# Risk Management - Problems I

Jakub Gizbert-Studnicki and Maciej A. Nowak

Mark Kac Complex Systems Research Center
Jagiellonian University
Kraków, Poland
Spring 2020
Jagiellonian University WFAIS.IF-Y491.0

**Probability Theory**

1. Calculate the characteristic function for the Gaussian distribution $X \sim N(\mu, \sigma)$. Then generalize the addition (stability) law for an arbitrary (let us say, $n$) number of independent Gaussian variables $X_i$ with arbitrary mean $\mu_i$ and variance $\sigma_i^2$.

2. Calculate the characteristic function for the Cauchy distribution
$$P(x) = \frac{1}{\pi(1+x^2)}$$
Check if addition (stability) law holds.

3. Derive the formula for the cdf (cumulative distribution function): $P_{\leq}(x)$ of the Gaussian $N(\mu, \sigma)$ and the quantile (functional inverse of the cdf): $P_{\leq}^{-1}(x)$. Check numerically what percentage of values is within the range of $1, 2, ..., 10$ $\sigma$ around the mean $\mu$, respectively.

4. Suppose we have normal distribution $N(\mu, \sigma)$. Calculate the 0.9 quantile for $\mu = 2, \sigma = 0.3$ and the 0.15 quantile for $\mu = 100, \sigma = 6$.

5. Calculate kurtosis $\kappa = \lambda_4 = C_4/\sigma^4$
   (a) for the exponential distribution $P(x) = \lambda \exp(-\lambda x)\Theta(x)$, where $\Theta(x)$ is a step (Heaviside) function, and
   (b) for the continuous uniform distribution on interval $[-1/2; 1/2]$.

   Interpret the signs of the results.

6. Consider lognormal distribution
$$P_{LN}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\ln^2(x/x_0)}{2\sigma^2}\right)$$
Calculate analytically the moments $m_n \equiv \langle x^n \rangle$. Write explicitly formulae for skewness and kurtosis.

7. Consider the family of distributions $P(x) = P_{LN}^*(x)[1 + \epsilon \sin(2\pi \ln x)]$, where $|\epsilon| \leq 1$ and $P_{LN}^*(x)$ is the lognormal distribution with $x_0 = \sigma = 1$. Show that the moments $m_n$ of such distributions are identical to the moments of lognormal $P_{LN}^*(x)$, i.e. are independent on $\epsilon$. Such case, where from the knowledge of all moments one cannot infer the corresponding distribution, is known as *indeterminate*.

8. Analyze how the maximum of $N$ iid Gaussian variables $N(0,1)$ (with mean = 0 and variance = 1) converges to the Gumbel distribution.

   (a) Use the exact formula for the CDF of the maximum of $N$ iid variables: $P_{\leq}^{max}(x) = (P_{\leq}(x))^N$ and draw the exact PDF of $X_{max}$: $p^{max}(x) = dP_{\leq}^{max}(x)/dx$ for $N = 10, 10^2, 10^3, 10^4$

   (b) On the same plot draw the PDF of the Gumbel distribution. Remember that the Gumbel PDF has to be appropriately rescaled with $N$, such that: $u = (X_{max} - a_N)/b_N$, where:

   $$a_N = P_{\leq}^{-1}(1 - 1/N) \quad , \quad b_n = P_{\leq}^{-1}(1 - 1/(Ne)) - a_N \ ,$$

   and that in order to get the correct normalization the probability (not the PDF) is conserved, i.e. $p(u)du = p(x)dx$.

   NOTE: If you have problems with computing: $p^{max}(x) = dP_{\leq}^{max}(x)/dx$ you can instead plot $P_{\leq}^{max}(x)$ vs (rescaled) CDF of the Gumbel distribution.

9. Consider multivariate random variables $X_i$ with a covariance matrix elements $C_{ij}^{(X)} \equiv Cov(X_i, X_j) \equiv \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$. Show that random variables $Y_\alpha$ being linear combinations of $X_i$, such that $Y_\alpha = \sum_i v_{\alpha i} X_i$, where $v_{\alpha i}$ are constant weights, have the covariance matrix:

   $$C_{\alpha\beta}^{(Y)} \equiv Cov(Y_\alpha, Y_\beta) = \sum_{i,j} v_{\alpha i} C_{ij} v_{\beta j}$$

   In the matrix form: $C_{\alpha\beta}^{(Y)} = \boldsymbol{v}_\alpha^T \boldsymbol{C}^{(X)} \boldsymbol{v}_\beta$, if $\boldsymbol{v}_\alpha$, $\boldsymbol{v}_\beta$ are column vectors.

10. Show that the (Gaussian) $d-$dimensional random vector $\boldsymbol{Y}_d \equiv \boldsymbol{X}_d/\sqrt{d}$, where $\boldsymbol{X}_d$ is drawn from a multivariate Gaussian distribution with zero mean vector ($m_i = 0$) and unit covariance matrix ($C_{ij} = \delta_{ij}$) converges to the (unit) $d-$dimensional sphere, i.e. $||\boldsymbol{Y}_d||^2 \equiv \boldsymbol{Y}_d^T \boldsymbol{Y}_d \to 1$ for $d \to \infty$.

    (a) Numerically: generate a (large, e.g. $N = 10^4$) sample of $\boldsymbol{Y}_d$ for $d = 1, 2^2, 3^2, ..., 10^2$ and compute the mean and variance of $||\boldsymbol{Y}_d||^2$ and draw them as a function of $d$. (HINT: as for $C_{ij} = \delta_{ij}$ the elements of the random vector $\boldsymbol{X}_d$ are independent, one can simply generate each of them separately from a one-dim Gaussian)

(b) Analytically: one can use the Central Limit Theorem (HINT: the sum of squares of $d$ iid standard Gaussian variables $X_i \sim N(0,1)$: $Q = \sum_{i=1}^{d} X_i^2$ has the Chi-Squared distribution with $d$ degrees of freedom.)

Note that this is also the case for any (normalized, i.e. shifted and rescaled: $\frac{\boldsymbol{X}_i - \boldsymbol{m}_i}{C_{ii}^{1/2}}$) correlated Gaussian variables, as introducing correlation is just a rotation of the sphere (see Principal Components).

11. (Contour) plot the joint PDF of the 2-dimensional probability distribution defined by the Gumbel Copula:

$$\text{CDF: } C(u,v) = \exp\left[\left(\ln\frac{1}{u}\right)^{\theta} + \left(\ln\frac{1}{v}\right)^{\theta}\right]^{1/\theta}$$

with standard Gaussian marginals $N(0,1)$ for the (coupling) parameter $\theta = 1, 1.5, 2$.

Note: for $\theta > 1$ this is not a multivariate Gaussian but it has Gaussian marginal distributions !

12. The file *dat.txt* contains sample bivariate data (each line is a 2-dim random vector with components separated by a space), generated using the Gaussian Copula with some (unknown) marginals.

(a) Compute numerically and plot CDFs of the marginal distributions

HINT: empirical CDF for the sample with $N$ elements:

$$P_{\leq}(x) = \frac{\text{number of elements in the sample} \leq x}{N} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{X_i \leq x}$$

(b) Estimate the (only) parameter of the 2-dim Gaussian Copula, i.e. the correlation coefficient $\rho$.

HINT: Using empirical marginal CDFs convert the sample data into the 2-dim uniform distribution and then (using inverse Gaussian CDF) into the 2-dim Gaussian and compute the covariance $\rho$, see (the opposite) example in the Lecture notes, p. 47.

13. Show that $S^2_{N-1} \equiv \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X}_N)^2$, where $\bar{X}_N \equiv \frac{1}{N}\sum_{i=1}^{N} X_i$ is the unbiased estimator of the population $X$ variance (if it exists), i.e.:

$$\langle S^2_{N-1}\rangle = \langle (X - \langle X\rangle)^2\rangle$$

14. Using the Maximum (log-) Likelihood method

    (a) find estimators of the lognormal distribution parameters ($x_0$ and $\sigma$)

    $$P_{LN}(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\ln^2(x/x_0)}{2\sigma^2}\right)$$

    (b) show that estimators of the parameters ($a$ and $b$) in the linear regression model: $y_i = a\,x_i + b + \epsilon_i$ with Gaussian white noise ($\epsilon_i$ are iid $\sim N(0,\sigma)$) are given by the Least Squares Method estimators.

15. The files *dat2a.txt* and *dat2b.txt* contain sample data generated from some probability distributions. Using the Kolmogorov-Smirnov test check if at $\alpha = 0.05$ significance level

    (a) *dat2a.txt* come from the Gaussian $N(5,2)$ distribution,

    (b) *dat2b.txt* come from the Gaussian $N(5,2)$ distribution,

    (c) *dat2a.txt* and *dat2b.txt* come from the same distribution.

    In each case compute the test statistic: $D_N$ and the p-value: $Q_{KS}(u)$.

    NOTE: KS formula for the p-value is valid only assymptoticaly, i.e. for $N \to \infty$. For finite (and not too large) $N$ in a single distribution KS test you can use a correction: $u \to u + \frac{1}{6\sqrt{N}} + \frac{u-1}{4N}$ to get a result closer to some software (e.g. Wolfram Mathematica) automatic functions.