# Modeling Extreme Events from Computer Simulations
# Part II

Emmanuel Vazquez

SUPELEC, Gif-sur-Yvette, France

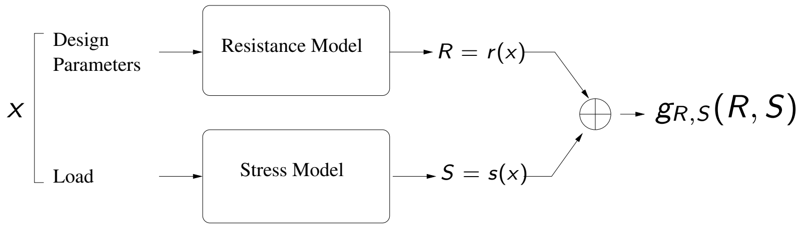
Summer School CEA-EDF-INRIA, 2011



## Outline of Part II

## Reliability approach in mechanics

- In mechanics, the term reliability describes the ability of a system to accomplish a required function
- Mechanical materials or structures are considered as systems comprising an input, a state and an output
- The classical point of view




- The success of a design (or serviceability) is seen in the verification of the inequality

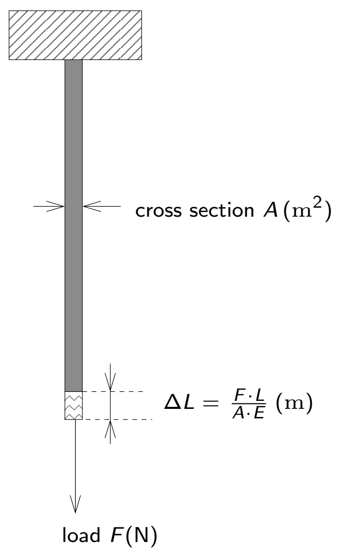
$$g_{R,S}(R, S) = R - S \geq 0$$

- The quantity $g_{R,S}(R, S) = R - S$ is called the (safety) margin
- $g_{R,S}$ is called the limit-state function.





## Example: deformation of a rod




- The yield point of a material is defined as the stress at which a material begins to deform plastically (prior to the yield point the material will deform elastically and will return to its original shape when the applied stress is removed)
- Stress: $\sigma = F/A$ (Pa)
- Yield limit: $\sigma_\mathrm{y}$ (Pa)
- Design rule: $F/A \leq \sigma_\mathrm{y}$
- Set

$$\begin{aligned} x &= (F, A, \sigma_\mathrm{y}) \in \mathbb{R}^3 \\ r(x) &= A\sigma_\mathrm{y} \\ s(x) &= F \end{aligned}$$

- Define $g_{R,S} : (R, S) \mapsto R - S$. The design must verify

$$g_{R,S}(r(x), s(x)) = A\sigma_\mathrm{y} - F \geq 0$$

## Failure region, Limit-state function

- The operating scenario is the availability of a resistance $R$ greater than the stress $S$, such that
$$g_{R,S}(R,S) = R - S \geq 0$$
- The failure scenario is
$$g_{R,S}(R,S) = R - S < 0$$
and the failure region is defined as the region of $\mathbb{R}_+^2$
$$\Gamma = \{(r,s) \in \mathbb{R}_+^2;\ g_{R,S}(r,s) < 0\}$$
- The limit state is defined as the region of $\mathbb{R}_+^2$
$$\partial\Gamma = \{(r,s) \in \mathbb{R}_+^2;\ g_{R,S}(r,s) = 0\}$$
(hence, the name limit-state function for $g_{R,S}$)



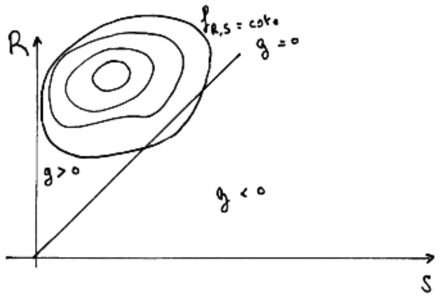


## Probability of failure

- In practice, the load applied on a system is unknown, and the design parameters are subjected to dispersions

$\rightarrow$ the parameter vector $x$ is uncertain, and can be modeled by a random vector
$$X \sim \mathrm{P}_{\mathbb{X}}$$
- Then, $R = r(X)$ and $S = s(X)$ are also random variables
- Let $f_{R,S}$ be the joint pdf of $(R,S)$ (wrt to the Lebesgue measure)
- The probability of failure of the system corresponds to
$$\alpha = \mathrm{P}\{R - S < 0\} = \int_{r-s<0} f_{R,S}(r,s)\mathrm{d}r\mathrm{d}s$$
- Reliability is defined as $1 - \alpha$

# Probability of failure




- Usually, $R$ and $S$ assumed to be independent
- Then,

$$\begin{aligned}\alpha &= \int_{r-s<0} f_S(s) f_R(r)\,\mathrm{d}r\mathrm{d}s \\ &= \int_{s\in[0,\infty[} f_S(s)\left(\int_{r\in[0,s]} f_R(r)\mathrm{d}r\right)\mathrm{d}s \\ &= \int_{s\in[0,\infty[} f_S(s) F_R(s)\mathrm{d}s\end{aligned}$$

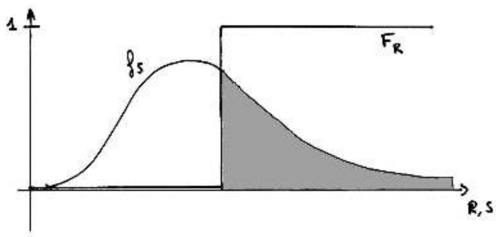
where $f_S$ and $f_R$ are the pdf of $S$ and $R$, and $F_R$ is the cdf of $R$





# Probability of failure: case of a known resistance

- When the resistance is deterministic: $R = r_0 \implies F_R(r) = \mathbb{1}_{r_0,\infty[}(r)$




- Then $\alpha = \mathrm{P}\{S > r_0\} = \mathrm{P}\{s(X) > r_0\}$

## Reliability indexes in structural reliability

- In the domain of structural reliability, the serviceability of a design is often quantified using the notion of a reliability index
- When the resistance and the stress are deterministic, the notion of reliability index may be defined arbitrarily as the numbers

$$\frac{R}{S} \text{ or } R - S$$

- When $R$ and $S$ are viewed as independent random variables, a notion reliability index may be defined as the number

$$\beta_C = \frac{\mathsf{E}(g_{R,S}(R,S))}{\mathrm{var}(g_{R,S}(R,S))^{1/2}} = \frac{m_R - m_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}$$

- However, the interpretation of $\beta_C$ is not simple and, above all, $\beta_C$ is generally not related to the probability of failure of the system





## Probability of failure expressed in terms of the uncertain factors

- Note that the point of view of structural reliability is not different from that presented in Part I
- In practice, $f_R$ and $f_S$ cannot be determined directly by the user $\rightarrow$ only the distribution $\mathsf{P}_{\mathbb{X}}$ of the vector of uncertain factors can
- Thus, it is generally easier to express a limit-state function $g_X$ in the factor space:

$$g_X : x \mapsto g_{R,S}(r(x), s(x))$$

so that the probability of failure can be expressed as

$$\alpha = \mathsf{P}\{g_X(X) < 0\}\,, \quad X \sim \mathsf{P}_{\mathbb{X}}$$

or

$$\alpha = \mathsf{P}_{\mathbb{X}}\{x \in \mathbb{X}; g_X(x) < 0\} = \mathsf{P}_{\mathbb{X}}\{g_X < 0\}$$

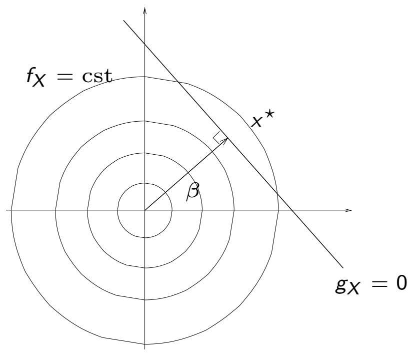
(The probability of failure is the volume of excursion of $g$ above zero.)

## Probability of failure for Gaussian random factors and affine limit-state function

- Assume $X \sim \mathrm{N}(0, \mathbb{I}_d) \in \mathbb{R}^d$
- Assume moreover that $g_X$ is affine: $\forall x \in \mathbb{R}^d$

$$g_X(x) = a_0 + a_1 x_{[1]} + \cdots + a_d x_{[d]} = a_0 + (a, x)$$

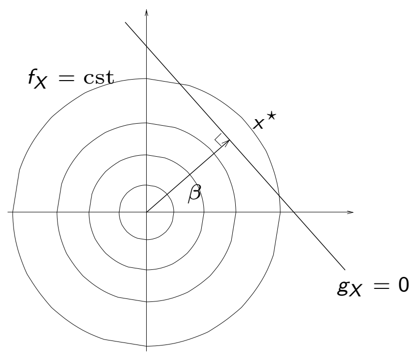
- Then, the limit-state $\partial\Gamma$ is the hyperplane defined by $a_0 + (a, x) = 0$








## Probability of failure for Gaussian random factors and affine limit-state function




- Let $x^\star$ be the nearest point of the hyperplane $\partial\Gamma = \{x \in \mathbb{R}^d; a_0 + (a, x) = 0\}$ to the origin
- $x \in \partial\Gamma \iff (x - x^\star, x^\star) = (x^\star, x) - \|x^\star\|^2 = 0$
- Thus, for $a_0 \neq 0$,

$$\frac{a}{a_0} = -\frac{x^\star}{\|x^\star\|^2} \implies \beta := \|x^\star\| = \frac{|a_0|}{\|a\|}$$

- Note that $U = (X, \eta)$, with $X \sim \mathrm{N}(0, \mathbb{I}_d)$ and $\eta = x^\star/\beta$, is a random variable with distribution $\mathrm{N}(0, 1)$

## Probability of failure for Gaussian random factors and affine limit-state function

- A failure corresponds to the event $\{U > \beta\} = \{(X, x^\star) > \beta^2\}$, which has probability

$$\alpha = 1 - \mathrm{P}\{U \leq \beta\} = 1 - \Phi(\beta) = \Phi(-\beta)$$

- Therefore, to compute a probability of failure in the case of standard normal factors and affine limit-state function $g_X$:
  1. Solve
  $$x^\star = \arg\min_{x \in \mathbb{R}^d} \|x\|$$
  $$\text{subject to } g_X(x) = 0$$
  2. The probability of failure is $\alpha = \Phi(-\beta)$, with $\beta = \|x^\star\|$
- In the literature of structural reliability, $x^\star$ is called a design point, or most central failure point





## Probability of failure for Gaussian random factors and affine limit-state function

- Consider again an affine limit-state function $g_X : x \mapsto a_0 + (a, x)$
- Define the *margin* as the random variable $Z = g_X(X)$
- Note that
$$E(Z) = a_0$$
  and
$$\mathrm{var}(Z) = a_1^2 + \cdots + a_d^2 = \|a\|^2$$
- Thus,
$$\beta = \frac{|a_0|}{\|a\|} = \frac{|E(Z)|}{\mathrm{var}(Z)^{1/2}}$$
- In the literature of structural reliability, the ratio $\frac{E(Z)}{\mathrm{var}(Z)^{1/2}}$ is interpreted as a reliability index
- $\beta$ is called the Hasofer-Lind reliability index

## Measure of the importance of the factors with respect to a failure

- The sensitivity of the probability of failure with respect to changes in the factors is an important information for the design of a system (makes it possible to understand which factor are most important to control)
- One possibility to define a notion of sensitivity is to look at the variation of the margin $z = g_X(x)$ as a function of $x$;
- We have

$$\nabla g_X(x) = a = -\frac{a_0}{\beta}\eta$$

  with $\eta = \frac{1}{\beta}x^\star$
- Thus, if the $i^{\text{th}}$ component $\eta_{[i]}$ of the unit vector $\eta$ is large, the margin will vary rapidly as we move along the $i^{\text{th}}$ direction

  ➟ $\eta_{[i]}$ accounts for the importance of the ith factor





## Extensions

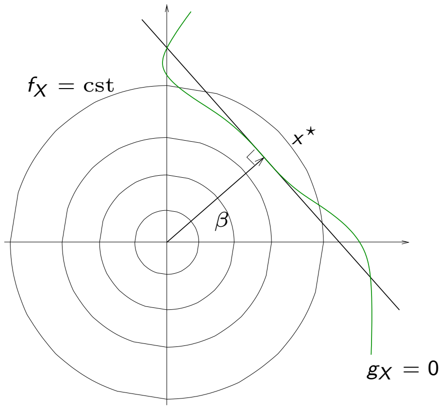
At this point, several extensions can be considered:

- Non-affine limit-state functions
- Gaussian non-standard random factors
- Non-Gaussian independent random factors
- Non-Gaussian non-independent factors

→ first-order reliability method (and related methods)

## Non-affine limit-state functions

- In some applications, it may be reasonable to approximate the limit state $\partial\Gamma$ by a geometric shape such as a hyperplane

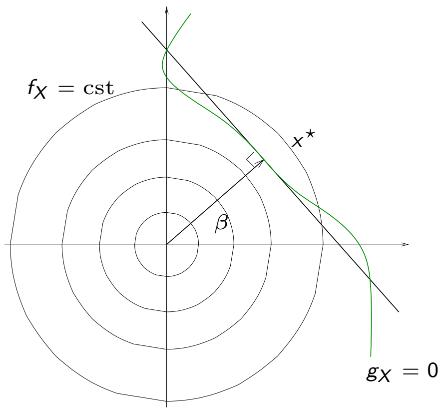



- Therefore, to approximate a probability of failure in the case of standard normal factors and a non-affine limit-state function $g_X$:
  1. Solve
  $$x^\star = \arg\min_{x \in \mathbb{R}^d} \|x\|$$
  $$\text{subject to } g_X(x) = 0$$
  2. An approximation of the probability of failure is $\widehat{\alpha} = \Phi(-\beta)$, with $\beta = \|x^\star\|$





## Non-affine limit-state functions: FORM




- If $g_X$ is smooth, note that $\nabla g_{X|x=x^\star} \propto \eta$, with $\eta = x^\star/\beta$
- Thus the equation of the approximating hyperplane is
$$(\nabla g_{X|x=x^\star}, x - x^\star) = 0$$
- Note that $x \mapsto (\nabla g_{X|x=x^\star}, x - x^\star)$ is the first-order approximation of $g_X$ in the neighborhood of $x^\star$
- In structural reliability, using a first-order approximation of $g_X$ at $x^\star$ in a standard Gaussian factor space is called First-Order Reliability Method (FORM)

## Finding the most central failure point

- Solving

$$x^\star = \arg\min_{x \in \mathbb{R}^d} \|x\| \quad \text{subject to } g_X(x) = 0$$

is a constrained optimization problem, for which many algorithms has been proposed in the literature:

  - Penalty methods
  - Augmented Lagrangian
  - Projected gradient
  - BFGS
  - SQP
  - ...

(they all are local optimization algorithms)

- In general, finding a good approximation of $x^\star$ can be done with a moderate number of evaluations of $g_X$
- However: $g_X$ generally need to be convex and differentiable
➟ If there exist (even small) numerical instabilities when computing $g_X$, as may be the case in models based on the numerical solution of some partial differential equations, then the optimization of $g_X$ using standard techniques can fail direly





## How much can go wrong with FORM?

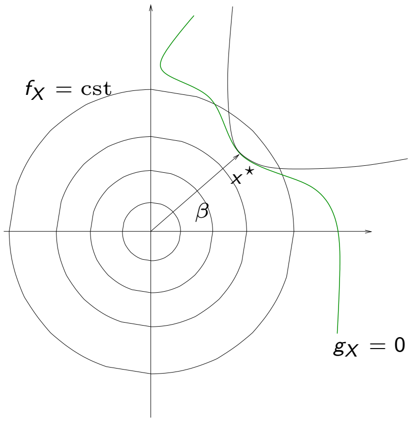
- For some applications, finding $x^\star$ can be done with only a few evaluations of $g_X$ $\rightarrow$ interesting when $g_X$ is expensive to evaluate
- However, having found $x^\star$ does not tell if the approximation $\alpha \approx \Phi(-\beta)$ is good or not
- The probability of failure can be overestimated or underestimated
- Consider, for instance the following domain of failure:

$$\Gamma = \{x \in \mathbb{R}^d; \|x\| > \beta_0\}$$

- Let $V = X_{[1]}^2 + \cdots X_{[d]}^2$, so that $V \sim \chi^2(d)$. The failure event is $\{V > \beta_0^2\}$.
- Hence, we have $\alpha = 1 - F_V(\beta_0^2)$ and $\widehat{\alpha} = 1 - \Phi(\beta_0)$
- Example: suppose $d = 20$ and $\beta_0 = 5$, we obtain $\widehat{\alpha} \approx 2.9 \cdot 10^{-7}$ but $\alpha \approx 0.2$!
- The FORM approximation should be used only when prior knowledge about the shape of $\partial\Gamma$ is available

## Quadratic approximation: SORM

- Instead of using a first-order approximation, one could think of using a second-order approximation




→ Second-Order Reliability Method (SORM)

- Such an approximation requires the approximation of $x^\star$ and the estimation of the curvature of the limit-state at $x^\star$
- A quadratic approximation is more expensive to obtain, but there is, in general, no guarantee that the approximation of the probability of failure will be better than the first-order approximation
- In fact, we can have

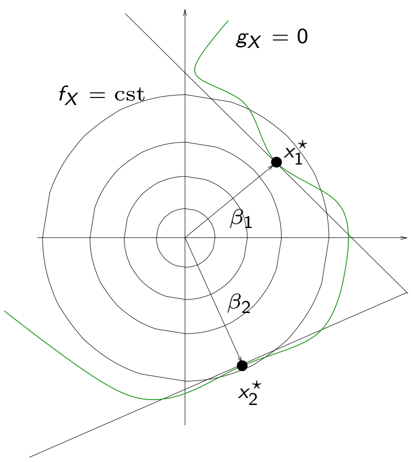
$$\alpha > \widehat{\alpha}_{\mathrm{FORM}} > \widehat{\alpha}_{\mathrm{SORM}}$$





## Multiple local first-order approximations: multi-FORM

- Another possibility proposed in the literature: compute several design points, that is, several local minimizers of $\|x\|$ subject to the constraint $g_X(x) = 0$




- Finding several local minimizers can be a difficult task
- Consider again the domain of failure:

$$\Gamma = \{x \in \mathbb{R}^d; \|x\| > \beta_0\}$$

and the approximation

$$\widehat{\Gamma} = \{x \in \mathbb{R}^d; \max_i |x_{[i]}| > \beta_0\}$$

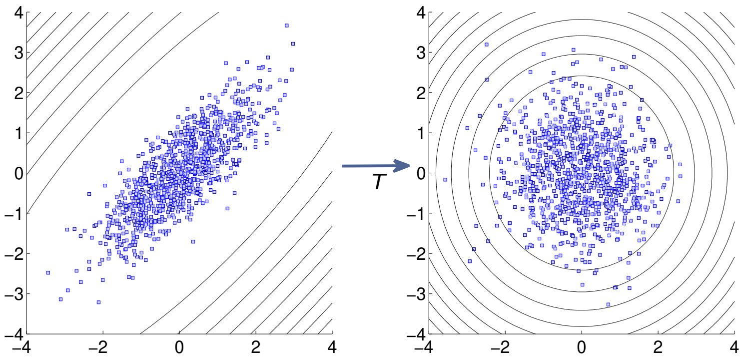
(obtained with $2d$ design points)

- Then,
$\widehat{\alpha} = \mathrm{P}\{X \in \widehat{\Gamma}\} = 1 - \prod_i (\Phi(\beta_0) - \Phi(-\beta_0)) = 1 - (1 - 2\Phi(-\beta_0))^d \approx 2d\Phi(-\beta_0)$

$\Longrightarrow$ for $d \gg 1$, $\alpha \gg \widehat{\alpha}$

# Non-standard Gaussian random factors

- Assume $X \sim \mathrm{N}(m, K) \in \mathbb{R}^d$, with $m \in \mathbb{R}^d$ and $K$ a $d \times d$ symmetric definite positive (SDP) matrix
- To apply the framework above, the idea is to search for a one-to-one whitening transformation $T : \mathbb{R}^d \to \mathbb{R}^d$ such that $U = T(X)$ is a standard Gaussian random vector
- Consider the eigendecomposition of $K$ such that $K = Q\Lambda Q^\mathsf{T}$, where $\Lambda$ is the diagonal matrix of the eigenvalues of $K$, and $Q$ is orthogonal
- Then, $T : X \mapsto \Lambda^{-1/2} Q^\mathsf{T}(X - m)$ is such that $U = T(X) \sim \mathrm{N}(0, \mathbb{I}_d)$
- Note that if $g_X$ is affine, then $g_U = g_X \circ T^{-1}$ is also affine in the standardized space








# Non-Gaussian independent random factors

- When the components of $X$ are independent but non Gaussian, the idea is to search for one-to-one transformations $T_i : \mathbb{R}^d \to \mathbb{R}^d$ such that for each $i = 1, \ldots, d$, $U_{[i]} = T_i(X_{[i]})$ is a standard Gaussian random vector
- Assume that the cdf $F_i$ of $X_{[i]}$ is continuous and strictly increasing, then

$$U_{[i]} = \Phi^{-1}(F_i(X_{[i]})) \sim \mathrm{N}(0, 1),$$

  where $\Phi^{-1}$ stands for the inverse (reciprocal) function of $\Phi$
  Indeed,

$$\begin{aligned} \mathrm{P}\{U_{[i]} \le u\} &= \mathrm{P}\{\Phi^{-1}(F_i(X_{[i]})) \le u\} \\ &= \mathrm{P}\{X_{[i]} \le F_i^{-1}(\Phi(u))\} \\ &= \Phi(u) \end{aligned}$$

- Thus, if the components of $X$ are independent but non Gaussian, the random vector

$$U = T(X) = (\Phi^{-1} \circ F_i(X_{[1]}), \ldots, \Phi^{-1} \circ F_d(X_{[d]})) \sim \mathrm{N}(0, \mathbb{I}_d)$$

## Non-Gaussian non-independent random factors

- As above, the idea is to transform the random vector of factors
- This case will be examined separately: in fact it raises two difficult issues:
  - how to specify the distribution of a random vector in the case of non-independent components?
  - what is the influence of the input distribution on the probability of failure?





## FORM and related methods: summing up

- Very popular methods in the domain of structural reliability
- FORM is easy to understand and to implement
- When a first-order approximation is relevant, a good approximation of the probability of failure can be found with a small number of function evaluations
- A major pitfall: the approximation of the probability of failure can be significantly wrong

## A personal perspective on the geometrical approximation approaches

- If $g_X$ is not expensive to evaluate:
  - Monte Carlo should be preferred over any geometrical approximation
- If $g_X$ is expensive:
  - a simple MC approach cannot be used
  - what can be the use of a geometrical approximation?
- It can be very wrong to approximate the limit state with a geometrical shape
  - can only be justified when it is known in advance that a given geometrical approximation is correct
- SORM can be though as a correction of FORM, but from a mathematical perspective, it is not $\rightarrow$ using SORM over FORM can only be justified when it is known in advance that $g_X$ is almost quadratic
- A multi-FORM approach seems preferable, but using it is to admit that the shape of the limit state is unknown, which is dangerous for a geometrical approximation approach
- Fortunately, in a large number of applications, the limit-state function is almost affine, which explains why FORM remains a very popular method





## Reliability of systems

- Until now, we have implicitly considered the case of the failure of a unique component
- In a real system, a failure can happen due to the failure of just one of its (possibly many) components
- The designer can also choose to have redundancy on critical components; in this case the failure of the system happens when all redundant components fail
- To deal with these issues, the domain of structural reliability generally introduces the notions of parallel and series systems

## Reliability of systems: series systems

- Let $x$ denotes the state of a system, and let $\Gamma$ be a domain of failure
- In structural reliability, a system is called a *series system* if the occurrence of one single failure event brings a failure on the whole system
- The classical example is that of a chain whose failure is related to any of its links
- In other words, it means that we can write

$$\Gamma = \bigcup_{i=1}^{I} \Gamma^{(i)}$$

- Assume that each domain of failure $\Gamma^{(i)}$ is characterized by a limit-state function $g_X^{(i)}$, so that the failure of the $i^{\text{th}}$ component corresponds to the event $\{g_X^{(i)}(X) < 0\}$
- Then the failure event for the whole system can be characterized by the limit-state function

$$g_X : x \mapsto g_X^{(1)}(x) \wedge \cdots \wedge g_X^{(I)}(x)$$

- Conclusion: the case of series systems can be dealt with using the framework we have exposed previously





## Reliability of systems: parallel systems

- In structural reliability, a system is called a *parallel system* if the failure of all events is necessary for the failure of the whole system
- A parallel system is a principle a redundancy
- Using the notations above, the domain of failure for a parallel system corresponds to

$$\Gamma = \bigcap_{i} \Gamma_i$$

- Again, the failure event for the whole system can be characterized by a single limit-state function

$$g_X : x \mapsto g_X^{(1)}(x) \vee \cdots \vee g_X^{(I)}(x)$$

- Conclusion: parallel systems can also be dealt with using the framework we have exposed previously

## Some references

- Ditlevsen O. and Madsen H.O. (1996), Structural reliability methods, Wiley
- Lemaire M., Chateauneuf A., and Mitteau J.C., (2009) Structural Reliability, Wiley





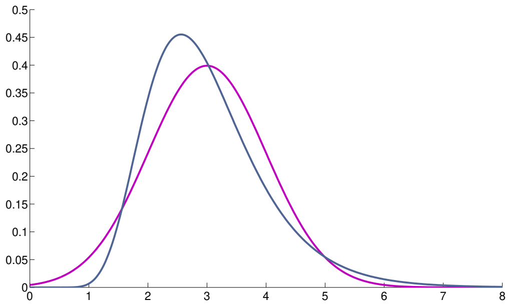
## Elicitation of subjective probability distributions

- A very difficult issue: to choose a probability distribution for modeling the uncertain factors of a system
- The process of choosing a distribution for the factors is called elicitation[1] in the literature of decision analysis
- Elicitation is particularly difficult when doing risk analysis about new and untried technologies, for which little data are available
- Very often, risk analysis relies on expert judgment
- Elicitation of subjective probability distributions is often subject to a number of serious biases, such as overconfidence in the ability to quantify uncertainty

[1] from Latin elicere: draw forth, bring out

# Elicitation of subjective probability distributions

- Assume that we are given an approximation of the mean and the standard deviation of a random variable
- An experiment:




- $X \sim \mathrm{N}(\mu, \sigma^2)$
- $Y \sim \log\mathrm{N}(m, s^2)$, with $s^2 = \log(\frac{\sigma^2}{\mu^2} + 1)$ and $m = \log \mu - s^2/2$.
- Then, $\mathsf{E}[X] = \mathsf{E}[Y] = \mu$ and $\mathsf{var}[X] = \mathsf{var}[Y] = \sigma^2$.
- Assume $\mu = 3$ and $\sigma^2 = 1$. Then, $\mathsf{P}(X > 7) = 3.1 \cdot 10^{-5}$ and $\mathsf{P}(Y > 7) = 2.8 \cdot 10^{-3}$

- For risk analysis, "knowing" only the mean and the standard deviation of a random variable is a very poor information
- In fact, in many applications, the occurrence of a failure is likely to be related to an extreme event in the factor space

  $\rightarrow$ it is probably more important to characterize the tail behavior of the factors than the central behavior (EVT can help)





# Elicitation of subjective probability distributions

- Elicitation gets even more complicated if we want to introduce dependence information between random variables. In fact, the question is: "how to measure dependence?"
- In probability theory and statistics, the dependence between random variables is described by the concept of copula (from Latin co-apere "join together")

## Copulas

- Consider a random vector $X = (X_1, \ldots, X_d) \in \mathbb{R}^d$. The dependence between the component random variables $X_1, \ldots, X_d$ is completely described by the joint cdf

$$F(x_1, \ldots, x_d) = \mathrm{P}\{X_1 \leq x_1, \ldots, X_d \leq x_d\}$$

- For simplicity, assume that the components $X_i$, $i = 1, \ldots, d$, have continuous, strictly increasing, marginal cdfs $F_i$
- The concept of copula: separate $F$ into a part that describes the dependence structure and parts which describe the marginal behavior only
- Transform $X$ component-wise to obtain standard-uniform marginal distributions $\mathrm{U}([0, 1])$

$$\begin{array}{rlcl} T: & \mathrm{dom}\, F & \to & [0,1]^d \\ & (x_1, \ldots, x_d) & \mapsto & (F_1(x_1), \ldots, F_d(x_d)) \end{array}$$

- The joint cdf of $U = T(X)$ is called the copula of the random vector $X$
- It follows that for $x \in \mathrm{dom}\, F$

$$F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$$

and for $u \in [0, 1]^d$

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



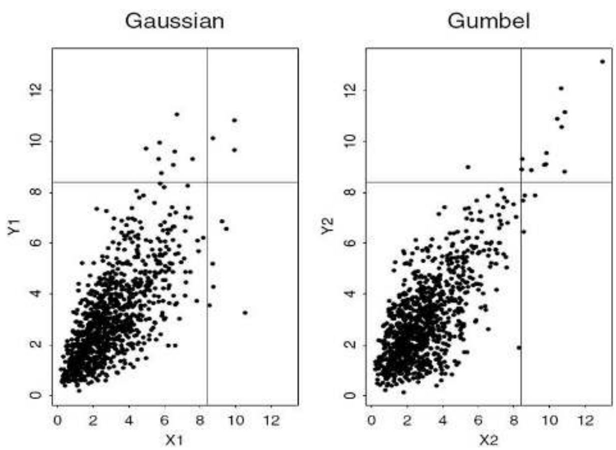


## Copulas

- Some copulas/dependence structures:
  - The independent copula: $C_{\mathrm{ind}}(u) = u_1 u_2 \cdots u_d$
  - The Gaussian copula

$$C_{\mathrm{g},R}(u_1, \ldots, u_d) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_d)} (2\pi)^{-d/2} (\det R)^{-1/2} \exp\Big(-\frac{1}{2} u^{\mathsf{T}} R^{-1} u\Big) \mathrm{d}u_1 \cdots \mathrm{d}u_d$$

  - The bivariate Gumbel copula $C_{\mathrm{Gu},\beta}(u, v) = \exp\left[-\left\{(-\log u)^{1/\beta} + (-\log v)^{1/\beta}\right\}^{\beta}\right]$




from Embrechts et al. (1999)

Realizations from two distributions with identical $\mathrm{Gamma}(3, 1)$ marginal distributions and identical correlation $\rho = 0.7$, but different dependence structures.

## Linear correlation (or Pearson's correlation)

- In some applications, choosing independent random factors is irrelevant
- However, choosing a dependence structure can be a difficult task because information may be scarce (e.g., no data available)
- Very often, the dependence structure of a random vector is summarized through (linear) correlations between components

Recall that for $(X, Y) \in \mathbb{R}^2$, a second-order random vector, the linear correlation coefficient is

$$\rho(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)\,\mathrm{var}(Y)}}$$

If $|\rho(X, Y)| = 1$, then there is a perfect linear dependence between $X$ and $Y$, i.e., $Y = aX + b$, for some $a, b \in \mathbb{R}$





## Nataf transformation

- The Nataf transformation $T_{\mathrm{Nataf}}$ is a one-to-one function which maps a random vector $X$ with a Gaussian copula to a random vector $U$ with standard Gaussian distribution
- Makes it possible to apply FORM for non-Gaussian non-independent random vectors
- Conversely, the Nataf transformation makes it possible to define implicitly the distribution $F$ of a random vector $X = (X_1, \ldots, X_d)$ with
  1. Gaussian copula
  2. prescribed continuous, strictly increasing, marginals $F_1, \ldots, F_d$
  3. prescribed correlation matrix $R_X = (\rho(X_i, X_j))_{i,j}$
- Considering a Gaussian copula for the dependence structure of $X$ should not be considered as a canonical choice $\rightarrow$ a comprehensive risk analysis procedure should assess the consequences of this particular choice

# Inverse Nataf transformation

- Let $U \sim \mathrm{N}(0, \mathbb{I}_d)$
- Given continuous strictly increasing marginals cdfs $F_1, \ldots, F_d$, and a $d \times d$ correlation matrix $R_X$, the inverse Nataf transformation is defined as
  $X = T_{\mathrm{Nataf}}^{-1}(U) = T_3 \circ T_2 \circ T_1(U)$, with

$$
\begin{array}{llll}
T_1 : & \mathbb{R}^d & \to & \mathbb{R}^d \\
 & u = (u_1, \ldots, u_d) & \mapsto & Cu \\
 & & & \\
T_2 : & \mathbb{R}^d & \to & [0,1]^d \\
 & v = (v_1, \ldots, v_d) & \mapsto & (\Phi(v_1), \ldots, \Phi(v_d)) \\
 & & & \\
T_3 : & [0,1]^d & \to & \mathbb{R}^d \\
 & w = (w_1, \ldots, w_d) & \mapsto & (F_1^{-1}(w_1), \ldots, F_d^{-1}(w_d))
\end{array}
$$

  where $C$ is a $d \times d$ matrix which is computed in such a way that $X$ has correlation matrix $R_X$
- Note that $V = T_1(U)$ is a Gaussian vector. Thus, it has a Gaussian copula. Since $T_2$ and $T_3$ are component-wise monotonic transformations, $T_2$ and $T_3$ are copula-invariant
- Note that the correlation matrix of $V = T_1(U)$ is $R_V = CC^\mathsf{T}$. In general, $R_V \neq R_X$.
- Note also that it is not always possible to prescribe any correlation coefficient (depending on the choice of the marginals)





# Dependence measures

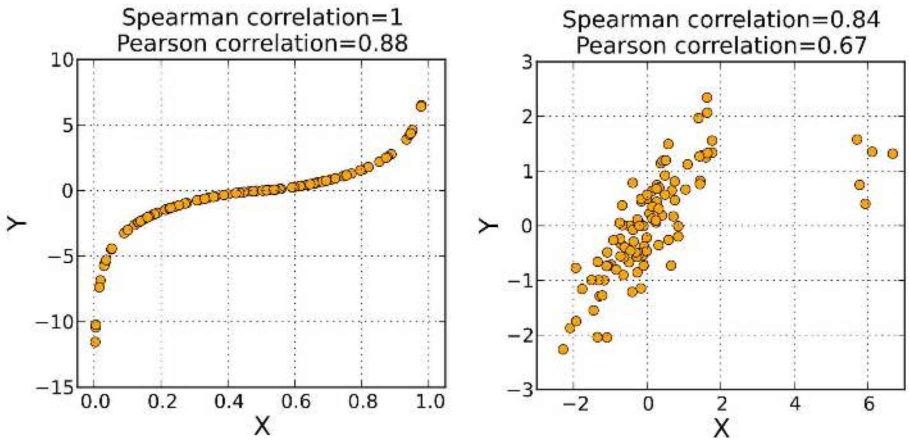
- As mentioned above, considering a Gaussian copula for the dependence structure of $X$ should not be considered as a canonical choice
- Moreover, measuring dependence based on correlation coefficients can be misleading $\to$ linear dependence should not be taken as a canonical dependence measure

Realizations from seven bi-variate distributions with zero correlation

- Other measures of the dependence?

## Dependence measures

- Non-linear dependence measures have been proposed in the literature
- In particular, rank correlation coefficients, such as Spearman's rank correlation coefficient $\rho$, and Kendall's rank correlation coefficient $\tau$ measure the extent to which two variables increase or decrease simultaneously
- For instance, Spearman's $\rho$ is defined as the Pearson's correlation coefficient between the ranked variables




from Wikipedia: article on Spearman's rank correlation coefficient





## Tail dependence

- In the context of risk analysis, it might be relevant to study the dependence between extreme values
- Let $X = (X_1, X_2)$ be a vector of continuous random variables with marginals $F_1$ and $F_2$. The coefficient of upper tail dependence of $(X_1, X_2)$ is

$$\lambda_U = \lim_{u \to 1} \mathrm{P}\{X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u)\}$$

  provided that the limit exists. If $\lambda_U$ is well-defined, $\lambda_U \in [0, 1]$
- The value of $\lambda_U$ is a property of the copula of $X$ only
- Examples:
  - Consider the Gaussian bi-variate copula, with correlation $\rho < 1$. Then, $\lambda_U = 0$; that is, a Gaussian copula with $\rho < 1$ does not have tail dependence
  - The Gumbel bi-variate copula, with parameter $\beta > 1$, has tail dependence $\lambda_U = 2 - 2^{1/\beta}$
- Choosing a copula with tail-dependence over the Gaussian copula can modify the probability of failure by several orders of magnitude—see, for instance, Dutfoy and Lebrun, Congrès Français de Mécanique (2007)

# Some references


- Embrechts P., McNeil A., Strautmann D. (1999), Correlation and dependence in risk management: properties and pitfalls
- Embrechts P., Lindskog F., McNeil A. (2001), Modelling Dependence with Copulas and Applications to Risk Management
- Lebrun R. and Dutfoy A. (2009), A generalization of the Nataf transformation to distributions with elliptical copula