ONLINE APPENDIX FOR "PARTIAL IDENTIFICATION OF LATENT CORRELATIONS WITH BINARY DATA"

STEFFEN GRØNNEBERG, JONAS MOSS, AND NJÅL FOLDNES

Besides this online appendix, the online supplementary material accompanying the paper "Partial identification of latent correlations with binary data" includes several `R`-scripts. These are described in the text-file `index.txt`.

1. Detailed algebraic verification of Theorem 1

For completeness, we here provide a complete algebraic verification of Theorem 1. The calculations are tedious but elementary.

The distribution of $Z = (Z_1, Z_2)$ is

$$\mathrm{P}(Z = (a,a)) = p_{11}, \qquad \mathrm{P}(Z = (b,-b)) = p_{10},$$
$$\mathrm{P}(Z = (-a,-a)) = p_{00}, \qquad \mathrm{P}(Z = (-b,b)) = p_{01}.$$

From this we compute

$$\mathrm{E}(Z_1 Z_2) = a^2(p_{11} + p_{00}) - b^2(p_{10} + p_{01}).$$

The marginal distributions of Z_1, Z_2 are

$$\mathrm{P}(Z_1 = a) = p_{11} = \mathrm{P}(Z_2 = a), \qquad \mathrm{P}(Z_1 = b) = p_{10} = \mathrm{P}(Z_2 = -b)$$
$$\mathrm{P}(Z_1 = -a) = p_{00} = \mathrm{P}(Z_2 = -a), \qquad \mathrm{P}(Z_1 = -b) = p_{01} = \mathrm{P}(Z_2 = b).$$

We therefore have

$$\begin{aligned}
\mathrm{E}(Z_1) &= ap_{11} + bp_{10} - ap_{00} - bp_{01} \\
&= a(p_{11} - p_{00}) + b(p_{10} - p_{01}), \\
\mathrm{E}(Z_2) &= ap_{11} - bp_{10} - ap_{00} + bp_{01} \\
&= a(p_{11} - p_{00}) - b(p_{10} - p_{01}) \\
&= \mathrm{E}\, Z_1 - 2b(p_{10} - p_{01}).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Cov}\,(Z_1, Z_2) &= \mathrm{E}(Z_1 Z_2) - \mathrm{E}(Z_1)\,\mathrm{E}(Z_2) \\
&= a^2(p_{11} + p_{00}) - b^2(p_{10} + p_{01}) \\
&\quad - [a(p_{11} - p_{00}) + b(p_{10} - p_{01})][a(p_{11} - p_{00}) - b(p_{10} - p_{01})] \\
&= a^2(p_{11} + p_{00}) - b^2(p_{10} + p_{01}) - a^2(p_{11} - p_{00})^2 + b^2(p_{10} - p_{01})^2 \\
&= a^2(p_{11} + p_{00} - (p_{11} - p_{00})^2) - b^2(p_{10} + p_{01} - (p_{10} - p_{01})^2).
\end{aligned}$$

We also have

$$\begin{aligned}\mathrm{E}(Z_1^2) &= \mathrm{E}(Z_2^2)\\ &= a^2 p_{11} + b^2 p_{10} + a^2 p_{00} + b^2 p_{01}\\ &= a^2(p_{11}+p_{00}) + b^2(p_{10}+p_{01}).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathrm{Var}\,(Z_1) &= \mathrm{E}(Z_1^2) - \mathrm{E}(Z_1)^2\\ &= a^2(p_{11}+p_{00}) + b^2(p_{10}+p_{01}) - [a(p_{11}-p_{00}) + b(p_{10}-p_{01})]^2\\ &= a^2(p_{11}+p_{00}) + b^2(p_{10}+p_{01}) - a^2(p_{11}-p_{00})^2\\ &\quad - 2ab(p_{11}-p_{00})(p_{10}-p_{01}) - b^2(p_{10}-p_{01})^2\\ &= a^2(p_{11}+p_{00}-(p_{11}-p_{00})^2) + b^2(p_{10}+p_{01}-(p_{10}-p_{01})^2) -\\ &\quad 2ab(p_{11}-p_{00})(p_{10}-p_{01}),\end{aligned}$$

and, using that $\mathrm{E}(Z_2) = \mathrm{E}(Z_1)$, and that $\mathrm{E}(Z_2) = \mathrm{E}(Z_1) - 2b(p_{10}-p_{01})$, we get

$$\begin{aligned}\mathrm{Var}\,(Z_2) &= \mathrm{E}(Z_2^2) - \mathrm{E}(Z_2)^2\\ &= \mathrm{E}(Z_1^2) - (\mathrm{E}(Z_1) - 2b(p_{10}-p_{01}))^2\\ &= \mathrm{E}(Z_1^2) - \mathrm{E}(Z_1)^2 + 4\,\mathrm{E}(Z_1) b(p_{10}-p_{01}) - 4b^2(p_{10}-p_{01})^2\\ &= \mathrm{Var}\,(Z_1) + 4[a(p_{11}-p_{00}) + b(p_{10}-p_{01})]\cdot b(p_{10}-p_{01}) - 4b^2(p_{10}-p_{01})^2\\ &= \mathrm{Var}\,(Z_1) + 4ab(p_{11}-p_{00})(p_{10}-p_{01}) + 4b^2(p_{10}-p_{01})^2 - 4b^2(p_{10}-p_{01})^2\\ &= \mathrm{Var}\,(Z_1) + 4ab(p_{11}-p_{00})(p_{10}-p_{01}).\end{aligned}$$

We want to calculate

$$\rho = \frac{\mathrm{Cov}\,(Z_1, Z_2)}{(\mathrm{Var}\,(Z_1)\,\mathrm{Var}\,(Z_2))^{1/2}}.$$

We first calculate the product $\mathrm{Var}\,(Z_1)\,\mathrm{Var}\,(Z_2)$. We now use $a = 1/b$. This simplifies the expressions to

$$\begin{aligned}\mathrm{Var}\,(Z_1) &= a^2(p_{11}+p_{00}-(p_{11}-p_{00})^2) + b^2(p_{10}+p_{01}\\ &\quad - (p_{10}-p_{01})^2) - 2ab(p_{11}-p_{00})(p_{10}-p_{01})\\ &= q - 2\Delta,\end{aligned}$$

where $q = a^2(p_{11}+p_{00}-(p_{11}-p_{00})^2)+b^2(p_{10}+p_{01}-(p_{10}-p_{01})^2)$ and $\Delta = (p_{11}-p_{00})(p_{10}-p_{01})$. Similarly, $\mathrm{Var}\,(Z_2) = q + 2\Delta$, and therefore,

$$\begin{aligned}\mathrm{Var}\,(Z_1)\,\mathrm{Var}\,(Z_2) &= (q-2\Delta)(q+2\Delta)\\ &= q^2 - 4\Delta^2\\ &= a^4c_1^2 + b^4c_2^2 + 2a^2b^2c_1c_2 - 4\Delta^2.\end{aligned}$$

Where $c_1 = p_{11} + p_{00} - (p_{11}-p_{00})^2$, and $c_2 = p_{10} + p_{01} - (p_{10}-p_{01})^2$. We note that c_1, c_2, Δ does not vary with a or b.

In terms of the introduced constants, we recognize that

$$\text{Cov}\left(Z_1, Z_2\right) = a^2 c_1 - b^2 c_2.$$

We therefore have

$$\begin{aligned}\rho &= \frac{\text{Cov}\left(Z_1, Z_2\right)}{\left(\text{Var}\left(Z_1\right)\text{Var}\left(Z_2\right)\right)^{1/2}} \\ &= \frac{a^2 c_1 - b^2 c_2}{\sqrt{a^4 c_1^2 + b^4 c_2^2 + 2a^2 b^2 c_1 c_2 - 4\Delta^2}}.\end{aligned}$$

Using $a = 1/b$, we see that

$$\rho = \frac{a^2 c_1 - b^2 c_2}{\sqrt{a^4 c_1^2 + b^4 c_2^2 + d}},$$

where $d = 2a^2 b^2 c_1 c_2 - 4\Delta^2 = d = 2c_1 c_2 - 4\Delta^2$ does not depend on a, b.

Case 1: Letting $b \to \infty$, giving the negative end-point. We use $a = 1/b$ and get

$$\begin{aligned}\rho &= \frac{a^2 c_1 - b^2 c_2}{\sqrt{a^4 c_1^2 + b^4 c_2^2 + d}} \\ &= \frac{b^{-2} c_1 - b^2 c_2}{\sqrt{b^{-4} c_1^2 + b^4 c_2^2 + d}} \\ &= \frac{b^{-4} c_1 - c_2}{\sqrt{b^{-4}(b^{-4} c_1^2 + b^4 c_2^2 + d)}} \\ &= \frac{b^{-4} c_1 - c_2}{\sqrt{b^{-8} c_1^2 + c_2^2 + b^{-4} d)}} \\ &\to \frac{-c_2}{|c_2|}.\end{aligned}$$

If $c_2 > 0$, this shows that $\rho \to -1$. We recall that $c_2^2 = (p_{10} + p_{01} - (p_{10} - p_{01})^2)^2 \geq 0$, and we only need to show that $c_2^2 \neq 0$. We have

$$\begin{aligned}p_{10} + p_{01} - (p_{10} - p_{01})^2 &= p_{10} + p_{01} - p_{10}^2 + 2p_{10}p_{01} - p_{01}^2 \\ &= (p_{10} - p_{10}^2) + (p_{01} - p_{01}^2) + 2p_{10}p_{01}.\end{aligned}$$

Since p_{01} and p_{10} are in $(0, 1)$, we have $p_{10}p_{01} > 0$. We have that $p_{10} > p_{10}^2$ and $p_{01} > p_{01}^2$, and therefore $p_{10} - p_{10}^2 > 0$ and $p_{01} - p_{01}^2 > 0$. Therefore, $c_2^2 \neq 0$.

Case 2: Letting $b \to 0^+$, giving the positive end-point. We use $b = 1/a$ and the exact same steps as above to get that

$$\begin{aligned}\rho &= \frac{a^2 c_1 - b^2 c_2}{\sqrt{a^4 c_1^2 + b^4 c_2^2 + d}} \\ &\to \frac{c_1}{|c_1|}.\end{aligned}$$

If $c_1 > 0$, this shows that $\rho \to 1$. We recall that $c_1^2 = (p_{11} + p_{00} - (p_{11} - p_{00})^2)^2 \geq 0$, and we only need to show that $c_1^2 \neq 0$. We have

$$\begin{aligned} p_{11} + p_{00} - (p_{11} - p_{00})^2 &= p_{11} + p_{00} - p_{11}^2 + 2p_{11}p_{00} - p_{00}^2 \\ &= (p_{11} - p_{11}^2) + (p_{00} - p_{00}^2) + 2p_{11}p_{00}. \end{aligned}$$

Since p_{00} and p_{11} are in $(0, 1)$, we have $p_{11}p_{00} > 0$. We have that $p_{11} > p_{11}^2$ and $p_{00} > p_{00}^2$, and therefore $p_{11} - p_{11}^2 > 0$ and $p_{00} - p_{00}^2 > 0$. Therefore, $c_1^2 \neq 0$.

Let ρ_b be the correlation of $Z(b) = Z(1/b, b)$ for $b > 0$. We recall

$$\rho_b = \frac{b^{-4}c_1 - c_2}{\sqrt{b^{-8}c_1^2 + c_2^2 + b^{-4}d)}}$$

and $c_1, c_2 > 0$. Since this is a continuous function with limits -1 and 1, every correlation in $(-1, 1)$ is attained by the intermediate value theorem.

Department of Economics, BI Norwegian Business School, Oslo, Norway 0484
Email address: steffeng@gmail.com

Department of Mathematics, University of Oslo, PB 1053, Blindern, NO-0316, Oslo, Norway
Email address: jonasmgj@math.uio.no

Department of Economics, BI Norwegian Business School, Stavanger, Norway 4014
Email address: njal.foldnes@gmail.com