Conditional distributions using copula function

Arcady Novosyolov
Siberian Federal University
School of Mathematics and Computer Science
Krasnoyarsk
mailto:arcady@novosyolov.com

Abstract. *The paper is devoted to description of component conditional distributions of arbitrary multivariate distribution assuming that dependence structure is driven by a Gaussian (normal) copula. Examples with Student t marginal distributions are provided, with application to financial assets study.*

Keywords. *Gaussian copula, conditional distribution, Student t distribution, financial portfolio management, trading strategy.*

MSC classes: 60E05, 60E99

1 Introduction

Given a multivariate random vector $X = (X_1, \ldots, X_d)$, we often need to know conditional distributions for part of its components, given other components fixed. In case of joint Gaussian distribution the problem is easily solved, see e.g. [1].

However often components possess distributions with heavy tails, so joint Gaussian distribution is not an appropriate choice. Nevertheless we still may use the elegant analytic Gaussian solution of the problem via the apparatus of copulas [2]. This approach is being developed in the present paper.

Specifically, we derive exact formulas for conditional density functions and distribution functions for arbitrary subset of components given other subset of components.

Section 2 introduces basic concepts and implements formal derivations. Here we describe multivariate distribution, copula and their interaction via marginal distributions, and a method for estimating Gaussian copula from data.

In the section 3 examples are provided for all concepts. We explain in detail working with Gaussian conditionals, and further extend to copula conditionals, and full conditionals with exponential and Student t marginal distributions.

Finally we present one technique for using in stock market statistical arbitrage strategies.

2 Multivariate distributions

2.1 Components partitioning and conditional distributions

Let $X = (X_1, \ldots, X_d)$ be a multivariate random vector with CDF

$$F(x_1, \ldots . x_d) = P\left(X_1 \le x_1, \ldots, X_d \le x_d\right) \tag{1}$$

and density function

$$f(x_1, \ldots, x_d) = \frac{\partial^d F(x_1, \ldots, x_d)}{\partial x_1 \ldots \partial x_d} \tag{2}$$

for $x = (x_1, \ldots, x_d) \in R^d$.

Denote $I, J \subseteq \{1, \ldots, d\}$ two disjoint subsets of indices, and let X_I, X_J be correspondent random subvectors. Fix $X_I = x_I^0$, where $a \in R^{|I|}$, then conditional distribution of X_J given $X_I = x_I^0$ has CDF

$$F_{J|I}(x_J|X_I = x_I^0) = P\left(X_J \le x_J|X_I = x_I^0\right) \tag{3}$$

for $x_J \in R^{|J|}$.

Conditional density for the latest conditional distribution is defined as usual as partial derivative with respect to all components involved:

$$f_{J|I}(x_J|X_I = x_I^0) = \frac{\partial^{|J|} F_{J|I}(x_J|X_I = x_I^0)}{\prod_{j\in J} \partial x_j} \tag{4}$$

2.2 Conditional distribution for multivariate Gaussian distribution

Consider a special case when X is multivariate Gaussian with zero mean and covariance matrix S. Denote S_{IJ} the submatrix of S with rows from index set I and columns from the index set J, and similar notations for S_{JI}, S_{II}, S_{JJ}; clearly here $S_{IJ} = S'_{JI}$.

It is well known [1] that in this case the conditional distribution of X_J given $X_I = x_I^0$ is joint normal with mean

$$\mu\left(x_I^0\right) = E(X_J|X_I = x_I^0) = S_{JI} S_{II}^{-1} x_I^0 \tag{5}$$

and covariance matrix

$$\begin{aligned} S\left(x_I^0\right) &= E((X_J - \mu(x_I^0))^2|X_I = x_I^0) \\ &= S_{JJ} - S_{JI} S_{II}^{-1} S_{IJ}. \end{aligned} \tag{6}$$

We see that in fact covariance matrix does not depend on the condition x_I^0.

2.3 Copula function

Copula C [2] represents dependence structure of X. Denote $F_1, \ldots, F_d$ marginal distribution functions of X. If marginal distributions are continuous, copula is a cumulative distribution function of the random vector $U = (U_1, \ldots, U_d)$ with components $U_1 = F_1(X_1), \ldots, U_d = F_d(X_d)$:

$$C(u_1, \ldots, u_d) = P(U_1 \le u_1, \ldots, U_d \le u_d)$$

for $u = (u_1, \ldots, u_d) \in [0,1]^d$. U possesses uniform marginals on $[0,1]$ and the following representation is true:

$$C(u_1, \ldots, u_d) = F(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)), \tag{7}$$

for $u = (u_1, \ldots, u_d) \in [0,1]^d$.

If the initial distribution has density f, then the copula function also possesses density, which equals to

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d} \tag{8}$$

Major mode of using copulas is linking together a simple copula function with arbitrary marginal distributions to come up to the joint distribution function as follows:

$$F(x_1, \ldots, x_d) = C\left(F_1(x_1), \ldots, F_d(x_d)\right), \tag{9}$$

for $x \in R^d$, with corresponding relation for densities

$$f(x_1, \ldots, x_d) = c\left(F_1(x_1), \ldots, F_d(x_d)\right) \prod_{k=1}^{d} f_k(x_k), \tag{10}$$

where f_k stands for the density of the k-th component. Since the product in the last expression represents joint density in the independent case, we may treat it as follows: actual density equals independent density

$$\prod_{k=1}^{d} f_k(x_k)$$

times correction multiplier

$$c\left(F_1(x_1), \ldots, F_d(x_d)\right).$$

In case of joint normal distribution of X, the copula function C is called normal or Gaussian, it has one parameter — correlation matrix of the basic distribution. Marginal distributions in this case are univariate normal.

2.4 Gaussian copula conditionals

Given a Gaussian copula C with parameter S, and a random vector U possessing the joint distribution function C; consider conditional distribution of U_I given $U_J = b$. Denote Φ the standard normal distribution function, and define a random vector $Y = \Phi^{-1}(U)$, where the function Φ^{-1} is applied component-wise. Clearly Y has joint normal distribution with covariance matrix S equal to the copula parameter, and standard normal components.

Thus we can write

$$\begin{aligned} C_{J|I}(u_J|U_I = u_I^0) &= P(U_J \le u_J|U_I = u_I^0) \\ &= P(Y_J \le \Phi^{-1}(u_J)|Y_I = \Phi^{-1}(u_I^0)) \\ &= F_{J|I}(\Phi^{-1}(u_J)|Y_I = \Phi^{-1}(u_I^0)). \end{aligned} \tag{11}$$

The latter as we know is an easily derived transform of the joint normal with mean

$$S_{JI}S_{II}^{-1}\Phi^{-1}(u_I^0)$$

and covariance matrix (6). In particular, the conditional density function takes the form

$$c_{J|I}(u_J|U_I = u_I^0) = \frac{f_{J|I}\left(\Phi^{-1}(u_J)|Y_I = \Phi^{-1}(u_I^0)\right)}{\varphi(\Phi^{-1}(u_J))}, \tag{12}$$

where $\varphi()$ with vector argument stands for the product of values of standard normal density φ applied to components of the vector:

$$\varphi(\Phi^{-1}(u_J)) = \prod_{j \in J} \varphi\left(\Phi^{-1}(u_j)\right).$$

2.5 Conditionals in initial space

Given a joint distribution of a random vector X with normal copula with parameter S and arbitrary marginal distribution functions $F_1, \ldots, F_d$, the conditional distribution function of X_J given $X_I = x_I^0$ formally is inversion of formulas (11) and (12), that is

$$\begin{aligned} F_{J|I}(x_J) &= P(X_J \le x_J|X_I = x_I^0) \\ &= C_{J|I}\left(F_J(x_J)|F_I(X_I) = F_I(a)\right),. \end{aligned} \tag{13}$$

Conditional density function takes the form

$$\begin{aligned} f_{J|I}(x_J|X_I = x_I^0) &= c_{J|I}(F_J(x_J)|F_I(X_I) = F_J(x_I^0)) \\ &\quad \times f_J(x_J) \end{aligned} \tag{14}$$

2.6 Estimating Gaussian copula from data

Let $X^{(i)},\ i=1,\dots,n$ be a sample of observations of the random vector X. We first estimate Spearman correlation matrix S_s from the data, and then transform it to Pearson correlation matrix using a well known correspondence between Spearman s and Pearson r correlations:

$$r = 2\sin\left(\frac{s\pi}{6}\right). \tag{15}$$

The resulting correlation matrix would serve a Gaussian copula parameter.

Note that in fact the two correlations differ not much in case of normal marginals. The plot of difference between the two is presented in fig. 1, maximum absolute difference equals 0.018.

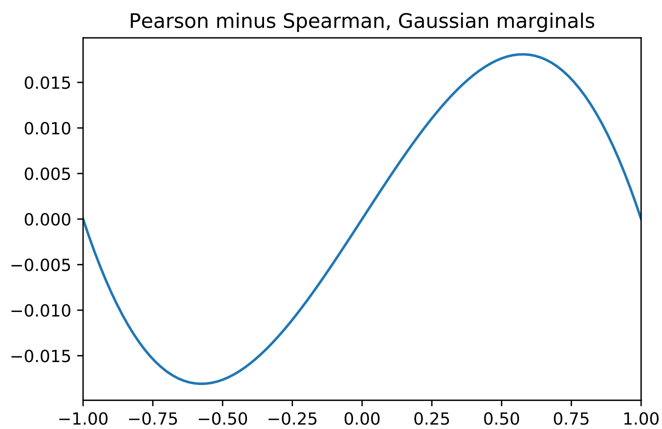

Figure 1: Difference between Pearson and Spearman correlations for Gaussian marginals.

Marginal distributions are estimated from data by any available tool; we'll keep notation $F_1,\dots,F_d$ for estimation of marginal distribution functions.

3 Examples

3.1 Gaussian conditionals

Let $d=4$ and

$$S = \begin{pmatrix} 1 & 0.5 & -0.1 & 0.3 \\ 0.5 & 1 & -0.5 & 0.2 \\ -0.1 & -0.5 & 1 & 0.3 \\ 0.3 & 0.2 & 0.3 & 1 \end{pmatrix}. \tag{16}$$

Next, let $I=\{1,3\}$, $J=\{2,4\}$, then

$$S_{II} = \begin{pmatrix} 1 & -0.1 \\ -0.1 & 1 \end{pmatrix},\quad S_{JJ} = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix},$$

$$S_{IJ} = \begin{pmatrix} 0.5 & 0.3 \\ -0.5 & 0.3 \end{pmatrix},\quad S_{JI} = \begin{pmatrix} 0.5 & -0.5 \\ 0.3 & 0.3 \end{pmatrix},$$

so that, given $x_I^0 = (1,2)'$, we get

$$\mu\left(x_I^0\right) = \begin{pmatrix} -0.455 \\ 1 \end{pmatrix},\quad S\left(x_I^0\right) = \begin{pmatrix} 0.545 & 0.2 \\ 0.2 & 0.8 \end{pmatrix}. \tag{17}$$

Thus conditional distribution of components (X_2, X_4), given $X_1 = 1$ and $X_3 = 2$, of a Gaussian random vector with zero mean and covariance matrix S, is Gaussian with mean and covariance matrix (17).

3.2 Conditional copula distributions

Now consider examples of conditional distributions of copula vector U components given other components fixed, as described in the section 2.4.

Let $d = 2$ and correlation matrix of the Gaussian copula of the random vector $U = (U_1, U_2)$ in this example be

$$S = \begin{pmatrix} 1 & 0.6 \\ 0.6 & 1 \end{pmatrix}. \tag{18}$$

Figures 2, 3 show plots of conditional density and distribution functions of U_1 given $U_2 = a$ fixed, for a number of values of a.

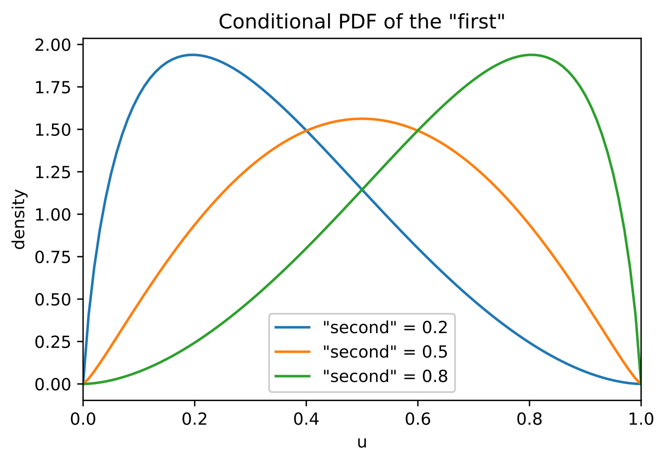

Figure 2: Conditional probability density functions of U_1 ("first") provided U_2 ("second") is fixed.

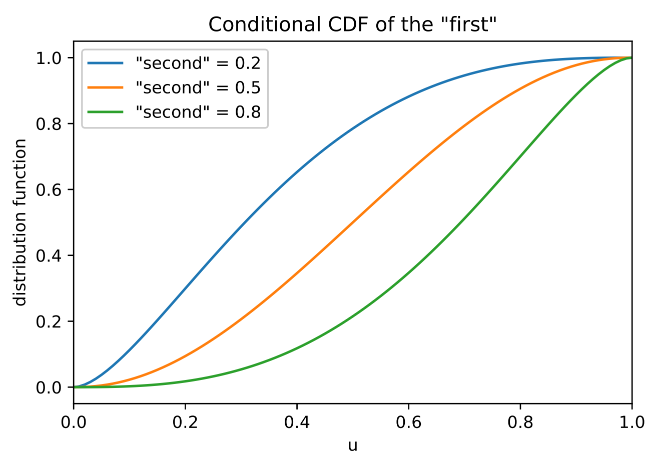

Figure 3: Conditional distribution functions of U_1 ("first") provided U_2 ("second") is fixed.

3.3 Conditional distribution with Gaussian copula and exponential marginal distributions

Let $d = 2$, and correlation matrix of the copula be defined in (18). Next, let marginal distributions of $X = (X_1, X_2)$ be exponential with mean 1.

Conditional density functions and distribution functions of X_1 given X_2 fixed are presented in the figures 4, 5.

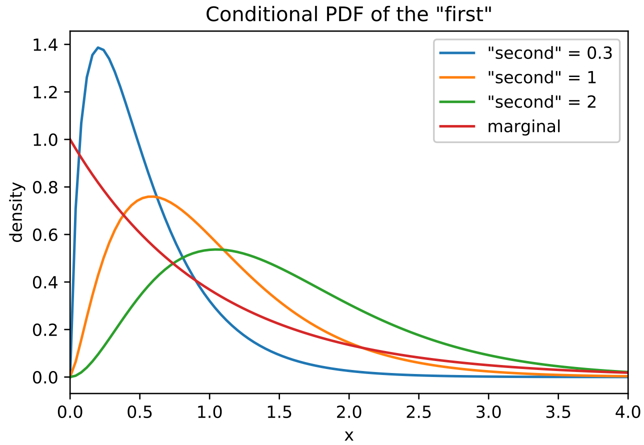

Figure 4: Conditional probability density functions of X_1 ("first") provided X_2 ("second") is fixed, and unconditional marginal density (exponential).

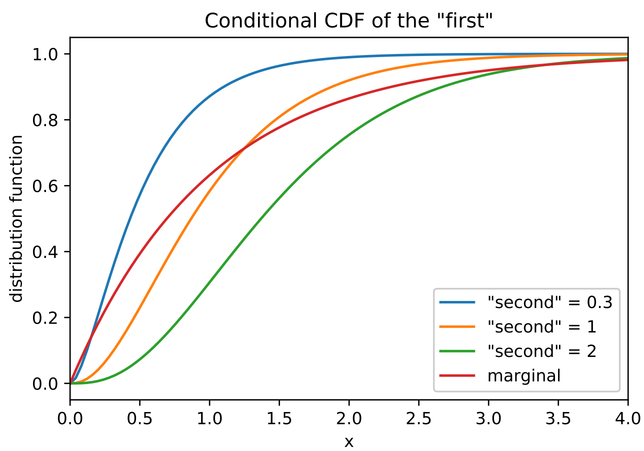

Figure 5: Conditional distribution functions of X_1 ("first") provided X_2 ("second") is fixed, and unconditional marginal distribution function (exponential).

3.4 Conditional distributions with Gaussian copula and t marginal distributions

Distributions of financial assets returns often closely follow Student t distribution with heavy tails, so using these distributions is of great practical importance.

Let $d = 4$ and correlation matrix of a Gaussian copula be defined in (16). Figures 6, 7 present conditional density and distribution function plots of X_1 given (X_2, X_3, X_4) fixed for a number of parameter values.

If the target dimension equals 2, we may use a coutour plot to present conditional density function. Figures 8, 9 show a couple examples of such plots in the same distribution model,

3.5 Financial assets example

Consider a number of market indices and a Apple stock, and look how conditional distributions may be applied in this case. First note that Student t distribution fits marginal distribution much better than normal one, so direct using of multivariate Gaussian case is indeed flawed.

Figure 10 presents plots of kernel density estimate of the stock returns distribution, together with best estimates from Gaussian and t classes.

Returns for other stocks exhibit similar behavior, so we would consider using Gaussian copula with Student t marginal distributions in our next example.

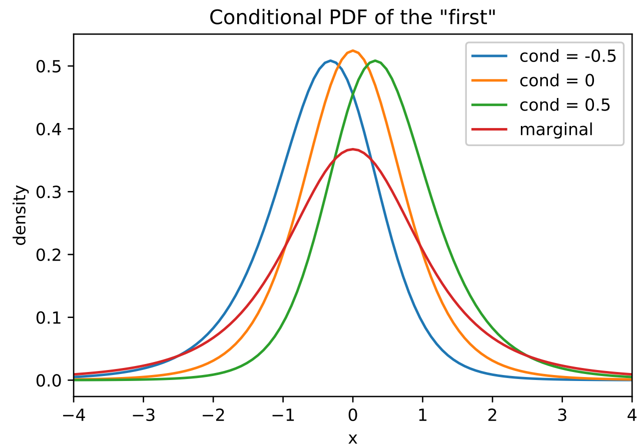

Figure 6: Conditional probability density functions of X_1 ("first") provided $(X_2, X_3, X_4) = (a, a, a)$ is fixed, for $a = -0.5, 0, 0.5$, and unconditional marginal density (Student t).

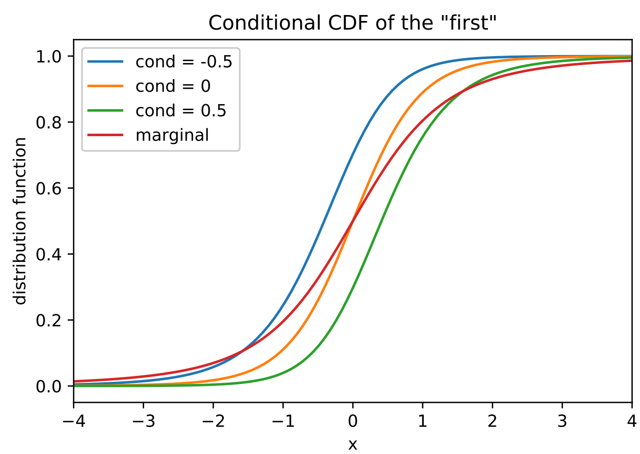

Figure 7: Conditional distribution functions of X_1 provided $(X_2, X_3, X_4) = (a, a, a)$ is fixed, for $a = -0.5, 0, 0.5$, and unconditional marginal distribution function (Student t).

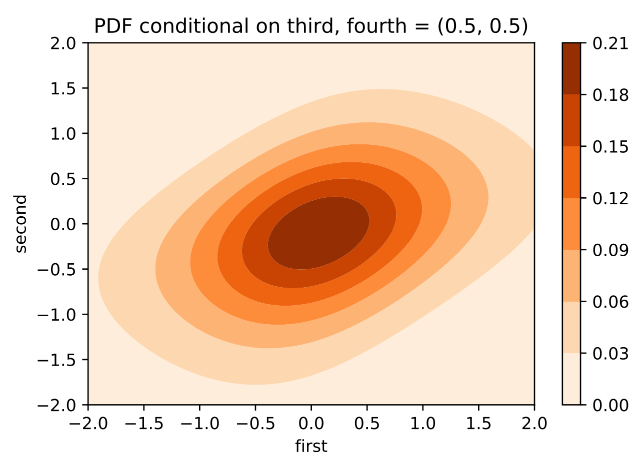

Figure 8: Conditional probability density contour plot of (X_1, X_2) ("first, second") provided $(X_3, X_4) = (0.5, 0.5)$ is fixed.

To build a factor model we take data of 5 market indices, specifically, DJIA, S&P500, NASDAQ, DAX, STOXX50, and add Apple stock quotes data, from January 2010 till November 2017, convert the data

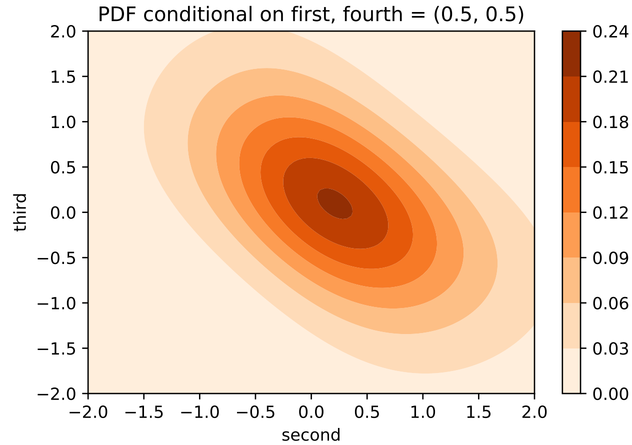

Figure 9: Conditional probability density contour plot of (X_2, X_3) ("second, third") provided $(X_1, X_4) = (0.5, 0.5)$ is fixed.

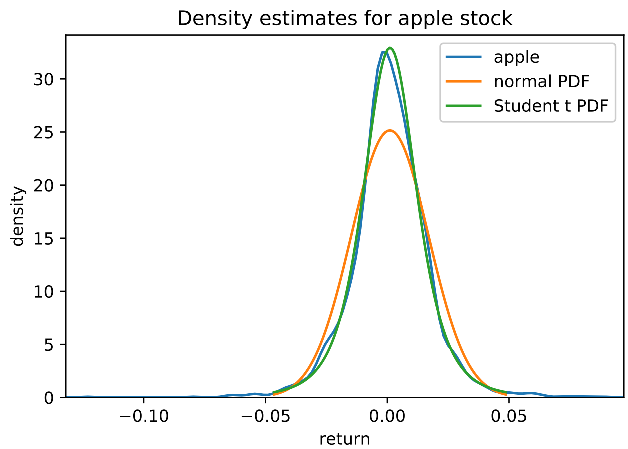

Figure 10: Conditional probability density contour plot of (X_1, X_2) ("first, second") provided $(X_3, X_4) = (0.5, 0.5)$ is fixed.

returns, and build a multivariate distribution model with Gaussian copula and Student t marginals, as described in the section 2.6. Heat map of the copula correlation matrix is presented in the figure 11.

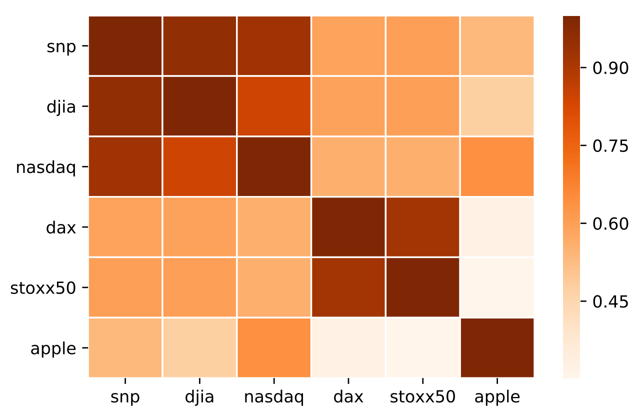

Figure 11: Heat map of the copula correlation matrix.

Figure 12 presents conditional density of Apple returns on November 27, 2017, and actual return of the stock on that date, we see that actual return is less than typical value.

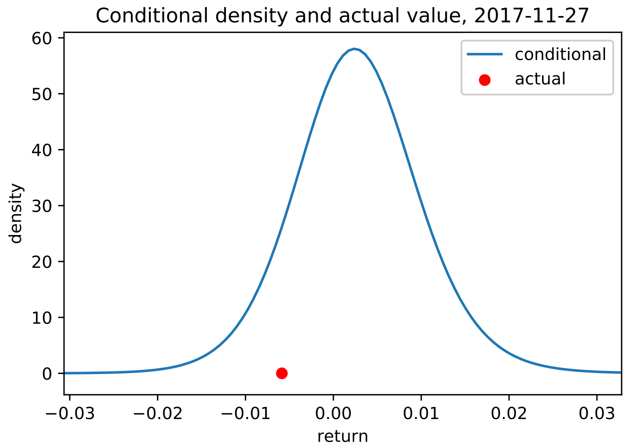

Figure 12: Conditional probability density of "apple"returns given factor returns at Nov 27, 2017, and actual "apple"return at the same date.

Next figure 13 shows a similar plot relating to November 21, 2017. Here actual return exceeds most of the typical return values.

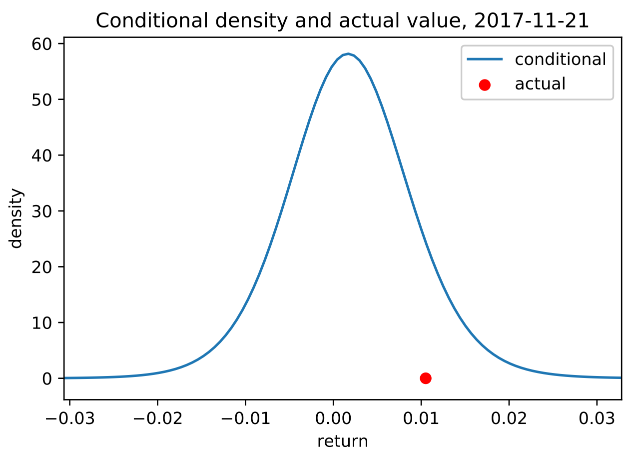

Figure 13: Conditional probability density of "apple"returns given factor returns at Nov 21, 2017, and actual "apple"return at the same date.

Comparison of actual returns with conditional distributions may be used for creating statistical arbitrage trading strategies.

4 Conclusion

We have shown that the elegant solution of multivariate conditional distributions problem, existing for joint Gaussian distributions, may be extended to arbitrary distributions with Gaussian copula, provided all the necessary formal description, and presented examples.

References

[1] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] R.B. Nelsen. *An Introduction To Copulas*. Springer, 1998.