Learning our ABCs

Nathan Cunningham Sherman Ip Ella Kaye

February 12th, 2016

Outline

- Approximate Bayesian computation
- ABC with conjugate priors
- ABC and semi-automatic ABC with the g-and-k distribution

Background and Motivation

Full Posterior

$$\pi(\theta|y_{obs}) = \frac{p(y_{obs}|\theta)\pi(\theta)}{p(y_{obs})}$$

Background and Motivation

Full Posterior

$$\pi(\theta|y_{obs}) = \frac{p(y_{obs}|\theta)\pi(\theta)}{p(y_{obs})}$$

Partial Posterior

$$\pi(\theta|s_{obs}) = \frac{p(s_{obs}|\theta)\pi(\theta)}{p(s_{obs})}$$

where $s_{obs} = S(y_{obs})$ is a vector of summary statistics (preferably sufficient) of lower dimension than the data

ABC approximation (rejection algorithm)

$\pi_\epsilon(\theta|y_{obs})$, is generated by repeatedly

- Generating θ' from the prior distribution $\pi(\cdot)$
- Generating y_i from the likelihood $p(\cdot|\theta')$
- Accepting θ' if $\rho\{S(y_i), S(y_{obs})\} \leq \epsilon$

ABC approximation (rejection algorithm)

$\pi_\epsilon(\theta|y_{obs})$, is generated by repeatedly

- Generating θ' from the prior distribution $\pi(\cdot)$
- Generating y_i from the likelihood $p(\cdot|\theta')$
- Accepting θ' if $\rho\{S(y_i), S(y_{obs})\} \leq \epsilon$

Requires tuning of

- $S(\cdot)$, the function to compute summary statistics
- ρ, the distance function
- ϵ, the threshold

ABC with local linear regression...

- Improve the previous algorithm by introducing a regression step to correct for discrepancy between $S(y_{obs})$ and $S(y_i)$
- For each simulated y_i, set

$$\theta_i = m(S(y_i)) + \eta_i$$

where m is the regression function and the η_i are centered random variables with common variance.

...and kernel

- Instead of accepting or rejecting a sample, assign each simulation $(\theta_i, S(y_i))$ a weight with a kernel

$$K[\rho(S(y_i), S(y_{obs}))]$$

- Higher weights given for proximity to the observation
- We use the Epanechnikov kernel and the Euclidean distance function

Principled summary statistic selection

- Best subset selection

Principled summary statistic selection

- Best subset selection
- Semi-automatic ABC
  1. Inital run of ABC to find a posterior region of non-negligible mass
  2. Simulate parameter values and summary statistics
  3. Run a linear regression of the summary statistics on each parameter. Use an information criterion such as BIC to determine which summary statistics can best be used to infer the parameter values
  4. Run ABC using this choice of summary statistics

Beta prior, binomial likelihood

$$\theta \sim \text{Beta}(1, 1) \tag{1}$$

$$X|\theta \sim \text{Bin}(12, \theta) \tag{2}$$

$$\theta|X = 2 \sim \text{Beta}(3, 11) \tag{3}$$

Beta prior, binomial likelihood

$$\theta \sim \text{Beta}(1,1) \tag{1}$$

$$X|\theta \sim \text{Bin}(12,\theta) \tag{2}$$

$$\theta|X = 2 \sim \text{Beta}(3,11) \tag{3}$$

ABC Algorithm

- Sample $\theta \sim \text{Beta}(1,1)$
- Sample $X|\theta \sim \text{Bin}(12,\theta)$
- Accept θ if $X|\theta = 2$, otherwise reject.

Beta prior, binomial likelihood

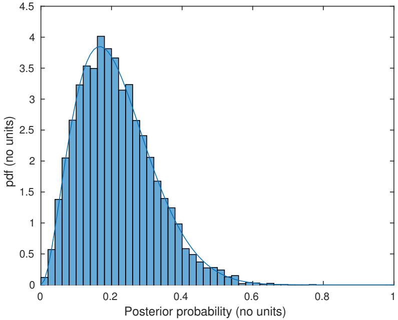

Figure: 10,000 samples using ABC.

p value for χ^2 goodness of fit test $(50 \pm 30)\%$
(10 ± 10) rejects per accepted sample

Normal-gamma prior, Normal likelihood

Prior

$$\tau \sim \text{Gamma}(\alpha_0, \beta_0) \tag{4}$$

$$\mu|\tau \sim \text{N}\left(\mu_0, 1/(\nu_0 \tau)\right) \tag{5}$$

Normal-gamma prior, Normal likelihood

Prior

$$\tau \sim \mathrm{Gamma}(\alpha_0, \beta_0) \tag{4}$$

$$\mu|\tau \sim \mathrm{N}\left(\mu_0, 1/(\nu_0\tau)\right) \tag{5}$$

Likelihood

$$X|\mu, \tau \sim \mathrm{N}(\mu, 1/\tau) \tag{6}$$

Normal-gamma prior, Normal likelihood

Prior

$$\tau \sim \text{Gamma}(\alpha_0, \beta_0) \tag{4}$$

$$\mu|\tau \sim \text{N}\left(\mu_0, 1/(\nu_0\tau)\right) \tag{5}$$

Likelihood

$$X|\mu,\tau \sim \text{N}(\mu, 1/\tau) \tag{6}$$

Joint Posterior

$$\mu,\tau|X \sim \text{NGamma}\left(\mu_1, \nu_1, \alpha_1, \beta_1\right) \tag{7}$$

Normal-gamma prior, Normal likelihood

Marginal Posterior

$$\mu|X = \sqrt{\frac{\beta_1}{\alpha_1 \nu_1}} T_{2\alpha_1} + \mu_1 \tag{8}$$

and

$$\tau|X \sim \text{Gamma}\left(\alpha_1, \beta_1\right) \tag{9}$$

where $T_{2\alpha} \sim t_{2\alpha}$.

Normal-gamma prior, Normal likelihood

ABC Algorithm

- Sample $\tau \sim \text{Gamma}(\alpha_0, \beta_0)$
- Sample $\mu|\tau \sim \text{N}\left(\mu_0, 1/(\nu_0\tau)\right)$
- Sample n times $Y|\mu, \tau \sim \text{N}(\mu, 1/\tau)$

Normal-gamma prior, Normal likelihood

ABC Algorithm

- Sample $\tau \sim \text{Gamma}(\alpha_0, \beta_0)$
- Sample $\mu|\tau \sim \text{N}\left(\mu_0, 1/(\nu_0\tau)\right)$
- Sample n times $Y|\mu, \tau \sim \text{N}(\mu, 1/\tau)$
- Conduct two tailed hypothesis tests, at some confidence level, on

$$\frac{\sqrt{n}\left(\bar{X} - \bar{Y}\right)}{\sqrt{S_X^2 + S_Y^2}} \sim \text{N}(0, 1) \tag{10}$$

$$\frac{S_X^2}{S_Y^2} \sim F_{n-1,n-1} \tag{11}$$

- Accept μ, τ if both null hypothesis are accepted, reject otherwise

Normal-gamma prior, Normal likelihood

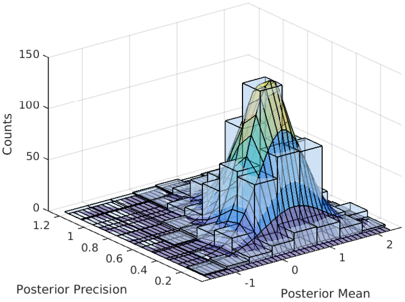

Figure: Joint histogram of 1000 samples from ABC. Confidence level set at 10%.

Normal-gamma prior, Normal likelihood

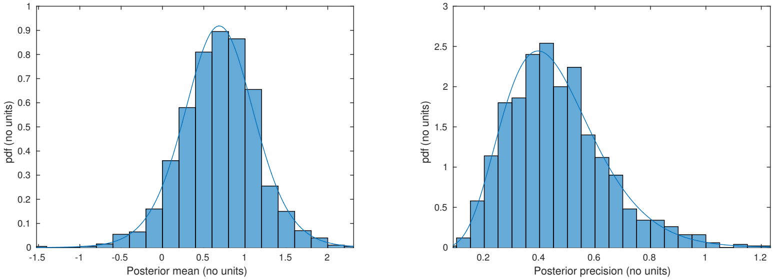

Figure: Marginal histogram of 100- samples from ABC. Confidence level set at 10%.

Normal-gamma prior, Normal likelihood

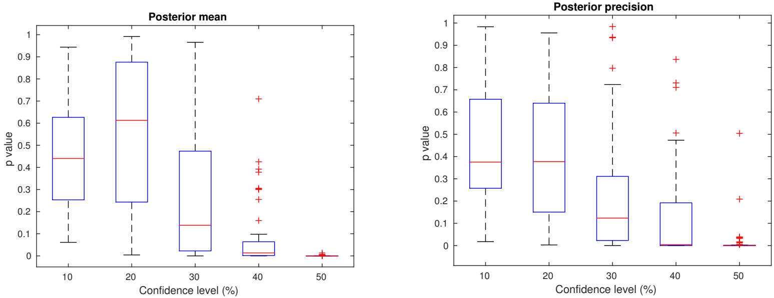

Figure: p values for the χ^2 goodness of fit test. The test was conducted 50 times with different observations.

Normal-gamma prior, Normal likelihood

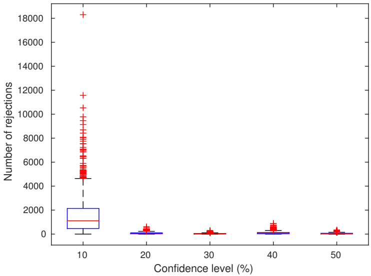

Figure: Rejections per accepted sample

- Threshold not too big, not too small.
- Ideally comparison through sufficient statistics.
- Posterior can be unknown.
- Unable to simulate from improper priors.

g-and-k distribution

- Easy to simulate from, but evaluating likelihood can be costly

$$F^{-1}(x,\theta) = A + B\left(1 + c\frac{1 - exp(-g.z(x))}{1 + exp(-g.z(x))}\right)(1 + z(x)^2)^k z(x)$$

- $\theta = (A, B, c, g, k)$
- c taken as fixed at 0.8
- Suitable choice of summary statistics not immediately obvious

Simulating data

- Generate 10^4 'observations' under 'true' $\theta = (A = 3, B = 1, g = 2, k = 0.5)$
- Simulate 10^5 parameters from $\mathcal{U}[0, 10]^4$
- Simulate 10^4 observations from g-and-k distribution for each simulated parameter vector
- Use order statistics as summary statistics

Summary statistic selection

m	l	$mean$
60	1	$24,016$
60	2	$9,470$
60	3	$1,090$
60	**4**	$\mathbf{-5,794}$
100	1	$25,253$
100	2	$10,835$
100	3	$2,946$
100	4	$-3,447$
140	1	$24,813$
140	2	$10,965$
140	3	$3,360$
140	4	$-2,734$

Table: The BIC resulting from inferring the parameter values from the order statistics for a variety of number of order statistics (m), and powers of these (l)

Fitted vs true parameter values

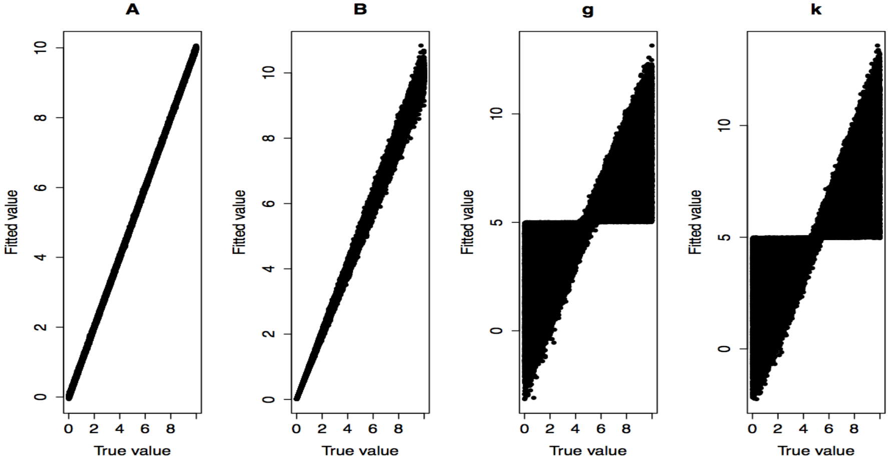

Figure: Inference of parameter values using order statistics s

Fitted vs true parameter values

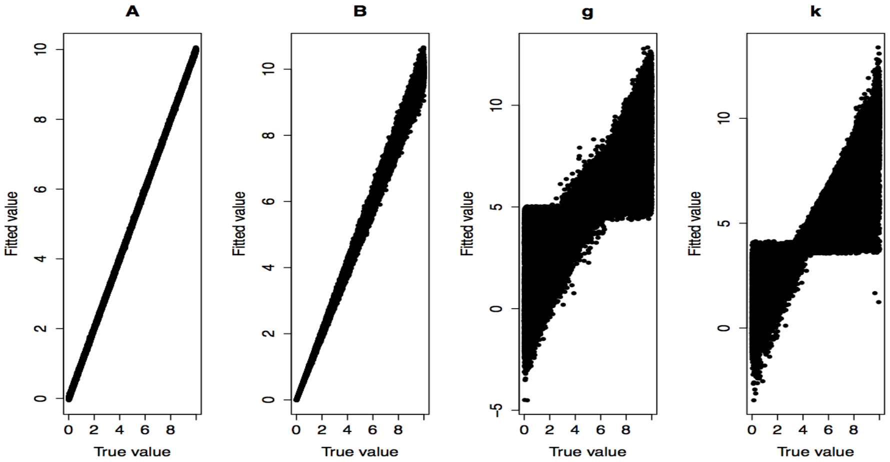

Figure: Inference of parameter values using order statistics on summary statistics and transformations (s, s^2, s^3, s^4)

Inference on A and B parameters

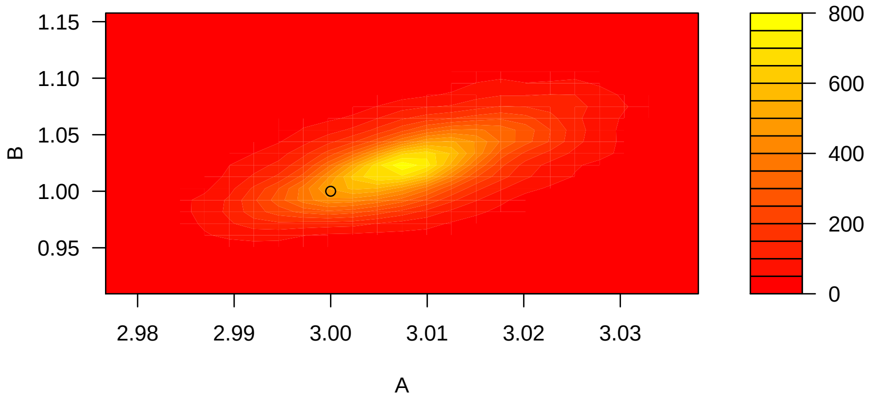

Figure: ABC with semi-automatic summary statistic selection, local linear regresson correction, tolerance 0.01 (chosen by cross-validation)

Inference on g and k parameters

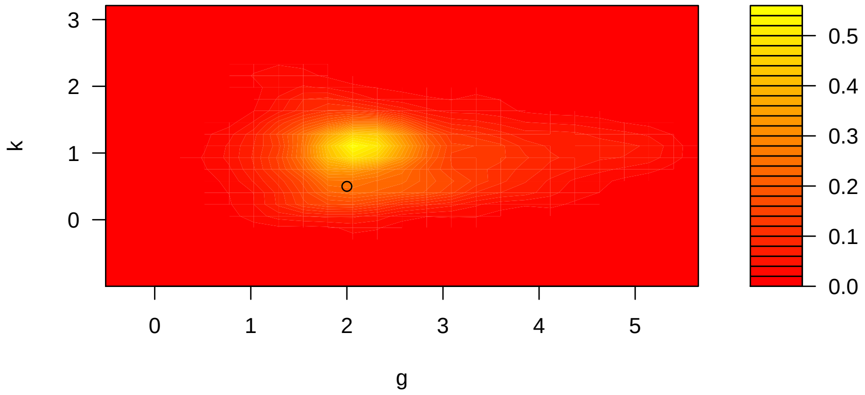

Figure: ABC with semi-automatic summary statistic selection, local linear regresson correction, tolerance 0.01 (chosen by cross-validation)

Inference on all parameters

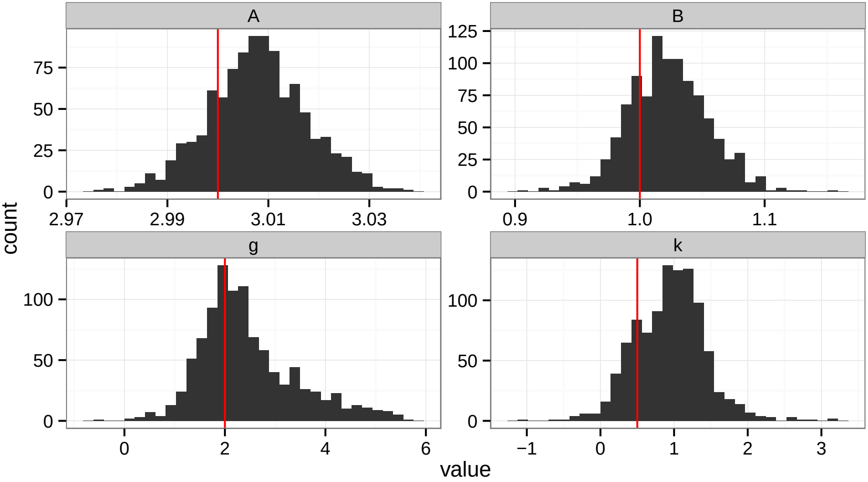

Figure: ABC with semi-automatic summary statistic selection, local linear regresson correction, tolerance 0.01 (chosen by cross-validation)

Comparison with standard ABC

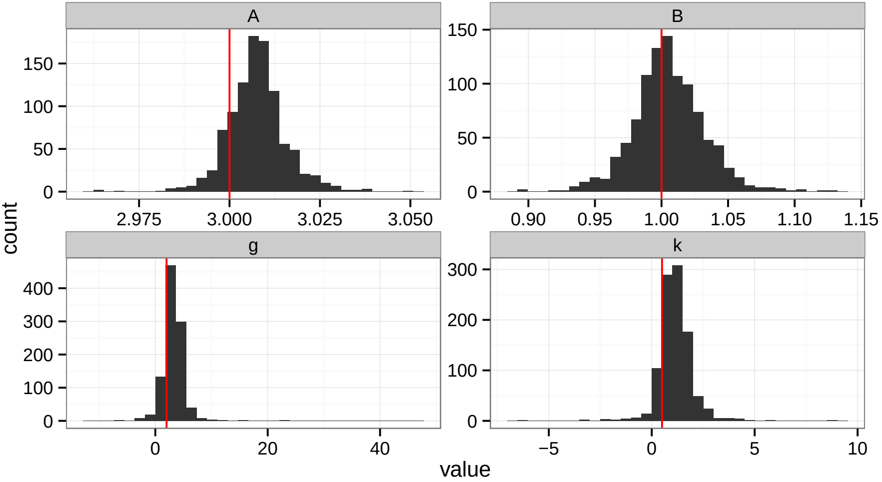

Figure: ABC using 60 order statistics as summary statistics, local linear regresson correction, tolerance 0.01 (chosen by cross-validation)

Conclusions

- Semi-automatic selection of summary statistics can be useful
- Not a panacea
  - Still need to choose summary statistics for step 1
  - Need to choose set of explanatory variables for linear model
  - Still need to tune ϵ, ρ (and K)
- Plenty more to explore . . . (over to Paul, Giuseppe and Marcin)