Multivariate Models

MFM Practitioner Module: Quantitative Risk Management

John Dodson

December 2, 2020

Application: Mixtures

Conditioning generally reduces entropy. Mixing has the opposite effect. This is useful if you want to moderate statistical hubris.

Say X is the parametric r.v. we are interested in.

1. Concatenate it with some function of the parameters to make a multivariate r.v. $X + \!\!\! + \Theta$.
2. Specify the marginal density of Θ and the conditional density of $X|\Theta$.
3. Integrate over the support of Θ to get the marginal density of X, the new mixture.

$$
\begin{aligned}
& f_{X+\!\!\!+\Theta}(x+\!\!\!+\theta) = f_{X|\Theta}(x;\theta) f_{\Theta}(\theta) \\
\implies\ & f_X(x) = \int_{\Omega(\Theta)} f_{X|\Theta}(x;\theta) f_{\Theta}(\theta)\, d\theta
\end{aligned}
$$

Review: Univariate Scalar Random Variables

Student's-t is a symmetric r.v. which exhibits leptokurtosis.

(Gosset's) Student's-t

Consider a normal r.v. with an unknown variance close to one. If the variance is a draw from an reciprocal Gamma,

$$\begin{aligned} X\big|\sigma^2 &\sim \mathcal{N}\left(0,\sigma^2\right) \\ \sigma^2 &\sim \text{Gamma}^{-1}\left(\frac{\nu}{2},\frac{\nu}{2}\right) \end{aligned}$$

the resulting unconditioned density is

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)}\frac{1}{\sqrt{\nu}}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

- The version with $\nu=1$ is the Cauchy
- The limit $\nu\to\infty$ is a normal

The version of the Student's-t above has a variance for $\nu > 2$, but it is not unity.

Standardized Student's-t

The standardized version can be useful for fitting residuals*. It has the density

$$f_X(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\Gamma\left(\frac{1}{2}\right)} \frac{1}{\sqrt{\nu-2}} \left(1 + \frac{x^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}$$

- *Note that $\mathrm{E}\, e^X \to \infty$ for any finite ν so Student's-t cannot be used with log-returns of asset prices.
- For historical reasons, if the parameter ν is an integer, it is termed the degrees of freedom.

Review: Univariate Scalar Random Variables

The most important parametric random variable with $\mathbb{R}^+$ support is the Generalized Inverse Gaussian

Generalized Inverse Gaussian (GIG)

$$f(x) = \frac{\chi^{-\lambda}\left(\sqrt{\chi\psi}\right)^{\lambda}}{2K_\lambda\left(\sqrt{\chi\psi}\right)} x^{\lambda-1} e^{-\frac{\chi}{2x}-\frac{\psi x}{2}}$$

for $x > 0$, where $K_\lambda(\cdot)$ is modified Bessel function of the second kind.

- This generalizes the Gamma and reciprocal Gamma
- There are several versions of parameterization in use
- Other members of this family include the inverse Gaussian and the reciprocal inverse Gaussian
- This can be generalized to positive-definite matrices

Review: Univariate Scalar Random Variables

The generalized hyperbolic family is a Normal mean / GIG variance mixture. The Student's-t is a special case (with $\lambda = -\nu/2$).

Normal / reciprocal inverse Gaussian (NRIG)

Another useful GH is the symmetric Normal / reciprocal inverse Gaussian mixture (with $\lambda = \frac{1}{2}$). The standardized version has the density

$$f_X(x) = \frac{1}{\pi} e^g \sqrt{1+g}\, K_0\left(\sqrt{g^2 + (1+g)x^2}\right)$$

for shape parameter $g \geq 0$. It is not obvious, but the limit $g \to \infty$ corresponds to the normal.

- As a model for the residuals of the log-returns of asset prices, this is superior to the Student's-t example from the text because $\mathrm{E}\, e^X$ is finite.

Spherical Random Variables

It is helpful to build up a theory of multivariate random variables from geometric principles. By definition, a spherical random variable is distributionally invariant to rotations,

$$U\boldsymbol{X} \stackrel{\mathrm{d}}{=} \boldsymbol{X}$$

where U is a square matrix representation of a rotation, which means that $U'U = I$.
Spherical random variables have two equivalent defining properties,

$$\boldsymbol{a}'\boldsymbol{X} \stackrel{\mathrm{d}}{=} \|\boldsymbol{a}\| X_1$$
$$\mathrm{E}\, e^{i\boldsymbol{t}'\boldsymbol{X}} = \psi\left(\boldsymbol{t}'\boldsymbol{t}\right)$$

for vectors $\boldsymbol{a}$ and $\boldsymbol{t}$. We term $\psi(\cdot)$ the characteristic generator of $\boldsymbol{X}$. We therefore write $\boldsymbol{X} \sim S_d(\psi)$ to denote a spherical random variable in d dimensions with characteristic generator $\psi(\cdot)$.

Elliptical Random Variables

An affine transformation of a spherical random variable is termed an elliptical random variable.

$$X \stackrel{d}{=} \boldsymbol{\mu} + A\boldsymbol{Y}$$

where $\boldsymbol{Y} \sim S_k(\psi)$ and A is a $d \times k$ matrix.
The distributional invariance of $\boldsymbol{Y}$ to rotations means that A is generally redundant. All we need to characterize $\boldsymbol{X}$ is $\boldsymbol{\mu}$, $\psi(\cdot)$, and $\Sigma = AA'$. But note that

$$E_d(\boldsymbol{\mu}, \Sigma, \psi(\cdot)) \stackrel{d}{=} E_d(\boldsymbol{\mu}, c\Sigma, \psi(\cdot/c))$$

for $c > 0$, so Σ may not necessarily be the covariance of $\boldsymbol{X}$.

- Note that Σ need not be full rank. In this case, the rank of Σ is at most $d \wedge k$.

Elliptical Random Variables

Some Properties

Say $\boldsymbol{X} \sim E_d\left(\boldsymbol{\mu}, \Sigma, \psi\right)$.

- linear combinations If B $k \times d$ and $\boldsymbol{b}$ $k \times 1$ constants, then
$$B\boldsymbol{X} + \boldsymbol{b} \sim E_k\left(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B', \psi\right)$$
- if Σ is full rank, then the non-negative scalar r.v.
$$R = \sqrt{(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})}$$
is independent of $S = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})/R$ and S is uniformly distributed on a unit sphere.
- convolutions If $\boldsymbol{Y} \sim E_d(\tilde{\boldsymbol{\mu}}, \Sigma, \tilde{\psi})$ independent of $\boldsymbol{X}$, then
$$\boldsymbol{X} + \boldsymbol{Y} \sim E_d\left(\boldsymbol{\mu} + \tilde{\boldsymbol{\mu}}, \Sigma, \psi \cdot \tilde{\psi}\right)$$

Elliptical Random Variables

Maximum Likelihood Estimator

if the density of an r.v. $X \in \mathbb{R}^M$ can be written in the form

$$f_{X|\mu,\Sigma}(x) = g\left(\mathrm{Ma}^2(x,\mu,\Sigma)\right)\sqrt{|\Sigma^{-1}|}$$

for some function $g(\cdot)$ where

$$\mathrm{Ma}(x,\mu,\Sigma) = \sqrt{(x-\mu)'\,\Sigma^{-1}\,(x-\mu)}$$

is the Mahalanobis distance, then the MLE based on a sample $\{x_1,\ldots,x_N\}$ solves the system

$$\hat{\mu} = \sum_{i=1}^{N} \frac{w_i}{\sum_j w_j} x_i \qquad \hat{\Sigma} = \sum_{i=1}^{N} \frac{w_i}{N}\,(x_i-\hat{\mu})\,(x_i-\hat{\mu})'$$

$$\text{with } w_i = \frac{-2g'\left(\mathrm{Ma}^2\left(x_i,\hat{\mu},\hat{\Sigma}\right)\right)}{g\left(\mathrm{Ma}^2\left(x_i,\hat{\mu},\hat{\Sigma}\right)\right)} \quad \forall i = 1,\ldots,N$$

Linear Factor Models

If $\boldsymbol{X}$ is a d-dim random variable, and we can write

$$\boldsymbol{X} = \boldsymbol{a} + B\boldsymbol{F} + \varepsilon$$

where $\boldsymbol{F}$ is a p-dim random vector with $p < d$ and $\operatorname{cov} F > 0$, B is a $d \times p$ matrix, the entries of ε are zero mean and uncorrelated, and $\operatorname{cov}(\boldsymbol{F}, \varepsilon) = 0$, we call $\boldsymbol{F}$ the common factors and B the factor loadings.
We would consider this a model or approximation if $d \gg p$.
Sometimes we have an idea about what the factors or loadings might be; they might even be observable.

- In macroeconomic factor models, we observe the factors.
- In fundamental factor models, we observe the loadings.
- In statistical or latent factor models, we observe neither the factors nor the loadings.

Linear Factor Models

Capital Asset Pricing Model

CAPM for investments is an example of a macroeconomic factor model. It is typically applied to traded equity securities and a risk-free deposit as canonical "capital assets". We will take $\boldsymbol{X}$ to be the (simple) return on each risky capital asset over some investment period.

If $\boldsymbol{X}$ is normal and investors allocate to maximize expected exponential utility, then we can express the equilibrium solution as a single-factor model where $\boldsymbol{F}$ is the return on a broad index of risky capital assets.

The factor loadings B_i can be determined by regression, and are termed the asset "betas".

The intercept components turn out to be $a_i = r\tau\,(1 - B_i)$ where r is the return rate on the risk-free asset.

Linear Factor Models

Fundamental Model

Sometimes it is useful to impose a classification scheme on the components of $\boldsymbol{X}$, for example an industry classification scheme or a geographic or demographic scheme. In this case, we generally know the non-zero loadings in B, but we do not observe the factors $\boldsymbol{F}$.

In this case, we can estimate timeseries for $\boldsymbol{F}$ in terms of timeseries for $\boldsymbol{X}$ according to ordinary least squares regression

$$\hat{\boldsymbol{F}}_t^{\mathrm{OLS}} = \left(B'B\right)^{-1} B'\boldsymbol{X}_t$$

if the variance of the residuals is the same (homoscedastic) or generalized least squares regression

$$\hat{\boldsymbol{F}}_t^{\mathrm{GLS}} = \left(B'\Upsilon^{-1}B\right)^{-1} B'\Upsilon^{-1}\boldsymbol{X}_t$$

if not.

Principal Components

Principal components analysis is inspired by the concept of a statistical factor model, but since it is entirely endogenous it is really a separate concept.

A covariance or correlation matrix Σ has the property of being positive semi-definite, which means that $\boldsymbol{x}'\Sigma\boldsymbol{x} \geq 0$ for all compatible vectors $\boldsymbol{x}$. Therefore, by the spectral decomposition theorem, we can write

$$\Sigma = \Gamma\Lambda\Gamma'$$

where Λ is a diagonal matrix with non-negative entries (the eigenvalues) and Γ is a square matrix whose columns (the eigenvectors) are orthonormal, which means $\Gamma\Gamma' = I$.

Principal Component Analysis

If Σ has full rank d, all of the eigenvalues will be positive. The potential for dimension reduction comes from partitioning the model into the largest $k < d$ eignevalues and eigenvectors, and relegating the remaining $d - k$ to the residual.

Principal Components as Factors

Let $d \times 1$ $\boldsymbol{Y} = \Gamma'(\boldsymbol{X} - \boldsymbol{\mu})$ where $\boldsymbol{\mu}$ is the mean of $\boldsymbol{X}$. Partition $\boldsymbol{Y}$ and Γ into $k \times 1$ $\boldsymbol{Y}_1$ and $(d-k) \times 1$ $\boldsymbol{Y}_2$ and $d \times k$ Γ_1 and $d \times (d-k)$ Γ_2 and let $\boldsymbol{\varepsilon} = \Gamma_2 \boldsymbol{Y}_2$, then

$$X = \boldsymbol{\mu} + \Gamma_1 \boldsymbol{Y}_1 + \boldsymbol{\varepsilon}$$

and ε *almost* satisfies the assumptions for a linear factor model.

Robustness

Non-Parametric Estimators

The term robustness in statistics can sometimes refer to non-parametric techniques that do not require assumptions about the characterization of the random variables involved.

- Such techniques usually lean on the Law of Large Numbers, and hence require very large samples to be effective.

Robust Estimators

A more precise meaning has evolved that focuses on estimators that may be based on parametric characterizations, but which can produce reasonable results for data that does not come from that class of characterizations or stress-test distributions.

- We can make this desire concrete in term of the the influence function associated with an estimator.

Robust Estimators

Influence Function

We have discussed estimators as functions of samples. If instead we consider the estimator as a functional of the density from which observations are drawn, we can consider its (functional) derivative with respect to an infinitesimal perturbation in the density given by

$$f_X(x) \to (1-\epsilon) f_X(x) + \epsilon \delta(x-y)$$

Thus, with $\tilde{\theta}$ the functional induced by the estimator $\hat{\theta}$,

$$\mathrm{IF}\left[y, f_X, \hat{\theta}\right] = \lim_{\epsilon \to 0} \frac{\tilde{\theta}\left[(1-\epsilon) f_X(x) + \epsilon \delta(x-y)\right] - \tilde{\theta}\left[f_X\right]}{\epsilon}$$

If this derivative is bounded for all possible displacements, y, we say the estimator is robust.

Robust Estimators

Robustness of the MLE

For the maximum likelihood estimator, the influence function turns out to be proportional to

$$\mathrm{IF}\left[y, f_X, \hat{\theta}\right] \propto \left.\frac{\partial \log f_{X|\theta}(y)}{\partial \theta}\right|_{\theta=\hat{\theta}}$$

For some characterizations, the parameter MLE's are robust. For some they are not.

- for $X \sim \mathcal{N}(\mu, \Sigma)$, $\hat{\mu}$ and $\hat{\Sigma}$ are not robust
- for $X \sim \mathrm{Cauchy}(\mu, \Sigma)$, they are

Even for the empirical characterization, the influence functions for the sample mean and the sample covariance are not bounded; therefore these sample estimators are *never* robust.

Robust Estimators

Recall the general elliptic location and dispersion MLE's,

$$\hat{\mu} = \sum_{i=1}^{N} \frac{w_i}{\sum_j w_j} x_i$$

$$\hat{\Sigma} = \sum_{i=1}^{N} \frac{w_i}{N} \left(x_i - \hat{\mu}\right) \left(x_i - \hat{\mu}\right)' \quad \text{with}$$

$$w_i \triangleq h\left(\mathrm{Ma}^2\left(x_i, \hat{\mu}, \hat{\Sigma}\right)\right) \quad \forall i = 1, \ldots, N$$

where the function $h(\cdot)$ is the value of a particular functional on the density.

M-Estimators

The idea with M-estimators is to choose $h(\cdot)$ exogenously in order to bound the influence function by design.

Robust Estimators

M-Estimators

We know that $h(\cdot) = 1$ corresponds to the MLE for normals and also to the sample estimators, which do not have bounded influence functions. A weighting function that goes to zero for large arguments is more likely to be robust. Some examples include

- Trimmed estimators, for which
$$h(z) = \begin{cases} 1 & z < z_0 = F^{-1}_{\chi^2_{\dim X}}(p) \\ 0 & \text{otherwise} \end{cases}$$
- Cauchy estimators for which $h(z) = \frac{1+\dim X}{1+z}$
- Schemes such as Huber's or Hampel's for which
$$h(z) = \begin{cases} 1 & z < z_0 = \left(\sqrt{2} + \sqrt{\dim X}\right)^2 \\ \sqrt{\frac{z_0}{z}}\, e^{-\frac{(\sqrt{z}-\sqrt{z_0})^2}{2b^2}} & \text{otherwise} \end{cases}$$

These estimators can be evaluated numerically by iterating to the fixed point.