Copulas for Uncertainty Analysis

Antonio Possolo

National Institute of Standards and Technology
U.S. Department of Commerce

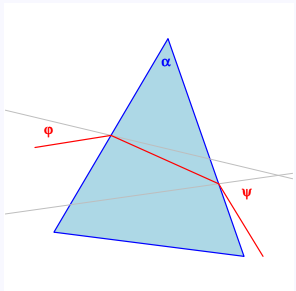
Jun 21st, 2010

Outline

- GUM
  - Refractive index
  - Shortcomings
- GUM SUPPLEMENT 1
  - Change-of-Variables Formula
  - Monte Carlo Method
- COPULAS
  - Definition & Illustrations
  - GUM Example H.2

Refractive Index

- Apex angle α, refractive index n, immersed in medium with refractive index m
- Light enters prism at angle φ, traverses prism's body, and exits at angle ψ
- As prism rotates about light's entrance point, $\delta = 2(\varphi + \psi - \alpha)$ decreases, reaches minimum δ_M, then increases



$$n = m\frac{\sin((\alpha + \delta_M)/2)}{\sin(\alpha/2)}$$

Refractive Index — Partial Derivatives

$$\dot{f}_m(m, \varphi, \psi, \alpha) = \frac{\sin(\varphi + \psi - \alpha/2)}{\sin(\alpha/2)}$$

$$\dot{f}_\varphi(m, \varphi, \psi, \alpha) = m\,\frac{\cos(\varphi + \psi - \alpha/2)}{\sin(\alpha/2)}$$

$$\dot{f}_\psi(m, \varphi, \psi, \alpha) = m\,\frac{\cos(\varphi + \psi - \alpha/2)}{\sin(\alpha/2)}$$

$$\dot{f}_\alpha(m, \varphi, \psi, \alpha) = -\frac{m}{2}\left[\frac{\cos(\varphi + \psi - \frac{\alpha}{2})}{\sin(\frac{\alpha}{2})} + \frac{\sin(\varphi + \psi - \frac{\alpha}{2})}{\sin(\frac{\alpha}{2})\tan(\frac{\alpha}{2})}\right]$$

Streamlining Computation

- Computation of standard uncertainty using GUM's Taylor approximation is tedious and error-prone, especially when measurement equation is non-linear or involves special functions
- Employ software capable of computing derivatives
  - Analytically (for example, Maple)
  - Numerically (for example, R)
    - R is freely available, and its source code is open: `http://www.r-project.org/`
    - metRology package (LGC / NIST)

GUM Approximation — Shortcomings

- If some first order partial derivatives of f are zero at values of input quantities, then GUM's approximation may be faulty
  - Radiant power $W = \kappa \cos(A)$
- GUM's approximation may be poor when f is markedly non-linear in neighborhood of values of input quantities
  - Need to study f's curvature is extra burden
  - If curvature is appreciable and influential, need higher-order approximation
    - However, examples can be constructed where first-order approximation is better than second-order approximation — Wang & Iyer (2005)

GUM Approximation — More Shortcomings

- Expanded uncertainties and coverage intervals involve coverage factor whose value depends on generally unverifiable assumption that output quantity has Gaussian or Student's t distribution
- Even when $Y \sim t_\nu$, Welch-Satterthwaite formula often yields inappropriate value for ν

Example (GUM H.1)

- $l = l_S + d - l_S(1 - \delta\alpha\,\theta - \alpha_S\,\delta\theta)$
- Welch-Satterthwaite formula yields $\nu_{\text{eff}} = 16$
- GUM's 99 % coverage interval **17 % longer** than it needs to be

GUM Supplement 1 — Problem

- Given
  - Measurement equation $Y = f(X_1, \ldots, X_n)$
  - Joint probability distribution of $X_1, \ldots, X_n$ (with density φ)
- Determine Y's probability distribution
  - Derive Y's density analytically
    — *Change-of-variables formula*
  - Sample Y's distribution

Change of Variables Formula

- Transformation $\tau = (f, \tau_2, \ldots, \tau_n)$
  - One-to-one, continuously differentiable
  - Inverse has non-vanishing Jacobian

$$\begin{aligned} Y &= f(X_1, \ldots, X_n) \\ Z_2 &= \tau_2(X_1, \ldots, X_n) \\ &\ldots \\ Z_n &= \tau_n(X_1, \ldots, X_n) \end{aligned}$$

- Y has probability density ψ

$$\psi(y) = \int_{\mathcal{Z}_2} \ldots \int_{\mathcal{Z}_n} \varphi(\tau^{-1}(y, z_2, \ldots, z_n)) \, |J_{\tau^{-1}}(y, z_2, \ldots, z_n)| \, \mathrm{d}z_2 \ldots \mathrm{d}z_n$$

GUM Supplement 1 — Step 1

PROPAGATING DISTRIBUTIONS (MONTE CARLO)

- STEP 1: Select probability distribution to describe state of knowledge about input quantities
  - If $X_1, \ldots, X_n$ can be regarded as independent random variables, then assign probability distributions to each one of them separately
  - Otherwise, assign multivariate distribution to all of them together **. . . using a copula**
  - In either case, mean and standard deviation of each X_j's distribution should be x_j and $u(x_j)$

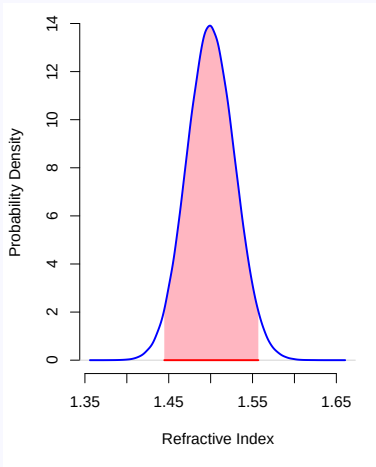
GUM Supplement 1 — Steps 2–4

PROPAGATING DISTRIBUTIONS (MONTE CARLO)

- STEP 2: Choose suitably large number K for size of sample to generate from probability distribution of output quantity
- STEP 3: Simulate K vectors of values of input quantities, and compute
  $y_k = f(x_{k1}, \ldots, x_{kn})$ for $k = 1, \ldots, K$
- STEP 4: Assign to $u(y)$ (standard uncertainty of output quantity) the value of the standard deviation of $y_1, \ldots, y_K$

GUM Supplement 1 — Refractive Index (CI)

- Use sample $\{y_1, \ldots, y_K\}$ to estimate probability density of prism's refractive index — most complete characterization of uncertainty
- Shaded region includes 95 % of total area under curve: footprint on horizontal axis is a *coverage interval* for measurand with 95% confidence



GUM Example H.2

INPUT QUANTITIES

- Amplitude V of sinusoidally-alternating potential difference across electrical circuit's terminals
- Amplitude I of alternating current
- Phase-shift angle ϕ of alternating potential difference relative to alternating current
- V, I, and ϕ are correlated

GUM Example H.2

OUTPUT QUANTITIES

- Resistance $R = \frac{V}{I} \cos \phi$
- Reactance $X = \frac{V}{I} \sin \phi$
- Impedance's magnitude $Z = \frac{V}{I}$

Problem & Solutions

PROBLEM

- Given probability distributions for input quantities, and correlations between them, how does one manufacture a *joint* probability distribution consistent with those margins and correlations?

SOLUTIONS

- There are infinitely many joint distributions that are consistent with given marginal distributions and correlations
- *Copulas* join univariate probability distributions into multivariate distributions and impose dependence structure

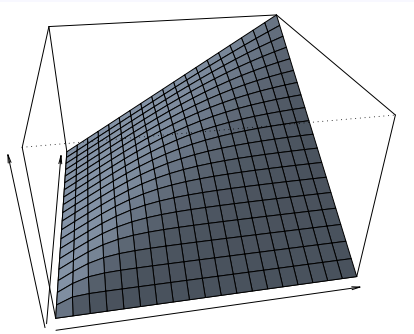
Copulas are not Cupolas

Cape Cod Weathervane Company
Hand Crafted Copper Weathervanes & more

Manufacturer mentioned solely to acknowledge image source, with no implied recommendation or endorsement by NIST that cupola portrayed is the best available for any particular purpose

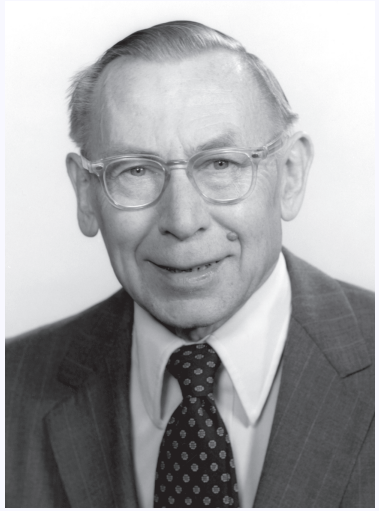
Copula — Definition

- A copula is the cumulative distribution function of a multivariate distribution on the unit hypercube all of whose margins are uniform



- Clayton copula inducing Kendall's $\tau = 0.6$

Wassily Hoeffding, 1914–1991



Sklar's (1959) Theorem

- If H is CDF of multivariate distribution whose margins have CDFs $F_1, \ldots, F_n$, then there exists copula C such that

$$H(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n))$$

- If $F_1, \ldots, F_n$ are continuous, then C is unique
- $C(p_1, \ldots, p_n) = H(F_1^{-1}(p_1), \ldots, F_n^{-1}(p_n))$

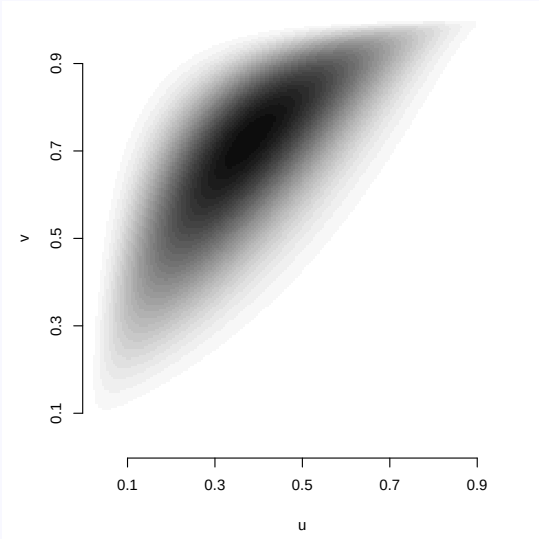
Copula — Example: Definition

Example (Gaussian Copula on Beta Margins)

- $U \sim \text{BETA}(3, 5)$ has CDF F
- $V \sim \text{BETA}(4, 2)$ has CDF G
- $\rho = \text{cor}(U, V) = 0.7$
- Copula C such that

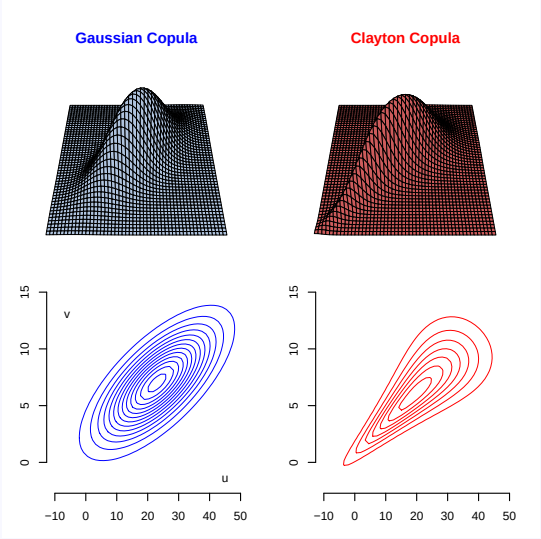
$$C(p, q) = \Phi_\rho(F^{-1}(p), G^{-1}(q)) \text{ for } 0 \leqslant p, q \leqslant 1$$

Copula — Example: Joint Density



Influential Copulas

- Joint distributions with same Gaussian margins and same correlation



GUM Example H.2

OUTPUT QUANTITIES

- Resistance $R = \frac{V}{I} \cos \phi$
- Reactance $X = \frac{V}{I} \sin \phi$
- Impedance's magnitude $Z = \frac{V}{I}$

GUM Example H.2 — Data

V (V)	I (mA)	ϕ (rad)
5.007	19.663	1.045 6
4.994	19.639	1.043 8
5.005	19.640	1.046 8
4.990	19.685	1.042 8
4.999	19.678	1.043 3

GUM Example H.2 — Evaluations

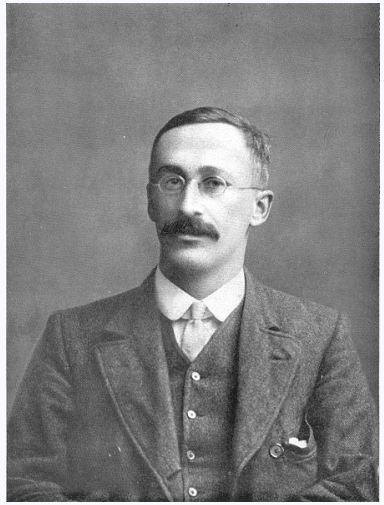
- Input quantity values estimated by averages $\overline{V}$, $\overline{I}$, and $\overline{\phi}$, of sets of five indications
- Uncertainties and correlations of input quantities assessed by Type A evaluations
  - $u(\overline{V}) = \mathrm{SD}(5.007, 4.994, 5.005, 4.990, 4.999)/\sqrt{5}$
  - Similarly for $u(\overline{I})$ and $u(\overline{\phi})$
  - $\mathrm{cor}(V, I)$, $\mathrm{cor}(V, \phi)$, and $\mathrm{cor}(I, \phi)$ estimated by correlations between paired sets of five indications

GUM Example H.2 — GUM Supplement 1

CHALLENGES

- Quality of estimates of $u(V)$, $u(I)$, $u(\phi)$, $\mathrm{cor}(V, I)$, $\mathrm{cor}(V, \phi)$, and $\mathrm{cor}(I, \phi)$ limited by very small number of indications they are based on
- Marginal distributions must be chosen for V, I and ϕ, and then linked using a copula that reproduces correlations between them

William Gosset, 1876–1937



GUM Example H.2 — GUM Supplement 1

SOLUTION

- Employ Student t_4 distributions for

$$\frac{\overline{V} - \mu_V}{u(\overline{V})}, \frac{\overline{I} - \mu_I}{u(\overline{I})}, \text{ and } \frac{\overline{\phi} - \mu_\phi}{u(\overline{\phi})}$$

- *Tune* Gaussian copula using *correlation supplicant*
- Apply copula
  - To sample correlation matrix
  - To use sampled correlation matrix in turn to sample output quantities

Correlation Supplicant

- Copula parameter θ determines dependence structure, in particular correlation matrix $\rho(\theta)$

Example

- STUDENT: θ comprises ν and Σ
- CLAYTON: $C(p,q) = (p^{-\theta} + q^{-\theta} - 1)^{-1/\theta}$, $\theta \geqslant -1$

- *Correlation Supplicant* selects θ that minimizes difference between ρ and $\rho(\theta)$
  - Minimization with respect to θ done under constraint that $\rho(\theta)$ must be *bona fide* correlation matrix

GUM Example H.2 — GUM Supplement 1

RESULTS

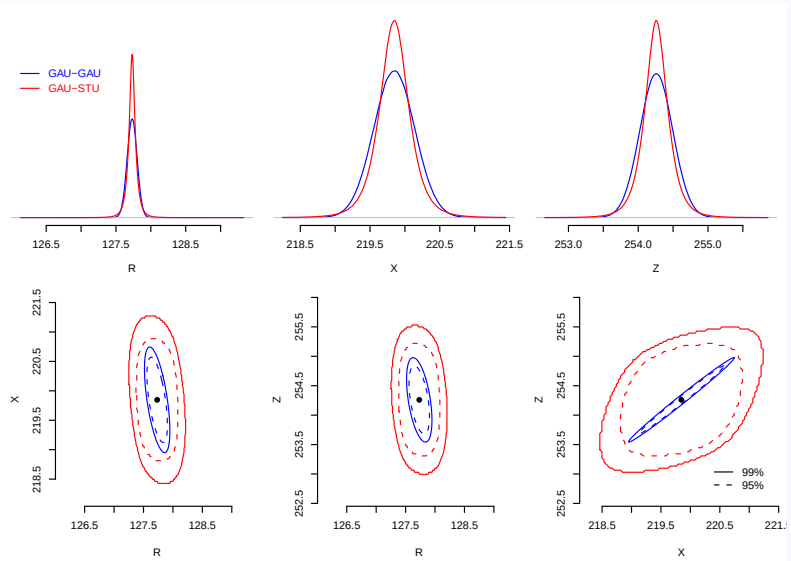
GAUSSIAN COPULA WITH GAUSSIAN MARGINS

	R (Ω)	X (Ω)	Z (Ω)
AVE	127.7	219.8	254.3
u	0.0711	0.296	0.236
$U_{99\%}$	0.18	0.76	0.61

GAUSSIAN COPULA WITH STUDENT MARGINS

	R (Ω)	X (Ω)	Z (Ω)
AVE	127.7	219.8	254.3
u	0.0828	0.289	0.232
$U_{99\%}$	0.29	0.93	0.75

GUM Example H.2 — Results



GUM Example H.2 — Limitation

- All usual copulas fail to capture fact that joint distribution of (R, X, Z) is degenerate

$$Z^2 = R^2 + X^2$$

  - However, this is largely irrelevant in this case because relative standard uncertainties are very small — all less than 0.1 %

Conclusions & Reference

- When input quantities are correlated, and their joint distribution is not determined otherwise, copula must be used to apply Monte Carlo Method of GUM Supplement 1
- Many different copulas can reproduce given correlations, yet lead to possibly very different uncertainty evaluations for output quantity
- A. Possolo (2010) Copulas for uncertainty analysis *Metrologia*, 47 (3): 262–271