Review Session 3

July 15, 2015

Estimation

- p34 4.1.2 Estimation of quantile
- p35 4.2.2 MLE for univariate elliptical variables
- p39 4.3.3 Sample covariance and eigenvalue dispersion
- p41 4.5.1 Influence function of sample mean
- p46 4.8.1 Sample mean
- p47 4.8.2 p-value analytical
- p52 4.8.5 Generalized t-tests, simulations

SYMMYS
Advanced Risk and Portfolio Management®

4.1.2 Estimation of quantile

Assume that we are interested in this functional:

$$G[f_X] \equiv (\mathcal{I}[f_X])^{-1}(p), \tag{165}$$

where $\mathcal{I}[\cdot]$ is the integration operator and $p \equiv 0.5$. Notice that the above is simply the quantile with confidence p, see (1.8) and (1.17) in Meucci (2005):

$$G[f_X] \equiv Q_X(p). \tag{166}$$

In particular, given that $p \equiv 0.5$, the above is the median.

Compute the non-parametric estimator $\widehat{q}_p$ of (165) defined by (4.36) in Meucci (2005). Assume knowledge of the parameters (153) and evaluate the performance of $\widehat{q}_p$ with respect to (165) as in the script `S_Estimator` by stress-testing the parameter μ_Y in the range $[0, 0.2]$.

Hint. Use the function `QuantileMixture`.

Evaluate the performance of the estimator (156) with respect to (165) as in the script `S_Estimator` by stress-testing the parameter μ_Y in the range $[0, 0.2]$.

Hint. Use the function `QuantileMixture`.

$$G[f_X] = \left(I[f_X]\right)^{-1}(p) = F_X^{-1}(p) = Q_X(p) \rightarrow \text{quantile with confidence } p$$

$p = 0.5$ (median)

- DISTRIBUTION OF X:

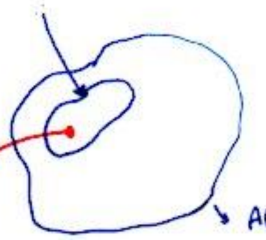
Y, Z, B indep.

$$B: \text{Bernoulli} \begin{cases} 0 & \text{prob } 1-\alpha = 0.2 \\ 1 & \text{prob } \alpha = 0.8 \end{cases}$$

$Y \sim N(\mu_Y, 0.2^2)$

$Z \sim \log N(0, 0.15^2)$

$\mu_Y \in [0, 0.2]$ STRESS TEST SET



$$\boxed{X = BY + (1-B)Z}$$

ASSUME THAT THE TRUE μ_Y IS 0.1

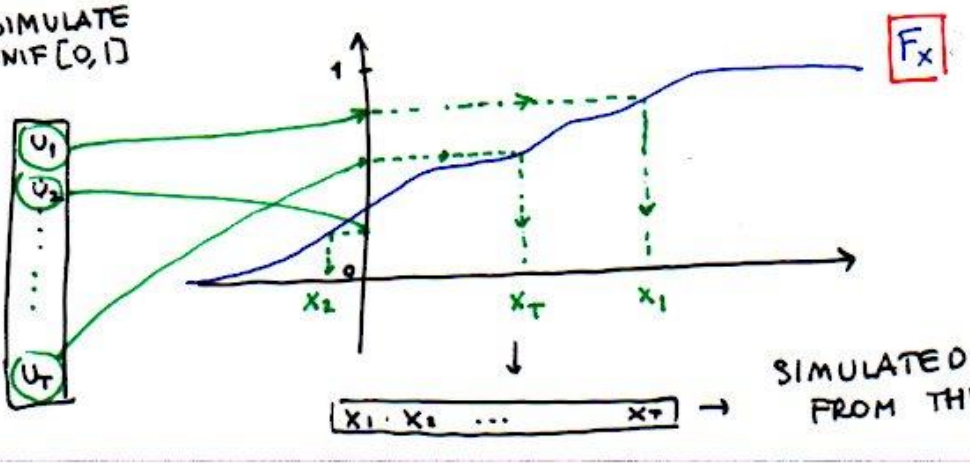
pdf: $f_X(x) = \alpha f_Y(x) + (1-\alpha) f_Z(x)$

cdf: $F_X(x) = \alpha F_Y(x) + (1-\alpha) F_Z(x)$

① GENERATE A SAMPLE FROM THE DISTRIBUTION OF X

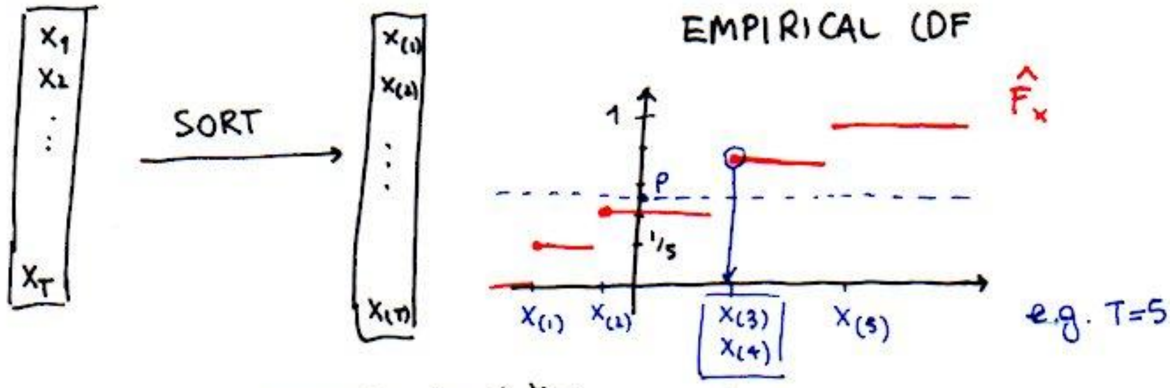
→ EITHER SIMULATE Y, Z, B INDEPENDENTLY $\rightsquigarrow X$

OR: SIMULATE $U \sim UNIF[0,1]$



→ SIMULATED SAMPLE FROM THE DIST. OF X

② GIVEN THE SAMPLE, COMPUTE THE NON-PARAM ESTIMATOR OF THE QUANTILE



$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(T)}$ $\qquad$ $\hat{Q}_x(p)$

$$\hat{Q}_x(p) = \hat{F}_X^{-1}(p) = \min \{ x \in \mathbb{R} : \hat{F}(x) \geq p \}$$

$p = 0.5 \qquad \hat{Q}_x(0.5) = x_{\lceil T \cdot 0.5 \rceil}$

$T/2$: POSITION OF $\hat{Q}$ IN THE SORTED SAMPLE

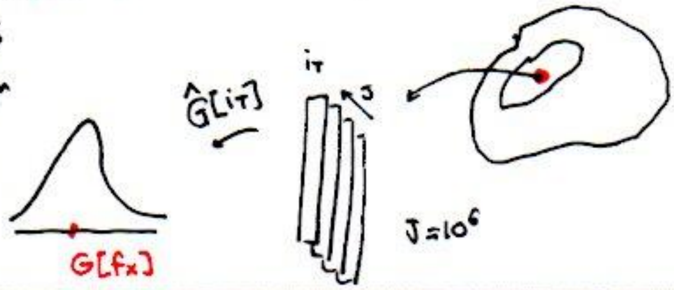
→ COMPARE TWO ESTIMATORS FOR THE MEDIAN

ORDER STATISTIC $p = 0.5 \qquad \hat{Q}_x(p) \rightarrow \hat{G}_e[i_T]$

SAMPLE MEAN $\qquad \dfrac{\sum_{t=1}^{T} x_t}{T} \rightarrow \hat{G}_b[i_T]$

- LUCK vs REPLICABILITY

Test the two estimators repeating simul + estim many times (J) and plotting the dist of the estimators



· STRESS TEST $\mu_Y \in [0, 0.2]$

Repeat the test on a grid of points for μ_Y

and compute error, bias, ineff

of the two estimators for each value

of μ_Y.

→ choose the one that behaves better in the worst case (order statistic)

4.2.2 MLE for univariate elliptical variables

Consider as in (1.28) in Meucci (2005) a symmetrical univariate random variable X. It is easy to check that such distributions are all and only the one-dimensional elliptical distributions. In other words, there exist two numbers μ and σ and a univariate function g such that:

$$X \sim \mathrm{El}\left(\mu, \sigma^2, g\right). \tag{172}$$

Assume that you know the functional form of g. Adapt the proof in the technical appendix www.4.2 at `symmys.com` > `Book` > `Downloads` to compute the maximum likelihood estimators $\widehat{\mu}$ and $\widehat{\sigma}^2$ of μ and σ^2 respectively.

$X \sim El(\mu, \sigma^2; g)$

↳ function : decay kernel

pdf: $f_X(x) = \frac{1}{\sigma} g\left(\left(\frac{x-\mu}{\sigma}\right)^2\right)$

e.g. NORMAL $g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x}$

CAUCHY $g(x) = \frac{1}{\pi} \frac{1}{1+x}$

STUDENT t ...

Suppose $X \sim E\ell(\mu, \sigma^2, g)$ with g known

MLE for μ and σ^2 ?

data : $x_1, \dots, x_T$ iid $\sim X$

$$L(x_1, \dots, x_T; \mu, \sigma) = \prod_{t=1}^{T} f_X(x_t, \mu, \sigma) \qquad \text{LIKELIHOOD}$$

$$\ln L(x_1, \dots, x_T; \mu, \sigma) = \sum_{t=1}^{T} \ln f_X(x_t, \mu, \sigma) \qquad \text{LOG-LIKELIHOOD}$$

$$= \sum_{t=1}^{T} \ln \left[\frac{1}{\sigma} g\left[\left(\frac{x_t - \mu}{\sigma} \right)^2 \right] \right] =$$

$$= \sum_{t=1}^{T} \ln \left(\frac{1}{\sigma} \right) + \sum_{t=1}^{T} \ln g \left[\left(\frac{x_t - \mu}{\sigma} \right)^2 \right] \qquad s = \frac{1}{\sigma}$$

$$= T \cdot \ln(s) + \sum_{t=1}^{T} \ln g\left(s^2 (x_t - \mu)^2 \right)$$

TO BE MAXIMIZED wrt μ, s

1st order criterion

$$w_t = -2 \left. \frac{g'}{g} \right|_{s^2(x_t-\mu)^2}$$

$$\frac{\partial \ln L}{\partial \mu} = \sum_{t=1}^{T} \frac{g'(s^2(x_t-\mu)^2)}{g(s^2(x_t-\mu)^2)} \cdot (-2) \, s^2 (x_t - \mu)$$

$$= \sum_{t=1}^{T} w_t \, s^2 (x_t - \mu) = \underbrace{\cancel{s^2}}_{>0} \sum_{t=1}^{T} w_t (x_t - \mu) \overset{!}{=} 0$$

$$\rightarrow \hat{\mu} = \frac{\sum_{t=1}^{T} w_t x_t}{\sum_{t=1}^{T} w_t}$$

(implicit: w_t depends on μ and s)

SOLVED RECOURSIVELY

$$-w_t$$

$$\frac{\partial \ln L}{\partial s} = T \cdot \frac{1}{s} + \sum_{t=1}^{T} \frac{g'(s^2(x_t-\mu)^2)}{g(s^2(x_t-\mu)^2)} \cdot 2 \, s (x_t - \mu)^2$$

$$= \frac{T}{s} - s \sum_{t=1}^{T} w_t (x_t - \mu)^2 \overset{!}{=} 0 \rightarrow \frac{1}{\hat{s}^2} \equiv \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} w_t (x_t - \mu)^2$$

4.3.3 Sample covariance and eigenvalue dispersion

Fix $N \equiv 50$, $\boldsymbol{\mu} \equiv \mathbf{0}_N$, $\boldsymbol{\Sigma} \equiv \mathbf{I}_N$. Reproduce the surface in Figure 4.15. You do not need to superimpose the true spectrum as in the figure

Hints: Determine a grid of values for the number of observations T in the time series. For each value of T

a) generate an i.i.d. time series

$$i_T \equiv \{\mathrm{x}_1, \ldots, \mathrm{x}_T\} \tag{190}$$

from

$$\mathbf{X} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{191}$$

b) compute the sample covariance $\widehat{\boldsymbol{\Sigma}}$.

c) perform the PC decomposition of $\widehat{\boldsymbol{\Sigma}}$ and store the sample eigenvalues (i.e. the sample spectrum)

d) perform a)-c) a large enough number of times ($\tilde{}$ 100 times)

e) compute the average sample spectrum

$X \sim N(0, I_{50})$ $\qquad$ $(N = 50)$

Consider $T = 50, 100, 150, \ldots, 500$

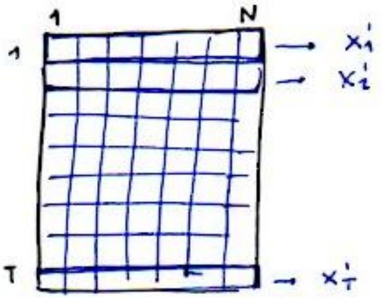
(i.e. $\frac{T}{N} = 1, 2, \ldots, 10$)

1

a) • for each T, simulate a sample from X

$(x_1, \ldots, x_T)$

SAMPLE:

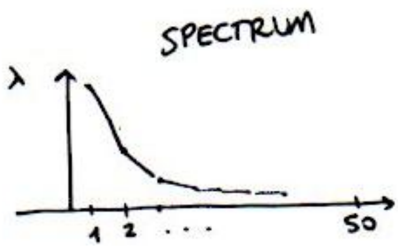


b) compute the sample cov.

$\tilde{x}_t = x_t - \hat{\mu}$ (centering) $\qquad$ $\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \tilde{x}_t \tilde{x}_t'$

c) • Compute the eigenvalues of $\hat{\Sigma}$

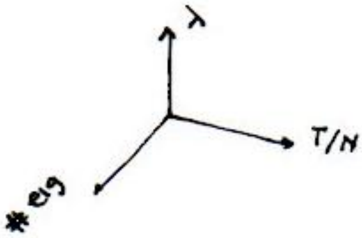
$\lambda_{(1)} \geq \lambda_{(2)} \geq \ldots \geq \lambda_{(50)}$



For the same T, repeat a)-b)-c) many times and plot the average spectrum

② Repeat 1 for all the other T's plot the surface

→ SCATTERING OF SAMPLE EIGENVALUES



4.5.1 Influence function of sample mean

Adapt the proof in the technical appendix www.4.7 at symmys.com > Book > Downloads to the univariate case to compute the influence function of the sample mean.

$$\underset{\text{pdf}}{f_X} \xrightarrow[]{\substack{\text{statistic}\\\text{functional}\\ G}} \underset{\text{number}}{G[f_X]}$$

most of the common statistic functionals are in form of expectations :

$$G_h[f] = \int h(x) f(x)\, dx = E[\overset{\text{integrable function}}{h(x)}]$$

↓

compute the I.F. of G

$$f_\varepsilon(x) = (1-\varepsilon) f_X(x) + \varepsilon\, \delta_{x_0}(x)$$

→ PERTURB THE pdf / cdf of X by a tiny point mass concentrating at x_0

$$IF_G(x_0, f) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\Big[G[f_\varepsilon(x)] - G[f] \Big]$$

$$= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} G[f_\varepsilon]$$

$$= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int h(x) \big((1-\varepsilon) f_x(x) + \varepsilon\, \delta_{x_0}(x)\big)\, dx$$

$$= \int h(x) \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \Big[(1-\varepsilon) f_x(x) + \varepsilon\, \delta_{x_0}(x)\Big]\, dx$$

$$= \int h(x) \Big[- f_x(x) + \delta_{x_0}(x)\Big]\, dx$$

$$= - \int h(x) f_x(x)\, dx + h(x_0)$$

$$= - E[h(x)] + h(x_0)$$

$G_h[f]$ IS ROBUST if IF_G is bounded

$\to$ h must be bounded

SAMPLE MEAN $\quad h(x) = x$

$$G[f] = \int x f(x)\, dx$$

$$IF_G(x_0, f) = - E[X] + x_0$$

$\to$ NOT BOUNDED

(SAMPLE MEAN IS NOT ROBUST)

4.8 Testing

4.8.1 Sample mean

Consider a time series of independent and identically distributed random variables

$$X_t \sim \mathrm{N}\left(\mu, \sigma^2\right), \quad t = 1, \ldots, T. \tag{223}$$

Consider the sample mean

$$\widehat{\mu} \equiv \frac{1}{T} \sum_{t=1}^{T} X_t. \tag{224}$$

Compute the distribution of $\widehat{\mu}$.

What is the probability that the sample mean (224) exceed a given value $\widetilde{\mu}$?

$X_t \sim N(\mu, \sigma^2) \quad t = 1 \ldots T \qquad (\text{iid})$

$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{N} X_t$

- DISTRIBUTION OF $\hat{\mu}$?

  sum of <u>iid</u> normal is normal

  $\Sigma X_t \sim N(T.\mu, T\sigma^2)$ → seen in the projection

$\rightarrow \frac{1}{T} \Sigma X_t \sim N\left(\frac{1}{\cancel{T}} \cdot (\cancel{T}\mu), \frac{1}{T^2}(\cancel{T}\sigma^2)\right)$

$\rightarrow \hat{\mu} \sim N\left(\mu, \frac{\sigma^2}{T}\right)$

- $P(\hat{\mu} > \tilde{\mu}) = 1 - P(\hat{\mu} \leq \tilde{\mu}) = 1 - P\left(\frac{\hat{\mu} - \mu}{\sqrt{\frac{\sigma^2}{T}}} \leq \frac{\tilde{\mu} - \mu}{\sqrt{\frac{\sigma^2}{T}}}\right)$

$= 1 - \phi\left(\frac{\tilde{\mu} - \mu}{\sqrt{\frac{\sigma^2}{T}}}\right)$

standard normal cdf

4.8.2 p-value analytical

Consider a normal invariant

$$X_t \sim \mathrm{N}\left(\mu, \sigma^2\right) \tag{230}$$

in a time series of length T. Consider the ML estimator $\widehat{\mu}$ of the location parameter μ. Suppose that you observe a value $\widetilde{\mu}$ for the estimator. Assume that you believe that

$$\mu \equiv \mu_0, \quad \sigma^2 \equiv \sigma_0^2. \tag{231}$$

The *p-value* of $\widehat{\mu}$ for $\widetilde{\mu}$ under the hypothesis (231) is the probability of observing a value as extreme as the observed value:

$$p \equiv \mathbb{P}\left\{\widehat{\mu} \lesseqgtr \widetilde{\mu}\right\}. \tag{232}$$

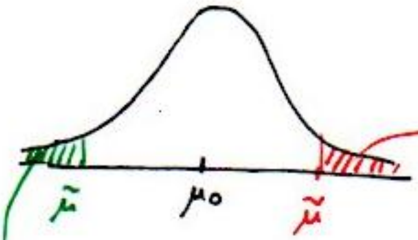
Compute the expression of the p-value in terms of the cdf of the estimator.

$X_t \sim N(\mu, \sigma^2)$

$x_1, \dots, x_T \rightarrow \hat{\mu}$ ML estimator (= sample mean in the normal case)

we believe $\mu = \mu_0$, $\sigma^2 = \sigma_0^2$ $\Rightarrow$ we believe $\hat{\mu} \sim N\left(\mu_0, \frac{\sigma_0^2}{T}\right)$

p-value of $\hat{\mu}$ for $\tilde{\mu}$ $\rightarrow$ prob. of observing a value as (or more) extreme as $\tilde{\mu}$



$P(\mu \geq \tilde{\mu}) = 1 - F^N_{\mu_0, \frac{\sigma_0^2}{T}}(\tilde{\mu})$

$P(\mu \leq \tilde{\mu}) = F^N_{\mu_0, \frac{\sigma_0^2}{T}}(\tilde{\mu})$

4.8.5 Generalized t-tests, simulations

Consider a joint model for the invariants: for $t = 1, \ldots, T$. The marginals are

$$X_t \sim \text{LogN}\left(\mu_X, \sigma_X^2\right), \tag{274}$$
$$F_t \sim \text{Ga}\left(\nu_F, \sigma_F^2\right); \tag{275}$$

and the copula is the copula of the diagonal entries of Wishart distribution

$$\mathbf{W}_t \sim \text{W}\left(\nu_W, \mathbf{\Sigma}_W\right). \tag{276}$$

Consider the coefficients that define the regression line (3.127)

$$\widetilde{X}_t \equiv \alpha + \beta F_t. \tag{277}$$

Compute the non-parametric estimators $\left(\widehat{\alpha}, \widehat{\beta}\right)$ of the regression coefficients.

Are $\left(\widehat{\alpha}, \widehat{\beta}\right)$ the maximum-likelihood estimators of the regression coefficients?

JOINT MODEL FOR THE INVARIANTS

marginals $X_t \sim \text{LogN}(\mu_X, \sigma_X^2)$ $t = 1..T$

$F_t \sim \text{Ga}(\nu_F, \sigma_F^2)$

copula $\rightarrow$ copula of the diagonal entries of a wishart dist. $W_t \sim W(\nu_W, \Sigma_W)$

$\tilde{X}_t = \alpha + \beta F_t$ $\rightarrow$ compute the non param. estimators $\hat{\alpha}, \hat{\beta}$

$$f_t = \begin{pmatrix} 1 \\ F_t \end{pmatrix}$$

$$(\hat{\alpha}, \hat{\beta}) = \left(\sum_t x_t f_t'\right)\left(\sum_t f_t f_t'\right)^{-1} = \hat{\Sigma}_{XF}\, \hat{\Sigma}_F^{-1}$$

IF THE REGRESSION MODEL IS CONDITIONALLY NORMAL $\hat{\alpha}, \hat{\beta} \equiv$ MLE estimators

IN THIS EXERCISE THIS IS NOT THE CASE

HOW TO SIMULATE A SAMPLE FROM THE DIST. of (X,F) ?

Recall WISHART DISTRIBUTION $W \sim W(\nu, \Sigma)$

$Y_1, Y_2, \dots, Y_\nu \sim N(0, \Sigma)$ indep

$Y_1 Y_1' + Y_2 Y_2' + \dots + Y_\nu Y_\nu' \sim W(\nu, \Sigma)$

In this example $\dim(Y) = 2$

assume $\nu = 3$ $\rightarrow Y_1 = \begin{bmatrix} Y_{11} \\ Y_{12} \end{bmatrix}, Y_2 = \begin{bmatrix} Y_{21} \\ Y_{22} \end{bmatrix}, Y_3 = \begin{bmatrix} Y_{31} \\ Y_{32} \end{bmatrix}$

$Y_i \sim N(0, \Sigma)$ $\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$

$$W = Y_1 Y_1' + Y_2 Y_2' + Y_3 Y_3' = \begin{bmatrix} \underbrace{Y_{11}^2 + Y_{21}^2 + Y_{31}^2}_{W_1} & \text{mixed prod} \\ \text{mixed prod} & \underbrace{Y_{12}^2 + Y_{22}^2 + Y_{32}^2}_{W_2} \end{bmatrix}$$

$W_1 \sim \text{Gamma}\left(\frac{\nu}{2}, 2\sigma_{11}\right)$

$W_2 \sim \text{Gamma}\left(\frac{\nu}{2}, 2\sigma_{22}\right)$

(SUM OF INDEP GAMMA IS GAMMA)

$\text{copula}(X,F) = \text{copula}(W_1, W_2)$

↓

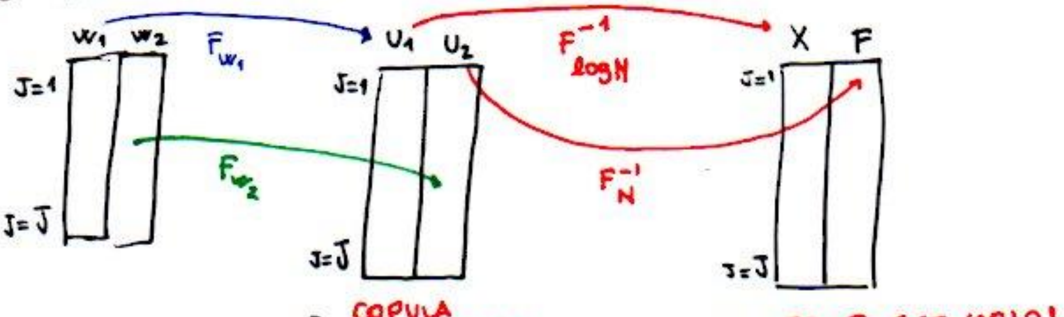
GENERATE SCENARIOS FOR THE COPULA ⟶ GENERATE JOINT SCEN FOR (X,F)

① SIMULATE $Y_1^{(J)}, Y_2^{(J)}, Y_3^{(J)}$ $J = 1 \dots 100000$

where each $Y_i \sim N(0, \Sigma_W)$

② CALCULATE $W^{(J)} = Y_1^{(J)} Y_1^{(J)'} + \dots + Y_3^{(J)} \cdot Y_3^{(J)'}$

③ RETRIEVE THE DIAGONAL ELEMENTS



④ COPULA SCENARIOS

⑤ JOINT SCENARIOS FOR (X,F)

ONCE OBTAINED SCENARIOS, COMPUTE $\hat{\alpha}, \hat{\beta}$

- REPEAT THE ABOVE STEPS (SIMULATIONS + ESTIMATION) MANY TIMES TO OBTAIN THE DISTRIBUTION (HIST) OF $\hat{\alpha}, \hat{\beta}$

Generate arbitrary values for the parameters in (274)-(276) and for the number of observations T and compute in simulation the distribution of the statistic

$$\widehat{G}_{\alpha} \equiv \sqrt{T-2}\frac{(\widehat{\alpha}-\alpha)}{\sqrt{\widehat{\sigma}^2 \widehat{\sigma}_{\alpha}^2}} \tag{278}$$

and the distribution of the statistic

$$\widehat{G}_{\beta} \equiv \sqrt{T-2}\frac{\widehat{\beta}-\beta}{\sqrt{\widehat{\sigma}^2 \widehat{\sigma}_{\beta}^2}}. \tag{279}$$

Compare the empirical distribution of (278) with the analytical distribution of (261) as well as the empirical distribution of (279) with the analytical distribution of (262) and comment.

statistics $\quad \hat{G}_{\alpha} = \sqrt{T-2}\,\dfrac{\hat{\alpha}-\alpha}{\sqrt{\hat{\sigma}^2_{RES}\,\hat{\sigma}^2_{\alpha}}}$

$$\hat{G}_{\beta} = \sqrt{T-2}\,\frac{\hat{\beta}-\beta}{\sqrt{\hat{\sigma}^2_{RES}\,\hat{\sigma}_{\beta}}}$$

compare their distribution with

$St(T-2, 0, 1)$

r

if X is $St(T-2, 0, 1)$ then $F_{St(T-2,0,1)}(X) \sim U$

$\rightarrow$ plot the hist of $F_{St}(\hat{G}_{\alpha})$ and compare it with U