ONLINE SUPPLEMENTARY MATERIAL FOR "ON IDENTIFICATION AND NON-NORMAL SIMULATION IN ORDINAL COVARIANCE AND ITEM RESPONSE MODELS"

NJÅL FOLDNES AND STEFFEN GRØNNEBERG

1. Technical proofs

Proof of Proposition 1. By the assumed conditional independence, we have for $x_1, x_2, \ldots, x_d \in \{0, 1\}$ that

$$P(\cap_{1\leq i\leq d}\{X_i = x_i\}) = \mathbb{E}P(\cap_{1\leq i\leq d}\{X_i = x_i\}|f) = \mathbb{E}\prod_{1\leq i\leq d} P(X_i = x_i|f).$$

By Assumption 1 (2), we have

$$\pi_i(f) = H(\zeta_i) = \int_{\mathbb{R}} I\{z \leq \zeta_i\}h(z)\,\mathrm{d}z. \tag{1}$$

For $x_i \in \{0, 1\}$, we have

$$P(X_i = x_i|f) = \begin{cases} \pi_i(f), & \text{if } x_i = 1 \\ 1 - \pi_i(f), & \text{if } x_i = 0 \end{cases}.$$

Using eq. (1) and that $\int_{\mathbb{R}} h(z)\,\mathrm{d}z = 1$, we see

$$1-\pi_i(f) = \int_{\mathbb{R}}[1-I\{z \leq \zeta_i\}]h(z)\,\mathrm{d}z = \int_{\mathbb{R}} I\{z > \zeta_i\}h(z)\,\mathrm{d}z = \int_{\mathbb{R}} I\{z \geq \zeta_i\}h(z)\,\mathrm{d}z$$

where the last equality follows since h is a density. With the standard convention that the product over the empty set is the multiplicative

Date: September 25, 2019.

identity, we therefore have that

$$\prod_{1\leq i\leq d} P(X_i = x_i|f) = \left(\prod_{i:x_i=1} \pi_i(f)\right)\left(\prod_{j:x_j=0}[1-\pi_j(f)]\right)$$
$$= \left(\prod_{i:x_i=1}\int_{\mathbb{R}} I\{z_i \leq \zeta_i\}h(z_i)\,\mathrm{d}z_i\right)\left(\prod_{j:x_j=0}\int_{\mathbb{R}} I\{z_j \geq \zeta_j\}h(z_j)\,\mathrm{d}z_j\right)$$
$$= \left(\prod_{i:x_i=1}\int_{\mathbb{R}} I\{(-1)^{1-x_i}z_i \leq (-1)^{1-x_i}\zeta_i\}h(z_i)\,\mathrm{d}z_i\right)$$
$$\times\left(\prod_{j:x_j=0}\int_{\mathbb{R}} I\{(-1)^{1-x_j}z_j \leq (-1)^{1-x_i}\zeta_j\}h(z_j)\,\mathrm{d}z_j\right)$$
$$= \prod_{i=1}^{d}\int_{\mathbb{R}} I\{(-1)^{1-x_i}z_i \leq (-1)^{1-x_i}\zeta_i\}h(z_i)\,\mathrm{d}z_i$$

where the second to last equality uses that for $x_j = 0$ we have $(-1)^{1-x_j}z_j \leq (-1)^{1-x_i}\zeta_j \iff (-1)z_j \leq (-1)\zeta_j \iff z_j \geq \zeta_j$.

Hence,

$$\mathbb{E}\prod_{1\leq i\leq d} P(X_i = x_i|f) = \mathbb{E}\prod_{i=1}^{d}\int_{\mathbb{R}} I\{(-1)^{1-x_i}z_i \leq (-1)^{1-x_i}\zeta_i\}h(z_i)\,\mathrm{d}z_i$$
$$= \mathbb{E}\int_{\mathbb{R}^d}[\prod_{i=1}^{d} I\{(-1)^{1-x_i}z_i \leq (-1)^{1-x_i}\zeta_i\}][\prod_{i=1}^{d} h(z_i)]\,[\prod_{i=1}^{d}\mathrm{d}z_i]$$
$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}[\prod_{i=1}^{d} I\{(-1)^{1-x_i}z_i \leq (-1)^{1-x_i}y_i\}][\prod_{i=1}^{d} h(z_i)]\,[\prod_{i=1}^{d}\mathrm{d}z_i]\mathrm{d}G(y_1,\ldots,y_d)$$

where G is the CDF of $(\zeta_1,\ldots,\zeta_d)'$. Recall that $Z_1, Z_2, \ldots, Z_d$ are independent from ζ, and further that $Z_1, Z_d, \ldots, Z_d$ are IID with marginal distribution H. Considering the above integral, we have

$$\mathbb{E}\prod_{1\leq i\leq d} P(X_i = x_i|f) = P(\cap_{1\leq i\leq d}\{(-1)^{1-x_i}Z_i \leq (-1)^{1-x_i}\zeta_i\})$$
$$= P(\cap_{1\leq i\leq d}\{(-1)^{1-x_i}Z_i - (-1)^{1-x_i}\zeta_i \leq 0\})$$
$$= P(\cap_{1\leq i\leq d}\{(-1)^{1-x_i}(Z_i - \zeta_i) \leq 0\}).$$

Now consider

$$X = (X_1, \dots, X_d)' = (I\{Z_1 - \zeta_1 \le 0\}, \dots, I\{Z_d - \zeta_d \le 0\})'.$$

We have

$$X_1 = 0 \iff Z_1 - \zeta_1 > 0 \iff (-1)(Z_1 - \zeta_1) < 0,$$
$$X_1 = 1 \iff Z_1 - \zeta_1 \le 0,$$

which shows that X is a stochastic representation of the IRT model. □

We now prove Lemma 2 and 3 in detail. Please note that the length of the following proofs is mainly due to our desire to give a as reader friendly presentation as possible. The material applies standard probability theory, and does not contain technical innovations.

Complete proof of Lemma 2. The start of the proof consists of setting up and understanding certain structures that will be used in constructing a density with the required properties. These structures are illustrated in a simple setting in an example given immediately after the present proof.

By Lemma 1, we may without loss of generality assume that $\xi = X$, and that the thresholds are $x_1, x_2, \dots, x_K$ for each coordinate. Let us here note, since it will be relevant later in the proof, that since we use the same thresholds for all coordinates, there may be certain combinations of values that occur with zero probability. For example, if e.g. X_1 is dichotomous, while X_2 is trichotomous, we would still have thresholds $x_1 = 0, x_2 = 1, x_3 = 2$, but there is no probability that $X_1 = 2$.

Each X_i may take on values in the set $S_X = \{x_1, x_2 \dots, x_K\}$, so that $X = (X_1, X_2, \dots, X_d)'$ only takes on values in the set S_X^d, the d-times Cartesian product of S_X, given by $S_X^d = \otimes_{j=1}^d S_X$. Note again that we do not assume that X takes on all values in S_X^d with positive probability. Despite this, we will refer to S_X^d as the support of X.

Consider the class of subsets of $\mathbb{R}^d$ given by

$$\mathcal{Q} = \{\otimes_{l=1}^d (x_{j_l}, x_{j_l+1}] : j_l \in \{0, 1, \dots, K-1\} \text{ for } l = 1, 2, \dots, d\},$$

where $x_0 = x_1 - 1$. That is, $\mathcal{Q}$ contains the hyper-rectangles contained between the points of the support S_X^d of X. Note that while it may be the case that $P(X \in Q_i) = 0$ for some i, such as in the above mentioned case when X_1 is dichotomous and X_2 is trichotomous, but all Q_i will have non-empty volume in $\mathbb{R}^d$ since it is assumed that $x_1 < x_2 < \cdots x_K$.

Notice that the finite collection of sets in $\mathcal{Q}$ are disjoint, and enumerate them by $Q_1, Q_2, \ldots, Q_N$. The union of the elements of $\mathcal{Q}$ equals $\cup_{j=1}^N Q_i = \otimes_{l=1}^d (x_0, x_K]$. Since $x_0 = x_1 - 1 < x_1$ we have that the support of X i.e., S_X^d is contained in $\cup_{j=1}^N Q_i$, i.e., $S_X^d \subseteq \cup_{j=1}^N Q_i$.

Please recall that these definitions are illustrated and the above facts are verified for a specific case in the upcoming Example 1.

We now define a density $\tilde{f}$, which smears the probability that X is in Q_i uniformly over each Q_i. I.e., we let

$$\tilde{f}(x) = \sum_{i=1}^N \frac{P(X \in Q_i)}{V_i} I\{x \in Q_i\},$$

where $I\{A\}$ is the indicator function of A, which is one if A is true and zero otherwise, and, for $i = 1, 2, \ldots, N$ we let $V_i = \int_{\mathbb{R}^d} I\{x \in Q_i\} \, \mathrm{d}x$ which is non-zero.

Hence $\tilde{f}$ is finite and non-negative and is therefore a density if it also integrates to one, which it is seen to do by

$$\begin{aligned}\int_{\mathbb{R}^d} \tilde{f}(x) \, \mathrm{d}x &= \int_{\mathbb{R}^d} \sum_{i=1}^N \frac{P(X \in Q_i)}{V_i} I\{x \in Q_i\} \, \mathrm{d}x \\ &= \sum_{i=1}^N \frac{P(X \in Q_i)}{V_i} \int_{\mathbb{R}^d} I\{x \in Q_i\} \, \mathrm{d}x = \sum_{i=1}^N P(X \in Q_i) \\ &\overset{(a)}{=} P(X \in \cup_{i=1}^N Q_i) = P(X \in \otimes_{j=1}^d (x_0, x_K]) \overset{(b)}{=} P(X \in S_X^d) = 1,\end{aligned}$$

where *(a)* uses that (Q_i) forms a disjoint sequence and additivity of probability measures, and *(b)* uses that X only takes values in S_X^d, which is contained in $\otimes_{j=1}^d (x_0, x_K]$.

Let $\tilde{\xi}$ have density $\tilde{f}$. Since $\tilde{\xi}$ has a density with respect to Lebesgue measure, it is a continuous random vector. We complete the proof by showing that $\tilde{\xi}$ has the same probability of being in an element in $\mathcal{Q}$ as X does. That is, we show that $P(\tilde{\xi} \in Q_k) = P(X \in Q_k)$ for $k = 1, 2, \ldots, N$. Then $\tilde{\xi}$ is discretize equivalent to $\xi = X$, since $\tilde{\xi}$ has the same probability of being in thresholds defined by the limits of the rectangles in Q_k for $k = 1, 2, \dots, N$.

Indeed, we have

$$\int_{Q_k} \tilde{f}(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} I\{x \in Q_k\} \sum_{i=1}^{N} \frac{P(X \in Q_i)}{V_i} I\{x \in Q_i\}\,\mathrm{d}x$$
$$= \sum_{i=1}^{N} \frac{P(X \in Q_i)}{V_i} \int_{\mathbb{R}^d} I\{x \in Q_k, x \in Q_i\}\,\mathrm{d}x.$$

Since (Q_i) is a disjoint sequence, we have that $I\{x \in Q_k, x \in Q_i\} = I\{x \in Q_k\}I\{i = k\}$, showing

$$\int_{Q_k} \tilde{f}(x)\,\mathrm{d}x = \sum_{i=1}^{N} I\{i = k\} \frac{P(X \in Q_i)}{V_i} \int_{\mathbb{R}^d} I\{x \in Q_k\}\,\mathrm{d}x$$
$$= \frac{P(X \in Q_k)}{I_k} \int_{\mathbb{R}^d} I\{x \in Q_k\}\,\mathrm{d}x = P(X \in Q_k).$$

Hence, $\tilde{\xi}$ has the same probability of being in Q_k as X and the proof is complete. □

Example 1. We here include an illustration of the quantities defined in Lemma 2 for the bivariate case $d = 2$ with dichotomous observations, i.e., $K = 2$, and, say, $x_1 = 0, x_2 = 1$. Then $\mathcal{Q}$ contains sets of the form $(x_{j_1}, x_{j_1+1}] \otimes (x_{j_1}, x_{j_1+1}]$ where $j_1 \in \{0, 1\}$ and $j_2 \in \{0, 1\}$. This means $\mathcal{Q} = \{Q_1, Q_2, Q_3, Q_4\}$ with

$$Q_1 = (-1, 0] \otimes (-1, 0], \qquad Q_2 = (-1, 0] \otimes (0, 1]$$
$$Q_3 = (0, 1] \otimes (-1, 0], \qquad Q_4 = (0, 1] \otimes (0, 1].$$

Note that (Q_i) forms a disjoint sequence. Indeed, suppose $x \in Q_1$. Then $x = (x_1, x_2)$ where $-1 < x_1 \le 0$ and $-1 < x_2 \le 0$. Clearly x cannot be in either Q_2, Q_3 or Q_4, since at least one of the coordinate values are incompatible with such a membership. Indeed, consider Q_2. While the value of x_1 is compatible with being in Q_2, the value of x_2 implies that it is impossible that x is a member of Q_2.

The support of X is $S_X^2 = \{0, 1\}^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, and we see the correspondence between $\mathcal{Q}$ and S_X^2 given by

$$(0, 0) \in Q_1, \quad (0, 1) \in Q_2, \quad (1, 0) \in Q_3, \quad (1, 1) \in Q_4.$$

It is further clear that $\cup_{i=1}^4 Q_i = (-1,1] \otimes (-1,1]$, and so that we also see that $S_X^2 \subseteq \cup_{i=1}^4 Q_i$. □

Complete proof of Lemma 3. We argue for each coordinate $1 \leq i \leq d$ separately. Since the upcoming argument holds simultaneously for all coordinates, the multivariate conclusion follows. By Lemma 2, we may assume that ξ is a continuous random vector. Hence, each F_i is continuous and increasing, though not necessarily strictly increasing. This necessitates the use of heavier theory than would be required if F_i was strictly increasing. In the strictly increasing case, our argument reduces to the one given the main paper.

Recall the definition of the left continuous inverse, as e.g., described in Rüschendorf (2009) and Shorack & Wellner (2009, Chapter 1), given by

$$F_i^{-1}(t) = \inf\{x : F_i(x) \geq t\} \quad \text{for } 0 < t < 1.$$

Define $U_i = F_i(\xi_i)$. Since F_i is continuous, Proposition 2.1 of Rüschendorf (2009) shows that $\xi_i = F_i^{-1}(U_i)$ almost surely, and that $U_i \sim U[0,1]$ where $U[0,1]$ is the uniform distribution on the unit interval. We may therefore assume that $\xi_i = F_i^{-1}(U_i)$, that is, $\xi_i = F_i^{-1}(F_i(\xi_i))$, a trivial equality in the strictly increasing case. We therefore have

$$X_i = \sum_{j=1}^{K} x_j I\{\tau_{i,j-1} < \xi_i \leq \tau_{i,j}\} = \sum_{j=1}^{K} x_j I\{\tau_{i,j-1} < F_i^{-1}(U_i) \leq \tau_{i,j}\}.$$

Theorem 1 and eq. (23) both in Shorack & Wellner (2009, Chapter 1) shows that for any cumulative distribution function G, any $0 < t < 1$ and numbers x_1, x_2 we have

$$x_1 < G^{-1}(t) \leq x_2 \iff G(x_1) < t \leq G(x_2). \tag{2}$$

Using this result for $G = F_i$ gives

$$X_i = \sum_{j=1}^{K} x_j I\{\tau_{i,j-1} < F_i^{-1}(U_i) \leq \tau_{i,j}\} = \sum_{j=1}^{K} x_j I\{F_1(\tau_{i,j-1}) < U_i \leq F_1(\tau_{i,j})\}.$$

We then again use eq. (2) but now with $G = \Phi$, the cumulative distribution function of the standard normal, and see that

$$\begin{aligned} X_i &= \sum_{j=1}^{K} x_j I\{F_1(\tau_{i,j-1}) < U_i \leq F_1(\tau_{i,j})\} \\ &= \sum_{j=1}^{K} x_j I\{\Phi^{-1}[F_1(\tau_{i,j-1})] < \Phi^{-1}[U_i] \leq \Phi^{-1}[F_1(\tau_{i,j})]\} \\ &= \sum_{j=1}^{K} x_j I\{\tilde{\tau}_{i,j-1} < \tilde{\xi}_i \leq \tilde{\tau}_{i,j}\} \end{aligned}$$

where $\tilde{\tau}_{i,j-1} = \Phi^{-1}(F_i(\tau_{i,j-1}))$, and $\tilde{\tau}_{i,j} = \Phi^{-1}(F_i(\tau_{i,j}))$, and where $\tilde{\xi}_i = \Phi^{-1}(U_i)$. Since $U_i \sim U[0,1]$ we have that $\Phi^{-1}(U_i)$ is standard normal, and the conclusion follows. We note that the only property of Φ we used was that it was a cumulative distribution function. By the above argument, the marginals can be transformed to any distribution function, including discrete or continuous distributions with atoms. This follows from the uniformity of U_i and Theorem 1 of Shorack & Wellner (2009, Chapter 1), where it is in general established that when $U_i \sim U[0,1]$, we have $G^{-1}(U_i) \sim G$ for any cumulative distribution function G. □

References

Rüschendorf, L. (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference* **139**, 3921–3927.

Shorack, G. R. & Wellner, J. A. (2009). *Empirical processes with applications to statistics*, vol. 59. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM). Originally published in 1986 by Wiley, New York.

Department of Economics, BI Norwegian Business School
E-mail address: njal.foldnes@bi.no

Department of Economics, BI Norwegian Business School
E-mail address: steffeng@gmail.com