Econ2146: Time series
Lecture 3

version 0.1

Neil Shephard

Economics & Statistics Departments,
Harvard University

November 2015

Generic filtering

Generic filtering and prediction density. Assume initial $f(\alpha_0|\mathcal{F}_0)$. Run $t = 1, 2, ..., T$, for HMM:

$$
\begin{aligned}
f\left(\alpha_{t+1}|\mathcal{F}_t\right) &= \int f(\alpha_{t+1}|\alpha_t, \mathcal{F}_t)\mathrm{d}F\left(\alpha_t|\mathcal{F}_t\right), \\
f(Y_{t+1}|\mathcal{F}_t) &= \int f(Y_{t+1}|\alpha_{t+1}, \mathcal{F}_t)\mathrm{d}F\left(\alpha_{t+1}|\mathcal{F}_t\right) \\
f\left(\alpha_{t+1}|\mathcal{F}_{t+1}\right) &\propto f(Y_{t+1}|\alpha_{t+1}, \mathcal{F}_t)f\left(\alpha_{t+1}|\mathcal{F}_t\right).
\end{aligned}
$$

- Modern subject starts with Gordon et al. (1993).
  - This is called "particle filtering" or "sequential Monte Carlo".
  - Introduction focusing on economics e.g. include Creal (2012). Also Doucet and Johansen (2011).
- I will talk about the most basic version.
- Large resources at `http://www.stats.ox.ac.uk/~doucet/smc_resources.html`

Particle filter

- Assume θ is known, like the Kalman and discrete filters.
- Like the Kalman filter, we progress by looping through time.
- Big idea: replace distributions with samples. Similar to a bootstrap.

Bootstrap particle filter

Particle filter: Run forward, $t = 1, 2, ..., T$, initialized with sample $\left\{ \left(\alpha_{0|0}^{(j)} \right), j = 1, ..., M \right\}$.

- *Propergate* K copies, that is $R = KM$ times,

$$\alpha_{t|t-1}^{(Mk+j)} \stackrel{L}{=} \alpha_t | \alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1}, \quad j = 1, 2, ..., M, \quad k = 0, ..., K-1.$$

- Compute:
  - weights $w_t^{(j)} = f(Y_t | \alpha_{t|t-1}^{(j)}, \mathcal{F}_{t-1})$, $j = 1, 2, ..., R$.
  - normalized weights: $\widetilde{w}_t^{(j)} = w_t^{(j)} / \sum_{i=1}^{R} w_t^{(i)}$.
- *Resample* (i.e. multinomial) M times from $\left\{ \widetilde{w}_t^{(j)}, \alpha_{t|t-1}^{(j)}, j = 1, 2, ..., R \right\}$ to produce the sample

$$\left\{ \alpha_{t|t}^{(j)}, j = 1, 2, ..., M \right\}.$$

Estimating the likelihood

- I usually use $K = 1$ in practice.
- Estimated prediction distribution

$$\widehat{f}(Y_t|\mathcal{F}_{t-1}) = \frac{1}{R}\sum_{i=1}^{R} w_t^{(i)}, \quad t = 1, 2, ..., T.$$

- Then the simulation estimator of the likelihood is, for any $R \in \mathbb{N}_{>0}$,

$$L_{t,R} = \widehat{f}(Y_1, ..., Y_t|\mathcal{F}_0) = \prod_{j=1}^{t} \widehat{f}(Y_t|\mathcal{F}_{t-1}).$$

Behavor of estimated likelihood

- If $f(\alpha_t|\alpha_{t-1},\mathcal{F}_{t-1})>0$ and $f(Y_t|\alpha_t,\mathcal{F}_{t-1})>0$ and

$$\frac{1}{2}\int \left|f(\alpha_t|\alpha_k,\mathcal{F}_t)-f(\alpha_t|\alpha_k',\mathcal{F}_t)\right| \mathrm{d}\alpha_t \leq \beta^{t-k}, \quad \beta\in[0,1)\,,$$

  uniformly over (α_k,α_k') if $t-k$ is large enough. Then under multinomial resampling

$$\mathrm{E}_u\left(\frac{L_{t,R}}{L_t}|\mathcal{F}_t\right)=1,$$

  while if $K=1$ then

$$\frac{M}{t}\times\mathrm{Var}_u\left(\frac{L_{t,R}}{L_t}|\mathcal{F}_t\right)=O(1).$$

  The asymptotic variance is computable, but I have never seen it used.
  - The proof of unbiasedness is difficult, it is due to Del Moral (2004). A more accessible version appears in Pitt et al. (2012).
  - Remarkable result, e.g. standard importance sampling's variance increases exponentially in t. Gain from sequential Monte Carlo.
  - Estimate $L_{t,R}$ is discontinuous in θ.

Compute of prediction and filtering on the fly

- Summaries of prediction and filtering samples, e.g.

$$
\begin{aligned}
\mathrm{E}_R\left(\alpha_t|\mathcal{F}_{t-1}\right) &= \int \alpha_t \mathrm{d}F_R(\alpha_t|\mathcal{F}_{t-1}) = \frac{1}{R}\sum_{i=1}^{R}\alpha_{t|t-1}^{(i)}, \\
\mathrm{E}_M\left(\alpha_t|\mathcal{F}_{t}\right) &= \int \alpha_t \mathrm{d}F_M(\alpha_t|\mathcal{F}_{t}) = \frac{1}{M}\sum_{i=1}^{M}\alpha_{t|t}^{(i)}.
\end{aligned}
$$

  Here $F_n(X)$ denotes the empirical distribution function of some X.
  - Under above assumptions then, for example,

    $$M \times \mathrm{Var}_u\left(\mathrm{E}_R\left(\alpha_t|\mathcal{F}_t\right)|\mathcal{F}_t\right) = O(1),$$

    and central limit theorems exist. Notice it does not involve t.
  - Likewise, e.g. for medians

    $$med_M(\alpha_t|\mathcal{F}_t) = med\left\{\alpha_{t|t}^{(1)}, ..., \alpha_{t|t}^{(M)}\right\}.$$

    Likewise covariances, other quantiles, etc.

Computational aspects

- Simulate through transition equation

$$\alpha_t | \alpha^{(j)}_{t-1|t-1}, \mathcal{F}_{t-1},$$

no need to compute density! Nice in some models where dynamics have to be computed, e.g. DSGE, continuous time.

- Evaluate measurement density

$$f(Y_t | \alpha^{(j)}_{t|t-1}, \mathcal{F}_{t-1}).$$

- Typically easier to code than Kalman filter.
- Computational load $O(RT)$. Does not depend on the dimension of the state.

Example: log-Gaussian Stochatsic volatility

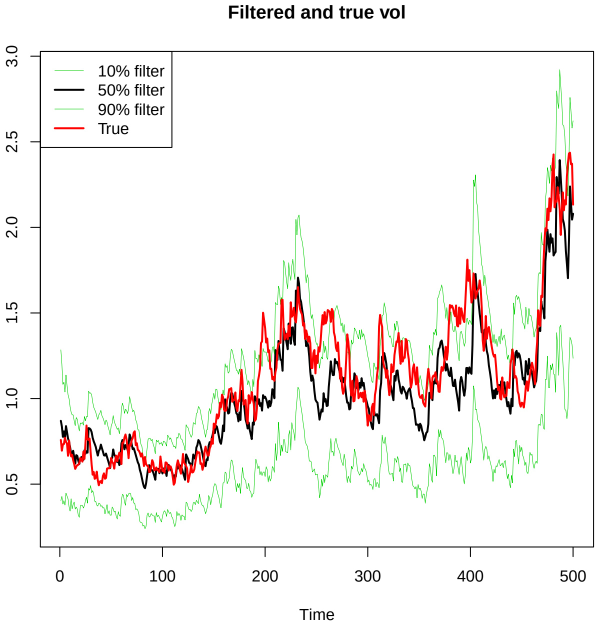
- Now analyze a stochastic volatility process:

$$
\begin{aligned}
y_t &= \varepsilon_t e^{\alpha_t/2}, \quad t = 1, 2, ..., T, \\
\alpha_{t+1} &= \mu + \phi(\alpha_{t+1} - \mu) + \eta_t.
\end{aligned}
$$

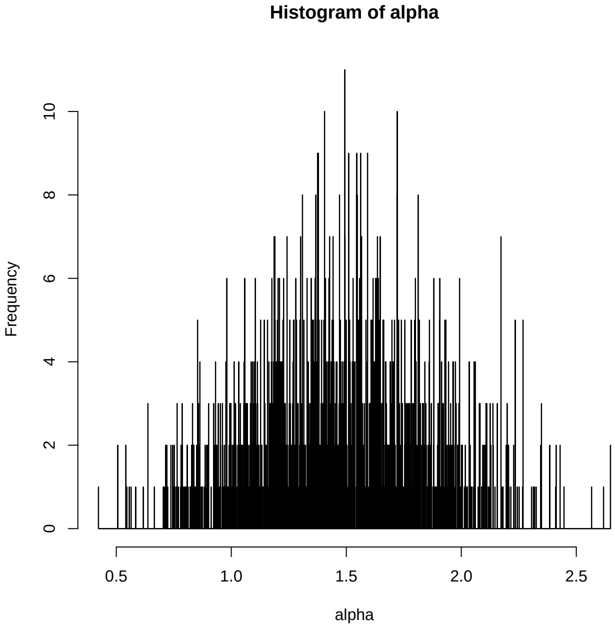
  Use simulated data with $T = 500$, $\mu = -0.5$, $\phi = 0.98$, $\sigma = 0.13$.
- e.g. Kim et al. (1998) and Pitt and Shephard (1999).

```
simSV <- function(T,mu,phi,sigma){
  y = array(0, dim=c(T,1))
  mAlpha = array(0, dim=c(T,1)); alpha = mu;
  for (i in (1:T)){
    alpha = mu + phi*(alpha-mu) + sigma*rnorm(1,0,1)
    y[i] = rnorm(1,0,1)*exp(0.5*alpha)
    mAlpha[i] = alpha
  }
  mRes = list(Y = y, logVol = exp(0.5*mAlpha))
}

T = 500; mu = -0.5; phi =0.98; sigma = 0.13;
lSV = simSV(T,mu,phi,sigma); y = lSV$Y;

M = 2000
mlogL = array(0, dim=c(T,1))
mAlpha= array(0,dim=c(T,3))

alpha = mu + rnorm(M,0.0,sigma/sqrt(1.0-(phi^2)))
```

```
for (t in (1:T)){
  alpha = mu + phi*(alpha-mu) + sigma*rnorm(M,0,1)
  logw = dnorm(y[t],0.0,exp(0.5*alpha),log=TRUE)
  w1 = exp(logw-max(logw))
  Wstar = w1/sum(w1)
  alpha = sample(alpha,M,replace=T,prob=Wstar)
  mAlpha[t,] = quantile(exp(0.5*alpha),probs=c(0.0,0.5,0.9))
  mlogL[t,1] = max(logw) + log(mean(w1))
}

pdf("SVfilter.pdf")
ts.plot(cbind(mAlpha,lSV$logVol),type="l",col=c(3,1,3,2),
   lwd=c(0.2,2.0,0.2,2.0),main="Filtered and true vol")
legend("topleft",legend=c("10% filter","50% filter",
   "90% filter","True"),col=c(3,1,3,2),lwd=c(0.2,2.0,0.2,2.0),
dev.off()

pdf("alpha_T_measure.pdf"); hist(alpha,breaks=M); dev.off();
```



10%, 50% and 90% quantiles of $e^{\alpha_t/2}|\mathcal{F}_t$.



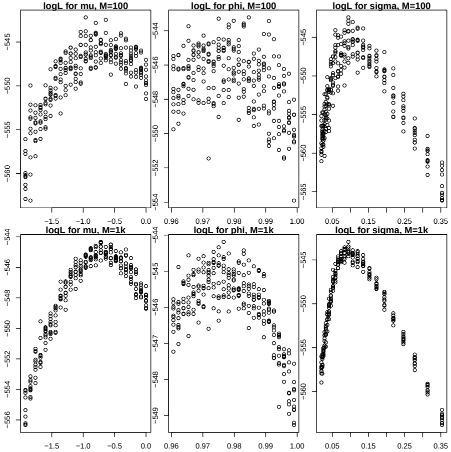
Histogram of samples from $e^{\alpha_T/2}|\mathcal{F}_T$. Many repeat particles.

```
particleF <- function(M,mY,mTheta){
 mu = mTheta[1]; phi = mTheta[2]; sigma = mTheta[3];
 T = dim(mY)[1]
 mLogL = array(0,dim=c(T,1))

 alpha = mu + rnorm(M,0.0,sigma/sqrt(1.0-(phi^2)))
 for (t in (1:T)){
   alpha = mu + phi*(alpha-mu) + sigma*rnorm(M,0,1)
   logw = dnorm(y[t],0.0,exp(0.5*alpha),log=TRUE)
   w1 = exp(logw-max(logw))
   Wstar = w1/sum(w1)
   alpha = sample(alpha,M,replace=T,prob=Wstar)
   mLogL[t] = max(logw) + log(mean(w1))
 }

 mLogL
}

iRep=10
```

```
jRep = 25
mlogLStore = array(0,dim=c(jRep,iRep))
mPara = array(0,dim=c(jRep,iRep))

pdf("particlelogL.pdf")
par(mfcol=c(2,3), mar=c(2,2,1.0,0.0), oma=c(1.5,2,1,1)) # make
for (k in (1:3)){
for (l in (1:2)){
 if (l==1) M = 100
 if (l==2) M = 1000
 mu = -0.5; phi =0.98; sigma = 0.13;
 for (j in (1:jRep)){
 if (k==1) mu = -2.0 + (j*(2.0/jRep))
 if (k==2) phi = 0.999*(0.96+((0.04*j/(jRep))))
 if (k==3) sigma = exp(log(0.13) - 2.0 + (j*(3.0/jRep)))

 for (i in (1:iRep)){
 mlogL = sum(particleF(M,y,c(mu,phi,sigma)))
 mlogLStore[j,i] = mlogL
```

```
 if (k==1) mPara[j,i] = mu
 if (k==2) mPara[j,i] = phi
 if (k==3) mPara[j,i] = sigma
# print(c(i,mu,mlogL))
 }
 }
 if (k==1 && l==1) plot(mPara,mlogLStore,main="logL for mu, M=
 if (k==2 && l==1) plot(mPara,mlogLStore,main="logL for phi, M
 if (k==3 && l==1) plot(mPara,mlogLStore,main="logL for sigma,
 if (k==1 && l==2) plot(mPara,mlogLStore,main="logL for mu, M=
 if (k==2 && l==2) plot(mPara,mlogLStore,main="logL for phi, M
 if (k==3 && l==2) plot(mPara,mlogLStore,main="logL for sigma,
 }
}
dev.off()
```

Drawing the estimated likelihood functions



Simulated data, with M=100 and M=1,000. Data simulated using $T = 500$, $\mu = -0.5$, $\phi = 0.98$, $\sigma = 0.13$.

Why? I

- Assume we have a sample from $\alpha_{t-1}|\mathcal{F}_{t-1}$

$$\alpha_{t-1|t-1}^{(j)}, \quad j = 1, 2, ..., M.$$

  Each draw is called a "particle". Think of M as $1,000$ or $10,000$ or much bigger.

- Then recall

$$\begin{aligned} f(\alpha_t|\mathcal{F}_{t-1}) &= \int f(\alpha_t, \alpha_{t-1}|\mathcal{F}_{t-1})\mathrm{d}\alpha_{t-1} \\ &= \int f(\alpha_t|\alpha_{t-1}, \mathcal{F}_{t-1})\mathrm{d}F(\alpha_{t-1}|\mathcal{F}_{t-1}) \end{aligned}$$

  so we could (uniformly, over α_t, unbiasedly) estimate $f(\alpha_t|\mathcal{F}_{t-1})$ as

$$\widehat{f}(\alpha_t|\mathcal{F}_{t-1}) = \frac{1}{M}\sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1}). \tag{1}$$

Why? II

- If we were to sample from this, we write them as

$$\alpha_{t|t-1}^{(j)}, \quad j = 1, 2, ..., M.$$

  Often called propergation.

- Then we might sample from

$$\widehat{f}(\alpha_t|\mathcal{F}_t) \propto f(Y_t|\alpha_t, \mathcal{F}_{t-1})\widehat{f}(\alpha_t|\mathcal{F}_{t-1}) \tag{2}$$

$$\propto f(Y_t|\alpha_t, \mathcal{F}_{t-1}) \sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1}) \tag{3}$$

  M times to produce:

$$\alpha_{t|t}^{(j)}, \quad j = 1, 2, ..., M.$$

  Often called learning or updating.

- Completes the interation. Now loop.

Likelihood estimation

- A biproduct of this is

$$f(Y_t|\mathcal{F}_{t-1}) = \int f(Y_t, \alpha_t|\mathcal{F}_{t-1})\mathrm{d}\alpha_t = \int f(Y_t|\alpha_t, \mathcal{F}_{t-1})\mathrm{d}F(\alpha_t|\mathcal{F}_{t-1})$$

which we (uniformly in Y_t) unbiasedly estimate as

$$\widehat{f}(Y_t|\mathcal{F}_{t-1}) = \frac{1}{M}\sum_{j=1}^{M} f(Y_t|\alpha_{t|t-1}^{(j)}, \mathcal{F}_{t-1}).$$

  - The computational load of this step is $O(M)$.

Unsolved problems

- **Problem 1.** How can I sample from

$$\widehat{f}(\alpha_t|\mathcal{F}_{t-1}) = \frac{1}{M}\sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1}).$$

  This is easy!

- **Problem 2.** How can I sample from

$$\widehat{f}(\alpha_t|\mathcal{F}_t) \propto f(Y_t|\alpha_t, \mathcal{F}_{t-1}) \sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1}) \tag{4}$$

  Generic, e.g. MCMC.

- **Problem 3.** Can we sample from (4) fast! Tricky.

Problem 1: propergation

- How can I sample $\left\{\alpha_{t|t-1}^{(j)}, j = 1, .., M\right\}$ from

$$\widehat{f}(\alpha_t|\mathcal{F}_{t-1}) = \frac{1}{M}\sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1})? \tag{5}$$

- Sample from a discrete mixture. Randomly select M versions of $j \in \{1, 2, ..., M\}$, $i_1, ..., i_M$, then assume we can sample from

$$\alpha_{t|t-1}^{(j)} \stackrel{L}{=} \alpha_t|\alpha_{t-1|t-1}^{(i_j)}, \mathcal{F}_{t-1}, \quad j = 1, 2, ..., M.$$

- Produces i.i.d. samples from (5).
- Sounds great, very general. But dumb.
- Better and easier to stratify.
    - Sample

    $$\alpha_{t|t-1}^{(j)} \stackrel{L}{=} \alpha_t|\alpha_{t-1|t-1}^{(j)}, \mathcal{F}_{t-1}, \quad j = 1, 2, ..., M.$$

    i.e. just propergate each particle once. Or if you want to sample $R = KM$ times, propergate each sample K times.

Problem 2: updating I

- The challange is to sample from

$$\begin{aligned} \widehat{f}(\alpha_t|\mathcal{F}_t) &\propto Mf(Y_t|\alpha_t,\mathcal{F}_{t-1})\widehat{f}(\alpha_t|\mathcal{F}_{t-1}) && (6) \\ &= f(Y_t|\alpha_t,\mathcal{F}_{t-1})\sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1|t-1}^{(j)},\mathcal{F}_{t-1}) && (7) \end{aligned}$$

  Can use generic methods, e.g. MCMC, Hamilton Monte Carlo etc. But each density evaluation costs $O(M)$. Deadend.

- We assume we have a sample $\left\{\alpha_{t|t-1}^{(j)}, j=1,...,R\right\}$ from $\widehat{f}(\alpha_t|\mathcal{F}_{t-1})$. Now you can think about the likelihood ratio of the filtering density to the prediction density as

$$L(\alpha_t) = \frac{\widehat{f}(\alpha_t|\mathcal{F}_t)}{\widehat{f}(\alpha_t|\mathcal{F}_{t-1})},$$

Problem 2: updating II

where

$$L(\alpha_t) = \frac{f(Y_t|\alpha_t, \mathcal{F}_{t-1})}{f(Y_t|\mathcal{F}_{t-1})} \quad (8)$$

$$= \frac{f(Y_t|\alpha_t, \mathcal{F}_{t-1})}{\int f(Y_t|\alpha_t, \mathcal{F}_{t-1})\widehat{f}(\alpha_t|\mathcal{F}_{t-1})\mathrm{d}\alpha_t}. \quad (9)$$

- Then the weighted population

$$\left\{\left(L(\alpha_{t|t-1}^{(j)}), \alpha_{t|t-1}^{(j)}\right), j = 1, ..., R\right\}, \quad (10)$$

has distribution function

$$\frac{1}{R}\sum_{j=1}^{R} L(\alpha_{t|t-1}^{(j)})1\left(\alpha_{t|t-1}^{(j)} < b\right). \quad (11)$$

Problem 2: updating III

So, averaging over the propergation simulation

$$
\begin{aligned}
& \mathrm{E}_u \left\{ \frac{1}{R} \sum_{j=1}^{R} L(\alpha_{t|t-1}^{(j)}) 1 \left(\alpha_{t|t-1}^{(j)} < b \right) \right\} \\
= & \int 1 \left(\alpha_t < b \right) \frac{\widehat{f}(\alpha_t | \mathcal{F}_t)}{\widehat{f}(\alpha_t | \mathcal{F}_{t-1})} \widehat{f}(\alpha_t | \mathcal{F}_{t-1}) \mathrm{d}\alpha_t \\
= & \int 1 \left(\alpha_t < b \right) \widehat{f}(\alpha_t | \mathcal{F}_t) \mathrm{d}\alpha_t ,
\end{aligned}
$$

so this distribution function (11) is simulation unbiased for $\widehat{f}(\alpha_t | \mathcal{F}_t)$.

- We can sample from it by resampling the weighted population (10).
- As $R \to \infty$ we remove the noise.

Problem 2: updating IV

- Only problem is do not know $L(\alpha_t)$ in (8). But

$$w_j = f(Y_t|\alpha_t^{(j)}, \mathcal{F}_{t-1}),$$

then

$$\frac{w_j}{\frac{1}{R}\sum_{i=1}^{R} w_i} \xrightarrow{p} L(\alpha_t^{(j)}).$$

So in practice we sample M times with probabilities proportional to

$$\widetilde{w}_j = \frac{w_j}{\sum_{i=1}^{R} w_i},$$

so weighted sample is

$$\left\{ \left(\widetilde{w}_j, \alpha_{t|t-1}^{(j)} \right), j = 1, ..., R \right\}.$$

Problem 2: updating V

- Produces
$$\left\{\alpha_{t|t}^{(j)}, j = 1, 2, ..., M\right\}.$$
- This is sometimes called the bootstrap filter. Computational cost is $O(R)$.

References I

Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews 31*, 245–296.

Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer.

Doucet, A. and A. Johansen (2011). A tutorial on particle filtering and smoothing: fifteen years later. In D. Crisan and B. Rozovsky (Eds.), *The Oxford Handbook of Nonlinear filtering*. Oxford University Press.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F 140*, 107–113.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

References II

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Pitt, M. K., R. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics 171*, 134–151.