FinM 331/Stat 339 Financial Data Analysis, Winter 2010

Floyd B. Hanson, Visiting Professor

Email: fhanson@uchicago.edu

Master of Science in Financial Mathematics Program
University of Chicago

Lecture 9

6:30-9:30 pm, 01 March 2010, Ryerson 251 in Chicago

7:30-10:30 pm, 01 March 2010 at UBS in Stamford

7:30-10:30 am, 02 March 2010 at Spring in Singapore

9. Bayesian Methods in Finance[a]:

● 9.1. Observed Variables, Unobserved Variables and Parameters:

In the previous lecture about Bayesian estimation, we had an observed vector variable $\vec{X}$ that was used to estimate a scalar or vector parameter $\vec{\Theta}$ that is unobserved. For instance in the basic log-normal model, the observables $\vec{X} = \mathbf{L}\vec{\mathbf{R}}$ are the log-returns of the asset prices and/or the $\vec{X} = \vec{C}$ option prices based upon the underlying asset. The parameters for $\vec{\Theta}$ are the $\{\mu, \sigma^2, r\}$ are the mean and variance coefficients and the spot rate, often a zero-coupon bond (ZCB) such that no interest is paid out until maturity satisfying the usual assumptions on r.

---

[a]For background to a good part of this lecture, see Michael Johannes and Nicholas Polson (2003), *MCMC Methods for Continuous-Time Financial Econometrics,* 96 pages, preprint for the Handbook of Financial Econometrics, edited by Yacine Ait-Sahalia and Lars Hansen.

○ *9.1.1. More Realistic Models:*

However, since the variance plays so much importance in options and other derivatives, it is important to recognize that it has its own dynamics and hence include that dynamics with the asset dynamics. Also, rare jump changes are very important because they can lead to severe losses or very profitable gains, in spite of their sparse population count among all the changes in price. Thus, our asset model might be a jump-diffusion in differential form,

$$dS(t) = S(t)(\mu dt + \sqrt{V(t)} dW_s(t) + (e^Q - 1)dP(t)), \quad (9.1)$$

where $\mathbf{E}[dP(t) = \lambda dt]$ in the mean number of jumps in $(0, dt]$ and $Q \in [a, b]$ is the underlying jump IID RV, and a stochastic volatility

$$dV(t) = \kappa_v(\theta_v - V(t))dt + \sigma_v\sqrt{V(t)} dW_v(t). \quad (9.2)$$

The diffusions are correlated by $\mathbf{E}[dW_s(t)dW_v(t)] = \rho_{s,v}dt$ and the can be implemented by another diffusion, $dW_\perp(t)$, independent of $dW_s(t)$ such that $dW v(t) = \rho_{s,v}dW_s(t) + \sqrt{1 - \rho_{s,v}^2}dW_\perp(t).$

In terms of observable units of time, assumed sufficiently small like one trading day, the difference equations for the log-return, $\mathbf{LR}_i = \mathbf{\Delta}\log(S_i) = \log(S(t_i)/S(t_{i-1}))$ for $i = 1:n$ **from** $n+1$ asset prices, are

$$\mathbf{LR}_i = \mu_{ld}\mathbf{\Delta}t + \sqrt{V_i}\mathbf{\Delta}W_{s,i} + \sum_{k=1}^{\mathbf{\Delta}P_i} Q_k, \qquad (9.3)$$

for $i = 1:n$, using the stochastic chain rule, where $\mu_{ld} = \mu - \sigma^2/2$ is the log-diffusion connected drift. In case of sufficiently small $\mathbf{\Delta}t$ such that the probability of more than one jump is very small than the sum term can be replaced by $Q_i\mathbf{\Delta}P_i$ and this will be assumed in the rest of the lecture.

The stochastic volatility or variance difference equation becomes

$$V_{i+1} = V_i + \kappa_v(\theta_v - V_i)\mathbf{\Delta}t + \sigma_v\sqrt{V_i}dW_{v,i}, \qquad (9.4)$$

forfor $i = 0:n-1$ given some V_0.

Thus, observables still are the log-returns

$$\vec{X} = \mathbf{L}\vec{\mathbf{R}}. \qquad (9.5)$$

The **dynamic parameters** are

$$\vec{\Theta} = \{\mu, \sigma^2, \lambda, a, b, \kappa_v, \theta_v, \sigma_v, \rho_{s,v}\}. \tag{9.6}$$

However, the are **unobserved variables**, not observed, except perhaps for some very large jumps, and they are

$$\vec{Y} = [\vec{V}; \vec{Q}; \vec{T}], \tag{9.7}$$

where $\vec{V} = [V_i]_{i=1}^n$ are the variances corresponding to each LR_i observation, $\vec{Q} = [Q_i]_{i=1}^n$ is the total jump-amplitude variable change between observations, and $\vec{T} = [T_i]_{i=1}^n$, if needed, is the jump-time if there is a jump, else both T_i and Q_i would be zero. The $\vec{Y} = [\vec{V}; \vec{Q}; \vec{T}]$ as unobserved variables are also called **latent variables or hidden variables** since there is either no data or incomplete data for them. Hence, the models for them such as (9.42) for $\vec{V}$ or the cumulative Poisson time between jumps distribution for $\vec{T}$ or the jump-amplitude RV distribution for $\vec{Q}$, all by simulation.

In the case of the **risk-neutral (RN) option pricing problem**, the stochastic-volatility jump-diffusion dynamics need to be revised to fit the assumption that the mean growth rate of the asset be constrained to grow at the spot rate r. Hence, we have

$$\mathbf{LR}_i \overset{\mathbf{rn}}{=} (r - \sigma^2/2)\Delta t + \sqrt{V_i}\Delta W_{s,i} + \Delta P_i Q_i - \mu_J \lambda \Delta t, \quad (9.8)$$

for $i = 1:n$, where $\mu_J = \mathbf{E}[Q]$. Note that the basic risk-neutral requirement is that

$$\mathbf{E}[\Delta S_i/S_i | S_i] = r\Delta t, \quad (9.9)$$

based on the underlying asset price and not its log-return.

The stochastic volatility or variance difference equation becomes

$$V_{i+1} \overset{\mathbf{rn}}{=} V_i + (\kappa_v(\theta_v - V_i) + \gamma_v V_i)\Delta t + \sigma_v \sqrt{V_i}\Delta W_{v,i}, \quad (9.10)$$

for $i = 0:n-1$ given some V_0, where γ_v is the **diffusive price of volatility risk**, which can be just be rescaled into RN values of $\{\kappa_v, \theta_v\}$.

The risk-neutral price of a European call option at current time t_i, maturing at $T = t_n$ with strike price K is the discounted, conditionally-expected call payoff,

$$C_i = C(S_i, t_i, V_i, \vec{\Theta}^{(\mathrm{rn})})$$

$$\overset{\mathrm{rn}}{=} e^{-r(T-t_i)} \mathbf{E}^{(\mathrm{rn})}[\max[S_i - K, 0]|S_i, V_i, \vec{\Theta}^{(\mathrm{rn})}], \tag{9.11}$$

where the purely risk-neutral parameters are

$$\vec{\Theta}^{(\mathrm{rn})} = \{r, K, T\}, \tag{9.12}$$

which are in addition to the dynamic parameters

$$\vec{\Theta} = \{\mu_{ld}\sigma^2, \lambda, a, b, \kappa_v, \theta_v, \sigma_v, \rho_v\} \tag{9.13}$$

and the total parameter set is $\vec{\Theta}^{(\mathrm{tot})} = [\vec{\Theta}; \vec{\Theta}^{(\mathrm{rn})}]$. Note, that the only genuine risk-neutral parameter is the spot rate r and that the estimate of the dynamical model parameter set (9.6) is still required. **Also, the data vectors are needed, i.e.,**

$$X = [\vec{S}; \vec{C}] = [S_i, C_i]_{i=1}^n. \tag{9.14}$$

- ● *9.2. Bayesian Formulation:*

The **starter Bayes rule** will include both parameters and latent variables, so is

$$f^{(\text{post})}_{\vec{\Theta}, \vec{Y} | \vec{X}} (\vec{\theta}, \vec{y} | \vec{x}) \propto f_{\vec{X} | \vec{\Theta}, \vec{Y}} (\vec{x} | \vec{\theta}, \vec{y}) \times f^{(\text{prior})}_{\vec{\Theta}, \vec{Y}} (\vec{\theta}, \vec{y}). \qquad (9.15)$$

The $\vec{x} = [x_i]^n_{i=1}$ are the observed data and the $\vec{y} = [y_i]^n_{i=1}$ are the unobserved, simulated data.

Already there are signs of difficulty since there are three vectors in the posterior distribution, parameters, latent variable and observation conditioning suggesting high dimensional computational and statistical complexity. Nonstandard distributions could easily arise. Monte Carlo (MC) simulation or the Markov chain Monte Carlo (MCMC=(MC)^2) simulation variant can reduce these complexities but using Markov chain approximations of the parameters.

Further, the joint density, especially in the posterior, make it inconvenient in finding the marginal density for the individual parameters.

Expanding the starter Bayes rule to separate out the parameters from the latent variables in the posterior yields,

$$f^{(\text{post})}_{\vec{\Theta},\vec{Y}|\vec{X}}(\vec{\theta},\vec{y}|\vec{x}) \propto f_{\vec{X}|\vec{\Theta},\vec{Y}}(\vec{x}|\vec{\theta},\vec{y}) \times f_{\vec{Y}|\vec{\Theta}}(\vec{y}|\vec{\theta}) \times f^{(\text{prior})}_{\vec{\Theta}}(\vec{\theta}). \quad (9.16)$$

Here, $f_{\vec{Y}|\vec{\Theta}}(\vec{y}|\vec{\theta})$ is the latent variable density and $f^{(\text{prior})}_{\vec{\Theta}}(\vec{\theta})$ is the prior parameter density.

The **full information likelihood** is

$$f_{\vec{X}|\vec{\Theta},\vec{Y}}\left(\vec{x}\,\middle|\,\vec{\theta},\vec{y}\right) \qquad (9.17)$$

and the **data-augmented likelihood** is the conditional marginal distribution, integrating out the latent variable,

$$f_{\vec{X}|\vec{\Theta}}\left(\vec{x}\,\middle|\,\vec{\theta}\right) = \int_{\mathcal{R}} f_{\vec{X},\vec{Y}|\vec{\Theta}}\left(\vec{x},\vec{y}\,\middle|\,\vec{\theta}\right) d\vec{y}$$
$$= \int_{\mathcal{R}} f_{\vec{X}|\vec{\Theta},\vec{Y}}\left(\vec{x}\,\middle|\,\vec{\theta},\vec{y}\right) f_{\vec{Y}|\vec{\Theta}}\left(\vec{y}\,\middle|\,\vec{\theta}\right) d\vec{y}. \qquad (9.18)$$

The application of these alternate forms depends on what information is available from the model, when the distribution is not available then simulation is necessary.

The **prior parameter density** $f_{\vec{\Theta}}^{(\mathrm{prior})}(\vec{\theta})$ is the place where the user puts in additional information that is beyond the probability inference of the prior.

- **9.3. *Bayesian Model Specification:***

Model specification is used to compare models using the posterior distribution, say for model $\mathcal{M}_i$,

$$f^{(\text{post})}_{\mathcal{M}_i|\vec{X}}(\mathcal{M}_i|\vec{x}) = f_{\vec{X}|\mathcal{M}_i}(\vec{x}|\mathcal{M}_i)\, f_{\mathcal{M}_i}(\mathcal{M}_i)/f_{\vec{X}}(\vec{x})\,, \qquad (9.19)$$

for $i = 1 : M$. Then the **posterior odds for model $\mathcal{M}_i$ versus $\mathcal{M}_j$** is

$$\frac{f^{(\text{post})}_{\mathcal{M}_i|\vec{X}}(\mathcal{M}_i|\vec{x})}{f^{(\text{post})}_{\mathcal{M}_j|\vec{X}}(\mathcal{M}_j|\vec{x})} = \frac{f_{\vec{X}|\mathcal{M}_i}(\vec{x}|\mathcal{M}_i)}{f_{\vec{X}|\mathcal{M}_j}(\vec{x}|\mathcal{M}_j)}\frac{f_{\mathcal{M}_i}(\mathcal{M}_i)}{f_{\mathcal{M}_j}(\mathcal{M}_j)}\,, \qquad (9.20)$$

where

$$\text{BF}[\mathcal{M}_i, \mathcal{M}_j] \equiv \frac{f_{\vec{X}|\mathcal{M}_i}(\vec{x}|\mathcal{M}_i)}{f_{\vec{X}|\mathcal{M}_j}(\vec{x}|\mathcal{M}_j)} \qquad (9.21)$$

is called the **Bayes factor**, such that if the Bayes is greater than one then model i is favored over model j and vice-versa.

● 9.4. *Joint Density and Bivariate Full Conditional Densities, the Hammersly-Clifford Theorem:*

However, the graph-theoretic theorem of Hammersly and Clifford, improved on by others, allows the relationship of the joint density to the full conditional densities of the parameters in the following bivariate form[a]

Theorem 9.4. Characterization of Joint Density by Full Conditional Densities:

The joint distribution with density $f_{X,Y}(x, y)$ associated with the conditional densities $f_{X|Y}(x|y)$ and $f_{Y|X}(y|x)$ has the density,

$$f_{X,Y}(x, y) = \frac{f_{Y|X}(y|x)}{\displaystyle\int_{\mathcal{R}} \frac{f_{Y|X}(z|x)}{f_{X|Y}(x|z)} dz}, \qquad (9.22)$$

assuming all densities exist (and satisfy positivity).

---

[a]C.P. Robert and G. Casella (2004), Monte Carlo Statistical Methods, Springer, p. 344. This book also clarifies a lot of the technical notions in Johannes and Polson (2003.2006).

Proof: By alternate application of the multiplicative expansion in conditional densities,

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = f_{X|Y}(x|y)f_Y(y) \qquad (9.23)$$

then

$$\int_{\mathcal{R}} \frac{f_{Y|X}(y|x)}{f_{X|Y}(x|y)}dy = \int_{\mathcal{R}} \frac{f_Y(y)}{f_X(x)}dy = \frac{1}{f_X(x)} \qquad (9.24)$$

so

$$f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x) = \frac{f_{Y|X}(y|x)}{\displaystyle\int_{\mathcal{R}} \frac{f_{Y|X}(z|x)}{f_{X|Y}(x|z)}dz}. \qquad (9.25)$$

☐

The validity of this theorem is related to global and local Markov properties, so is useful for the Markov models we use in finance. The best news that this theorem helps break down the high dimensional demands of the more realistic financial models.

● 9.5. *Marginal Densities for Single Parameters:*

For estimation or simulation, it is sometimes necessary to focus on a single parameter, say θ_i, so it necessary to find the **single parameter posterior marginal distribution** by integration over the parameter complement,

$$\vec{\theta}_{(-i)} \equiv [\theta_j]_{j \neq i}, \tag{9.26}$$

for for $i = 1 : p$ parameters, so

$$f^{(\text{post})}_{\Theta_i | \vec{Y}}(\theta_i | \vec{y}) = \int_{\mathcal{R}} f^{(\text{post})}_{\Theta_i, \vec{\Theta}_{(-i)}, \vec{X} | \vec{Y}} \left(\theta_i, \vec{\theta}_{(-i)}, \vec{x} \,\middle|\, \vec{y} \right) d\vec{x} d\vec{\theta}_{(-i)}. \tag{9.27}$$

This marginal is useful for means, modes and and risk estimation, such as standard errors, quantiles and **credible sets (Bayes proper name for Bayesian confidence intervals)**.

● 9.6. *Joint Density and Multivariate Full Conditional Densities, the General Hammersly-Clifford Theorem:*

Like the Gauss-Seidel method of successive approximations, some important simulation methods will update another parameter at each successive iteration, so that at the end of the ith iteration

$$\vec{\theta}_i^{(1)} = [\theta_j^{(1)}]_{j=1}^i \text{ and } \vec{\theta}_i^{(0)} = [\theta_j^{(0)}]_{j=i+1}^p (1 - \delta_{i,p}), \qquad (9.28)$$

denotes the updated part and the unupdated part, respectively.

Theorem 9.6. Characterization of Joint Density by Full Conditional Densities in Successive Approximations:

The joint distribution with density $f_{\vec{\Theta}}\left(\vec{\theta}\right)$, associated with the conditional densities $f_{\Theta_i}\left(\theta_i | \vec{\theta}_{(-i)}\right)$, has the density,

$$f_{\vec{\Theta}}(\vec{\theta}^{(1)}) \propto \prod_{i=1}^p \frac{f_{\Theta_i}\left(\theta_i^{(1)} | \vec{\theta}_{i-1}^{(1)}, \vec{\theta}_{i+1}^{(0)}\right)}{f_{\Theta_i}\left(\theta_i^{(0)} | \vec{\theta}_{i-1}^{(1)}, \vec{\theta}_{i+1}^{(0)}\right)}, \qquad (9.29)$$

assuming admissible $\theta_i^{(0)}$, all densities exist (and satisfy positivity).

This multi-parameter version of the bi-parameter version in Theorem 9.4, but is less general than a similar version in Robert and Casella (2004) and there exist even stronger versions. Dropping positivity takes much stronger assumptions, but for simulation, you have to remember that random simulation values are almost always on open intervals, but that excludes the normal.

• *9.7. Gibbs Sampler, a Simple MCMC Algorithm*[a]*:*

Given an initial parameter-latent variable guestimate $[\vec{\Theta}^{(0)}, \vec{Y}^{(0)}]$, the **Gibbs sampler, in the full block form**, alternately draws samples from the posterior full conditionals of the parameters and the latent variables distributions,

$$\vec{\theta}^{(g)} \overset{\text{dist}}{\sim} f^{(\text{post})}_{\vec{\Theta}|\vec{Y},\vec{X}}\left(\vec{\theta}\,\middle|\,\vec{y}^{(g-1)}, \vec{x}\right) \tag{9.30}$$

$$\vec{y}^{(g)} \overset{\text{dist}}{\sim} f^{(\text{post})}_{\vec{Y}|\vec{\Theta},\vec{X}}\left(\vec{y}\,\middle|\,\vec{\theta}^{(g)}, \vec{x}\right), \tag{9.31}$$

for $g = 1:G$ Gibbs iterations or **stopping for convergence at g^* according to a specified stopping criterion**, and if convergence then a sample is drawn from the convergent limit

$$f^{(\text{post}),g^*}_{\vec{\Theta},\vec{Y}|\vec{X}}\left(\vec{\theta}, \vec{y}\,\middle|\,\vec{x}\right), \tag{9.32}$$

with possible kernel smoothing. However, if the block form is not usable, then the scalar parameter form can be used.

---

[a]See Rice (2007), p.297ff or see Johannes and Polson (2003,2006), pp. 14-15 or better Robert and Casella (2004).

In the **parameter by parameter form**, relying on joint density decomposition of Theorem 9.6, then, given the $[\vec{\Theta}^{(0)}, \vec{Y}^{(0)}]$ block start, the Gibbs sampler iteration is

$$\theta_i^{(1)} \overset{\text{dist}}{\sim} f_{\vec{\Theta}|\vec{Y},\vec{X}}^{(\text{post})}\left(\theta_i \middle| \vec{\theta}_{i-1}^{(1)}, \vec{\theta}_{i+1}^{(0)}, \vec{y}, \vec{x}\right) \qquad (9.33)$$

for $i = 1:p$ and then the latent variables states are draw if possible in block form or else component by component as with the parameters.

● *9.8. Metropolis and Hastings MCMC Algorithm:*

When the Gibbs Sampler is not convenient or practical, the classical Monte Carlo method of Metropolis-Hastings[a] can be used with the more general acceptance and rejection technique of von Neumann.

In the simple case, the focus is on a single parameter θ_i and suppressing other dependence, let

$$\pi(\theta_i) \equiv f_{\Theta_i}\left(\theta_i \middle| \vec{\theta}_{(-i)}, \vec{y}, \vec{x}\right). \tag{9.34}$$

For generating iteration samples of $\pi(\theta_i)$, it is necessary to propose a candidate conditional **transition** density $q\left(\theta_i^{(g+1)} \middle| \theta_i^{(g)}\right)$ for each iteration $g+1 = 1:G$, again suppressing dependence on necessary parameters. The acceptance criterion for each new sample is the threshold,

$$\alpha(\theta_i^{(g)}, \theta_i^{(g+1)}) = \min\left[\frac{\pi\left(\theta_i^{(g+1)}\right) \middle/ q\left(\theta_i^{(g+1)} \middle| \theta_i^{(g)}\right)}{\pi\left(\theta_i^{(g)}\right) \middle/ q\left(\theta_i^{(g)} \middle| \theta_i^{(g+1)}\right)}, 1\right]. \tag{9.35}$$

---

[a]The Monte Carlo method is often called is often called just the Metropolis method or algorithm from Metropolis et al. (1953) and Metropolis and Ulam (1949), but Hastings (1970) made some later improvements.

The Metropolis-Hastings MCMC algorithm in this case is

$$\text{Draw proposed } \widehat{\theta} \overset{\text{dist}}{\sim} q\left(\theta \mid \theta_i^{(g)}\right); \qquad (9.36)$$

$$\text{Draw acceptor } u \overset{\text{dist}}{\sim} \text{Uniform}(0,1); \qquad (9.37)$$

$$\text{If } u < \alpha(\theta_i^{(g)}, \widehat{\theta}), \quad \text{Then accept } \theta_i^{(g+1)} = \widehat{\theta}; \qquad (9.38)$$

$$\text{Else reject}, \quad \theta_i^{(g+1)} = \theta_i^{(g)}; \qquad (9.39)$$

for $g + 1 = 1 : G$.

If the sample draws are **independent between iterations** g, then let
$q\left(\theta_i^{(g+1)} \mid \theta_i^{(g)}\right) = \widetilde{q}\left(\theta_i^{(g+1)}\right)$ and

$$\alpha(\theta_i^{(g)}, \theta_i^{(g+1)}) = \min\left[\frac{\pi\left(\theta_i^{(g+1)}\right) \Big/ \widetilde{q}\left(\theta_i^{(g+1)}\right)}{\pi\left(\theta_i^{(g)}\right) \Big/ \widetilde{q}\left(\theta_i^{(g)}\right)}, 1\right]. \qquad (9.40)$$

It is customary to discard the initial iterations from the distribution iterations and is called **burn-in** since they are too crude to be used.

• 9.9. Bayesian Formulation for Discretized Stochastic Asset and Volatility/Variance:

As an example, the stochastic asset and variance equations, (9.3) and (9.42), are rewritten for Bayesian implementation, asset prices, are

$$X_i = \mathbf{LR}_i = \mu_{ld}\Delta t + \sqrt{V_i\Delta t}\,Z_{s,i} + Q_iN_i, \qquad (9.41)$$

and

$$V_{i+1} = V_i + \kappa_v(\theta_v - V_i)\Delta t$$
$$+ \sigma_v\sqrt{V_i\Delta t}(\rho_{s,v}Z_{s,i} + \sqrt{1-\rho_{s,v}^2}\,Z_{\perp,i}), \qquad (9.42)$$

where we set $\Delta W_{j,i} = \sqrt{\Delta t}Z_{j,i}$ for $j = s:\perp$ with $[Z_{s,i}; Z_{\perp,i}] \stackrel{\mathbf{dist}}{\sim} \mathcal{N}(\vec{0}_2, I_2)$ and $\Delta P_i = N_i$. For the jump-process, $N_i \stackrel{\mathbf{dist}}{\sim} \mathcal{B}ern(\lambda_s\Delta t)$, a Bernoulli for a $(0-1)$ Poisson, and $Q_i \stackrel{\mathbf{dist}}{\sim} \Pi(a, b)$, a log-jump-amplitude distribution, openly specified.

Note that although the $X_i = \mathbf{LR}_i$ are given as IID through the IIND $Z_{s,i}$, the variance V_{i+1} depends directly on the previous value V_i, so are thus interdependent.

Consequently, the **likelihood is decomposable**,

$$f_{\vec{X}|\vec{Y},\vec{\Theta}}\left(\vec{x}|\vec{y},\vec{\theta}\right) = \prod_{i=1}^{n} f_{X_i|Y_i\vec{\Theta}}\left(x_i|y_i,\vec{\theta}\right), \qquad (9.43)$$

but is still a multivariate normal.

However, the **latent state dynamics** has variance coupling since

$$\vec{Y} = [\vec{V}, \vec{N}, \vec{Q}], \qquad (9.44)$$

so

$$f_{\vec{Y}|\vec{\Theta}}\left(\vec{y}|\vec{\theta}\right) = \prod_{i=1}^{n} f\left(V_i, N_i, Q_i|V_{i-1},\vec{\theta}\right)$$

$$= \prod_{i=1}^{n} f\left(V_i|V_{i-1}, N_i, Q_i,\vec{\theta}\right) f\left(N_i, Q_i|\vec{\theta}\right) \qquad (9.45)$$

$$= \prod_{i=1}^{n} f\left(V_i|V_{i-1}, N_i, Q_i,\vec{\theta}\right) f\left(Q_i|N_i,\vec{\theta}\right) f\left(N_i|\vec{\theta}\right),$$

after multiple applications of Bayesian rules.

For the prior parameter $f_{\vec{\Theta}}(\vec{\theta})$, it is strongly suggested that **conjugate priors** in coordination with the likelihood function be used and that a flat prior be use so as not to strongly bias the posterior to a strongly desired value. For instance, a normal prior could be used for the mean μ_{ld} of a pure diffusive log-return so that the the full conditional would be $f(\mu_{ld}|\sigma_s^2, \vec{x})$.

For the precision or reciprocal variance $\xi = 1/\sigma_s^2$, we found previously that the gamma distribution was a good conjugate prior. However, working directly with the variance means that the appropriate conjugate prior is the inverted or so-called inverse gamma function, $f_X^{(ig)}$, which is derivable from the gamma function,

$$f_X^{(g)}(x; \lambda, \alpha) = \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}. \tag{9.46}$$

By the transformation rule for densities, letting $y = 1/x$, yields the inverted gamma density[a],

$$f_X^{(ig)}(x; \lambda, \alpha) = f_X^{(g)}(1/x; \lambda, \alpha)|d(1/x)/dx|$$
$$= \frac{\lambda^\alpha x^{-\alpha-1}}{\Gamma(\alpha)} e^{-\lambda/x}, \tag{9.47}$$

There does not appear to by an inverted gamma RNG, but it is not needed by the RV relationship the gamma distribution RNG can be used,

$$X^{(ig)} = 1/X^{(g)} \tag{9.48}$$

and

$$\text{Prob}[X^{(ig)} \leq x] = \text{Prob}[X^{(g)} \geq 1/x]. \tag{9.49}$$

Getting the Bayesian estimate of the variance rather than the precision allows direct use of the posterior distribution of the variance, the prime variable of interest.

---

[a]Caution: Johannes and Polson (2003) have a sign typo on p. 40 in the exponent of the power term.

Since **MATLAB** defines the **gamma probability density function**, **gampdf(x,a,b)**, as

$$f(x|a,b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-x/b}, \tag{9.50}$$

then the inverted gamma density, called say

$$\texttt{y=igampdf(x,lambda,alpha);} \tag{9.51}$$

is computed by

$$\texttt{igampdf=gampdf(1/x,alpha,1/lambda)}/x^2; \tag{9.52}$$

In theory this would need the qualification that $x > 0$, but technically the random number generation is always on the open interval. The inverted gamma complementary cumulative distribution function (CCDF) is given by

$$\texttt{igamccdf=gamcdf(1/x,alpha,1/lambda);} \tag{9.53}$$

Similarly, inverted gamma random samples can be generated by

$$\texttt{igamrnd=1/gamrnd(alpha,1/lambda,m,n);} \tag{9.54}$$

• 9.10. Bayesian Asset Pricing:

○ 9.10.1. Geometric Brownian Motion (GBM) or Pure Diffusion Example:

In the case to GBM with constant coefficients, the log-return has a simple solution,

$$X_i = \mathrm{LR}_i = \mu_{ld}\Delta t + \sigma_s\sqrt{\Delta t}\,Z_{s,i}, \tag{9.55}$$

where $Z_{s,i} \overset{\mathrm{dist}}{\sim} \mathcal{N}(0,1)$. The Bayes rule for the posterior is simply,

$$f^{(\mathrm{post})}_{\vec{\Theta}|\vec{X}}\left(\vec{\theta}|\vec{y}\right) \propto f_{\vec{X}|\vec{\Theta}}\left(\vec{x}|\vec{\theta}\right) \times f^{(\mathrm{prior})}_{\vec{\Theta}}\left(\vec{\theta}\right). \tag{9.56}$$

The likelihood is continuously compounded log-returns, assuming a constant trading day Δt, $\widetilde{\mu}_{ld} = \mu_{ld}\Delta t$ and $\widetilde{\sigma}_s^2 = \sigma_s^2\Delta t$,

$$
\begin{aligned}
f_{\vec{X}|\vec{\Theta}}\left(\vec{x}|\vec{\theta}\right) &= f(\vec{X}|\widetilde{\mu}_{ld},\widetilde{\sigma}_s^2) = \prod_{i=1}^{n} f^{(\mathrm{n})}(x_i;\widetilde{\mu}_{ld},\widetilde{\sigma}_s) \\
&= (2\pi\widetilde{\sigma}_s^2)^{-n/2}\exp\left(-\frac{1}{2\widetilde{\sigma}_s^2}\sum_{i=1}^{n}(x_i-\widetilde{\mu}_{ld})^2\right).
\end{aligned} \tag{9.57}
$$

The posterior parameter distribution is

$$f_{\vec{\Theta}|\vec{X}}^{(\text{post})}\left(\vec{\theta}|\vec{x}\right) = f_{\vec{\Theta}|\vec{X}}\left(\widetilde{\mu}_{ld}, \widetilde{\sigma}_s^2|\vec{x}\right). \tag{9.58}$$

Although the posterior is available in closed form, we will proceed to demonstrate more general methods for more general problems.

By the Hammersley-Clifford theorems, we need only consider separated posterior full conditionals assuming parameter independence,

$$f_{\Theta_1}^{(\text{post})}\left(\widetilde{\mu}_{ld}|\widetilde{\sigma}_s^2, \vec{x}\right) \propto f_{\vec{X}}(\vec{x}|\widetilde{\mu}_{ld}, \widetilde{\sigma}_s^2) \times f_{\Theta_1}^{(\text{prior})}(\widetilde{\mu}_{ld});$$

$$f_{\Theta_2}^{(\text{post})}\left(\widetilde{\sigma}_s^2|\widetilde{\mu}_{ld}, \vec{x}\right) \propto f_{\vec{X}}(\vec{x}|\widetilde{\mu}_{ld}, \widetilde{\sigma}_s^2) \times f_{\Theta_2}^{(\text{prior})}(\widetilde{\sigma}_s^2). \tag{9.59}$$

Next assuming **conjugate priors**, using normal for the mean parameter and inverted gamma for the variance parameter,

$$f_{\Theta_1}^{(\text{prior})}(\widetilde{\mu}_{ld}) \stackrel{\text{dist}}{\sim} \mathcal{N}(m_0, v_0);$$

$$f_{\Theta_2}^{(\text{prior})}(\widetilde{\sigma}_s^2) \stackrel{\text{dist}}{\sim} \mathcal{IG}(lam_0, alp_0); \tag{9.60}$$

for prior parameters $\{m_0, v_0, lam_0, alp_0\}$, selected for distribution flatness, but not too flat.

The MCMC Gibbs sampler version uses the successive sampling updates from the posterior conditionals (9.59) of similar distributions to the priors (9.86),

$$\widetilde{\mu}_{ld}^{(g+1)} \overset{\text{dist}}{\sim} f^{(\text{post})}\big(\widetilde{\mu}_{ld}|(\widetilde{\sigma}_s^2)^{(g)}, \vec{x}\big) \propto \mathcal{N}(m_1^{(g)}, v_1^{(g)});$$

$$(\widetilde{\sigma}_s^2)^{(g+1)} \overset{\text{dist}}{\sim} f^{(\text{post})}\big(\widetilde{\sigma}_s^2|\widetilde{\mu}_{ld}^{(g+1)}, \vec{x}\big) \propto \mathcal{IG}(lam_1^{(g+1)}, alp_1). \tag{9.61}$$

Upon completing the square for the prior mean normal with the likelihood normal and similarly combining the pair of gamma distributions[a],

$$m_1^{(g)} = \frac{nv_0\overline{X}_n + m_0(\widetilde{\sigma}_s^2)^{(g)}}{nv_0 + (\widetilde{\sigma}_s^2)^{(g)}}; \quad v_1^{(g)} = \frac{v_0(\widetilde{\sigma}_s^2)^{(g)}}{nv_0 + (\widetilde{\sigma}_s^2)^{(g)}};$$

$$alp_1 = \frac{n}{2} + alp_0; \quad lam_1^{(g+1)} = \frac{n}{2}\left(S_n^2 + \left(\overline{X}_n - \widetilde{\mu}_{ld}^{(g+1)}\right)^2\right) + lam_0. \tag{9.62}$$

---

[a]See similar results on Lecture 8, pp. 50-52 for a similar problem using a proper conjugate pair in the normal distribution example. See also Fink (1997) and Wikipedia page.

○ 9.10.2. Black Scholes European Option Pricing Example:

Option prices, though a mixed set of data are observed. In 1973 Black and Scholes finally published their famous paper with Merton publishing a lengthy justification and generalization of it at the same time. Their model uses a geometric Brownian motion representation of the asset underling the option. For the call European option with the possibility to buy the asset at a strike price K and the time of maturity T from the beginning of the contract results in the formula of Black and Scholes,

$$C(t) = \mathrm{BS}(S(t), t; K, T, \sigma_s, r)$$
$$= S(t) F^{(n)}(d_1(S(t), t); 0, 1) \qquad (9.63)$$
$$- e^{-r(T-t)} K F^{(n)}(d_1(S(t), t) - \sigma_s \sqrt{T-t}; 0.1),$$

where the argument of the first standard normal distribution is

$$d_1(S(t), t) = \frac{\log(S(t)/K) + (r + \sigma_s^2/2)(T-t)}{\sigma_s \sqrt{T-t}} \qquad (9.64)$$

and it is assumed that the spot rate of interest r is known.

There is one technical problem them here which is called the **stochastic singularity** that arises with the solution simultaneous equation where there are some imperfection, so in the spirit of least square errors and kernel smoothing we will **add a small normal error to the discretized call option price**

$$C_i = \text{BS}(S_i, t_i; K, T, \sigma_s, r) + \sigma_c Z_{c,i}, \qquad (9.65)$$

where $Z_{c,i} \overset{\text{dist}}{=} \mathcal{N}(0,1)$, and couple this with the log-return equation

$$\text{LR}_i = \log(S_i/S_{i-1}) = \mu_{ld}\Delta t + \sigma_s\sqrt{\Delta t}\, Z_{s,i} = \widetilde{\mu}_{ld} + \widetilde{\sigma}_s Z_{s,i} \quad (9.66)$$

for $i = 1 : n$, where Δt is known, $t_i = i\Delta t$ and $t_n = n\Delta t = T$. Note that the call price is a conditional normal mixture and not a simple normal due to the nonlinear Black-Scholes function BS. The **likelihood function is joint with asset and option likelihoods**,

$$f_{s,c}\left(\vec{S}, \vec{C} \mid \widetilde{\mu}_{ld}, \widetilde{\sigma}_s^2, \sigma_c^2\right) = \prod_{i=1}^{n} f_c\left(C_i \mid S_i, \widetilde{\sigma}_s^2, \sigma_c^2\right)$$
$$\times f_s\left(\log(S_i/S_{i-1}) \mid \widetilde{\mu}_{ld}, \widetilde{\sigma}_s^2\right). \qquad (9.67)$$

Due to hedging, $\widetilde{\mu}_{ld}$ only appears in the **asset part of the likelihood distribution** f_s and this is the same as previously in (9.57) with $x_i = LR_i$.

The **option part of the likelihood** f_c inherits the normal distribution from the heuristic error $\sigma_c Z_{c,i}$,

$$f_c\left(C_i | S_i, \widetilde{\sigma}_s^2, \sigma_c^2\right) \propto \exp\left(-\frac{1}{2\sigma_c^2}(C_i - \widetilde{\text{BS}}_i(S_i; \widetilde{\sigma}_s))^2\right). \qquad (9.68)$$

where $\widetilde{r} = r\Delta t$,

$$\widetilde{\text{BS}}_i(S_i; \widetilde{\sigma}_s) = S_i F^{(n)}\left(\widetilde{d}_{1,i}(S_i; \widetilde{\sigma}_s); 0, 1\right)$$
$$-e^{-\widetilde{r}(n-i)} F^{(n)}\left(\widetilde{d}_{1,i}(S_i; \widetilde{\sigma}_s) - \widetilde{\sigma}_s\sqrt{n-i}; 0, 1\right) \qquad (9.69)$$

and

$$\widetilde{d}_{1,i}(S_i; \widetilde{\sigma}_s) = (\log(S_i/K) + (\widetilde{r} + \widetilde{\sigma}_s^2/2)(n-i))/(\widetilde{\sigma}_s\sqrt{n-i}). \quad (9.70)$$

For the MCMC simulation algorithm, again be can sample the joint posterior distribution $f_{\vec{\Theta}}^{(\text{post})}(\widetilde{\mu}_{ld}, \widetilde{\sigma}_s^2, \sigma_c^2 | \vec{S}, \vec{C})$ by the Hammersley-Clifford theorem through the full posterior conditionals $f_{\Theta_1}^{(\text{post})}(\widetilde{\mu}_{ld} | \widetilde{\sigma}_s^2, \vec{S})$, $f_{\Theta_2}^{(\text{post})}(\widetilde{\sigma}_s^2 | \widetilde{\mu}_{ld}, \vec{S})$ and $f_{\Theta_3}^{(\text{post})}(\sigma_c^2 | \widetilde{\sigma}_s^2, \vec{S}, \vec{C})$.

The full conditional priors are assumed to be independent, such that

$$f_{\Theta_1}^{(\text{prior})}(\widetilde{\mu}_{ld}) \overset{\text{dist}}{\sim} \mathcal{N}(m_s, v_s);$$

$$f_{\Theta_2}^{(\text{prior})}(\widetilde{\sigma}_s^2) \overset{\text{dist}}{\sim} \mathcal{IG}(\text{lam}_s, \text{alp}_s); \qquad (9.71)$$

$$f_{\Theta_3}^{(\text{prior})}(\sigma_c^2) \overset{\text{dist}}{\sim} \mathcal{IG}(\text{lam}_c, \text{alp}_c).$$

For the **posterior MCMC sampling**, the full conditional posteriors for both $\widetilde{\mu}_{ld}$ and σ_c^2 will be conjugate to the prior, but since $\widetilde{\sigma}_s^2$ is strongly coupled to the option pricing formula of unknown distribution, the **more general Metropolis-Hastings algorithm will be needed**. Thus, the successive iterations are of the form

$$\widetilde{\mu}_{ld}^{(g+1)} \overset{\text{dist}}{\sim} f^{(\text{post})}\left(\widetilde{\mu}_{ld}|(\widetilde{\sigma}_s^2)^{(g)}, \vec{S}\right) \propto \mathcal{N}(\widetilde{m}_s^{(g)}, \widetilde{v}_s^{(g)});$$

$$(\sigma_c^2)^{(g+1)} \overset{\text{dist}}{\sim} f^{(\text{post})}\left(\sigma_c^2|(\widetilde{\sigma}_s^2)^{(g)}, \vec{S}, \vec{C}\right) \propto \mathcal{IG}(\widetilde{\lambda}_c^{(g+1)}, \widetilde{\alpha}_c). \quad (9.72)$$

$$(\widetilde{\sigma}_s^2)^{(g+1)} \overset{\text{dist}}{\sim} f^{(\text{post})}\left(\widetilde{\sigma}_s^2|\widetilde{\mu}_{ld}^{(g+1)}, (\sigma_c^2)^{(g+1)}, \vec{S}, \vec{C}\right) \propto \ ??$$

An alternate way of sampling $(\widetilde{\sigma}_s^2)^{(g+1)}$ is to use the independence version of the Metropolis algorithm which includes von Neumann's acceptance and rejection technique and was discussed on pp. 19-20 of this lecture.

○ 9.10.2. Merton's Jump-Diffusion Model Example:

This is a multivariate extension of Merton's (1976) pioneering jump-diffusion paper. The asset price vector is $\vec{S} = [S_j]_{j=1}^{K}$ and it satisfies the **geometric Brownian-Poisson motion**,

$$d\vec{S}(t) = \vec{\mu}.\ast\vec{S}(t)dt + \vec{\sigma}.\ast\vec{S}(t).\ast d\vec{W}(t)$$
$$+ \vec{S}(t).\ast(e^{\vec{Q}} - \vec{1})dP(t), \tag{9.73}$$

where elemental **MATLAB** dot-multiplication is used, the $\{\vec{\mu}, \vec{W}(t), \vec{Q}, \vec{1}, P(t), \vec{\lambda}\}$ are all K dimensional vectors with, except for $dP(t)$, otherwise, the usual assumptions. The diffusion matrix is $\Sigma \equiv \vec{\sigma}\vec{\sigma}^{\top}$ Although is is assumed that the Poisson jumps are common jumps so $dP(t)$ **is a scalar**, the log-jump-amplitudes $\vec{Q} \overset{\text{dist}}{=} \mathcal{N}(\vec{\mu}_J, \Sigma_J)$ are a multivariate normal with mean vector $\vec{\mu}_J$ and covariance matrix Σ_J. There can be correlated log-jump-amplitudes.

Transforming the component form of (9.73) to obtain the **log-returns,** assuming that more than one jump per trading day is insignificant,

$$\mathbf{LR}_j(t) = \log(S_j(t)/S_j(t-\Delta t))$$
$$= \mu_j \Delta t + \sigma_j \Delta W_j(t-\Delta t) + Q_j \Delta P(t-\Delta t), \quad (9.74)$$

for $j = 1:K$. Next descretizing to obtain the **daily log-returns**,

$$\mathbf{LR}_{j,i} = \mu_j \Delta t + \sigma_j \sqrt{\Delta t} Z_{j,i} + Q_{j,i} N_i, \quad (9.75)$$

for $i = 1:n$, with obvious substitution for certain terms. Finally, converting back to vector form,

$$\vec{\mathbf{LR}}_i = \vec{\mu}_s + \vec{\sigma}_s . * \vec{Z}_i + \vec{Q}_i N_i, \quad (9.76)$$

where $\vec{\mu}_s \equiv [\mu_j \Delta t]_{j=1}^K$ and $\vec{\sigma}_s \equiv \left[\sigma_j \sqrt{\Delta t}\right]_{j=1}^K$, $\Sigma_s \equiv \vec{\sigma}_s \vec{\sigma}_s^\top = \Sigma \Delta t$, and $\mathbf{E}[N_i] = \Lambda_J \equiv \lambda \Delta t$, absorbing the time-step. Once Λ_J, the average jump count, is found, then the validity of the Bernoulli **0-1** jump law can be demonstrated by **MATLAB** computation if

$$\texttt{P2=1-poisscdf(2,LamJ)} \quad (9.77)$$

is negligible, i.e., two or more jumps are unlikely.

For the hybrid **MCMC approximation**, the **unknown parameters** are

$$\Theta = [\vec{\mu}_s; \Sigma_s, \Lambda_J, \vec{\mu}_J, \Sigma_J], \tag{9.78}$$

a structure of a scalar, vectors and matrices. The **latent or hidden variables**, due to the compound Poisson jump process, are

$$Y = [\vec{N}; Q] = \left[\vec{N}; [\vec{Q}_i]_{i=1}^n\right]. \tag{9.79}$$

The observation asset prices is an array of K-vectors,

$$X = [\vec{X}_i]_{i=1}^n = [\mathbf{L}\vec{R}_i]_{i=1}^n = [[\mathbf{L}R_{j,i}]_{j=1}^K]_{i=1}^n. \tag{9.80}$$

Assuming the that the log-returns are IID, then the **full likelihood** distribution is decomposable,

$$f_{X|\Theta,Y}(x|\theta, y) = \prod_{i=1}^n f_{\vec{X}_i|\Theta,\vec{Y}_i}\left(\vec{x}_i|\theta, N_i, \vec{Q}_i\right), \tag{9.81}$$

where

$$f_{\vec{X}_i|\Theta,\vec{Y}_i}\left(\vec{x}_i|\theta, N_i, \vec{Q}_i\right) \propto |\Sigma_s|^{-0.5}\exp\left(-0.5\left(\vec{x}_i - \vec{\mu}_s - N_i\vec{Q}_i\right)^\top \right.$$
$$\left. \Sigma_s^{-1}\left(\vec{x}_i - \vec{\mu}_s - N_i\vec{Q}_i\right)\right). \tag{9.82}$$

The linear algebra complexity grows with dimension.

The **observed likelihood** in the **(0-1)** jump-case, is the conditional marginal distribution

$$f_{\vec{X}_i|\Theta}(\vec{x}_i|\theta) \simeq \sum_{k=0}^{1} \frac{\delta_{k,0} + \Lambda_J \delta_{k,1}}{1 + \Lambda_J} \int_{\mathcal{R}^K} f_{\vec{X}_i|\Theta,\vec{Y}_i}(\vec{x}_i|\theta, k, \vec{q}_i)$$
$$\cdot f^{(n)}(\vec{q}_i; \vec{\mu}_J, \Sigma_J) d\vec{q}_i, \qquad (9.83)$$

will still be a complicated mixture of multivariate normals.

By Hammersley-Clifford, the posterior parameter full conditionals characterizing the full parameter posterior $f_{\Theta|Y,X}^{(\text{post})}(\theta|y,x)$ are

$$f_{\vec{\Theta}_s|\vec{N},Q,X}^{(\text{post})}(\vec{\mu}_s; \Sigma_s|\vec{n}, q, x);$$
$$f_{\vec{\Theta}_J|\vec{N},Q}^{(\text{post})}(\vec{\mu}_J; \Sigma_J|\vec{n}, q); \qquad (9.84)$$
$$f_{\Lambda_J|\vec{N}}^{(\text{post})}(\Lambda_J|\vec{n}).$$

The latent state full conditionals characterizing the full latent state posterior $f_{Y|\Theta,X}^{(\text{post})}(y|\theta,x)$ are

$$f_{N_i|\vec{\Theta},\vec{Q}_i,\vec{X}_i}^{(\text{post})}(n_i|\theta,\vec{q}_i,\vec{x}_i);$$

$$f_{\vec{Q}_i|\vec{\Theta},N_i,\vec{X}_i}^{(\text{post})}(\vec{q}_i|\theta,n_i,\vec{x}_i);$$

(9.85)

for $i=1:n$.

The standard conjugate priors are

$$f_{\Theta_1}^{(\text{prior})}(\vec{\mu}_s) \overset{\text{dist}}{\sim} \mathcal{N}(\vec{m}_s, \Sigma_s^{(\text{prior})});$$

$$f_{\Theta_2}^{(\text{prior})}(\Sigma_s) \overset{\text{dist}}{\sim} \mathcal{IW}(\Psi_s, \text{dof}_s);$$

$$f_{\Theta_3}^{(\text{prior})}(\vec{\mu}_J) \overset{\text{dist}}{\sim} \mathcal{N}(\vec{m}_J, \Sigma_J^{(\text{prior})});$$

(9.86)

$$f_{\Theta_4}^{(\text{prior})}(\Sigma_J) \overset{\text{dist}}{\sim} \mathcal{IW}(\Psi_J, \text{dof}_J);$$

$$f_{\Theta_5}^{(\text{prior})}(\Lambda_J) \overset{\text{dist}}{\sim} \mathcal{B}(a_J, b_J);$$

where $\mathcal{B}(a, b)$ is the **beta** distribution introduced in (8.8;L8-p5) and the $\mathcal{IW}(\Psi, \text{dof})$ is the inverse Wishart distribution, also denoted by $\mathcal{W}^{-1}$.

The **Wishart distribution** is the distribution of the sample covariance matrix for samples from multivariate normals, so the inverse Wishart is conjugate to the multivariate normal[a]. The Wishart and inverse Wishart distribution are related by

$$\text{If } \Sigma \overset{\text{dist}}{\sim} \mathcal{W}^{-1}(\Psi, \text{dof}); \quad \text{Then } \Sigma^{-1} \overset{\text{dist}}{\sim} \mathcal{W}(\Psi^{-1}, \text{dof}); \quad (9.87)$$

The Wishart and inverse Wishart densities are given by

$$f_W(W|\Sigma, \text{dof}) \propto \frac{|W|^{(\text{dof}-K-1)/2}}{|\Sigma|^{\text{dof}/2}} \exp\left(-0.5\text{Tr}[\Sigma^{-1}W]\right);$$

$$\quad (9.88)$$

$$f_{W^{-1}}(W|\Sigma, \text{dof}) \propto \frac{|\Sigma|^{\text{dof}/2}}{|W|^{(\text{dof}+K+1)/2}} \exp\left(-0.5\text{Tr}[\Sigma\, W^{-1}]\right);$$

where $|\Sigma| = \text{Det}[\Sigma]$ denotes the determinant of Σ. Note that the forms are very similar except for interchanging the two $K \times K$ symmetric positive semi-definite matrices Σ and W.

---

[a]See Wikipedia on the three related articles *Wishart Distribution*, *Inverse Wishart Distribution* and *Estimation of Covariance Matrices*.

Why are the exponents is the form of the **trace**, $\mathrm{Tr}[AB] = \sum_{i=1}^{K} \sum_{j=1}^{K} A_{i,j} B_{i,j}$ of the product of two **symmetric matrices**, rather than the normal quadratic form of the likelihood (9.82)? The reason is that the trace of two matrices in easier to handle in the Bayesian relations and do sum over the data using the trace form, making it closer to the case to the inverted gamma distribution, which happens to be the inverse Wishart distribution in $K = 1$ dimension. If, in the quadratic form in the exponent of (9.82), we let $\vec{D}_i = (\vec{x}_i - \vec{\mu}_s - N_i \vec{Q}_i)$, be the deviation, $A = \vec{D}_i \vec{D}_i^{\top}$, $B = \Sigma_s^{-1}$ and noting that both are symmetric by definition,

$$\vec{D}_i B \vec{D}_i^{\top} = \sum_{k=1}^{K} \sum_{j=1}^{K} D_{i,k} B_{k,j} D_{i,j} = \sum_{k=1}^{K} \sum_{j=1}^{K} A_{k,j} B_{k,j} \quad (9.89)$$
$$= \mathrm{Tr}[AB] = \mathrm{Tr}[\vec{D}_i \vec{D}_i^{\top} \Sigma_s^{-1}].$$

The trace has some nice properties for establishing that prior-posterior distribution pairs are conjugate to the likelihood distribution. The **trace is symmetric** in its two arguments:
$$\text{Tr}[AB] = \text{Tr}[BA]. \tag{9.90}$$
The **trace is linear** in each of its arguments:
$$\text{Tr}[AB] + \text{Tr}[CB] = \text{Tr}[(A+C)B]. \tag{9.91}$$
Now, suppose that the likelihood distribution for the covariance parameter Σ and a zero-mean, K-multivariate normal distribution $\mathcal{N}(0, \Sigma)$ with a sample of size n, so
$$p(X|\Sigma) \propto |\Sigma|^{-n/2} \exp(-0.5 X^\top \Sigma^{-1} X))$$
$$\propto |\Sigma|^{-n/2} \exp(-0.5 \text{Tr}[A\Sigma^{-1}])), \tag{9.92}$$
$X = [[X_{j,i}]_{j=1}^K]_{i=1}^n$ is the $K \times n$ data and $A = XX^\top$ is roughly related to n times the sample covariance. Then selecting a potential inverse Wishart conjugate prior with parameter $K \times K \Psi$ with m degrees of freedom,
$$p(\Sigma) \stackrel{\text{dist}}{=} \mathcal{W}^{-1}(\Psi, m) \tag{9.93}$$

By Bayes rule,

$$p(\Sigma|X) \stackrel{\text{dist}}{\propto} p(X|\Sigma) \times p(\Sigma)$$

$$\propto |\Sigma|^{-n/2} \exp\left(-0.5\text{Tr}[A\Sigma^{-1}]\right))$$

$$\cdot |\Psi|^{m/2} |\Sigma|^{-(m+K+1)/2} \exp\left(-0.5\text{Tr}[\Psi\Sigma^{-1}]\right) \qquad (9.94)$$

$$\propto |\Psi|^{m/2} |\Sigma|^{-(m+n+K+1)/2} \exp\left(-0.5\text{Tr}[(A+\Psi)\Sigma^{-1}]\right)$$

$$\stackrel{\text{dist}}{\propto} \mathcal{W}^{-1}(A+\Psi, m+n),$$

so the inverse Wisthart posterior parameters are $\Psi^{(\text{post})} = A + \Psi$ and $m^{(\text{post})} = m + n$.

MATLAB only has random number generators in the Statistics Toolbox, **W=wishrnd(sigma,dof)** is the Wishart RNG and **IW=iwishrnd(sigma,dof)** for the inverse Wishart RNG[a].

---

[a]However, David Shera has a nice collection of MCMC and related codes in *mcmc at MATLAB Central,* including codes for RNGs, densities, Metropolis-Hastings and a demo. showing how to use the $\mathcal{W}^{-1}$ RNG and get basic statistics.

The **full conditional conjugate posteriors** are as follows:

The posterior for the Bayesian calculation of the **average jump count parameter** Λ_J is

$$f^{(\text{post})}(\Lambda_J|\vec{N}) \propto \prod_{i=1}^{n} f(N_i|\Lambda_J) \times f^{(\text{prior})}(\Lambda_J)$$

$$\propto \Lambda_J^{\sum_{i=1}^{n} N_i} \times \Lambda_J^{a_J-1}(1-\Lambda_J)^{b_J-1} \overset{\text{dist}}{\propto} \mathcal{B}(a_J^*, b_J^*), \qquad (9.95)$$

where we use a simple **0-1** law

$$f(N_i|\Lambda) \overset{1}{=} \frac{1}{(1+\Lambda)^{1-N_i}} \frac{\Lambda^{N_i}}{(1+\Lambda)^{N_i}} = \Lambda^{N_i}, \qquad (9.96)$$

so $a_J^* = a_J + \sum_{i=1}^{n} N_i$ and $b_J^* = b_J = 1$, since it seems to be redundant with the simpler form. However, Johannes and Polson (2006) use alternate form $f(N_i|\Lambda) \overset{2}{=} (1-\Lambda)^{1-N_i}\Lambda^{N_i}$ which is closer to a full beta conjugate prior $f_{\mathcal{B}}(\Lambda_J|a,b)$ and $b^* = b + n - \sum_{i=1}^{n} N_i$, but less valid as a Poisson distribution when Λ is near one.

The full conditional conjugate posterior for the **log-jump-amplitude vector** $\vec{Q}_i$ is

$$f^{(\mathbf{post})}(\vec{Q}_i | \vec{X}_i, N_i, \Theta) \propto \exp\Big(-0.5\Big(\vec{D}_i^{\top} \Sigma_s^{-1} \vec{D}_i\Big)$$

$$+ \Big(\vec{Q}_i - \vec{\mu}_J\Big)^{\top} \Sigma_s^{-1} \Big(\vec{Q}_i - \vec{\mu}_J\Big)\Big) \qquad (9.97)$$

$$\stackrel{\mathbf{cts}}{\propto} \exp\Big(-0.5\Big(\vec{Q}_i - \vec{m}_i\Big)^{\top} V_i^{-1} \Big(\vec{Q}_i - \vec{m}_i\Big)\Big),$$

where

$$V_i^{-1} = \Big(\Sigma_{s,i}^{(\mathbf{post})}\Big)^{-1} = N_i \Sigma_s^{-1} + \Sigma_J^{-1};$$

$$\vec{m}_i = \vec{\mu}_{J,i}^{(\mathbf{post})} = \Sigma_J^{-1}\Big(N_i \Sigma_s^{-1}(\vec{x}_i - \vec{\mu}) + \Sigma_J^{-1} \vec{\mu}_J\Big). \qquad (9.98)$$

The **jump-counter** N_i is discrete with Bernoulli $0 - 1$ values and so must be handled as such. The probability of a jump is

$$\mathbf{Prob}[N_i\!=\!1|\vec{Q}_i, \vec{X}_i, \theta] \propto f(\vec{x}_i|N_i\!=\!1, \vec{Q}_i, \theta)\mathbf{Prob}[N_i\!=\!1|\theta]$$
$$\propto \frac{\Lambda_J}{1+\Lambda_J}\exp\left(-0.5\vec{D}_{i,1}^{\top}\Sigma_s^{-1}\vec{D}_{i,1}\right), \tag{9.99}$$

where when $N_i\!=\!1$ then $\vec{D}_{i,1}\!=\!\vec{x}_i\!-\!\vec{\mu}_s\!-\!\vec{Q}_i$.

$$\mathbf{Prob}[N_i\!=\!0|\vec{Q}_i, \vec{X}_i, \theta] \propto f(\vec{x}_i|N_i\!=\!0, \vec{Q}_i, \theta)\mathbf{Prob}[N_i\!=\!0|\theta]$$
$$\propto \frac{1}{1+\Lambda_J}\exp\left(-0.5(\vec{D}_{i,0})^{\top}\Sigma_s^{-1}\vec{D}_{i,0}\right), \tag{9.100}$$

where when $N_i = 0$ then $\vec{D}_{i,0}\!=\!\vec{x}_i\!-\!\vec{\mu}_s$.

That completely specifies the MCMC algorithm of the multivariate asset Merton jump-diffusion problem.

Using the Gibbs sampler iterations, the full conditonal posterior distributions can be sampled directly, given (9.96,9.99,9.100) for N_i and (9.97) for $\vec{Q}_i$,

$$f_{\Theta_{s1}}^{(\text{post})}(\vec{\mu}_s|\Sigma_s, N, Q, X) \overset{\text{dist}}{\sim} \mathcal{N}(\vec{m}_s, V_s);$$

$$f_{\Theta_{s2}}^{(\text{post})}(\Sigma_s|\vec{\mu}_s, N, Q, X) \overset{\text{dist}}{\sim} \mathcal{IW}(\Psi_s, \text{dof}_s);$$

$$f_{\Theta_{J1}}^{(\text{post})}(\vec{\mu}_J|\Sigma_J, N, Q, X) \overset{\text{dist}}{\sim} \mathcal{N}(\vec{m}_J, V_J);$$

$$f_{\Theta_{J2}}^{(\text{post})}(\Sigma_J|\vec{\mu}_J, N, Q) \overset{\text{dist}}{\sim} \mathcal{IW}(\Psi_J, \text{dof}_J);$$ (9.101)

$$f_{Y_1}^{(\text{post})}(\vec{Q}_i|N_i, \vec{X}_i, \theta) \overset{\text{dist}}{\sim} \mathcal{N}(\mu_J^{(\text{post})}, (\Sigma_J)^{(\text{post})});$$

$$f_{Y_2}^{(\text{post})}(N_i|\vec{Q}_i, \vec{X}_i, \theta) \overset{\text{dist}}{\sim} \mathcal{N}(\mu_J,^{(\text{post})}, (\Sigma_J)^{(\text{post})}).$$

For instance, by Bayes rule, the computational form for $\vec{\mu}_s$ is

$$f_{\Theta_{s1}}^{(\text{post})}(\vec{\mu}_s|\Sigma_s, N, Q, X) \propto f_X(X|\vec{\mu}_s, \Sigma_s, N, Q)$$

$$\times f_{\Theta_{s1}}^{(\text{post})}(\vec{\mu}_s).$$ (9.102)

Table 1: Parameter estimates for the bi-variate jump-diffusion model for daily S&P 500 and Nasdaq 100 returns from 1986-2000.

	Prior		Posterior		
	Mean	Std	Mean	Std	(5,95)% Credible Set
μ_1	0	$\sqrt{1000}$	0.1417	0.0229	0.1065, 0.1797
μ_2	0	$\sqrt{1000}$	0.0839	0.0148	0.0589, 0.1082
σ_1	1.7770	0.9155	1.2073	0.0191	1.1778, 1.2396
σ_2	1.7770	0.9155	0.7236	0.0369	0.6903, 0.7599
ρ	0	0.1713	0.6509	0.0115	0.6317, 0.6690
λ	0.0476	0.0147	0.0799	0.0081	0.0663, 0.0933
$\mu_{1,J}$	0	$\sqrt{1000}$	-0.5747	0.2131	-0.9320, -0.2351
$\mu_{2,J}$	0	$\sqrt{1000}$	-0.3460	0.1765	-0.6537, -0.0648
$\sigma_{1,J}$	2.1113	1.1715	2.9666	0.1647	2.7073, 3.2435
$\sigma_{2,J}$	2.1113	1.1715	2.5873	0.1458	2.3540, 2.8233
ρ_J	0	0.1519	0.5190	0.0490	0.4360, 0.5986

Figure 9.1: **MCMC Table 1, page 47, results from Johannes and Polson (2006)**for $K = 2$ index asset example, the daily S and P500 (subscript 1) and Nasdaq 100 (subscript 2; if no 1 or 2 subscript then is value in common) index returns from 1986-2000 (note: includes 1987).
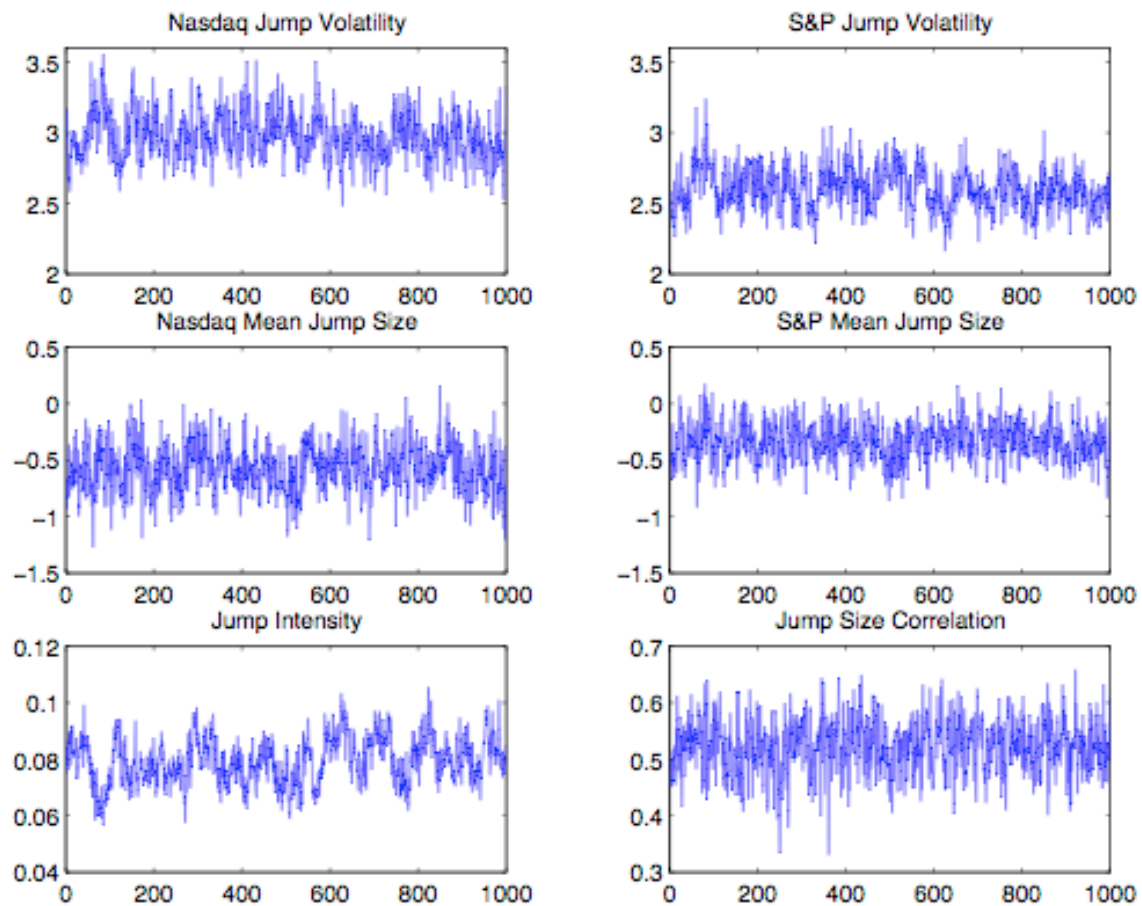
Figure 2: Parameter trace plots for the jump parameters. Each panel shows $\{\Theta^{(g)}\}_{g=1}^{G}$ for the individual parameters.

Figure 9.2: **MCMC Figure 2, page 48, trajectories of final Gibbs iterations from Johannes and Polson (2006)** for $K = 2$ index asset example, the daily S and P500 and Nasdaq 100 2 index returns from 1986-2000 (note: includes 1987).

Reminder: Lecture 9 Homework Posted in Chalk Assignments,
due in PDF by Lecture 10 in Chalk Assignments!

* Summary of Lecture 9:

1. **Bayesian Estimation Estimation in Finance**

2. **More Realistic Models**

3. **Bayesian Formulation Summary**

4. **Bayesian Model Specification**

5. **Hammersley-Clifford and Full Conditionals**

6. **Marginal Densities**

7. **Hammersley-Clifford Multivariate Case**

8. **Gibbs Sampling**

9. **Metropolis MCMC Algorithm**

10. **Bayesian Estmation for SVJD e**

11. **Bayesian Estmation for Options**

12. **Bayesian Estmation for Diffusion Asset Pricing Example**

13. **Bayesian Estmation for Black-Scholes Example**

14. **Bayesian Estmation for Multivariate Merton JD Example**