Estimation of the correlation coefficient with left censored and repeated measures data

Agustin Calatroni
Cindy Visness
Herman Mitchell

Rho Federal Systems Division, Inc.

Abstract

Cytokines are released by cells in the immune system and inform other cells to generate an immune response. Different cytokines are released depending on the specific immune response required. The correlations among cytokine responses are of interest to immunologists. It is common for laboratory assays of cytokine response to be performed in duplicates or triplicates for quality control purposes. These assays normally have a limit below which the cytokine response cannot be measured, resulting in left censoring.

In order to account simultaneously for left censoring and repeated measures, a maximum likelihood method is proposed to estimate the correlation coefficient using SAS® NLMIXED. The approach is illustrated with data from the Urban Environment and Childhood Asthma Study (URECA).

Objective

- Estimation of the correlation with left-censored data on both variables (derive the bivariate normal distribution with left-censored data).
- Estimation of the correlation with left-censored data and repeated measures.
- Put everything together with an example form the Urban Environment and Childhood Asthma Study (URECA).

Correlation with left-censored data

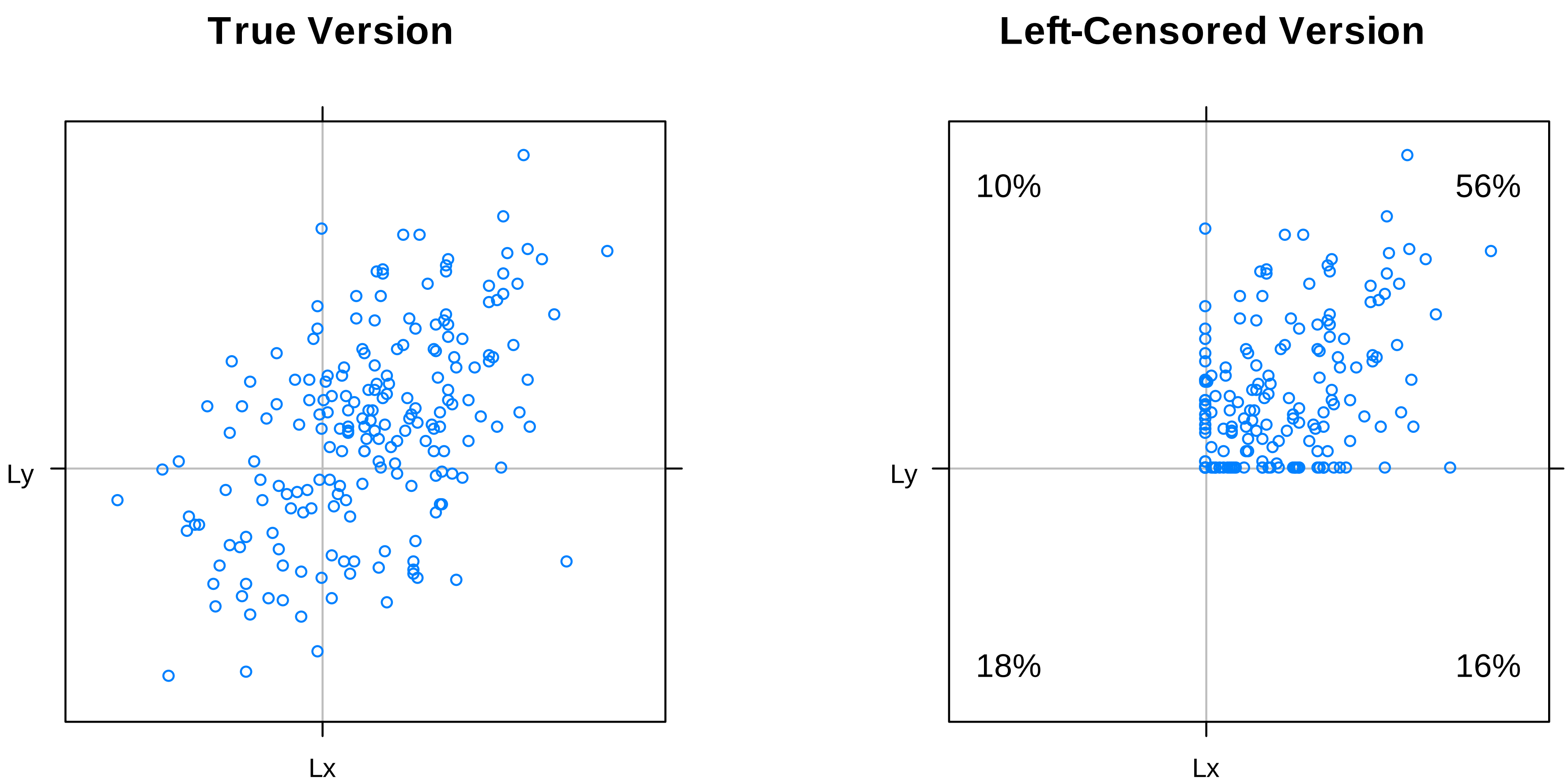
- (X_i, Y_i) normalized variables from the i^{th} subject
- (X_i, Y_i) left-censored bivariate normal distribution:
  - mean (μ_X, μ_Y)
  - covariance matrix $\sum = \begin{bmatrix} \sigma_X^2 & \sigma_X \sigma_Y \rho_{XY} \\ \sigma_X \sigma_Y \rho_{XY} & \sigma_y^2 \end{bmatrix}$
  - $L_X$and L_Y fixed lower limits of detection

Correlation with left-censored data

To derive the likelihood function we note that there are four possible types of observed pairs:

(1) Pairs with both X and Y observed
(2) Pairs with X observed and $Y < L_Y$
(3) Pairs with Y observed and $X < L_X$
(4) Pairs with $X < L_X$ and $Y < L_Y$

Graphical Illustration



Pairs with both X and Y observed

Pairs of type 1 contribute the likelihood

$$f(y_i, x_i) = f(y_i | x_i) f(x_i)$$

$$\left(2\pi\sigma_X\sigma_{Y|X}\right)^{-1} \exp\left\{-0.5\left[\frac{\left(y_i - \mu_{Y|X}\right)^2}{\sigma_{Y|X}^2} + \frac{(x_i - \mu_X)^2}{\sigma_X^2}\right]\right\}$$

$$\mu_{Y|X} = \mu_Y + \left(\frac{\rho_{XY}\sigma_Y}{\sigma_X}\right)(x_i - \mu_X)$$

$$\sigma_{Y|X}^2 = \sigma_Y^2(1 - \rho_{XY}^2)$$

Pairs with X observed and $Y < L_Y$

Pairs of type 2 contribute the likelihood

$$\Pr(Y < L_Y | X = x_i) \times f(x_i)$$

$$(2\pi\sigma_X^2)^{-1/2} \exp\left[-0.5\,\frac{(x_i - \mu_x)^2}{\sigma_X^2}\right] \times \Phi\left(\frac{\mathrm{L_Y} - \mu_{\mathrm{Y|X}}}{\sigma_{Y|X}^2}\right)$$

$$\mu_{Y|X} = \mu_Y + \left(\frac{\rho_{XY}\sigma_Y}{\sigma_X}\right)(x_i - \mu_X)$$

$$\sigma_{Y|X}^2 = \sigma_Y^2(1 - \rho_{XY}^2)$$

Pairs with Y observed and $X < L_X$

Pairs of type 3 contribute the likelihood

$$\Pr(X < L_X | Y = y_i) \times f(y_i)$$

$$(2\pi\sigma_Y^2)^{-1/2} \exp\left[-0.5\,\frac{(y_i - \mu_Y)^2}{\sigma_Y^2}\right] \times \Phi\left(\frac{\mathrm{L_Y} - \mathrm{\mu_{X|Y}}}{\sigma_{X|Y}^2}\right)$$

$$\mu_{X|Y} = \mu_X + \left(\frac{\rho_{XY}\sigma_X}{\sigma_Y}\right)(y_i - \mu_Y)$$

$$\sigma_{X|Y}^2 = \sigma_X^2(1 - \rho_{XY}^2)$$

Pairs with $X < L_X$ and $Y < L_Y$

Pairs of type 4 contribute the likelihood

$$\Pr\left[(\mathrm{Y} < \mathrm{L_Y}) \cap (\mathrm{X} < \mathrm{L_X})\right]$$

$$\int_{-\infty}^{L_Y} \Phi\left\{\frac{\mathrm{L_x} - \left[\mu_\mathrm{x} + \frac{\rho_\mathrm{XY}\sigma_\mathrm{X}(\mathrm{y} - \mu_\mathrm{Y})}{\sigma_\mathrm{Y}}\right]}{\sigma_\mathrm{X}\sqrt{1 - \rho_\mathrm{XY}}}\right\} \times (2\pi\sigma_Y^2)^{-1/2} \exp\left[-0.5\,\frac{(y - \mu_Y)^2}{\sigma_Y^2}\right] dy$$

I used a Riemann integral for a close numerical approximation of this integral.

SAS® NLMIXED

```
proc nlmixed data=URECA_Data;
 parms […]; bounds […];
mu_yx = mu_y + (rho*sigma_y) * ((x-mu_x)/sigma_x);
 sigma_yx = sigma_y*sqrt( (1-rho**2) );
  mu_xy = mu_x + (rho*sigma_x/sigma_y)*(y-mu_y);
 sigma_xy = sigma_x*sqrt( (1-rho**2) );

 if x^=lx and y^=ly then
  like = (1/sqrt(2*constant('PI')*(sigma_x*sigma_x)))*exp(-0.5*((x-mu_x)**2)/(sigma_x*sigma_x)) *
         (1/sqrt(2*constant('PI')*(sigma_yx*sigma_yx)))*exp(-0.5*((y-mu_yx)**2)/(sigma_yx*sigma_yx));
 if x=lx and y^=ly then
   like = probnorm((x-mu_x)/(sigma_x*sigma_x)) *
          (1/sqrt(2*constant('PI')*(sigma_yx*sigma_yx)))*exp(-0.5*((y-mu_yx)**2)/(sigma_yx*sigma_yx));
 if x^=lx and y=ly then
   like = (1/sqrt(2*constant('PI')*(sigma_x*sigma_x)))*exp(-0.5*((x-mu_x)**2)/(sigma_x*sigma_x)) *
          probnorm((y-mu_yx)/(sigma_yx*sigma_yx));
 if x=lx and y=ly then do;
   like = 0;
  do y = 0 to ly by 0.01;
    like_p = probnorm((x-mu_xy)/(sigma_xy*sigma_xy)) *
               (1/sqrt(2*constant('PI')*(sigma_y*sigma_y)))*exp(-0.5*((y-mu_y)**2)/(sigma_y*sigma_y)) / 2;
    like = like + like_p*0.01;
   end;
  end;
ll = log(like);
 model ll ~ general(ll); run;
```

Simulation Study

% left-censored (X,Y)	ρ_{XY}	Method	Mean $\widehat{\rho_{XY}}$	95% CI
(30,35)	0.50	$\frac{1}{2}$ LOD	0.42	(0.40, 0.43)
		$\frac{1}{\sqrt{2}}$ LOD	0.44	(0.43, 0.45)
		ML	0.50	(0.49, 0.52)

100 simulations with 250 pairs per simulation

Correlation with repeated measures

- (X_{ij}, Y_{ij}) normalized variables from the i^{th} subject at the j^{th} repeated observation
- $\left(X_{ij}, Y_{ij}\right)$ left-censored bivariate normal distribution:
  - mean (μ_X, μ_Y)
  - covariance matrix $\sum = \begin{bmatrix} \sigma_X^2 & \sigma_X \sigma_Y \rho_{XY} \\ \sigma_X \sigma_Y \rho_{XY} & \sigma_y^2 \end{bmatrix}$
  - $L_X$and L_Y fixed lower limits of detection

Correlation with repeated measures

Ad hoc methods:

- Ignore the repeated measures and treat the data as a simple random sample.
- Obtain mean response for each variable and each subject.
- Compute weighted correlation using the number of observations per subject as the weights.

Correlation with repeated measures

Relationship between the $X's$ and $Y's$

$$
\begin{array}{ccccc}
Time_{ij} & & & & Time_{ij`} \\
X_{ij} & \leftarrow & \rho_X & \rightarrow & X_{ij`} \\
\uparrow & & & \nearrow & \uparrow \\
\rho_{XY} & & \delta\rho_{XY} & & \rho_{XY} \\
\downarrow & \swarrow & & & \downarrow \\
Y_{ij} & \leftarrow & \rho_Y & \rightarrow & Y_{ij`}
\end{array}
$$

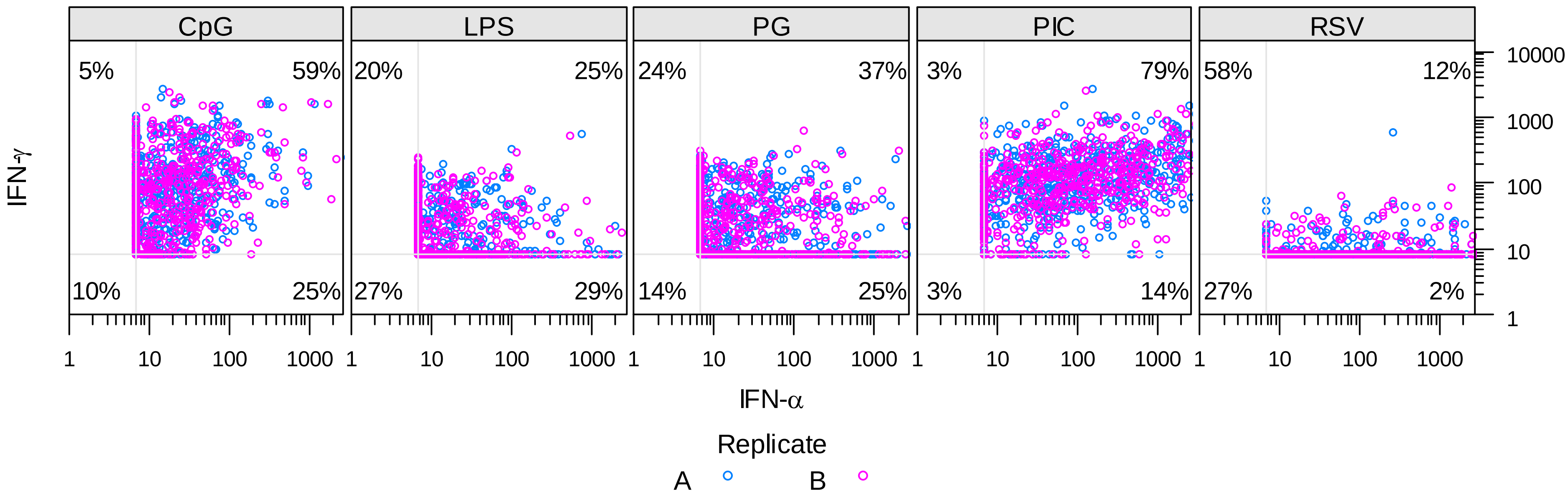
URECA Example

The Urban Environment and Childhood Asthma (URECA) study is a birth cohort study designed to explore the immunologic and environmental factors leading to the development of asthma among children at risk.

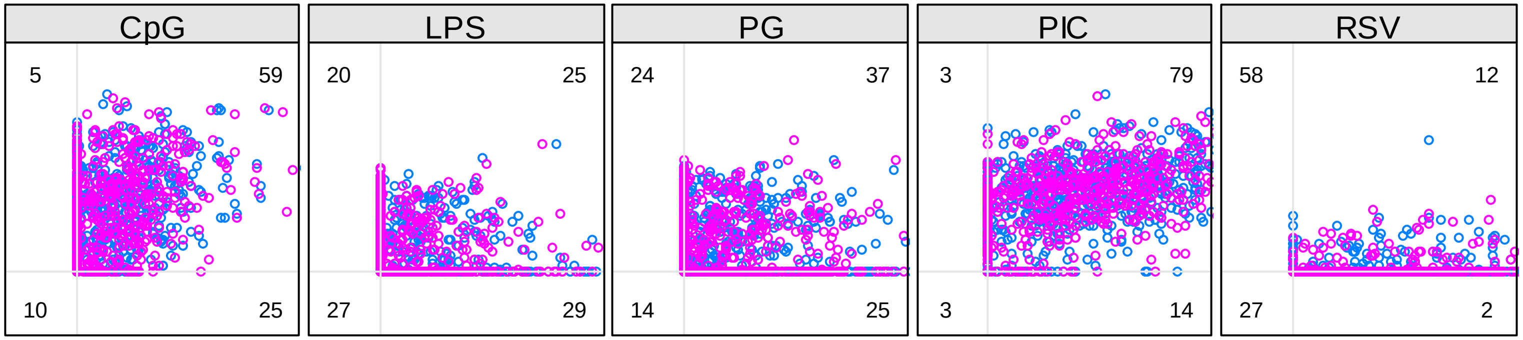
Cytokines are messenger proteins of the immune system that are released by cells as part of the immune response.

Cytokine responses are measured in the lab in response to stimulants that represent bacteria, viruses, allergens, etc.

URECA Example



URECA Example



	CpG (36,15)%	LPS (55,46)%	PG (39,38)%	PIC (17,7)%	RSV (29,86)%
½ NRep	0.39 (0.34,0.44)	0.00 (-0.06,0.05)	-0.11 (-0.17,-0.05)	0.46 (0.42,0.51)	0.17 (0.11,0.23)
ML NRep	0.25 (0.19,0.31)	-0.08 (-0.15,-0.01)	-0.06 (-0.14,0.02)	0.46 (0.41,0.52)	0.62 (0.58,0.67)
½ Rep	0.48 (0.44,0.51)	0.34 (0.31,0.37)	0.46 (0.39,0.53)	0.45 (0.41,0.48)	0.76 (0.75,0.77)
ML Rep	0.27 (0.23,0.31)	0.03 (0.00,0.07)	0.10 (0.05,0.14)	0.61 (0.58,0.63)	0.49 (0.44,0.54)

Extensions

- Right-censored, interval-censored data
- Calculate the concordance correlation
- Bootstrap Confidence Intervals
- Markov chain Monte Carlo (MCMC) simulation procedure with SAS® MCMC (Version 9.2)

References

1. Correlating two viral load assays with known detection limits Lyles RH, Williams JK, Chuachoowong R. *Biometrics* 2001; 57:1238-1244.

2. Mixed Models for Assessing Correlation in the Presence of Replication Hamlett A, Ryan L, Serrano-Trespalacios P, Wolfinger, R. *Journal of the Air & Waste Management Association* 2003, 53:442-450.

3. SAS Institute, Inc. (2004). SAS/STAT User's Guide, Version 9.1, Vol. 1-9. Cary, NC: SAS Institute, Inc.