A Replication Study

Bayesian Inference for Agreement Measures

Lingjuan Qi[1] Matthew Aaron Looney[2]

[1]Department of Mechanical Engineering
[2]Department of Applied Economics
Texas Tech University

July 3, 2018

Outline

- Introduction and Background
- Motivating Example - Part I
- Accessing Agreement
- Bayesian Model
- Simulation Results
- Motivating Example - Part II (Kidney Data Study)
- Conclusion

Introduction and Background

Goals of this Research

Replicate the results of:

- **Bayesian Inference for Agreement Measures.**[1]
  By: Ignacio Vidal and Mário de Castro.
  Journal of biopharmaceutical statistics, 2016 p. 809-823.

[1] Vidal, I., & de Castro Journal of biopharmaceutical statistics, M. (2017). Bayesian inference for agreement measures. Taylor & Francis , 27(5), 809–823. http://doi.org/10.1080/10543406.2016.1226323

Goals of this Research

The main ideas behind the paper:

- The agreement between a set of paired observations is important.
- Frequentist approach is commonly used.
- Frequentist approach yields point estimates; CI and SE are not easily available.
- A method which is general ,simple, effective, and doesn't require MCMC is presented.

Goals of this Research

Our Goals

- Study the method of agreement measures
- Implement different prior distributions
- Replicate simulation
- Real data study

Motivating Example - Part I

A Motivating Example with Real World Data:

Renal lithiasis (Kidney Stones) can be defined as the consequence of an alteration of the normal crystallization conditions of urine in the urinary tract. In a healthy individual, during the residence time of urine in the urinary tract, crystals either do not form or are so small they are eliminated uneventfully. Clinicians often employ diagnostic imaging techniques to quantify the size, shape, and likelihood the stones will pass through the system without incident.

A Motivating Example with Real World Data:

There are two main imaging tools used by clinicians to assess renal lithiasis:

- High-speed or dual energy Computerized Tomography (CT).
- Intravenous Urography, which involves injecting dye into an arm vein and taking X-rays (intravenous pyelogram).

A Motivating Example with Real World Data:

- The imaging costs associated with Computerized Tomography far out weigh the costs of imaging with alternative methods, including Intravenous Urography.
- The question is: do these two imaging modalities provide equivalently adequate results in terms of image quality, accuracy and precession.
- We need a way to assess how much agreement there is between measurements taken by Urography and Tomography.

A Motivating Example with Real World Data:

- We have data on the inferior pelvic infundibular angle (IPIA) for 52 kidneys, evaluated by means of computerized tomography and urography.
- We will use this data to determine the level of agreement between the two techniques.
- We will return to this example later...

Accessing Agreement

Assessing Agreement Between Measurements:

- A way to assess agreement between two random variables X and Y is the mean squared deviation (MSD).

$$MSD = E\left[(X - Y)^2\right] = (\mu_x - \mu_y)^2 + \sigma_x^2 + \sigma_y^2 - 2\sigma_{xy}$$

Assessing Agreement Between Measurements:

- Another measure of agreement between two random variables is the Concordance Correlation Coefficient (CCC).

$$\rho_c = 1 - \frac{MSD}{MSD|_{\sigma_{xy}=0}} = \frac{2\sigma_{XY}}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2}$$

Assessing Agreement Between Measurements:

- Two additional methods for assessing agreement that were explored by the authors are called the accuracy coefficient (χ_a) and the precision coefficient (ρ).

$$\chi_a = \frac{2}{\bar{\omega} + \frac{1}{\bar{\omega}} + v^2}, \bar{\omega} = \frac{\sigma^2{}_Y}{\sigma_X^2}, v^2 = \frac{(\mu_X - \mu_Y)^2}{\sigma_X \sigma_Y}$$

$$\rho = \frac{\rho_c}{\chi_a}$$

where ρ is simply the Pearson correlation coefficient ($\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$).

Bayesian Model

Likelihood

$$p(X|\mu,\Sigma) \propto \Sigma^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)\right\}$$

$$p(X|\mu,\Sigma) \propto \Sigma^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} tr(\Sigma^{-1}S)\right\}$$

where S is the matrix of sum of squares (sometimes called the scatter matrix).

$$S = \sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$$

Prior

The natural conjugate prior is normal-inverse-wishart:

$$\Sigma \sim \mathrm{IW}_{v0}\left(\Lambda_0^{-1}\right)$$

$$\mu|\Sigma \sim N(\mu_0, \Sigma/k_o)$$

$$p(\mu, \Sigma) = NIW(\mu_0, k_0, \Lambda_0, v_0)$$

$$p(\mu, \Sigma) = \frac{1}{Z}\Sigma^{-\Omega}\exp\left\{-\frac{1}{2}tr(\Lambda_0\Sigma^{-1}) - \frac{k_0}{2}(\mu-\mu_0)'\Sigma^{-1}(\mu-\mu_0)\right\}$$

where $z = \frac{2^{v_0 df/2}\Gamma_{df}(v_0/2)(2\pi/k_0)^{df/2}}{\Lambda_0{}^{v_0/2}}$; and $\Omega = [(v_0 + df)/2 + 1]$

Prior

Vidal and de Castro consider five priors in their research; Jeffreys' Prior, Independence Jefferys' Prior, right-Haar prior, and two reference priors, one for ρ and one for σ. We focus on the following three prior choices:

Jeffreys' Prior

Independence Jefferys' Prior

Reference Prior for ρ

Prior

General formula for the objective prior family:

$$\pi_{ab}(\mu, \Sigma) \propto \frac{1}{\sigma_x^{3-a}\sigma_y^{2-b}(1-\rho^2)^{2-b/2}}$$

Prior

Jeffreys' Prior

Jeffreys' Prior$\propto \pi_J(\mu, \Sigma) \propto \pi_{10}(\mu, \Sigma)$

$$\pi_J(\mu, \Sigma) \propto \frac{1}{\sigma_x^2 \sigma_y^2 (1 - \rho^2)^2}$$

Prior

Independence Jefferys Prior

Independence Jeffreys' Prior $\propto \pi_{IJ}(\mu, \Sigma) \propto \pi_{21}(\mu, \Sigma)$

$$\pi_{IJ}(\mu, \Sigma) \propto \frac{1}{\sigma_x \sigma_y (1 - \rho^2)^{-3/2}}$$

Prior

Reference Prior for ρ

$$\pi_{R\rho}(\mu, \Sigma) \propto \frac{1}{\sigma_x \sigma_y (1 - \rho^2)}$$

Posterior

$$p(\mu, \Sigma | X, \mu_0, k_0 \Lambda_0, v_0) = NIW(\mu, \Sigma | \mu_n, k_n, \Lambda_n, v_n)$$

$$\propto \Sigma^{-\{(v_n + df)/2\}+1} \exp\left\{ -\frac{1}{2} tr(\Phi_n \Sigma^{-1}) - \frac{k_n}{2} (\mu - \mu_n)' \Sigma^{-1} (\mu - \mu_n) \right\}$$

The posterior distribution is:
$p(\mu, \Sigma | X) \sim Normal - InverseWishart(\mu_n, k_n^{-1}, \Phi_n^{-1}, v_n)$

Or more succinctly,

$$\Sigma | x, y \sim IW_2 \left(S^{-1}, n \right)$$

$$\mu | \Sigma, x, y \sim \; N_2 \left[(\bar{x}, \bar{y})', n^{-1} \Sigma \right]$$

Posterior

Drawing from the posterior distribution:

Jeffreys' Prior:

$$\mu|\Sigma, x, y \sim N_2[(\bar{x}, \bar{y})', n^{-1}\Sigma] \quad \text{and} \quad \Sigma|x, y \sim IW_2(S^{-1}, n)$$

Independence Jefferys Prior:

$$\mu|\Sigma, x, y \sim N_2[(\bar{x}, \bar{y})', n^{-1}\Sigma] \quad \text{and} \quad \Sigma|x, y \sim IW_2(S^{-1}, n-1)$$

Posterior

Drawing from the posterior distribution:

Reference Prior for ρ

Must employ acceptance-rejection algorithm:

Simulation step. Generate $(\sigma_x, \sigma_y, \rho)$ from the IW_2 using $\pi_{IJ}(\mu, \Sigma)$ and independently generate u from $Uniform(0, 1)$.

Rejection step. If $u \leq \pi_{R\rho}(\mu, \Sigma)/\pi_{IJ}(\mu, \Sigma)$, then accept $(\sigma_x, \sigma_y, \rho)$ else, return to the simulation step. Upon acceptance, generate (μ_x, μ_y) from the N_2 distribution under the independence Jeffreys' prior.

Simulation Results

Generating Simulation Data

For each parameter combination given in the below Table, we generate 1000 data sets with sample size $n = (30, 100)$ and estimate MSD, CCC, precision (ρ), and accuracy (χ_a). Simulation data was drawn from the multivariate normal distribution.

Combination	μ_x	μ_y	σ_x^2	σ_y^2	ρ
1	100	100	100	100	0.99
2					0.90
3					0.80
4					0.40
5	100	100	100	125	0.99
6					0.90
7					0.80
8					0.40

Simulation Results

Averages of the posterior mean of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

		Jeffreys Prior			
		Original		Replication	
Comb	Measure	n=30	n=100	n=30	n=100
1	MSD	2.197	2.043	7.334	2.041
	CCC	0.989	0.990	0.986	0.989
	Precision	0.990	0.990	0.986	0.989
	Accuracy	0.999	1.000	0.999	0.999
6	MSD	26.008	24.328	8.266	2.300
	CCC	0.882	0.890	0.783	0.800
	Precision	0.90	0.899	0.803	0.820
	Accuracy	0.983	0.991	0.999	0.975

Simulation Results

Averages of the posterior mean of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

		Independence Jeffreys			
		Original		Replication	
Comb	Measure	n=30	n=100	n=30	n=100
1	MSD	2.275	2.063	7.614	2.062
	CCC	0.989	0.990	0.993	0.993
	Precision	0.990	0.990	0.993	0.993
	Accuracy	0.999	1.000	0.999	0.999
6	MSD	26.930	24.572	8.584	2.324
	CCC	0.882	0.890	0.805	0.810
	Precision	0.900	0.899	0.825	0.831
	Accuracy	0.983	0.991	0.975	0.975

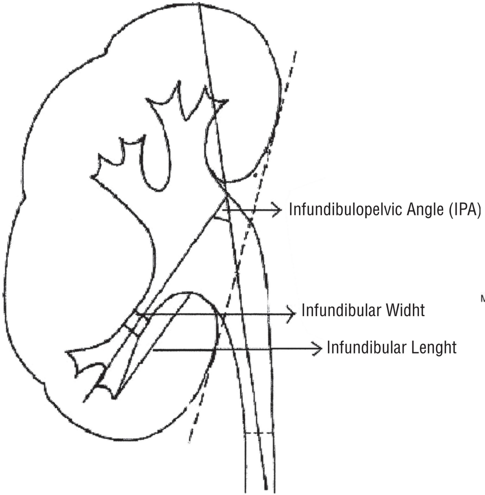
Simulation Results

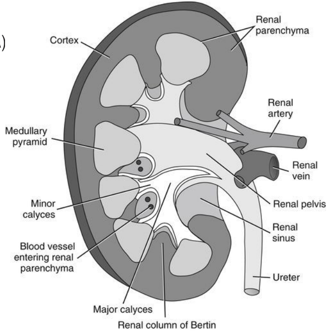
Averages of the posterior mean of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

		Reference Prior ρ			
		Original		Replication	
Comb	Measure	n=30	n=100	n=30	n=100
1	MSD	.	.	7.325	2.062
	CCC	.	.	0.400	0.971
	Precision	.	.	0.400	0.971
	Accuracy	.	.	0.999	0.999
6	MSD	.	.	8.218	2.324
	CCC	.	.	0.276	0.790
	Precision	.	.	0.283	0.810
	Accuracy	.	.	0.975	0.975

Motivating Example - Part II (Kidney Data Study)

Inferior Pelvic Infundibular Angle Data





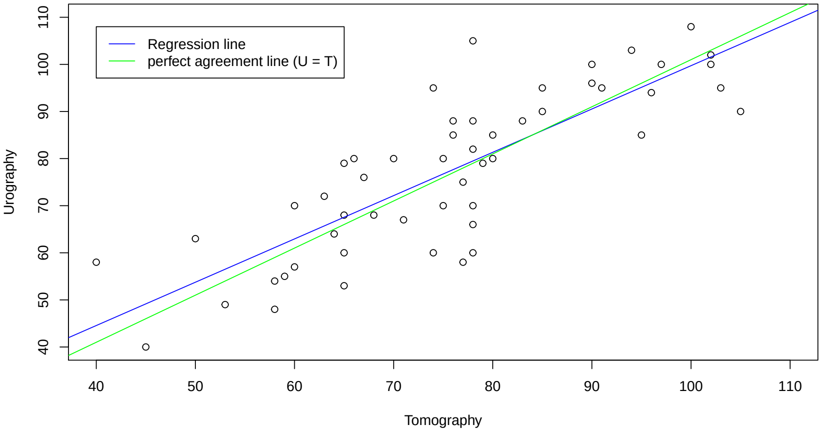
Data Summary

Vidal and de Castro reported different values for the range of the data and the means and standard deviations: Urography *range*$(40^\circ - 105^\circ)$; $\mu = 75.8^\circ$; $sd = 15.3^\circ$ and Tomography *range*$(40^\circ - 108^\circ)$; $\mu = 79.5^\circ$; $sd = 17.2^\circ$. The original data reported values consistent with below table.

	Urography	Tomography
N.Valid	52.00	52.00
Mean	77.46	75.79
Median	79.50	76.50
Std.Dev	17.17	15.30
Min	40.00	40.00
Max	108.00	105.00

A visual assessment

- Regression line plot versus perfect agreement line (U = T).



Kidney Data - Bayesian Inference for:

- Mean Squared Deviation (MSD).
- Concordance Correlation Coefficient (CCC).
- Accuracy coefficient (χ_a).
- Precision coefficient (ρ).

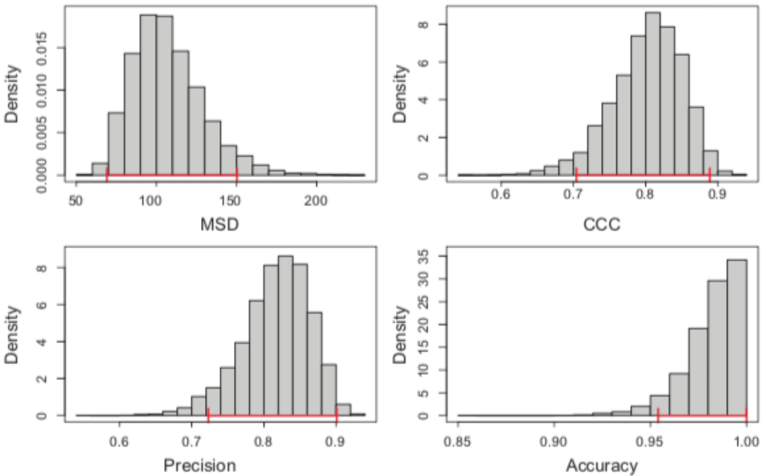
Kidney Data - Bayesian Inference

Results obtained using the Jeffreys' prior with 5000 posterior draws over 1000 replications. The average values over the Posterior distribution are reported below. Equal-Tail CI are reported in brackets.

Measure	Original	Replication
MSD	104.407	4.803
	(69.062, 150.126)	(4.792, 4.814)
CCC	0.808	0.6465
	(0.704, 0.889)	(0.644, 0.648)
Precision	0.823	0.832
	(0.723, 0.901)	(0.831, 0.833)
Accuracy	0.985	0.776
	(0.954, 1.000)	(0.775, 0.777)

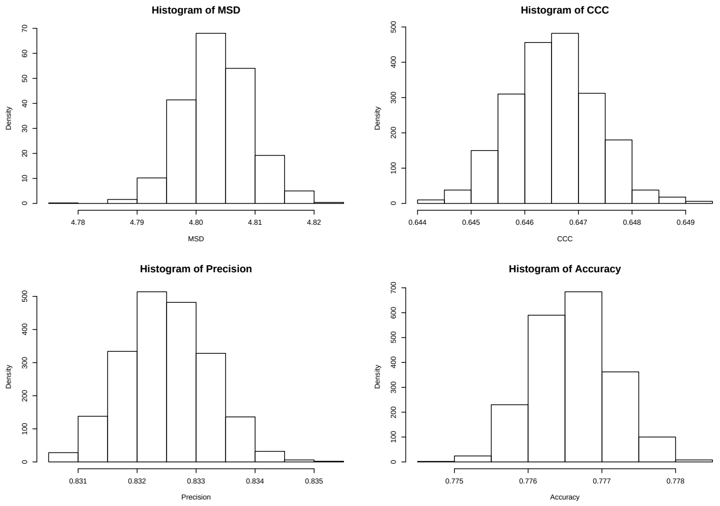
Kidney Data - Bayesian Inference

Authors results:



Kidney Data - Bayesian Inference

From replication study:



Conclusion

Conclusion

- We have successfully replicated a large portion of Vidal's and de Castro's results.
- There is a sizable discrepancy between several of our results and the authors.
- The results from several simulation studies match the authors reported results closely.
- However, several simulation results did not agree.
- The authors reported summary statistics for the real world data study do not agree with the original data source.
- The replication results from the real world data study do not agree.
- It is difficult to know for sure if we are working on the exact data as the authors.

Replication Results

Generating Simulation Data

For each parameter combination given in the below Table, we generate 1000 data sets with sample size $n = (30, 100)$ and estimate MSD, CCC, precision (ρ), and accuracy (χ_a). Simulation data was drawn from the multivariate normal distribution.

Combination	μ_x	μ_y	σ_x^2	σ_y^2	ρ
1	100	100	100	100	0.99
2					0.90
3					0.80
4					0.40
5	100	100	100	125	0.99
6					0.90
7					0.80
8					0.40

Generating Simulation Data

Combination	μ_x	μ_y	σ_x^2	σ_y^2	ρ
9	100	105	100	125	0.99
10					0.90
11					0.80
12					0.40
13	135	135	88	88	0.99
14					0.90
15					0.80
16					0.40

Simulation Results

Averages of the posterior mean of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

		Jeffreys Prior			
		Original		Replication	
Comb	Measure	n=30	n=100	n=30	n=100
1	MSD	2.197	2.043	7.334	2.041
	CCC	0.989	0.990	0.986	0.989
	Precision	0.990	0.990	0.986	0.989
	Accuracy	0.999	1.000	0.999	0.999
6	MSD	26.008	24.328	8.2666	2.300
	CCC	0.882	0.890	0.783	0.800
	Precision	0.90	0.899	0.803	0.820
	Accuracy	0.983	0.991	0.975	0.975

Simulation Results

Averages of the posterior mean of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

		Independence Jeffreys			
		Original		Replication	
Comb	Measure	n=30	n=100	n=30	n=100
1	MSD	2.275	2.063	7.614	2.062
	CCC	0.989	0.990	0.993	0.993
	Precision	0.990	0.990	0.993	0.993
	Accuracy	0.999	1.000	0.999	0.999
6	MSD	26.930	24.572	8.584	2.324
	CCC	0.882	0.890	0.805	0.810
	Precision	0.900	0.899	0.825	0.831
	Accuracy	0.983	0.991	0.975	0.975

Simulation Results

Averages of the posterior mean of some measures from 1000 replications and 5000 samples from the posterior distribution under selected parameter combinations (Comb.) and priors.

		Reference Prior ρ			
		Original		Replication	
Comb	Measure	n=30	n=100	n=30	n=100
1	MSD	.	.	7.325	2.062
	CCC	.	.	0.400	0.971
	Precision	.	.	0.400	0.971
	Accuracy	.	.	0.999	0.999
6	MSD	.	.	8.218	2.324
	CCC	.	.	0.276	0.790
	Precision	.	.	0.283	0.810
	Accuracy	.	.	0.975	0.975

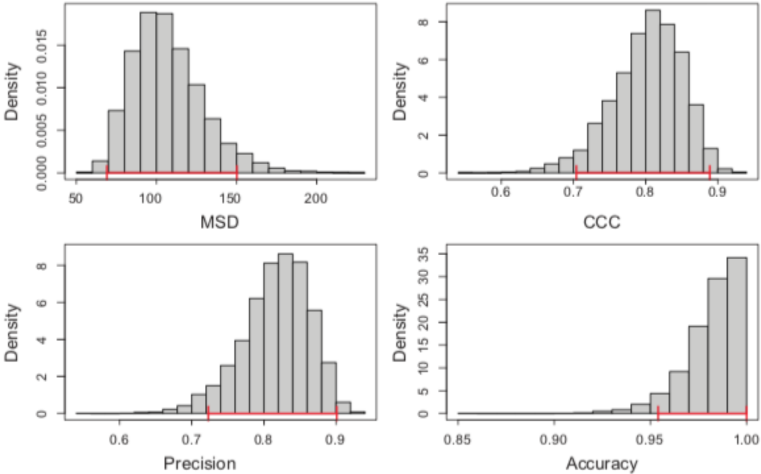
Kidney Data - Bayesian Inference

Results obtained using the Jeffreys' prior with 5000 posterior draws over 1000 replications. The average values over the Posterior distribution are reported below. Equal-Tail CI are reported in brackets.

Measure	Original	Replication
MSD	104.407	4.803
	(69.062, 150.126)	(4.792, 4.814)
CCC	0.808	0.6465
	(0.704, 0.889)	(0.644, 0.648)
Precision	0.823	0.832
	(0.723, 0.901)	(0.831, 0.833)
Accuracy	0.985	0.776
	(0.954, 1.000)	(0.775, 0.777)

Kidney Data - Bayesian Inference

Authors results:



Kidney Data - Bayesian Inference

From replication study: