Making Prediction in State Space Models with Copulas

Jack Liu
Department of Statistics & Data Science
University of Texas, Austin
Austin, TX 78712
jackwl@utexas.edu

Abstract

In estimating state space models with non-Gaussian densities, particle filters methods have been popular, but it often underestimates variance of the posterior distribution of the hidden variable, especially when transition and/or emission densities are not symmetric. In this report, we introduce a novel way of estimating Hidden Markov Models with copulas.

1 Introduction

In this report, we introduce a new way of estimating Hidden Markov Models with the use of copulas. First, under Gaussian distribution assumptions, we use the Gaussian copula to estimate distribution of the hidden variable and compare the estimation with that of a Kalman Filter. Then, we deviate from the Gaussian distribution assumption, and demonstrate that Gaussian copula still provides good estimation when the transition and emission densities are symmetric. Finally, we show that with skewed transition and emission densities, a mixture of Independent and Gaussian copulas are useful in density estimations.

The rest of this report shall proceed as follows. Section 2 provides background on state space models and copulas. Section 3 shows how the Gaussian copula can be used to estimate distribution of the state space when 1) the densities are normal, and 2) densities are not Gaussian but are symmetric. Section 4 shows how the mixture copula can be used to estimate distribution of state space in the more general case when transition and emission densities are not symmetric. Section 5 concludes.

2 Background

2.1 State Space Models

Consider the evolution of the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$ of a target given by:

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \tag{1}$$

where $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_x}$ is a possibly nonlinear function of the state $\mathbf{x}_{k-1}$, $\{v_{k-1}, k \in \mathbb{N}\}$ is an iid process noise sequence, n_x, n_v are dimensions of the state and process noise vectors. The objective is to recursively estimate $\mathbf{x}_k$ from measurements

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k) \tag{2}$$

where $h_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_n} \to \mathbb{R}^{n_z}$ is a possibly nonlinear function, $\{n_k, k \in \mathbb{N}\}$ is an iid measurement noise sequence, n_z, n_n are dimensions of the measurement and measurement noise vectors. In particular, we seek filtered estimates of $\mathbf{x}_k$ based on the set of all available measurements $\mathbf{z}_{1:k} = \{\mathbf{z}_i, i = 1, ..., k\}$ up to time k.

Preprint. Under review.

It is assumed that the exactly form of Equations (1) and (2) are known. Equation (1) is referred as the transition density, while Equation (2) is refereed as the emission density. Usually in State Space Models, the objective is to estimate the conditional density $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$. But in this report, our objective is to estimate the one-step ahead predictive density $p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1})$.

Suppose the density function $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ at time $k-1$ is available. the prediction stage involves using the system model (1) to obtain the prior density of the state at time k via the Chapman-Kolmogorov equation:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \tag{3}$$

Note that (3) uses the fact that $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_{1:k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$ as (1) describes a Markov process of order one. At time step k, a measurement $\mathbf{z}_k$ becomes available, and this may be used to update the prior via the Bayes' rule:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{\int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k} \tag{4}$$

The normalizing constant of (4) depends on the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$ defined by the measurement model (2) and the known statistics of $\mathbf{n}_k$. In the update stage (4), the measurement $\mathbf{z}_k$ is used to modify the prior density to obtain the required posterior density of the current state. The recurrence relations (3) and (4) form the basis for the optimal Bayesian solution. The optimal Bayesian solution solves the problem of recursively calculating the exact posterior density. An optimal algorithm is a method for deducing this solution.

In this report, we deviate from the traditional method of making inference on $\mathbf{x}_k$ with observations $\mathbf{z}_{1:k}$. Instead, we make inference on $\mathbf{x}_k$ with observations $\mathbf{z}_{1:k-1}$ only. The reader will see in the next section, that this allows use of copulas in the estimation of predictive density $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, which will be computed recursively from $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-2})$. Apply the Chapman-Kolmogorov equation from (3) and then apply Bayes' rule (4) to $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$:

$$\begin{aligned}
p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \\
&= \frac{\int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{z}_{k-1}|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-2})d\mathbf{x}_{k-1}}{\int p(\mathbf{z}_{k-1}|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-2})d\mathbf{x}_{k-1}} && (5) \\
&= \frac{\int p(\mathbf{x}_k, \mathbf{z}_{k-1}|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-2})d\mathbf{x}_{k-1}}{\int p(\mathbf{z}_{k-1}|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-2})d\mathbf{x}_{k-1}} && \text{conditional ind. bw } \mathbf{x}_k \text{ and } \mathbf{z}_{k-1} \\
&= \frac{p(\mathbf{x}_k, \mathbf{z}_{k-1} \mid \mathbf{z}_{1:k-2})}{p(\mathbf{z}_{k-1} \mid \mathbf{z}_{1:k-2})} && (6)
\end{aligned}$$

From Equation (5), we see $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is proportional to the expected value of the product of transmission density and emission density. When transition and emission densities are both Gaussian, the posterior density for x_k is also Gaussian, then Kalman filters can be applied. When the Gaussian assumption does not hold, usually particle filter algorithms are applied to estimate Equation (5).

2.2 Copulas Models

For simplicity, we will only consider bivariate copulas in this report. According to Sklar's Theorem, any joint CDF can be expressed as a copula:

$$\begin{aligned}
F_{XZ}(x,z) &= \mathbb{P}(X \leq x, Z \leq z) = \mathbb{P}(F_X(X) \leq F_X(x), F_Z(Z) \leq F_Z(z)) \\
&= \mathbb{P}(u_1 \leq F_X(x), u_2 \leq F_Z(z)) \\
&= C(F_X(x), F_Z(z)) && (7)
\end{aligned}$$

Taking derivative on $C(F_X(x), F_Z(z))$ with respect to x and z, we get:

$$\begin{aligned}
& \frac{\partial^2 F_{XZ}(x,z)}{\partial x \partial z} = f_{XZ}(x,z) \\
\Rightarrow \quad & \frac{\partial^2 C(F_X(x), F_Z(z))}{\partial x \partial z} = c(F_X(x), F_Z(z)) f_X(x) f_Z(z) = f_{XZ}(x,z) \\
\Rightarrow \quad & c(F_X(x), F_Z(z)) = \frac{f_{XZ}(x,z)}{f_X(x) f_Z(z)} && (8)
\end{aligned}$$

where $c(u, v)$ is the copula density function. Equation (8) tells us that any joint density divided by the component's marginal density has a corresponding copula. In many cases, the joint density function is not known, but we can approximate it by using a copula function. Now, multiply Equation (6) by $1 = \frac{p(x_k|z_{1:k-2})}{p(x_k|z_{1:k-2})}$, we can express the predictive conditional density as a product of copula and unconditional density:

$$
\begin{aligned}
p(x_k \mid z_{1:k-1}) &= \frac{p(x_k, z_{k-1} \mid z_{1:k-2})}{p(z_{k-1} \mid z_{1:k-2})} \frac{p(x_k \mid z_{1:k-2})}{p(x_k \mid z_{1:k-2})} \\
&= c(P(x_k \mid z_{1:k-2}), P(z_{k-1} \mid z_{1:k-2}))p(x_k \mid z_{1:k-2}) \\
&= c(P(x_k), P(z_{k-1}))p(x_k)
\end{aligned}
\tag{9}
$$

To ease notation burden, in the last line we take off conditional notation so equations are easier to read. Note

$$P(z_{k-1} \mid z_{1:k-2}) = \int H(z_{k-1}|x_{k-1})p(x_{k-1}|z_{1:k-2})dx_{k-1} \tag{10}$$

$$p(x_k|z_{1:k-2}) = \int f(x_k|x_{k-1})p(x_{k-1}|z_{1:k-2})dx_{k-1} \tag{11}$$

$$P(x_k|z_{1:k-2}) = \int F(x_k|x_{k-1})p(x_{k-1}|z_{1:k-2})dx_{k-1} \tag{12}$$

By assumption, emission and transmission densities from Equations (1) and (2) are known. The integrals in Equations (10) to (12) can be evaluated numerically if we have sufficient particles representing x_{k-1}.

Based on Equation (9), we can also relate the predictive cumulative distribution to copula:

$$
\begin{aligned}
P(x \mid z_{1:k-1}) &= \int_{-\infty}^{x} c(P(x_k), P(z_{k-1}))p(x_k)dx_k \\
&= \frac{\partial C(P(x), P(z_{k-1}))}{\partial P(z_{k-1})} = C_v(P(x))
\end{aligned}
\tag{13}
$$

3 Gaussian Copula Estimation

We obtain the ***Gaussian copula*** if $F_{XZ}(x, z)$ in equation (7) is constructed with bivariate normal CDF $\Phi_\rho(s, t)$ with zero means, unit variances, and correlation ρ. Let $\Phi(t)$ be the standard normal CDF. We have:

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \tag{14}$$

Denoting $s = \Phi^{-1}(u),\ t = \Phi^{-1}(v)$, thus $\Phi(s) = u$, $\phi(s)ds = du$, $\frac{ds}{du} = \frac{1}{\phi(s)}$. The density of the Gaussian copula can be written as:

$$c_\rho(u, v) = \frac{\partial^2 C_\rho(u, v)}{\partial u \partial v} = \frac{\partial^2 \Phi_\rho(s, t)}{\partial s \partial t} \cdot \frac{\partial s}{\partial u} \cdot \frac{\partial t}{\partial v} = \frac{\phi_\rho(s, t)}{\phi(s)\phi(t)} \tag{15}$$

where $\phi_\rho(s, t)$ is the density function for $\Phi_\rho(s, t)$ and $\phi(t)$ is the density function for $\Phi(t)$. The explicit expression for (15) is:

$$c_\rho(s, t) = \frac{1}{\sqrt{1 - \rho^2}}\, exp\left(-\frac{1}{2}\frac{\rho^2 s^2 - 2\rho st + \rho^2 t^2}{1 - \rho^2}\right) \tag{16}$$

IN modeling joint distributions, we can combine Gaussian copula with any marginal distributions $u = F_X(x)$ and $v = F_Z(z)$. The idea of Gaussian copula is to transform two random variables X and Y with respective CDFs F and G into standard normal variables $S = \Phi^{-1}(F(X))$ and $T = \Phi^{-1}(G(Y))$. Then the dependence between X and Y is expressed in terms of the dependence structure of their normal transformations S and T, reducing the dependence to linear correlation.

In this section, we will generate hidden variable **x** and observed variable **z** based on Gaussian densities. Then we will use the copula method to estimate the distribution of hidden variable **x**. At the same, we will also use the classical particle filters to estimate the distribution of $\mathbf{x}_t$ at each time step t. Since all densities are Gaussian, samples from the classical particle filters would represent the true density. We will compare particles generated from the Gaussian copula method against particles from the classical method (sequential importance sampling).

From Equation (13), $\frac{\partial C(u,v))}{\partial v} = C_v(u)$. For Gaussian copula (denote $\Phi(s) = u$, $\Phi(t) = v$) the equation becomes:

$$
\begin{aligned}
C_v(u) &= \frac{\partial \Phi}{\partial t}\frac{\partial t}{\partial v} \\
&= \phi(t)\Phi(s \mid \rho t, \sqrt{1-\rho^2})\frac{1}{\phi(t)} \\
&= F(\Phi^{-1}(u)) \qquad (17)
\end{aligned}
$$

where F is the CDF function for a normal distribution with parameters $(\rho\Phi^{-1}(v), \sqrt{1-\rho^2})$.

For particle generation, we will use inverse transform method. This will involve numerical algorithms to solve for locations of particles. Although the inverse transform method can be a slow algorithm, it appears to be more reliable than other algorithms such as accept-rejection sampling at the time of writing this report.

Here are the steps for particle generation based on the copula method:

- Generate N uniform random numbers between 0 and 1.
- Find numerical integral $F_Z(z)$ from Equation (10) based on existing particles.
- Based on Equation (17), Numerically find x such that $C(F_X(x), F_Z(z)) = u$, where u comes from step (1). With $F_Z(z)$ known, the joint bivariate normal distribution function C now reduces to univariate normal distribution function. Thus, produce u' according to the formula:

$$
\begin{aligned}
&u' = \Phi(\Phi^{-1}(u)\sigma + \mu) && (18) \\
&\sigma = \sqrt{1-\rho^2} \qquad\qquad \mu = \rho\Phi^{-1}(F_Z(z))
\end{aligned}
$$

- find x such that:

$$u' = F(x_k|z_{1:k-2}) = \int F(x|x_{k-1})p(x_{k-1}|z_{1:k-2})dx_{k-1} \qquad (19)$$

  This is the most time-consuming part of the algorithm.
- repeat the process for all time-steps.

3.1 Gaussian Densities

3.2 Symmetric Densities

4 Mixture Copula Estimation

In this section, we move away from the symmetric density assumptions. We will first show that the Gaussian copula method would no longer be able to generate particles close to the true density. Instead, we will use a mixture of Gaussian and independent copulas.

$$C_m = \alpha(uv) + (1-\alpha)\Phi(\Phi^{-1}(u), \Phi^{-1}(v)) \qquad (20)$$

Taking derivative with respect to v, we get

$$C_{m_v} = \alpha u + (1-\alpha)F(\Phi^{-1}(u)) \qquad (21)$$

Next, we run exactly the same algorithm as in the case of Gaussian copula. The only difference is Equation (18). Instead of solving for $F(\Phi^{-1}(u')) = u$, we need to solve for $\alpha u' + (1-\alpha)F(\Phi^{-1}(u')) = u$.

5 Conclusion

Acknowledgments

Use unnumbered third level headings for the acknowledgments. All acknowledgments go at the end of the paper. Do not include acknowledgments in the anonymized submission, only in the final paper.

References

References follow the acknowledgments. Use unnumbered first-level heading for the references. Any choice of citation style is acceptable as long as you are consistent. It is permissible to reduce the font size to `small` (9 point) when listing the references. **Remember that you can use more than eight pages as long as the additional pages contain *only* cited references.**

[1] Alexander, J.A. & Mozer, M.C. (1995) Template-based algorithms for connectionist rule extraction. In G. Tesauro, D.S. Touretzky and T.K. Leen (eds.), *Advances in Neural Information Processing Systems 7*, pp. 609–616. Cambridge, MA: MIT Press.

[2] Bower, J.M. & Beeman, D. (1995) *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System.* New York: TELOS/Springer–Verlag.

[3] Hasselmo, M.E., Schnell, E. & Barkai, E. (1995) Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in rat hippocampal region CA3. *Journal of Neuroscience* **15**(7):5249-5262.