Chaos, Complexity, and Inference (36-462)

Lecture 19: Inference from Simulations 2

Cosma Shalizi

24 March 2009

Inference from Simulations 2, Mostly Parameter Estimation

Reading: Smith (forthcoming) is comparatively easy to read; Gouriéroux *et al.* (1993) and (especially) Gouriéroux and Monfort (1996) are harder to read but more detailed; Kendall *et al.* (2005) is a nice application which does *not* require knowing any econometrics

Method of Simulated Moments

1. Pick your favorite test statistics T (“generalized moments”)
2. Calculate from data, t_{obs}
3. Now pick a parameter value θ
    1. simulate multiple times
    2. calculate average of $T \approx \mathbf{E}_{\theta}[T]$
4. Adjust θ so expectations are close to t_{obs}

The last step is a “stochastic approximation” problem Robbins and Monro (1951); Nevel’son and Has’minskiĭ (1972/1976)

Works if those expectations are enough to characterize the parameter

Why expectations rather than medians, modes, …?

Basically: easier to prove convergence

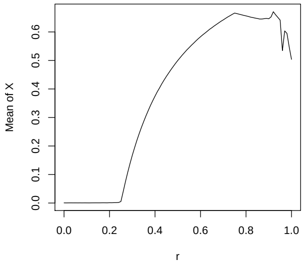
The mean is *not* always the most probable value!

Practicality: much faster & easier to optimize if the same set of random draws can be easily re-used for different parameter values

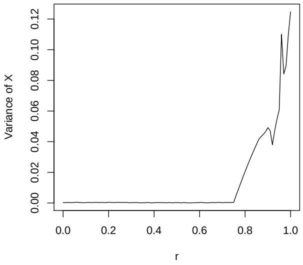
Example: use mean and variance for logistic map; chose r where simulated moments are closest (Euclidean distance) to observed

$$\widehat{r}_{MSM} = \underset{r \in [0,1]}{\text{argmin}} \left((m - \widehat{\mu}_r)^2 + (s^2 - \widehat{\sigma}^2_r)^2 \right)$$

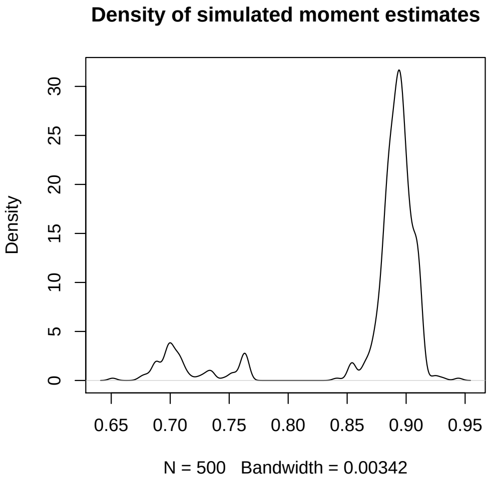
No particular reason to weight both moments equally



$\widehat{\mu}_r$



$\widehat{\sigma}^2_r$

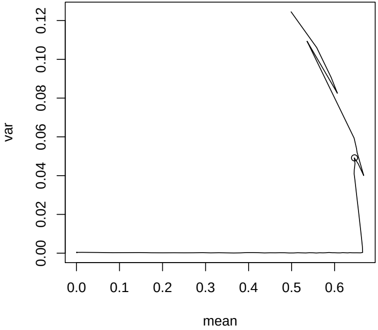


Distribution of $\widehat{r}_{MSM}$, time series length 100, true $r = 0.9$

Kinks in the curve of the moments: potentially confusing to optimizer, reduces sensitivity

big change in parameter leads to negligible change in moments

curve crossing itself $\Rightarrow$ non-identifiability



The Progress of Statistical Methods

First stage calculate likelihood, solve explicitly for MLE

Second stage can't solve for MLE but can still write down likelihood, calculate it, and maximize numerically

Third stage even calculating the likelihood is intractable

Outstanding example: hidden or latent variables $Y_1, Y_2, \ldots$ plus observed $X_1, X_2, \ldots$

Why Finding the Likelihood Becomes Hard

Likelihood become an integral/sum over all possible combinations of latent variables compatible with observations:

$$
\begin{aligned}
\Pr_\theta\left(X_1^n = x_1^n\right) \\
&= \int dy_1^n \Pr_\theta\left(X_1^n = x_1^n, Y_1^n = y_1^n\right) \\
&= \int dy_1^n \Pr_\theta\left(Y_1^n = y_1^n\right) \prod_{i=1}^{n} \Pr_\theta\left(X_i = x_i | Y_1^n = y_1^n, X_1^{i-1} = x_1^{i-1}\right)
\end{aligned}
$$

Evaluating this sum-over-histories is, itself, a hard problem
One approach: Expectation-Maximization algorithm, try to simultaneously estimate latent variables and parameters (Neal and Hinton, 1998)
Standard, clever, often messy

Indirect Inference

We have a model with parameter θ from which we can simulate also: data y

Introduce an **auxiliary model** which is *wrong* but *easy to fit*

Fit auxiliary to data, get parameters $\widehat{\beta}$

Simulate from model to produce y^S_θ — different simulations for different values of θ

Fit auxiliary to simulations, get $\widehat{\beta}^S_\theta$

Pick θ such that $\widehat{\beta}^S_\theta$ is as close as possible to $\widehat{\beta}$

Improvement: do several simulation runs at each θ, average $\widehat{\beta}^S_\theta$ over runs

What's going on here?

The auxiliary model says: the data has these sorts of patterns
Pick parameters which come as close as possible to matching those parameters
For this to work, those patterns must be enough to pin down the original parameter, requires at a minimum that $\dim \beta = \dim \theta$

A More Formal Statement

Auxiliary objective function ψ, depends on data and β

$$\widehat{\beta}_T \equiv \underset{\beta}{\operatorname{argmax}}\, \psi_T(\beta) \quad (1)$$

$$\widehat{\beta}_{T,S,\theta} \equiv \underset{\beta}{\operatorname{argmax}}\, \psi_{T,S,\theta}(\beta) \quad (2)$$

$$\widehat{\theta}_{II} \equiv \underset{\theta}{\operatorname{argmin}}\, (\widehat{\beta}_{T,S,\theta} - \widehat{\beta}_T)'\Omega(\widehat{\beta}_{T,S,\theta} - \widehat{\beta}_T) \quad (3)$$

Ω some positive definite matrix

which one doesn't matter asymptotically

Optimal choice gives most weight to the most-informative auxiliary parameters

(Gouriéroux and Monfort, 1996, §4.2.3)

identity matrix is usually OK

Assume:

1. As $T \to \infty$, $\psi_{T,S,\theta}(\beta) \to \psi(\beta, \theta)$, uniformly in β and θ.
2. For each θ, the limiting objective function has a unique optimum in β, call this $b(\theta)$.
3. As $T \to \infty$, $\widehat{\beta}_T \to b(\theta_0)$.
4. The equation $\beta = b(\theta)$ has a unique solution, i.e., b^{-1} is well-defined.

then as $T \to \infty$,

$$\widehat{\theta}_{II} \to \theta_0$$

in probability

Asymptotic Distribution of Indirect Estimates

(Gouriéroux and Monfort, 1996, §4.2.3)

Under additional (long, technical) regularity conditions, $\widehat{\theta}_{II} - \theta_0$ is asymptotically Gaussian with mean 0
Variance $\propto \frac{1}{T}\left(1 + \frac{1}{S}\right)$
Variance depends on something *like* the Fisher information matrix, only with $\partial b / \partial \theta$ in the role of $\partial p_\theta / \partial \theta$
basically, how sensitive is the auxiliary parameter to shifts in the underlying true parameter?

Checking Indirect Inference

Given real and auxiliary model, will indirect inference work, i.e., be consistent?

Do the math Provides proof; often hard (because the simulation model leads to difficulty-to-manipulate distributions)

Simulate some more Simulate from model for a particular θ, apply II, check that estimates are getting closer to θ as simulation grows, repeat for multiple θ
Not as fool-proof but just requires time (you have all the code already)

Autoregressive Models

Like it sounds: regress X_t on its past $X_{t-1}, X_{t-2}, \ldots$

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \ldots \beta_p X_{t-p} + \epsilon_t, \ \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Common as auxiliary models for time series (as well as models in their own right)
Auxiliary objective function is residual sum of squares over p
`R` command: `ar`

Example: Logistic Map + Noise

Take logistic map and add Gaussian noise to each observation

$$
\begin{aligned}
x_t &= y_t + \epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, \sigma^2) \\
y_{t+1} &= 4ry_t(1 - y_t)
\end{aligned}
$$

Any sequence x_1^T *could* be produced by any r

```
logistic.noisy.ts <- function(timelength,r,
                              initial.cond=NULL,noise.sd=0.1) {
  x <- logistic.map.ts(timelength,r,initial.cond)
  return(x+rnorm(timelength,0,noise.sd))
}
```

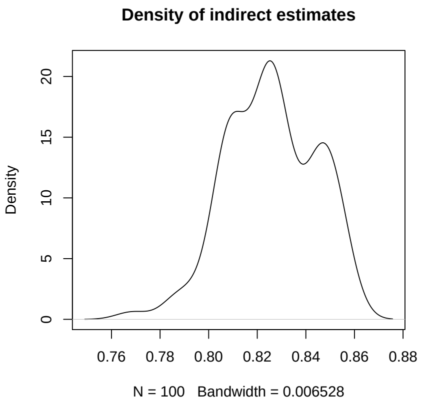
Assume that σ^2 is known — simplifies plotting if only one unknown parameter! Set it to $\sigma^2 = 0.1$

1. fix p for AR model
2. Fit AR(p) to data, get $\widehat{\beta} = (\widehat{\beta}_1, \ldots \widehat{\beta}_p)$
3. Simulate S sample trajectories with parameter r, calculate $(\widehat{\beta}_1, \ldots \widehat{\beta}_p)$ for each, average over trajectories to get $\widehat{\beta}^S_r$
4. Minimize $\|\widehat{\beta} - \widehat{\beta}^S_r\|$

```
logistic.map.II <- function(y,order=2,S=10) {
  T <- length(y)
  ar.fit <- function(x) {
    return(ar(x,aic=FALSE,order.max=order)$ar)
  }
  beta.data <- ar.fit(y)
  beta.discrep <- function(r) {
    beta.S <- mean(replicate(S,ar.fit(logistic.noisy.ts(T,r))))
    return(sum((beta.data - beta.S)^2))
  }
  return(optimize(beta.discrep,lower=0.75,upper=1))
}
```

To see how well this does, simulate it:

```
y <- logistic.noisy.ts(1e3,0.8)
plot(density(replicate(100,
                       logistic.map.II(y,order=2)$minimum)),
     main="Density of indirect estimates")
```



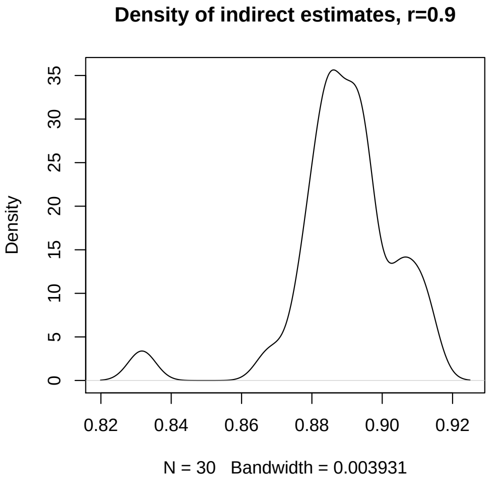
Some bias (here upward) but it shrinks as T grows, and it's pretty tight around the true value ($r = 0.8$)
Notice: fixed data set, all variability is from simulation

Also: $p = 2$ is arbitrary, can use more simulation to pick good/best

$r = 0.8$ is periodic, what about chaos, say $r = 0.9$?

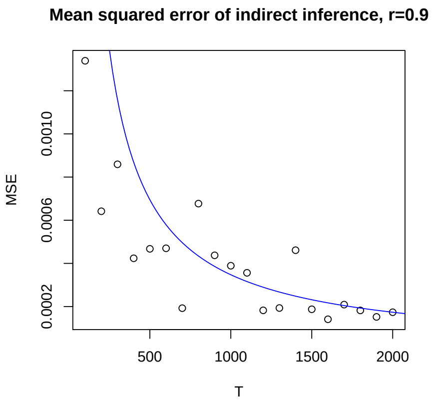
```
plot(density(replicate(30,
     logistic.map.II(logistic.noisy.ts(1e3,r=0.9),
                     order=2)$minimum)),
     main="Density of indirect estimates, r=0.9")
```

re-generate data each time


Density of indirect estimates, r=0.9
Density
0
5
10
15
20
25
30
35
0.82
0.84
0.86
0.88
0.90
0.92
N = 30 Bandwidth = 0.003931

I promised to check that the inference is working my seeing that the errors are shrinking:

```
mse.logistic.II <- function(T,r=0.9,reps=30,order=2,S=10) {
  II <- replicate(reps,logistic.map.II(logistic.noisy.ts(T,r),
                  order=order,S=S)$minimum)
  II.error = II - r # Uses recycling
  return(mean(II.error^2))
}
```



Black: mean squared error $\widehat{\theta}_{II}$, $S = 10$, average of 30 replications each of length T, all with $r = 0.9$; blue: curve $\propto T^{-1}$, fitted through last data point

The Correct Line on Inference from Complex Models

MORE SCIENCE, FEWER *F*-TESTS

- Craft a *really good* scientific model
  - represent your actual knowledge/assumptions/guesswork
  - “it’s in my regression textbook” isn’t a *scientific* justification
  - *must* be able to simulate it
- Pick a *reasonable* auxiliary model
  - Works on your observable data
  - Easy to fit
  - Predicts well is nice but not necessary
- Estimate parameters of complex model by indirect inference
- Test hypotheses by indirect inference as well (Gouriéroux *et al.*, 1993; Kendall *et al.*, 2005)

Gouriéroux, Christian and Alain Monfort (1996). *Simulation-Based Econometric Methods*. Oxford, England: Oxford University Pres.

Gouriéroux, Christian, Alain Monfort and E. Renault (1993). "Indirect Inference." *Journal of Applied Econometrics*, **8**: S85–S118. URL http://www.jstor.org/pss/2285076.

Kendall, Bruce E., Stephen P. Ellner, Edward Mccauley, Simon N. Wood, Cheryl J. Briggs, William W. Murdoch and Peter Turchin (2005). "Population Cycles in the Pine Looper Moth: Dynamical Tests of Mechanistic Hypotheses." *Ecological Monographs*, **75**: 259–276. URL http://www.eeb.cornell.edu/Ellner/pubs/CPDBupalusEcolMonog05.pdf.

Neal, Radford M. and Geoffrey E. Hinton (1998). "A View of the EM Algorithm that Justifies Incremental, Sparse, and Other Variants." In *Learning in Graphical Models* (Michael I. Jordan, ed.), pp. 355–368. Dordrecht: Kluwer Academic. URL `http://www.cs.toronto.edu/~radford/em.abstract.html`.

Nevel'son, M. B. and R. Z. Has'minskiĭ (1972/1976). *Stochastic Approximation and Recursive Estimation*, vol. 47 of *Translations of Mathematical Monographs*. Providence, Rhode Island: American Mathematical Society. Translated by the Israel Program for Scientific Translations and B. Silver from *Stokhasticheskaia approksimatsia i rekurrentnoe otsenivanie*, Moscow: Nauka.

Robbins, Herbert and Sutton Monro (1951). "A Stochastic Approximation Method." *Annals of Mathematical Statistics*, **22**: 400–407. URL `http://projecteuclid.org/euclid.aoms/1177729586`.

Smith, Jr., Anthony A. (forthcoming). "Indirect Inference." In *New Palgrave Dictionary of Economics* (Stephen Durlauf and

Lawrence Blume, eds.). London: Palgrave Macmillan, 2nd edn. URL
http://www.econ.yale.edu/smith/palgrave7.pdf.