curve.

• With that $j \equiv j_1$ we look for the smallest $p \equiv p_{j_1}$ satisfying (25).

• Next, find a $B_{(j_2)}$ on the B-curve closest to p_{j_1}.

• This gives a new $j \equiv j_2$ and the previous steps are repeated until convergence occurs and we keep getting the same p.

• This is a point estimate of the true p obtained from the iterative process. It is not the final estimate.

In symbols:

$$B_{(j_1)} \to p_{(j_1)} \to B_{(j_2)} \to \cdots B_{(j_k)} \to p_{j_k} \to B_{(j_{k+1})} \to p_{j_k} \to B_{(j_{k+1})} \to p_{j_k} \cdots$$

until p_{j_k} keeps giving the same $B_{(j_{k+1})}$

• More succinctly,

$$j_1 \to p_{(j_1)} \to j_2 \to p_{(j_2)} \to \cdots j_k \to p_{j_k} \to j_{k+1} \to p_{j_k} \to j_{k+1} \to p_{j_k} \cdots$$

• *Under some computational conditions this iterative process results in a contraction in a neighborhood of the true p.*

Proposition (K & Wang 2018)

Assume that the samples size n_0 of $\boldsymbol{X}_0$ is large enough, and that the number of fusions n is sufficiently large so that $B_{(1)} < p < B_{(n)}$. Consider the **smallest** $p_j \in (0,1)$ *which satisfies the inequality*

$$P(B_{(j)} > p_j) = \sum_{k=0}^{j-1} \binom{n}{k} [F_B(p_j)]^k [1 - F_B(p_j)]^{n-k} \leq 0.95. \tag{26}$$

Then, iterating between (26) and the corresponding B-curve produces "down" and "up" sequences depending on the $B_{(j)}$ relative to p_j. In particular, in a neighborhood of the true tail probability p, with a high probability, there are "down" sequences which converge from above and "up" sequences which converge from below to points close to p.

Computation

• The iteration process depends on n and the increments of p at which (26) is evaluated.

• Get F_B from a large number of B's, say, 10,000.

• Sample at random 1000 B's to obtain an approximate B-curve.

• The binomial coefficients $\binom{n}{k}$ are replaced by $\binom{1000}{k}$.

• We iterate between an approximate B-curve and approximate (26) with $n = 1000$ until a “down-up" convergence occurs, in which case an estimate $\hat{p}$ for p is obtained.

• This procedure can be repeated many times by sampling repeatedly many different sets of 1000 B's to obtain many point estimates $\hat{p}$ from which interval estimates can then be constructed, as well as variance estimates.

Illustration of ROSF and Iterative Method

Evaluate p along p-increments of order $\mathcal{O}(\bar{B})$, $p = 0.001$, $n_0 = n_1 = 100$, $h(x) = (x, \log x)$. In (26) $n = 1000$.

- **LN(1,1)**, F_B from 10,000 fusions with $X_1 \sim \mathrm{Unif}(0, 80)$. $\bar{B} = 0.00031$.

$1000 \to 0.003 \to 995 \to 0.0024 \to 991 \to 0.002 \to 986 \to 0.0018 \to 977 \to 0.0016 \to 965 \to 0.0014 \to 954 \to 0.0012 \to 941 \to 0.001 \to 923 \to 0.001 \cdots$

- **LN(0,1)**, F_B from 1,000,000 fusions with $X_1 \sim \mathrm{Unif}(0, 40)$. $\bar{B} = 0.000065$.

$1000 \to 0.001 \to 1000 \to 0.001 \to 1000 \to 0.001 \cdots$

- **Positive t(3)**, F_B from 10,000 fusions with $X_1 \sim \mathrm{Unif}(0, 20)$. $\bar{B} = 0.0005744416$.

$1000 \to 0.0038 \to 996 \to 0.003 \to 992 \to 0.0028 \to 986 \to 0.0024 \to 977 \to 0.0022 \to 964 \to 0.0020 \to 956 \to 0.0018 \to 939 \to 0.0016 \to 910 \to 0.0014 \to 882 \to 0.0012 \to 850 \to 0.001 \to 815 \to 0.001 \cdots$

• **Mercury**, F_B from 1,000,000 fusions with $X_1 \sim \mathrm{Unif}(0, 40)$. $\bar{B} = 0.00096$.

1000 $\to$ 0.0052 $\to$ 996 $\to$ 0.0046 $\to$ 991 $\to$ 0.0042 $\to$ 981 $\to$ 0.0038 $\to$ 966 $\to$ 0.0034 $\to$ 949 $\to$ 0.0032 $\to$ 942 $\to$ 0.0030 $\to$ 911 $\to$ 0.0026 $\to$ 895 $\to$ 0.0024 $\to$ 879 $\to$ 0.0022 $\to$ 851 $\to$ 0.0020 $\to$ 829 $\to$ 0.0018 $\to$ 801 $\to$ 0.0016, $\to$ 768 $\to$ 0.0014 $\to$ 732 $\to$ 0.0014 $\cdots$

• **Lead**, F_B from 10,000 fusions with $X_1 \sim \mathrm{Unif}(0, 40)$. p-increment 0.0001.

400 $\to$ 0.0017 $\to$ 371 $\to$ 0.0016 $\to$ 351 $\to$ 0.0015 $\to$ 327 $\to$ 0.0014 $\to$ 302 $\to$ 0.0013 $\to$ 278 $\to$ 0.0012 $\to$ 252 $\to$ 0.0011 $\to$ 229 $\to$ 0.0011 $\cdots$

If we start close to true p=0.001.

Convergence is upward:

$$201 \to 0.001 \to 203 \to 0.001 \cdots$$

Convergence is downward:

$$205 \to 0.001 \to 203 \to 0.001 \cdots$$

Lead Intake: a Higher Probability.

- $p = 0.01$, $T = 10$, $n_0 = n_1 = 100$. $\max(\boldsymbol{X}_0) = 6.875607 < T$, $h(x) = (x, \log x)$.
- F_B from 10,000 fusions with $\boldsymbol{X}_1 \sim \mathrm{Unif}(0, 20)$, $20 > T$.
- Sampling 1000 $B_{(j)}$'s from 10,000 $B_{(j)}$'s, the IM iterative (j, p_j) sequence along p-increments of 0.001 ($\bar{B} = 0.0035$) is:

$$1000 \to 0.01 \to 999 \to 0.009 \to 998 \to 0.009 \cdots$$

so that $\hat{p} = 0.009$.

Illustrations of Down-Up Convergence

Lognormal(1,1)

Table: $\mathbf{p} = \mathbf{0.001}$, $\boldsymbol{X}_0 \sim LN(1, 1)$, $\boldsymbol{X}_1 \sim Unif(0, 80)$, $\max(\boldsymbol{X}_0) = 32.36495$, $T = 59.75377$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
1000	0.001199466	21	Down
950	0.001099466	13	Down
900	0.000999465	10	Down
800	0.000999465	5	Down
750	0.000999465	3	Down
700	0.000999465	2	Down
680	0.000999465	2	Up
680	0.000999465	2	Up
670	0.000999465	2	Up

A sensible estimate of $p = 0.001$ is the average from the last 6 entries which gives $\hat{p} = 0.000999465$ with absolute error of 5.35×10^{-07}.

Lognormal(1,1)

Table: $\mathbf{p} = \mathbf{0.0001}$, $\boldsymbol{X}_0 \sim LN(1,1)$, $\boldsymbol{X}_1 \sim Unif(0,130)$, $\max(\boldsymbol{X}_0) = 44.82807$, $T = 112.058$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.000015.

Starting j	Convergence to	Iterations	
800	0.0001945544	23	Down
500	0.0001795544	10	Down
300	0.0001345544	5	Down
200	0.0001195544	2	Down
170	0.0001045544	2	Down
160	0.0001045544	2	Down
155	0.0001045544	2	Up
152	0.0001045544	2	Up
150	0.0001045544	2	Up

A sensible estimate of $p = 0.0001$ is the average from the last 5 entries which gives $\hat{p} = 0.0001045544$ with absolute error of 4.5544×10^{-06}.

Lognormal(0,1)

Table: $\mathbf{p} = \mathbf{0.001}$, $\boldsymbol{X}_0 \sim LN(0,1)$, $\boldsymbol{X}_1 \sim Unif(0,50)$, $\max(\boldsymbol{X}_0) = 11.86797$, $T = 21.98218$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
1000	0.001099445	19	Down
900	0.001099445	5	Down
820	0.001099445	2	Down
800	0.000999444	3	Down
790	0.000999444	2	Down
780	0.000999444	2	Up
770	0.000999444	2	Up
760	0.001099445	4	Up

A sensible estimate of $p = 0.001$ is the average from the last 5 entries which gives $\hat{p} = 0.001019444$ with absolute error of 1.9444×10^{-05}.

Lognormal(0,1)

Table: $\mathbf{p} = \mathbf{0.0001}$, $\boldsymbol{X}_0 \sim LN(0,1)$, $\boldsymbol{X}_1 \sim Unif(0,70)$, $\max(\boldsymbol{X}_0) = 13.77121$, $T = 41.22383$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.000015.

Starting j	Convergence to	Iterations	
900	0.0002392241	28	Down
800	0.0001042241	25	Down
700	0.0001042241	18	Down
500	0.0001192241	6	Down
360	0.0001042241	2	Down
355	0.0001042241	2	Up
350	0.0001042241	2	Up
350	0.0001042241	2	Up

A sensible estimate of $p = 0.0001$ is the average from the last 4 entries which gives $\hat{p} = 0.0001042241$ with absolute error of 4.2241×10^{-06}.

f(2,7)

Table: $\mathbf{p} = \mathbf{0.001}$, $\boldsymbol{X}_0 \sim f(2,7)$, $\boldsymbol{X}_1 \sim Unif(0,50)$, $\max(\boldsymbol{X}_0) = 12.25072$, $T = 21.689$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
500	0.001103351	10	Down
450	0.001003351	9	Down
400	0.001003351	7	Down
300	0.001003351	4	Down
210	0.001003351	2	Up
190	0.000903350	2	Up
180	0.000903350	2	Up

A sensible estimate of $p = 0.001$ occurs at the down-up shift which gives $\hat{p} = 0.001003351$ with absolute error of 3.351×10^{-06}.

f(2,7)

Table: $\mathbf{p} = \mathbf{0.0001}$, $\boldsymbol{X}_0 \sim f(2,7)$, $\boldsymbol{X}_1 \sim Unif(0,70)$, $\max(\boldsymbol{X}_0) = 14.62357$, $T = 45.13234$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.000015.

Starting j	Convergence to	Iterations	
750	0.0001341104	3	Down
740	0.0001041104	5	Down
730	0.0001041104	4	Down
700	0.0001341104	3	Up
660	0.0001041104	2	Down
650	0.0001041104	2	Up
645	0.0001041104	2	Up
640	0.0001041104	3	Up

A sensible estimate of $p = 0.0001$ occurs at the down-up shift which gives $\hat{p} = 0.0001041104$ with absolute error of 4.1104×10^{-06}.

Weibull(0.8,2)

Table: $\mathbf{p} = \mathbf{0.001}$, $\boldsymbol{X}_0 \sim \textit{Weibull}(0.8, 2)$, $\boldsymbol{X}_1 \sim \textit{Unif}(0, 40)$, $\max(\boldsymbol{X}_0) = 8.081707$, $T = 22.39758$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
1000	0.001899263	3	Down
1000	0.001099263	8	Down
950	0.000999262	2	Immediate
950	0.000999262	2	Up
940	0.001099263	4	Up
940	0.000999262	3	Up

In the 3rd entry there was an immediate convergence. A sensible estimate of $p = 0.001$ is the average from the last 5 entries which gives $\hat{p} = 0.001039261$ with absolute error of 3.9261×10^{-05}.

Weibull(0.8,2)

Table: $\mathbf{p} = \mathbf{0.0001}$, $\boldsymbol{X}_0 \sim Weibull(0.8, 2)$, $\boldsymbol{X}_1 \sim Unif(0, 50)$, $\max(\boldsymbol{X}_0) = 12.20032$, $T = 32.09036$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.000015.

Starting j	Convergence to	Iterations	
700	0.0002096393	21	Down
400	0.0001196393	11	Down
300	0.0001946393	2	Down
200	0.0001046393	5	Down
130	0.0001046393	2	Down
125	0.0001046393	2	Up
120	0.0001046393	2	Up
115	0.0001046393	2	Up

A sensible estimate of $p = 0.0001$ is the average from the last 5 entries which gives $\hat{p} = 0.0001046393$ with absolute error of 4.6393×10^{-06}.

NHANES: URX3TB Trichlorophenol

2604 observations of which the proportion exceeding $T = 9.5$ is $p = 0.001152074$.

The 3rd quartile from 10,000 B's is 0.001225: Reasonable guess of p.

Table: **p** $=$ **0.001152074**, $\boldsymbol{X}_0$ a trichlorophenol sample. $\boldsymbol{X}_1 \sim Unif(0, 30)$, $\max(\boldsymbol{X}_0) = 3$, $T = 9.5$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
840	0.001099096	8	Down
800	0.000999095	7	Down
760	0.000999095	4	Down
755	0.001099096	2	Down
750	0.001099096	2	Up
740	0.000999095	2	Up
735	0.000999095	2	Up
732	0.001099096	4	Up

The 8 estimates in Table 9 with $\max(\boldsymbol{X}_0) = 3$ seem to be in a neighborhood of the true $p = 0.001152074$. Their average is $0.001049096 \approx p$ with standard deviation of $0.5345278 \times 10^{-05}$.

NOAA: Mercury (mg/kg)

8,266 observations. Proportion exceeding $T = 22.41$ is $p = 0.001088797$.

Table: $\mathbf{p = 0.001088797}$, $\boldsymbol{X}_0$ a mercury sample. $\boldsymbol{X}_1 \sim Unif(0, 50)$, $\max(\boldsymbol{X}_0) = 7.99$, $T = 22.41$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
800	0.001099352	14	Down
700	0.001199352	8	Down
600	0.000999351	5	Down
500	0.000999351	2	Down
490	0.000999351	2	Up
480	0.000999351	2	Up
470	0.000999351	2	Up

Table: Do again with different mercury sample $\boldsymbol{X}_0$. $\boldsymbol{X}_1 \sim Unif(0, 50)$, $\max(\boldsymbol{X}_0) = 11.9$, $T = 22.41$, $n_0 = n_1 = 100$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
800	0.001199501	15	Down
700	0.001199501	12	Down
500	0.001199501	6	Down
400	0.001099501	2	Down
390	0.001099501	2	Up
380	0.001099501	2	Up
375	0.001199501	3	Up
360	0.001099501	3	Up

Mercury Larger Sample

NOAA: Mercury (mg/kg)

8,266 observations. Proportion exceeding $T = 22.41$ is $p = 0.001088797$.

Table: $\mathbf{p} = \mathbf{0.001088797}$, $\boldsymbol{X}_0$ a mercury sample. $\boldsymbol{X}_1 \sim \mathit{Unif}(0, 50)$, $\max(\boldsymbol{X}_0) = 13.8$, $T = 22.41$, $n_0 = n_1 = 200$, $h = (x, \log x)$, p-increment 0.0001.

Starting j	Convergence to	Iterations	
775	0.002792137	18	Down
600	0.002092137	16	Down
300	0.001492137	9	Down
200	0.001192137	7	Down
100	0.001192137	2	Down
90	0.001192137	2	Up
85	0.001092137	2	Down
84	0.001092137	2	Up
83	0.001092137	2	Up
81	0.001092137	2	Up
80	0.001092137	2	Up

A sensible estimate of $p = 0.00108879$ is $\hat{p} = 0.001092137$ with absolute error of 3.347×10^{-6}.

Much Smaller $p = 0.00001$

LN(1,1)

Table: $X_0 \sim \mathbf{LN}(1,1) : p = 1 - G(T) = 0.00001$, $\max(\boldsymbol{X}_0) = 56.53902$, $T = 193.4252$, $X_1 \sim$ Unif(0,250), $n_0 = n_1 = 500$, $h(x) = (x, \log x)$. p-increment 0.000001.

Starting j	Convergence to	Iterations	
950	0.0000140213	12	Down
900	0.0000108643	8	Down
850	0.0000108643	4	Down
800	0.0000108643	1	Down
770	0.0000105312	1	Up
760	0.0000105312	2	Up

Variability of Point Estimates

• For example: $p = 0.001$.

• Take different B-samples of size 1,000 taken from, say, 10,000 $B's$, to produce tail probability estimates as above from which variance approximations can be obtained.

• With $n_0 = n_1 = 100$ and $n_0 = n_1 = 200$, in all cases $\sigma_{\hat{p}} = O(10^{-4})$.

ROSF/IM vs POT (K & Wang 2018)

- F_B from 1000 fusions.
- Starting $B_{(j)}$ approx 3rd Quartile of observed 1000 B_i.
- From ROSF/IM we get N $\hat{p}$'s and construct CI for p as $(\min(\hat{p}), \max(\hat{p}))$.
- Mean absolute error (MAE) from 500 runs: $\sum(|\hat{p}_i - p|)/500$.

Table: $X_0 \sim \mathbf{t}_{(1)} > 0 : p = 1 - G(T) = 0.001$, $T = 631.8645$, $X_1 \sim$ Unif(0,800), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.0001.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	63.2%	0.00372	0.00149	72.1%	0.00292	0.00122
ROSF & IM	**50**	98.2%	0.00213	0.00061	100%	0.00193	0.00051
	100	100%	0.00264	-	100%	0.00241	-

Table: $X_0 \sim$ **Pareto**$(1, 4)$: $p = 1 - G(T) = 0.001$, $T = 5.623413$,
$X_1 \sim$ Unif(1,8), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.0001.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	81.8%	0.00419	0.00121	84.5%	0.00337	0.00070
ROSF/IM	50	96.2%	0.00232	0.00052	97.8%	0.00231	0.00041
	100	100%	0.00272	-	100%	0.00269	-

Table: $X_0 \sim$ **IG**$(2, 40)$: $p = 1 - G(T) = 0.001$, $T = 3.835791$,
$X_1 \sim$ Unif(0,8), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.00005.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	69.6%	0.00324	0.00123	82.3%	0.00316	0.00092
ROSF/IM	50	100%	0.00289	0.00047	100%	0.00206	0.00041
	100	100%	0.00332	-	100%	0.00313	-

Table: $X_0 \sim$ Mercury : $p = 1 - G(T) = 0.001$, $T = 22.41$, $X_1 \sim$ Unif(0,50), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.0001.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	85.3%	0.00455	0.00130	88.6%	0.00398	0.00122
ROSF/IM	50	97.5%	0.00215	0.00048	100%	0.00197	0.00045
	100	100%	0.00259	-	100%	0.00238	-

Table: $X_0 \sim$ Lead Intake : $p = 1 - G(T) = 0.001$, $T = 25$, $X_1 \sim$ Unif(0,30), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.0001.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	84.7%	0.00555	0.00142	87.7%	0.00536	0.00125
ROSF/IM	50	100%	0.00247	0.00066	100%	0.00229	0.00058
	100	100%	0.00289	-	100%	0.00268	

Table: $X_0 \sim \mathbf{F}(2,12) : p = 1 - G(T) = 0.0001$, $T = 21.84953$, $X_1 \sim$ Unif(0,25), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.00001.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	71.4%	0.00062	0.00052	81.6%	0.00053	0.000045
ROSF/IM	50	95.2%	0.00059	0.00022	96.3%	0.00052	0.000019
	100	100%	0.00082	-	100%	0.00069	-

Table: $X_0 \sim$ Mercury : $p = 1 - G(T) = 0.0001$, $T = 39.60$, $X_1 \sim$ Unif(0,80), $n_0 = n_1$, $h(x) = (x, \log x)$. p-increment 0.00001.

		$n_0 = 100$			$n_0 = 200$		
Method	N	Coverage	CI Length	MAE	Coverage	CI Length	MAE
POT	-	62.4%	0.00059	0.00049	73.4%	0.00051	0.000042
ROSF/IM	50	95.2%	0.00056	0.00023	100%	0.00054	0.000019
	100	100%	0.00083	-	100%	0.00079	-