Corrections to: Multivariate normal distribution approaches for dependently truncated data

Pan, Chi-Hung and Emura, Takeshi

Graduate Institute of Statistics, National Central University

9 August 2014

Corrections to: Multivariate normal distribution approaches for dependently truncated data

Chi-Hung Pan • Takeshi Emura

Graduate Institute of Statistics, National Central University, Taiwan

Corresponding to: T. Emura

Graduate Institute of Statistics, National Central University, Taiwan

e-mail: emura@stat.ncu.edu.tw

We provide corrections for Emura and Konno (2010). We also numerically verify the corrected formulae. Appendix gives a real data used for numerical analysis.

1. Correction in the score function [p.138, definition of $\mathbf{U}_i^*(\boldsymbol{\theta})$]

For $\boldsymbol{\theta}' = (\mu_L, \mu_X, \sigma_L^2, \sigma_X^2)$, the corrected score function is

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -n \frac{\dot{c}^*(\boldsymbol{\theta})}{c^*(\boldsymbol{\theta})} + \sum_i \mathbf{U}_i^*(\boldsymbol{\theta}), \tag{1}$$

where

$$c^*(\boldsymbol{\theta}) = \Phi\left(\frac{\mu_X - \mu_L}{\sqrt{\sigma_X^2 + \sigma_L^2}} \right), \quad \mathbf{U}_i^*(\boldsymbol{\theta}) = \begin{bmatrix} (L_i - \mu_L)/\sigma_L^2 \\ (X_i - \mu_X)/\sigma_X^2 \\ -/(2\sigma_L^2) + (L_i - \mu_L)^2/(2\sigma_L^4) \\ -/(2\sigma_X^2) + (X_i - \mu_X)^2/(2\sigma_X^4) \end{bmatrix}. \tag{2}$$

The error occurred in the third and fourth components of $\mathbf{U}_i^*(\boldsymbol{\theta})$.

To confirm that Equations (1) and (2) are correct, we focus on the third component of Equation (1):

$$\frac{\partial l(\boldsymbol{\theta})}{\partial(\sigma_L^2)} = -n \frac{1}{c^*(\boldsymbol{\theta})} \frac{\partial c^*(\boldsymbol{\theta})}{\partial(\sigma_L^2)} + \sum_i \left\{ -\frac{1}{2\sigma_L^2} + \frac{(L_i - \mu_L)^2}{2\sigma_L^4} \right\}. \tag{3}$$

We compute the score functions using a real data from Appendix. Then, compare Equation (3) with the numerical derivative

$$\{\, l(\mu_L, \mu_X, \sigma_L^2 + h, \sigma_X^2, 0) - l(\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, 0) \,\}/h,$$

where $h = 10^{-7}$. The results are given in Table 1. We see that there is virtually no difference between the corrected formula and the numerical derivative. On the other hand, the values of the formula of Emura and Konno (2010) are remarkably different from those of the numerical derivative. R codes for these calculations are given in Appendix B.

Table 1: Calculations of the score function using the three methods. The score function is computed by using a read data in Appendix.

$(\mu_L, \mu_X, \sigma_L^2, \sigma_X^2)$	Numerical Derivative	Corrected formula	Incorrect formula
(10,20,30,40)	2.86657155	2.86657152	-137.0459
(20,40,60,80)	0.04255583	0.04255583	-279.9137
(30,60,90,120)	-0.10249465	-0.10249456	-420.0733
(40,80,120,160)	-0.02560967	-0.02560980	-560.0037

- Numerical derivative: $\{\, l(\mu_L, \mu_X, \sigma_L^2 + h, \sigma_X^2, 0) - l(\mu_L, \mu_X, \sigma_L^2, \sigma_X^2, 0) \,\}/h$, where $h = 10^{-7}$.
- Corrected formula = $\dfrac{\partial l(\boldsymbol{\theta})}{\partial(\sigma_L^2)} = -n \dfrac{1}{c^*(\boldsymbol{\theta})} \dfrac{\partial c^*(\boldsymbol{\theta})}{\partial(\sigma_L^2)} + \sum_i \left\{ -\dfrac{1}{2\sigma_L^2} + \dfrac{(L_i - \mu_L)^2}{2\sigma_L^4} \right\}$.
- Incorrect formula of Emura and Konno (2010).

2. Correction in the function $\dot{w}(c)$ [p.139]

Emura and Konno (2010) considered a function $w(\cdot)$: $(0,1) \to [0,1]$, defined as

$$w(c) = \frac{\Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c} + \frac{\phi\{\Phi^{-1}(c)\}^2}{c^2}.$$

They showed that w is strictly decreasing, reflecting the decreasing loss of information at inclusion probability c. However, they do not give the formula of $\dot{w}(c) = dw(c)/dc$, and their claim $\dot{w}(1/2) = \sqrt{2/\pi}(1 - 4/\pi)$ is incorrect.

Here we provide an explicit derivative given by

$$\dot{w}(c) = \frac{\{\, 1 - \Phi^{-1}(c)^2 \,\}c - \Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c^2} - \frac{2\phi\{\Phi^{-1}(c)\}}{c^3}[c\Phi^{-1}(c)) + \phi\{\Phi^{-1}(c)\}] \quad (4)$$

With this formula, one has

$$\dot{w}(1/2) = \frac{1}{1/2} - \frac{1}{2\pi(1/2)^4} = 2 - \frac{8}{\pi} \cong \text{-}0.5464791. \quad (5)$$

We have confirmed the correctness of Equations (4) and (5) in Table 2. R codes for the above calculation are given in Appendix B.

Table 2: Numerical calculations of the functions $w(c)$ and $\dot{w}(c)$.

c	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$w(c)$	0.830	0.781	0.735	0.688	0.636	0.577	0.507	0.416	0.287
$\dot{w}(c)$ Numerical	-0.550	-0.464	-0.460	-0.489	-0.546	-0.638	-0.785	-1.046	-1.631
$\dot{w}(c)$ Equation (4)	-0.550	-0.464	-0.460	-0.489	-0.546	-0.638	-0.785	-1.046	-1.631

- We define $w(c)=\dfrac{\Phi^{-1}(c)\phi\{\Phi^{-1}(c)\}}{c}+\dfrac{\phi\{\Phi^{-1}(c)\}^2}{c^2}$.
- Numerical derivative is $\dot{w}(c)=\{w(c+h)-w(c)\}/h$, where $h=10^{-7}$.
- Equation (4) gives expression of $\dot{w}(c)$.

Appendix A: Data on scores from examination and homework (Chung, 2013)

The example is due to Chung (2013). Table A records two evaluation items: X^O = *score from examination* and Y^O = *score from homework* are recorded for 10 students. Define a weighted mean $Z^O=0.75X^O+0.25Y^O$. Students receive a "fair" if $Z^O\geq 60$ and a "fail" if $Z^O<60$, respectively. Suppose that a teacher can obtain samples for those students with $Z^O\geq 60$. Then, $Z^O\geq 60$ is equal to $X^O\geq L^O$, where $L^O=80-Y^O/3$. Accordingly, the observed data is (L_j, X_j), subject to $L_j\leq X_j$, for $j=1,2,\ldots,7$ in which 7 out of 10 samples are included.

Table A: Test score data for two items X^O = score from examination, and Y^O = score from homework for 10 students.

Student ID	1	2	3	4	5	6	7	8	9	10
X^O	95	100	40	45	60	75	55	60	70	90
Y^O	65	30	100	50	70	80	40	70	85	85

NOTE: The scores for ID=1, 2 and 3 are taken from Chung (2013). The scores for ID = 4 ~ 10 are randomly generated.

Appendix B: R codes for calculations

```
X0 = c(95,100,40,45,60,75,55,60,70,90) ## examination score ##
Y0=c(65,30,100,50,70,80,40,70,85,85) ## homework score ##
Z0=0.75*X0+0.25*Y0
L0=80-Y0/3
X = X0[(X0>=L0)]
L = L0[(X0>=L0)]
n=length(X)

cf = function(muX,muL,varX,varL,covLX){
  pnorm((muX-muL)/sqrt(varX+varL-2*covLX))
}

dot_cf = function(muX,muL,varX,varL,covLX){
  dnorm((muX-muL)/sqrt(varX+varL-2*covLX))*(-(muX-muL)/2*(varX+varL-2*covLX)^(-3/2))
}

l = function(n,muX,muL,varX,varL,covLX){
  D=1/(varL*varX-covLX^2)*(varX*(L-muL)^2-2*covLX*(L-muL)*(X-muX)+varL*(X-muX)^2)
  -n*log(cf(muX,muL,varX,varL,covLX))-n*log(2*pi)-n/2*log(varL*varX-covLX^2)-1/2*sum(D)
}

h = 10^-7
l_dot_numerical = c()
l_dot_corrected = c()
l_dot_incorrect = c()

for( i in 1:4 ){
  l_dot_numerical[i] = (l(n,10*i,20*i,30*i,40*i+h,0)-l(n,10*i,20*i,30*i,40*i,0))/h
  l_dot_corrected[i] = -n*dot_cf(10*i,20*i,30*i,40*i,0)/cf(10*i,20*i,30*i,40*i,0)+
    sum(-1/(40*i*2)+(L-20*i)^2/(2*(40*i)^2))
  l_dot_incorrect[i] = -n*dot_cf(10*i,20*i,30*i,40*i,0)/cf(10*i,20*i,30*i,40*i,0)+
    sum(-40*i/2+(L-20*i)^2/(2*(40*i)^2))
}
l_dot_numerical
l_dot_corrected
l_dot_incorrect

##########################################################################################
w = function(c){                                    # This is w functon in paper.
  qnorm(c)*dnorm(qnorm(c))/c+dnorm(qnorm(c))^2/c^2
}

c = seq(0,1,by = 0.001)
```

```
plot(c,w(c),type = "l",main="Fig.1")

w_dot_a = function(c){                              #This function is first order of w by approximate.
   (w(c+h)-w(c))/h
}

w_dot_b = function(c){                              #This function is first order of w by directly differential.
   ((1-qnorm(c)^2)*c-qnorm(c)*dnorm(qnorm(c)))/c^2-
      2*dnorm(qnorm(c) )/c^3*( c*qnorm(c)+ dnorm(qnorm(c)) )
}

h = 10^(-7)
w_dot_1 = c()
w_dot_2 = c()
for( i in 1:9)
{
   w_dot_1[i] = w_dot_a(i*0.1)
   w_dot_2[i] = w_dot_b(i*0.1)
}
w_dot_1
w_dot_2
w(seq(0.1,0.9,by = 0.1))
```

ACKNOWLEDGMENTS This work was financially supported by the National Science Council of Taiwan (NSC101-2118-M008-002-MY2).

REFERENCES

Chung H (2013) Application of approximate reasoning using triangular and sine-curved membership functions, V.E. Balas et al. (Eds.): *New Concepts and Applications in Soft Computing*, pp. 141–155, Springer-Verlag Berlin Heidelberg.

Emura T, Konno Y (2012) Multivariate normal distribution approaches for dependently truncated data. *Stat Papers* 53:133-149