# Review: Useful Distributions

**1. The multivariate normal distribution**

A 𝑘−*k*−dimensional random vector 𝑥=(𝑥1,⋯ ,𝑥𝑘)𝑇**x**=(*x*1​,⋯,*xk*​)*T* that follows a multivariate normal distribution with mean 𝜇***μ*** and variance-covariance matrix 𝛴***Σ***, 𝑥∼𝑁(𝜇,𝛴)**x**∼*N*(***μ***,***Σ***) or 𝑝(𝑥)=𝑁(𝑥∣𝜇,𝛴)*p*(**x**)=*N*(**x**∣***μ***,***Σ***), has a density function given by

p(**x**)=(2π)−k/2|**Σ**|−1/2exp[−12(**x**−***μ***)T**Σ**−1(**x**−***μ***)]

**2. The gamma and inverse-gamma distribution**

A random variable 𝑥*x* that follows a gamma distribution with shape parameter 𝛼*α* and inverse scale parameter 𝛽*β*, 𝑥∼𝐺(𝛼,𝛽)*x*∼*G*(*α*,*β*), or 𝑝(𝑥)=𝐺(𝑥∣𝛼,𝛽)*p*(*x*)=*G*(*x*∣*α*,*β*), has a density of the form

p(x)=βαΓ(α)xα−1e−βx,x>0

where Γ(⋅)Γ(⋅) is the gamma function. In addition, 𝐸(𝑥)=𝛼𝛽*E*(*x*)=*βα*​ and 𝑉𝑎𝑟(𝑥)=𝛼𝛽2*Var*(*x*)=*β*2*α*​.

If 1𝑥∼𝐺(𝛼,𝛽)*x*1​∼*G*(*α*,*β*), then 𝑥*x* follows an inverse-gamma distribution, 𝑥∼𝐼𝐺(𝛼,𝛽)*x*∼*IG*(*α*,*β*), or 𝑝(𝑥)=𝐼𝐺(𝑥∣𝛼,𝛽)*p*(*x*)=*IG*(*x*∣*α*,*β*) with

p(x)=βαΓ(α)x−(α+1)e−β/x,x>0

In this case, 𝐸(𝑥)=𝛽𝛼−1*E*(*x*)=*α*−1*β*​ for 𝛼>1*α*>1 and 𝑉𝑎𝑟(𝑥)=𝛽2(𝛼−1)2(𝛼−2)*Var*(*x*)=(*α*−1)2(*α*−2)*β*2​ for 𝛼>2*α*>2.

**3. The multivariate Student-t distribution**

A random vector 𝑥**x** of dimension 𝑘*k* follows a multivariate Student-t distribution with 𝜈*ν* degree of freedom, location 𝜇***μ***, and scale matrix 𝛴***Σ***, 𝑥∼𝑇𝜈(𝜇,𝛴)**x**∼*Tν*​(***μ***,***Σ***), if its density is given by

p(**x**)=Γ(ν+k2)Γ(ν2)(νπ)k/2|**Σ**|−1/2[1+1ν(**x**−***μ***)T**Σ**−1(**x**−***μ***)]−(ν+k)/2

𝐸(𝑥)=𝜇*E*(**x**)=***μ*** for 𝜈>1*ν*>1 and 𝑉𝑎𝑟(𝑥)=𝜈𝛴𝜈−2*Var*(**x**)=*ν*−2*ν****Σ***​, for 𝜈>2*ν*>2.

**Posterior Distribution Derivation**

We derive the posterior distributions for 𝜙***ϕ*** and 𝜈*ν*. Recall the model is

**y**∼N(**F**T***ϕ***,ν**I**n),***ϕ***∼N(**m**0,ν**C**0),ν∼IG(n02,d02)

Using Bayes theorem, we have

p(***ϕ***,ν|**y**)∝p(**y**|***ϕ***,ν)p(***ϕ***|ν)p(ν)∝ν−n/2exp(−(**y**−**F**T***ϕ***)T(**y**−**F**T***ϕ***)2ν)×ν−p/2exp(−(***ϕ***−**m**0)T**C**−10(***ϕ***−**m**0)2ν)×ν−(n02+1)exp(−d02ν)∝ν−(n∗2+1)exp(−d∗2ν)×ν−p/2exp(−(***ϕ***−**m**)T**C**−1(***ϕ***−**m**)2ν)∝p(ν|**y**)p(***ϕ***|ν,**y**)

where

**e**=**y**−**F**T**m**0,**Q**=**F**T**C**0**F**+**I**n,**A**=**C**0**FQ**−1**m**=**m**0+**Ae**,**C**=**C**0−**AQA**Tn∗=n+n0,d∗=(y−**F**T**m**0)T**Q**−1(y−**F**T**m**0)+d0

Therefore, we have the posterior distribution of 𝜙***ϕ*** and 𝜈*ν* as

p(***ϕ***,ν|**y**)=p(***ϕ***|ν,**y**)p(ν|**y**)=N(***ϕ***|**m**,ν**C**)IG(ν|n∗2,d∗2)

To get the 𝑠*s*th sample (𝜙(𝑠),𝜈(𝑠))(***ϕ***(*s*),*ν*(*s*)) from the joint posterior distribution of 𝜙***ϕ*** and 𝜈*ν*, we first sample 𝜈(𝑠)*ν*(*s*) from 𝐼𝐺(𝑛∗2,𝑑∗2)*IG*(2*n*∗​,2*d*∗​), then sample 𝜙(𝑠)***ϕ***(*s*) from 𝑁(𝑚,𝜈(𝑠)𝐶)*N*(**m**,*ν*(*s*)**C**).

One can also obtain posterior samples of 𝜙***ϕ*** by directly sampling from its marginal posterior distribution. Integrate out 𝜈*ν* from the joint posterior distribution, we have 𝑝(𝜙∣𝑦)=𝑇𝑛∗(𝑚,𝑑∗𝐶𝑛∗)*p*(***ϕ***∣**y**)=*Tn*∗​(**m**,*n*∗*d*∗**C**​).

Denote the *s*th posterior sample of 𝜙***ϕ*** and 𝜈*ν* as 𝜙(𝑠)***ϕ***(*s*) and 𝜈(𝑠)*ν*(*s*), we can obtain the corresponding *i*th posterior sample of 𝑦(𝑠)**y**(*s*) by sampling from 𝑁(𝐹𝑇𝜙(𝑠),𝜈(𝑠))*N*(**F***T****ϕ***(*s*),*ν*(*s*)). Therefore, we can obtain the posterior point and interval estimates for the time series.

**Location and scale mixture of AR model**

In this part, we will extend the location mixture of AR models to the location and scale mixture of AR models. We will show the derivation of the Gibbs sampler for the model parameters as well as the R code for the full posterior inference.

**The Model**

The location and scale mixture of AR(𝑝)(*p*) model for the data can be written hierarchically as follows:

yt∼∑k=1KωkN(**f**Tt***β***k,νk),**f**Tt=(yt−1,⋯,yt−p)T,t=p+1,⋯,Tωk∼Dir(a1,⋯,ak),***β***k∼N(**m**0,νk**C**0),νk∼IG(n02,d02)

Introducing latent configuration variable 𝐿𝑡∈{1,2,⋯ ,𝐾}*Lt*​∈{1,2,⋯,*K*} such that 𝐿𝑡=𝑘*Lt*​=*k* if and only if 𝑦𝑡∼𝑁(𝑓𝑡𝑇𝛽𝑘,𝜈𝑘)*yt*​∼*N*(**f***tT*​***β****k*​,*νk*​), and denote 𝛽=(𝛽1,⋯ ,𝛽𝐾),𝜔=(𝜔1,⋯ ,𝜔𝐾),𝐿=(𝐿1,⋯ ,𝐿𝑇)***β***=(***β***1​,⋯,***β****K*​),***ω***=(*ω*1​,⋯,*ωK*​),**L**=(*L*1​,⋯,*LT*​), we can write the full posterior distribution as

p(***ω***,***β***,ν,**L**|**y**)∝p(**y**|***ω***,***β***,ν,**L**)p(**L**|***ω***)p(***ω***)p(***β***)p(ν)∝∏k=1K∏{t:Lt=k}N(yt|**f**Tt***β***k,ν)∏k=1Kω∑Tt=1**1**(Lt=k)k∏k=1Kωak−1k∏k=1K(N(***β***k|**m**0,**C**0)IG(νk|n02,d02))

The full conditional distributions of all model parameters are therefore given as follows:

1. For 𝜔***ω***, we have

***ω***|⋯∼Dir(a1+∑t=1T**1**(Lt=1),⋯,aK+∑t=1T**1**(Lt=K))

1. For 𝐿𝑡*Lt*​, we have

p(Lt=k|⋯)∝ωkN(yt|**f**Tt***β***k,ν)

Therefore, 𝐿𝑡*Lt*​ follows a discrete distribution on {1,⋯ ,𝐾}{1,⋯,*K*} with probobaility that 𝐿𝑡*Lt*​ taking 𝑘*k* proportional to 𝜔𝑘𝑁(𝑦𝑡∣𝑓𝑡𝑇𝛽𝑘,𝜈)*ωk*​*N*(*yt*​∣**f***tT*​***β****k*​,*ν*).

1. For 𝜈𝑘*νk*​ and 𝛽𝑘***β****k*​, denote 𝑦~𝑘:={𝑦𝑡:𝐿𝑡=𝑘}**y~**​*k*​:={*yt*​:*Lt*​=*k*}, 𝐹~𝑘**F~***k*​ as the design matrix corresponding to 𝑦~𝑘**y~**​*k*​, and 𝑛𝑘=∑𝑡=1𝑇1(𝐿𝑡=1)*nk*​=∑*t*=1*T*​**1**(*Lt*​=1), we have 𝜈𝑘∣⋯∼𝐼𝐺(𝑛𝑘∗2,𝑑𝑘∗2)*νk*​∣⋯∼*IG*(2*nk*∗​​,2*dk*∗​​) and 𝛽𝑘∼𝑁(𝑚𝑘,𝜈𝑘𝐶𝑘)***β****k*​∼*N*(**m***k*​,*νk*​**C***k*​), where

**e**k=**y~**k−**F~**Tk**m**0,**Q**k=**F~**Tk**C**0**F~**k+**I**nk,**A**k=**C**0**F~**k**Q**−1k**m**k=**m**0+**A**k**e**k,**C**k=**C**0−**A**k**Q**k**A**Tkn∗k=n0+nk,d∗k=d0+**e**Tk**Q**−1k**e**k

Now we have the full conditional distributions for all model parameters. We proceed to implement the model in R with a simulate dataset.