Supplementary material for the article "Inference on a distribution function from ranked set samples" (Lutz Dümbgen and Ehsan Zamanzade)

A Further proofs and technical details

Proof of Lemma 1. Continuity of $L_n(x,\cdot) : [0,1] \to [-\infty, 0]$ follows essentially from continuity of $\log : [0,1] \to [-\infty, 0]$. For $p \in (0,1)$,

$$L_n'(x,p) \ = \ \sum_{r=1}^k N_{nr} \Bigl[\frac{\beta_r}{B_r}(p) \widehat{F}_{nr}(x) - \frac{\beta_r}{1 - B_r}(p)(1 - \widehat{F}_{nr}(x)) \Bigr].$$

It follows from the formula $B_r(p) = \sum_{i=r}^k \binom{k}{i} p^i (1-p)^{k-i}$ that

$$\frac{\beta_r}{B_r}(p) \ = \ C_r \Big/ \sum_{i=r}^k \binom{k}{i} p^{i+1-r} (1-p)^{r-i}$$

and

$$\frac{\beta_r}{1 - B_r}(p) \ = \ C_r \Big/ \sum_{i=0}^{r-1} \binom{k}{i} p^{i+1-r} (1-p)^{r-i}$$

are strictly decreasing and strictly increasing in $p \in (0,1)$, respectively. Consequently, the derivative $L_n'(x,\cdot)$ is continuous and strictly decreasing on $(0,1)$.

Elementary algebra yields the alternative formula

$$L_n'(x,p) \ = \ \sum_{r=1}^k N_{nr} w_r(p) \bigl[\widehat{F}_{nr}(x) - B_r(p) \bigr]$$

with the auxiliary function

$$w_r(p) \ = \ \frac{\beta_r}{B_r(1 - B_r)}(p) \ = \ \frac{\beta_r(p)}{B_r(p) B_{k+1-r}(1-p)}.$$

The latter equation follows from the relation $1 - B_r(p) = B_{k+1-r}(1-p)$ and is highly recommended to avoid rounding errors in case of p being close to 1. Note also that

$$w_r(p) \ = \ \begin{cases} \dfrac{r + o(1)}{p} & \text{as } p \to 0, \\ \dfrac{k + 1 - r + o(1)}{1 - p} & \text{as } p \to 1. \end{cases} \tag{10}$$

This implies that the limits of $L_n'(x,\cdot)$ at the boundary of $(0,1)$ satisfy

$$\begin{aligned} L_n'(x,0) \ &= \ +\infty \quad \text{if } x \ge X_{(1)}, \\ L_n'(x,1) \ &= \ -\infty \quad \text{if } x < X_{(n)}, \end{aligned}$$

because $x \ge X_{(1)}$ implies that $\widehat{F}_{nr}(x) > 0 = B_r(0)$ for at least one r, while $x < X_{(n)}$ implies that $\widehat{F}_{nr} < 1 = B_r(1)$ for at least one r. $\square$

Proof of Lemma 7. As to part (a), note that w_r is a rational and strictly positive function on $(0,1)$. Hence $\widetilde{w}_r(t) := t(1-t)w_r(t)$ defines a function with these properties, too. Moreover, it follows from (10) that $\lim_{t\downarrow 0}\widetilde{w}_r(t) = r$ and $\lim_{t\uparrow 1}\widetilde{w}_r(t) = k-r+1$. Hence $\widetilde{w}_r$ may be viewed as a rational and strictly positive function on a neighborhood of $[0,1]$. In particular, $\widetilde{w}_k$ is continuously differentiable on $[0,1]$.

It remains to show that $1 \le \widetilde{w}_r \le \max(r, k+1-r)$ on $[0,1]$. The upper bound follows from the fact that for $0 < t < 1$,

$$\begin{aligned}
\widetilde{w}_r(t) &= \frac{t(1-t)\beta_r(t)}{B_r(t)} + \frac{t(1-t)\beta_r(t)}{B_{k-r+1}(1-t)} \\
&= \frac{t^r(1-t)^{k-r+1}}{\int_0^t u^{r-1}(1-u)^{k-r}\,du} + \frac{t^r(1-t)^{k-r+1}}{\int_0^{1-t} u^{k-r}(1-u)^{r-1}\,du} \\
&\le \frac{t^r(1-t)^{k-r+1}}{\int_0^t u^{r-1}\,du\,(1-t)^{k-r}} + \frac{t^r(1-t)^{k-r+1}}{\int_0^{1-t} u^{k-r}\,du\,t^{r-1}} \\
&= (1-t)\,r + t\,(k-r+1) \\
&\le \max(r, k-r+1).
\end{aligned}$$

The lower bound is equivalent to the claim that $\beta_r(t) \ge B_r(t)(1-B_r(t))/(t(1-t))$ for any $t \in (0,1)$. Since $\log\beta_r(u) = \log C_r + (r-1)\log u + (k-r)\log(1-u)$ is concave in $u \in (0,1)$, this assertion follows from Lemma 8 below.

For proving part (b), note first that $|p-t| \le ct(1-t)$ implies the inequalities $p \le (1+c)t$ and $1-p \le (1+c)(1-t)$. Moreover, since $\big|p(1-p) - t(1-t)\big| \le |p-t|$, we may conclude that $p(1-p) \ge (1-c)t(1-t)$. Consequently,

$$\begin{aligned}
\Big|\frac{w_r(p)}{w_r(t)} - 1\Big| &= \frac{\big|\widetilde{w}_r(p)t(1-t) - \widetilde{w}_r(t)p(1-p)\big|}{\widetilde{w}_r(t)p(1-p)} \\
&\le \frac{\big|\widetilde{w}_r(p) - \widetilde{w}_r(t)\big|t(1-t) + \widetilde{w}_r(t)\big|t(1-t) - p(1-p)\big|}{\widetilde{w}_r(t)p(1-p)} \\
&\le \frac{\big|\widetilde{w}_r(p) - \widetilde{w}_r(t)\big|/4 + C_w|t-p|}{c_w(1-c)t(1-t)} \\
&\le \frac{c_w'/4 + C_w}{c_w(1-c)}\,\frac{|p-t|}{t(1-t)},
\end{aligned}$$

where $c_w' := \max_{1\le r\le k, u\in[0,1]}|\widetilde{w}_r'(u)|$. Moreover, for $\min(t,p) \le \xi \le \max(t,p)$,

$$\frac{|\beta_r'(\xi)|}{\beta_r(\xi)} = \frac{|r-1-(k-1)\xi|}{\xi(1-\xi)} \le \frac{k-1}{(1-c)t(1-t)} \quad \text{and} \quad \frac{\beta_r(\xi)}{\beta_r(t)} \le (1+c)^{k-1}.$$

Hence Taylor's formula shows that for a suitable such ξ,

$$\Big|\frac{B_r(p) - B_r(t)}{\beta_r(t)(p-t)} - 1\Big| = \frac{|\beta_r'(\xi)||p-t|}{2\beta_r(t)} \le \frac{(k-1)(1+c)^{k-1}}{c-1}\,\frac{|p-t|}{t(1-t)}. \qquad \square$$

In the proof of Lemma 7 we referred to the following general inequality which is possibly of independent interest:

Lemma 8. *Let β be a strictly positive probability density on $(0,1)$ such that $\log\beta$ is concave. Then its distribution function $B : [0,1] \to [0,1]$ satisfies the following inequalities: For any $t \in (0,1)$,*

$$\beta(t) \ \geq \ \frac{B(t)(1-B(t))}{t(1-t)}$$

with equality if, and only if, $\beta \equiv 1$.

Proof of Lemma 8. For $a \in \mathbb{R}$ let $G_a : [0,1] \to [0,1]$ be the distribution function given by

$$G_a(x) \ := \ \begin{cases} (e^{ax}-1)/(e^a-1) & \text{if } a \neq 0, \\ x & \text{if } a = 0. \end{cases}$$

Then G_a has log-linear density

$$g_a(x) := G_a'(x) \ = \ e^{ax-c(a)}$$

with $c(0) = 0$ and $c(a) = \log((e^a-1)/a)$ for $a \neq 0$. For fixed $t \in (0,1)$, $G_a(t)$ is continuous in $a \in \mathbb{R}$ with $\lim_{a\geq\infty} G_a(t) = 0$ and $\lim_{a\to-\infty} G_a(t) = 1$. Hence for a suitable $a = a(t) \in \mathbb{R}$,

$$B(t) \ = \ G_a(t).$$

If we fix this value a, then the previous equality implies that $\beta(s) \geq g_a(s)$ for some $s \in (0,t)$ and $\beta(u) \geq g_a(u)$ for some $u \in (t,1)$. But then concavity of $\log\beta$ and linearity of $\log g_a$ yield the inequality $\beta(t) \geq g_a(t)$. Moreover, if $\beta(t) = g_a(t)$, then $\beta \leq g_a$, and this implies that $\beta \equiv g_a$. Hence it suffices to prove the claim in case of $\beta \equiv g_a$ for some $a \in \mathbb{R}$.

Since $g_0 \equiv 1$ and $G_0(t) = t$, the asserted inequality is an equality in case of $a = 0$. Hence it remains to show that $G_a(t)(1-G_a(t)) < t(1-t)g_a(t)$ in case of $a \neq 0$. Indeed,

$$\begin{aligned} \frac{G_a(t)(1-G_a(t))}{t(1-t)g_a(t)} \ &= \ \frac{(e^{at}-1)(e^a-e^{at})}{t(1-t)e^{at}a(e^a-1)} \\ &= \ \frac{e^{at}-1}{at}\cdot\frac{e^{a(1-t)}-1}{a(1-t)}\Big/\frac{e^a-1}{a} \ = \ \exp\bigl(h(at)+h(a-at)-h(a)\bigr), \end{aligned}$$

where $h(x) := \log((e^x-1)/x)$ for $x \neq 0$. In case of $a > 0$ it follows from $\lim_{x\to0} h(x) = 0$ that

$$h(at)+h(a(1-t))-h(a) \ = \ \int_0^{at} \bigl(h'(u)-h'(a(1-t)+u)\bigr)\,du \ < \ 0,$$

because $h''(x) = x^{-2} - (e^x+e^{-x}-2)^{-1} > 0$, so h' is strictly increasing. In case of $a < 0$, it follows from $h(x) = x + h(-x)$ that

$$h(at)+h(a(1-t))-h(a) \ = \ h(|a|t)+h(|a|(1-t))-h(|a|) \ < \ 0$$

as well. □

Details about asymptotic variances and the function ρ in case of $k = 2$. In the special case $k = 2$, elementary calculations reveal that

$$\begin{aligned}
\beta_1(t) &= 1-u, \quad B_1(1-B_1)(t) = K(t)\frac{3-4u+u^2}{4}, \quad w_1(t) = \frac{4}{K(t)(3-u)},\\
\beta_2(t) &= 1+u, \quad B_2(1-B_2)(t) = K(t)\frac{3+4u+u^2}{4}, \quad w_2(t) = \frac{2}{K(t)(3+u)},
\end{aligned}$$

where $u := 2t-1 \in [-1,1]$ and $K(t) := K(t,t) = t(1-t)$. In particular,

$$\widetilde{w}_1(t) = \frac{4}{3-u}, \quad \widetilde{w}_2(t) = \frac{4}{3+u} \quad \text{and} \quad \frac{\rho(t)+\rho(t)^{-1}+2}{4} = \frac{9}{9-u^2}.$$

Moreover, with $\Delta := \pi_2 - \pi_1$ these formulae entail that

$$\begin{aligned}
K^{\mathrm{S}}(t) &= \frac{K(t)}{4}\frac{3+u^2-4u\Delta}{(1-\Delta^2)},\\
K^{\mathrm{M}}(t) &= \frac{K(t)}{4}\frac{3+u^2+4u\Delta}{(1+u\Delta)^2},\\
K^{\mathrm{L}}(t) &= \frac{K(t)}{4}\frac{9-u^2}{3-u^2+2u\Delta}.
\end{aligned}$$

The top left panel in Figure 8 shows for $\pi_1 = \pi_2 = 1/2$ the asymptotic variances $K^{\mathrm{M}}(t) = K^{\mathrm{S}}(t) > K^{\mathrm{L}}(t)$ as well as the variances $K(t)$ for simple random sampling. In the top right and lower panels one sees for $\pi_1 = 1 - \pi_2 = 1/2, 5/8, 3/4$ the relative asymptotic efficiencies $E^{\mathrm{M}}(t) = K^{\mathrm{M}}(t)/K^{\mathrm{L}}(t)$ and $E^{\mathrm{S}}(t) = K^{\mathrm{S}}(t)/K^{\mathrm{L}}(t)$ of $\widehat{B}_n^{\mathrm{L}}$ with respect to $\widehat{B}_n^{\mathrm{M}}$ and $\widehat{B}_n^{\mathrm{S}}$, respectively. In each panel the gray dotted line depicts the upper bound $E_{\max}^{M}(t) = (\rho(t)+\rho(t)^{-1})/4 \le 1.125$ for $E^{\mathrm{M}}(t)$. Note that $E^{\mathrm{S}}(t)$ can get arbitrarily large.

Proof of (3) **and** (4)**.** Let $(\widehat{p}_n)_n, (\widehat{q}_n)_n$ be random sequences in $[0,1]$ converging to $p := F(x)$ in probability. It follows from Lindeberg's Central Limit theorem, applied to convolutions of binomial distributions, that

$$\sup_{y\in\mathbb{R}} \Big| G_{\boldsymbol{N}_n,\widehat{p}_n}(y) - \Phi\Big(\frac{\sqrt{n}}{\sigma(p)}\Big(\frac{y}{n} - \mu_n(\widehat{p}_n)\Big)\Big)\Big| \to_p 0,$$

where

$$\begin{aligned}
\mu_n(q) &:= \sum_{r=1}^k \frac{N_{nr}}{n} B_r(q) \quad \text{for } q \in [0,1],\\
\sigma(p) &:= \Big(\sum_{r=1}^k \pi_r B_r(p)(1-B_r(p))\Big)^{1/2}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mu_n(\widehat{q}_n) - \mu_n(\widehat{p}_n) &= \sum_{r=1}^k \frac{N_{nr}}{n}\big(B_r(\widehat{q}_n) - B_r(\widehat{p}_n)\big)\\
&= \Big(\sum_{r=1}^k \pi_r\beta_r(p) + o_p(1)\Big)(\widehat{q}_n - \widehat{p}_n)\\
&= \Big(\frac{\sigma(p)}{K^{\mathrm{M}}(F(x))^{1/2}} + o_p(1)\Big)(\widehat{q}_n - \widehat{p}_n).
\end{aligned}$$

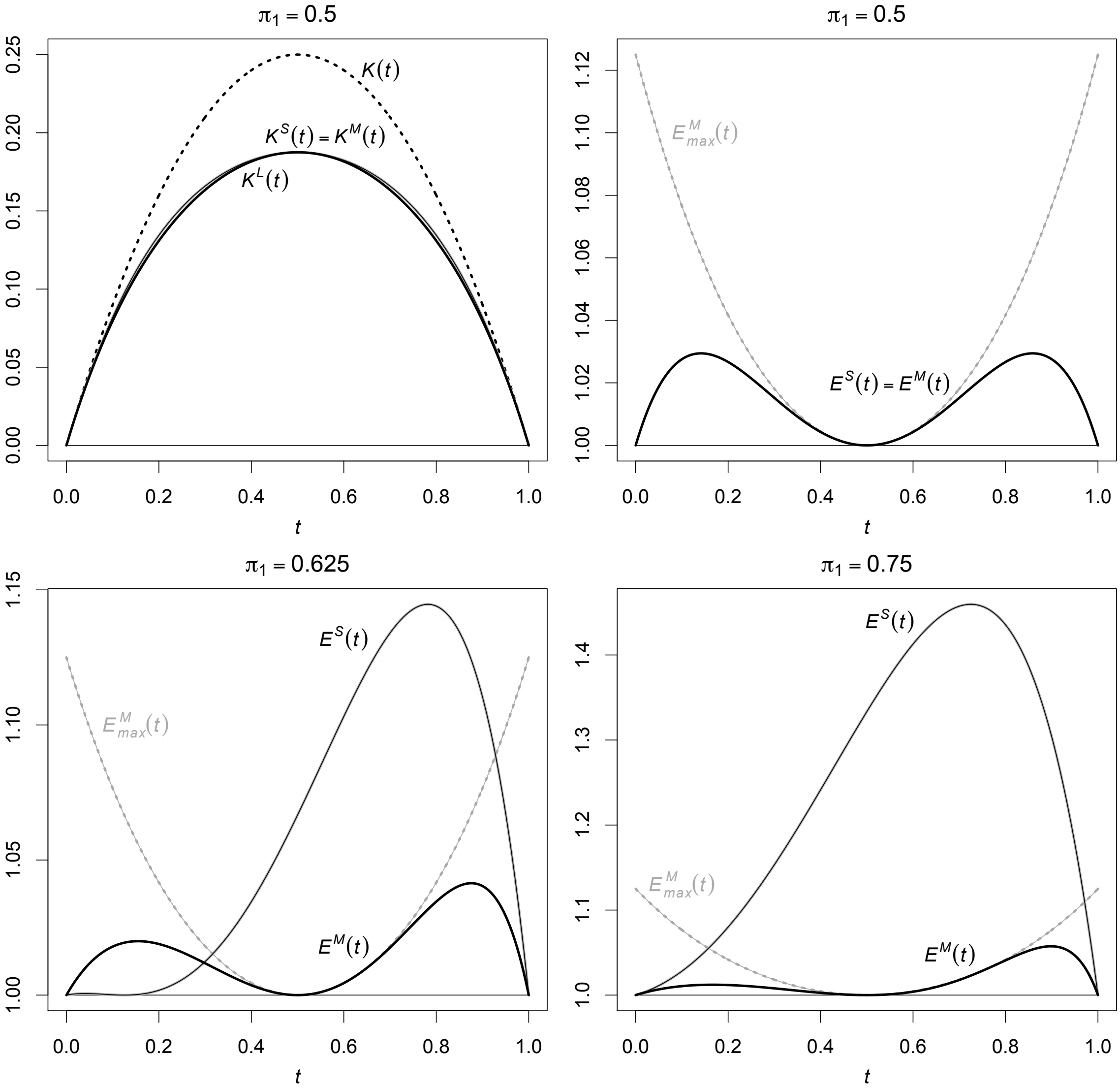

Figure 8: Asymptotic variances and relative efficiencies for $k = 2$.

Now we apply these findings to

$$\widehat{p}_n \;:=\; \widehat{F}_n^{\mathrm{M}}(x) + \frac{\Delta}{\sqrt{n}} \quad \text{and} \quad \widehat{q}_n \;:=\; \widehat{F}_n^{\mathrm{M}}(x)$$

with $\Delta \in \mathbb{R}$ to be specified later. Note that $\mu_n(\widehat{q}_n) = \widehat{F}_n(x)$ by definition of $\widehat{F}_n^{\mathrm{M}}(x)$. Hence for $c = 0, 1$,

$$\begin{aligned} G_{\boldsymbol{N}_n,\widehat{p}_n}(n\widehat{F}_n(x) - c) &= \Phi\Big(\frac{\sqrt{n}}{\sigma(p)}\big(\widehat{F}_n(x) + O(n^{-1}) - \mu_n(\widehat{p}_n)\big)\Big) + o_p(1) \\ &= \Phi\Big(\frac{\sqrt{n}}{\sigma(p)}\big(\mu_n(\widehat{q}_n) - \mu_n(\widehat{p}_n)\big)\Big) + o_p(1) \\ &\to_p \Phi\Big(\frac{-\Delta}{K^{\mathrm{M}}(F(x))^{1/2}}\Big). \end{aligned}$$

If we choose Δ strictly smaller or strictly larger than $K^{\mathrm{M}}(F(x))^{1/2}\Phi^{-1}(1-\alpha)$, then the limit of $G_{\boldsymbol{N}_n,\widehat{p}_n}(n\widehat{F}_n(x))$ is strictly larger or strictly smaller than α, respectively. This proves (4). If we choose Δ strictly smaller or strictly larger than $-K^{\mathrm{M}}(F(x))^{1/2}\Phi^{-1}(1-\alpha)$, then the limit of $G_{\boldsymbol{N}_n,\widehat{p}_n}(n\widehat{F}_n(x) - 1)$ is strictly larger or strictly smaller than $1-\alpha$, respectively, which proves (3). □

B Computer code

From the web pages of Lutz Dümbgen (www.stat.unibe.ch/duembgen) one can download specific computer programs for the methods and examples presented here. All code is for the statistical computing environment R (R Core Team 2013). The files are:

- Estimation.R: Computation of the point estimators $\widehat{F}_n^{\mathrm{S}}$, $\widehat{F}_n^{\mathrm{M}}$ and $\widehat{F}_n^{\mathrm{L}}$.
- Simulations.R: Simulation of RSS and JPS data sets, including sampling from the Dell–Clutter model.
- ConfBands.R: Computing pointwise and simultaneaous confidence bands for F.
- MonteCarlo.R: Monte Carlo estimation of the estimators' bias and RMSE; simulating sampling from a finite population as in Section 4.1.
- Municip_CH_2015.txt: Data for Section 4.1.
- MainScript.R: Main script file with examples for all procedures coded in the previous R files.