Non-parametric and semi-parametric estimation of Archimedean copula parameter

Application to real time series - MOPI daily observations. Standard error estimation included.

Initial settings

```
<< Statistics`MultiDescriptiveStatistics`
<< Statistics`StatisticsPlots`
<< Statistics`NonlinearFit`
<< Graphics`Graphics3D`
```

■ system settings

```
SetDirectory["d:\\Math\\Analyza CR\\copula"];
SetOptions[ListPlot, PlotJoined → True, PlotRange → All, DisplayFunction → Identity];
SetOptions[{Histogram, Plot, QuantilePlot, ContourPlot, Plot3D}, PlotRange → All, DisplayFunction → Identity];
Off[General::spell1];
```

■ commonly-used functions

fShow causes visibility of graphic objects, that are set to DisplayFunction→Identity. fNShow, on the contrary, sets this option to prevent visibility. These functions come useful when grouping several graphic objects.

```
fShow[plot__, options___] := Show[plot, DisplayFunction :> $DisplayFunction, options];
```

■ data

Setting `y=-y` causes positive dependence between "x" and "y". Next, the couples are being ordered according to increasing values of "y" for later easy handling with extremals if needed.

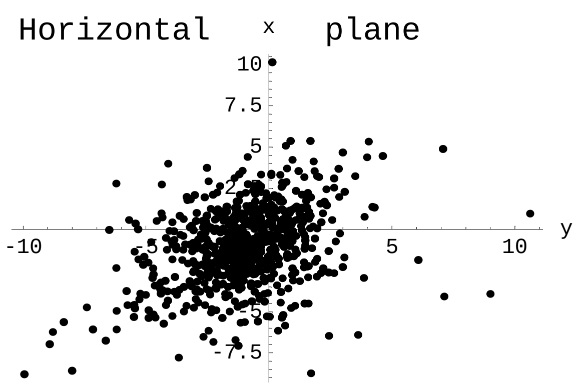
```
{x, y} = Transpose[ReadList["data\\mopi_nt.dat", Number, RecordLists → True]];
{y = -y};
```

```
Ordering[y];
x = Part[x, %];
y = Part[y, %%];
xy = Transpose[{x, y}];
n = Length[xy]
```

```
728
```

First look

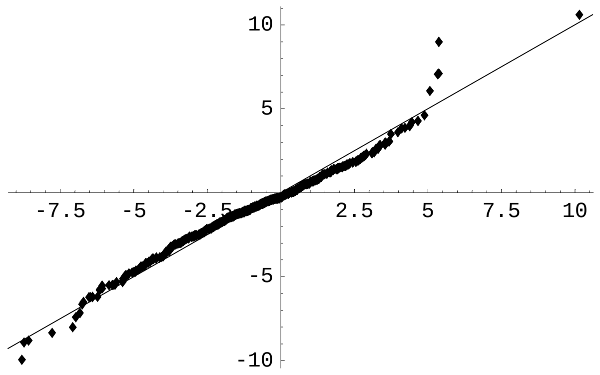
```
ListPlot[Transpose[{y, x}], PlotLabel → StyleForm["Horizontal      plane      ", FontSize → 15],
   AxesLabel → {"y", "x"}, PlotJoined → False, PlotStyle → {PointSize[0.015]}] // fShow;
```



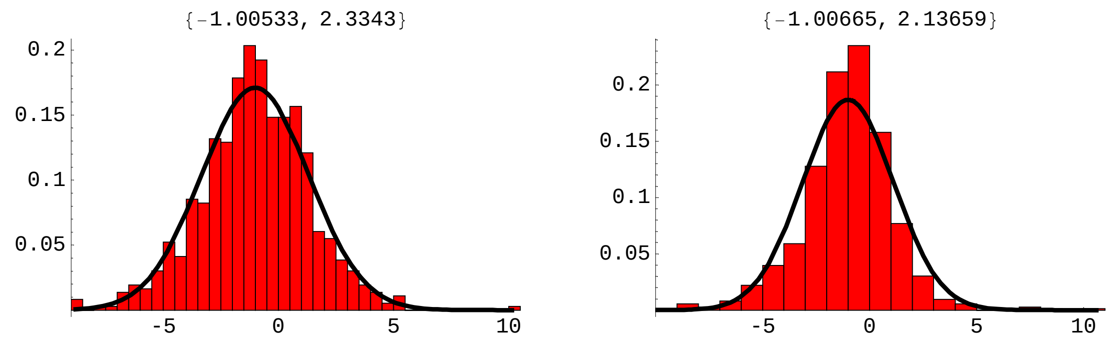

```
{corr = Correlation[x, y], ρ = SpearmanRankCorrelation[x, y], τ = KendallRankCorrelation[x, y]} // N
```

```
{0.367013, 0.331438, 0.234636}
```

```
QuantilePlot[x, y] // fShow;
```




```
Module[{μ, σ}, GraphicsArray[ Show[
       Histogram[#, HistogramScale → 1],
       Plot[PDF[NormalDistribution[μ = Mean[#], σ = StandardDeviation[#]], x],
        {x, Min[#], Max[#]}, PlotStyle → {Thickness[0.01]}],
       PlotLabel → StyleForm[{μ , σ}]
      ] & /@ {x, y}]] // fShow;
```



Nonparametric estimation of copula parameter

Procedure by Genest & Rivest (1993). Described in Frees & Valdez (1998) and Abid & Naifar (2005) }

■ Nonparametric estimate Kn

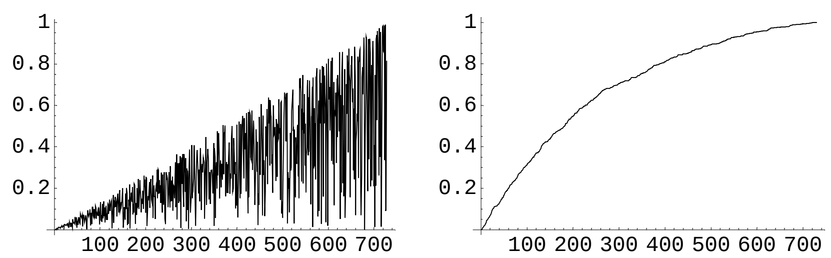
unobserved variable Z=H(x,y)

```
fZ[i_] := Sum[If[x⟦j⟧ < x⟦i⟧ && y⟦j⟧ < y⟦i⟧, 1., 0], {j, n}] / (n - 1)
Z = Table[fZ[i], {i, n}];
```

distribution function of Z

```
fKn[z_] := Sum[If[Z⟦i⟧ ≤ z, 1, 0], {i, n}] / n
Kn = Table[fKn[z], {z, 0, 1, 1 / n}];
```

```
GraphicsArray[{ListPlot[Z], ListPlot[Kn]}] // fShow;
```



■ Parametric estimate Kϕ

$$K_\phi(z) = z - \frac{\phi(z)}{\phi'(z)} \quad \text{using relation} \quad \tau = 1 + 4\int_0^1 \frac{\phi(t)}{\phi'(t)}\, dt$$

Procedure: $\tau \to \Theta \to \phi \to K_\phi$.

Independence copula (no parameter)

```
fKi[t_] := If[t ≠ 0, t (1 - Log[t]), 0];
Ki = Table[fKi[z], {z, 0, 1, 1 / n}];
```

Gumbel copula

```
θg = t /. NSolve[τ == (t - 1) / t] ⟦1⟧;
fKg[t_] := If[t ≠ 0, t - t Log[t] / θg, 0];
```

Clayton copula

```
θc = t /. NSolve[τ == t / (t + 2)] ⟦1⟧;
fKc[t_] := t - (t^(θc+1) - t) / θc;
```

Frank copula

```
fD1[x_] := 1 / x Integrate[t / (Exp[t] - 1), {t, 0, x}];
θf = Re[t /. FindRoot[τ == 1 + 4 / t (fD1[t] - 1), {t, 2.2}]];
fKf[t_] := If[t == 0, 0, t - Log[(Exp[-θf t] - 1) / (Exp[-θf] - 1)] * (Exp[θf t] - 1) / θf]
```

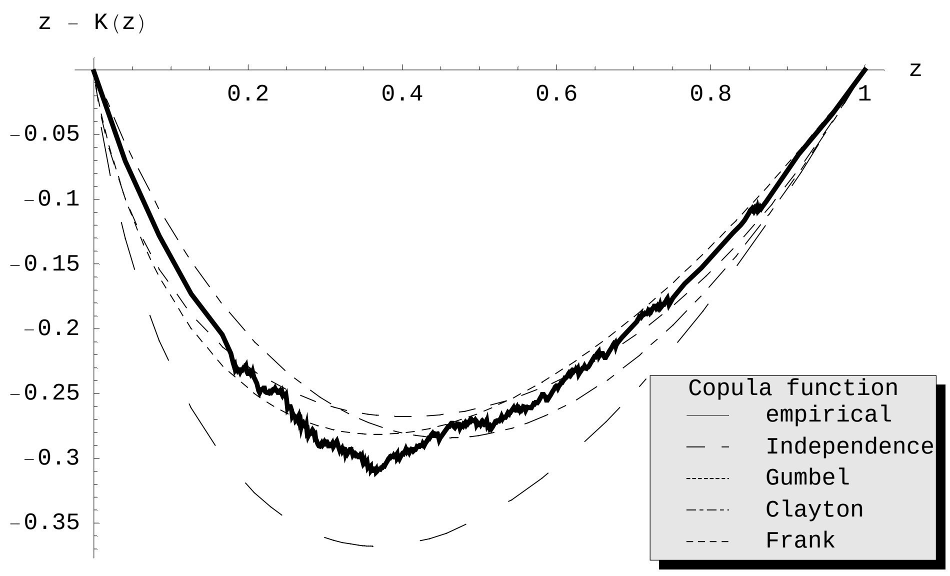
```
{Ki, Kg, Kc, Kf} = Table[#[z], {z, 0, 1, 1 / n}] & /@ {fKi, fKg, fKc, fKf};
{θg, θc, θf}
```

```
{1.30657, 0.613135, 2.21169}
```

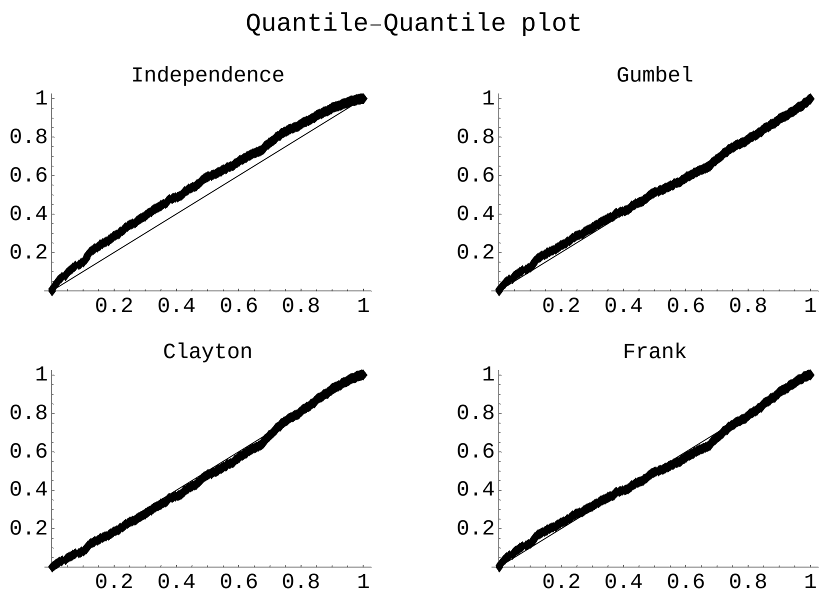
■ Comparing K's

▫ Graphically

```
gK = Plot[{z - fKn[z], z - fKi[z], z - fKg[z], z - fKc[z], z - fKf[z]}, {z, 0, 1},
    PlotStyle -> {Thickness[0.006], Dashing[{.06}], Dashing[{.01}], Dashing[{.03, 0.015, 0.01, 0.015}], Dashing[{.02}]},
    TextStyle → {FontSize → 11}, AxesLabel → {"z", "z - K(z)"}, LegendLabel -> "Copula function",
    PlotLegend -> {"empirical", "Independence", "Gumbel", "Clayton", "Frank"},
    LegendSize → {0.62, 0.4}, LegendPosition -> {.4, -.6}, LegendTextSpace → 2.0, LegendLabelSpace → 0.8,
    LegendOrientation → Vertical, LegendBackground -> GrayLevel[.9], LegendShadow -> {.02, -.02}] // fShow;
```




```
gQQ = GraphicsArray[{
     {QuantilePlot[Kn, Ki, PlotLabel → "Independence"], QuantilePlot[Kn, Kg, PlotLabel → "Gumbel"]},
     {QuantilePlot[Kn, Kc, PlotLabel → "Clayton"], QuantilePlot[Kn, Kf, PlotLabel → "Frank"]}},
    PlotLabel -> StyleForm["Quantile-Quantile plot", FontSize → 12]] // fShow;
```



▫ Numerically

L2 norm distance $\sqrt{(K_\phi - K_n)^2}$

```
Norm[# - Kn] & /@ {Ki, Kg, Kc, Kf} // N
```

```
{1.72375, 0.445571, 0.54291, 0.49252}
```

Semi-parametric estimation of copula parameter

Distribution and copula functions

Empirical marginal distribution function (rescaled)

CDF(x) = P(X≤x)

$$\texttt{fCDFe[X\_, x\_] := } \frac{1}{n+1} \texttt{ Sum[If[X[[i]]} \le \texttt{x, 1, 0], \{i, 1, n\}]}$$

empirical "probability integral transform" vectors of original univariate data

```
xT = Table[fCDFe[x, x[[i]]], {i, 1, n}];
yT = Table[fCDFe[y, y[[i]]], {i, 1, n}];
```

ranks

```
xR = Ordering[x] / (n + 1);
yR = Ordering[y] / (n + 1);
```

Empirical copula

Deheuvels (1979), $C_n[u, v] = \frac{1}{n} \sum_{i=1}^{n} 1_{(F_n[x_i] \le u)}\, 1_{(F_n[y_i] \le v)}$ or $C_n[u, v] = \frac{1}{n} \sum_{i=1}^{n} 1_{(xR_i \le u)}\, 1_{(yR_i \le v)}$

empirical copula function (illustrative but slow)

$$\texttt{fCe[u\_, v\_] := } \frac{1}{n} \texttt{ Sum[If[xT[[i]]} \le \texttt{u} \wedge \texttt{yT[[i]]} \le \texttt{v, 1, 0], \{i, 1, n\}]}$$

empirical copula; regular grid

```
xRM = Table[If[xR[[ii]] ≤ u, 1, 0], {u, 0, 1, 1 / n}, {ii, 1, n}];
yRM = Table[If[yR[[jj]] ≤ v, 1, 0], {v, 0, 1, 1 / n}, {jj, 1, n}];
Timing[Ce = Table[1/n Sum[xRM[[i, k]] * yRM[[j, k]], {k, 1, n}], {i, 1, n + 1}, {j, 1, n + 1}];]
```

```
{1605.57 Second, Null}
```

storing or loading large data

```
Ce >> Cempir.txt
```

```
Ce = (<< "Cempir.txt");
```

Archimedean copula

$$C(u, v) = \phi^{-1}[\phi(u) + \phi(v)]$$

$$\texttt{fCg[u\_, v\_, }\theta\texttt{\_] = } e^{-((-\mathrm{Log}[u])^{\theta} + (-\mathrm{Log}[v])^{\theta})^{1/\theta}};$$

$$\texttt{fCc[u\_, v\_, }\theta\texttt{\_] = } (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta};$$

$$\texttt{fCf[u\_, v\_, }\theta\texttt{\_] = } -\frac{\mathrm{Log}\left[\frac{(e^{-\theta v}-1)\,(e^{-\theta u}-1)}{e^{-\theta}-1} + 1\right]}{\theta};$$

Copula density functions:

```
fci[u_, v_] = D[fCi[u, v], u, v];
fcg[u_, v_, θ_] = Simplify[D[fCg[u, v, θ], u, v]];
fcc[u_, v_, θ_] = D[fCc[u, v, θ], u, v];
fcf[u_, v_, θ_] = D[fCf[u, v, θ], u, v];
```

Maximum likelihood

pseudo log-likelihood

```
fLg[θ_] = Sum[Log[fcg[ xT[[i]], yT[[i]], θ]], {i, 1, n}];
fLc[θ_] = Sum[Log[fcc[ xT[[i]], yT[[i]], θ]], {i, 1, n}];
fLf[θ_] = Sum[Log[fcf[ xT[[i]], yT[[i]], θ]], {i, 1, n}];
```

maximizing → parameters, AIC

```
Transpose[{
   LLg = FindMaximum[fLg[θ], {θ, 1}],
   LLc = FindMaximum[fLc[θ], {θ, 0.1}, AccuracyGoal → 7],
   LLf = FindMaximum[fLf[θ], {θ, 0.1}, AccuracyGoal → 7]}];
{θ1g, θ1c, θ1f} = ((θ /. #) & /@ %[[2]])
vAIC = -2 %%[[1]] + 2
```

```
{1.30445, 0.563803, 2.3153}
```

```
{-106.235, -109.035, -90.7313}
```

comparing to empirical copula (L2-norm distance)

```
Cg = Table[fCg[i, j, θ1g], {i, 0, 1, 1 / n}, {j, 0, 1, 1 / n}];
Table[fCc[i, j, θ1c], {i, 1 / n, 1, 1 / n}, {j, 1 / n, 1, 1 / n}];
Cc = Transpose[Prepend[ Transpose[Prepend[%, Table[0, {n}]]], Table[0, {n + 1}] ]];
Cf = Table[fCf[i, j, θ1f], {i, 0, 1, 1 / n}, {j, 0, 1, 1 / n}];
```

```
Norm[Flatten[Ce - #]] & /@ {Cg, Cc, Cf}
```

```
{5.66526, 6.3704, 6.00337}
```

■ Non-linear fit

full specified copula; regular grid without borders
(if borders are added, i.e. {i,1,n+1} and {j,1,n+1}, then CeXYZ=N[CeXYZ/.{0.->10^(-15),1.→(1-10^(-15))}] to preserve stability)

```
CeXYZ = N[Flatten[Table[{(i - 1) / n, (j - 1) / n, Ce[[i, j]]}, {i, 2, n}, {j, 2, n}], 1]];
```

Gumbel

```
regg = NonlinearRegress[CeXYZ, fCg[u, v, θ], {u, v}, {θ, 1.3},
  RegressionReport → {BestFitParameters, EstimatedVariance, ParameterCITable}]
```

```
{BestFitParameters → {θ → 1.29136}, EstimatedVariance → 0.0000595846,
 ParameterCITable →      Estimate     Asymptotic SE     CI
                    θ    1.29136      0.000126859       {1.29111, 1.29161}}
```

Clayton

```
regc = NonlinearRegress[CeXYZ, fCc[u, v, θ], {u, v}, {θ, 0.61},
  RegressionReport → {BestFitParameters, EstimatedVariance, ParameterCITable}]
```

```
{BestFitParameters → {θ → 0.559712}, EstimatedVariance → 0.0000767524,
 ParameterCITable →      Estimate     Asymptotic SE     CI
                    θ    0.559712     0.000275767       {0.559172, 0.560253}}
```

Frank

```
regf = NonlinearRegress[CeXYZ, fCf[u, v, θ], {u, v}, {θ, 2.2},
  RegressionReport → {BestFitParameters, EstimatedVariance, ParameterCITable}]
```

```
{BestFitParameters → {θ → 2.02452}, EstimatedVariance → 0.0000515414,
 ParameterCITable →      Estimate     Asymptotic SE     CI
                    θ    2.02452      0.000683983       {2.02318, 2.02586}}
```

parameters summary

```
{θ2g, θ2c, θ2f} = (θ /. (BestFitParameters /. #)) & /@ {regg, regc, regf}
```

```
{1.29136, 0.559712, 2.02452}
```

comparing to empirical copula (L2-norm distance)

```
√((Length[CeXYZ] - 1) (EstimatedVariance /. #)) & /@ {regg, regc, regf}
```

```
{5.61178, 6.36913, 5.2193}
```

Linear convex combination

■ Non-linear fit

Clayton-Gumbel

```
regcg = NonlinearRegress[CeXYZ, α * fCc[u, v, θ2c] + (1 - α) * fCg[u, v, θ2g],
  {u, v}, {α, 0.5}, RegressionReport → {BestFitParameters, EstimatedVariance, ParameterCITable}]
```

```
{BestFitParameters → {α → 0.443701}, EstimatedVariance → 0.0000295681,
 ParameterCITable →      Estimate     Asymptotic SE     CI
                    α    0.443701     0.000605742       {0.442513, 0.444888}}
```

Clayton-Frank

```
regcf = NonlinearRegress[CeXYZ, α * fCc[u, v, θ2c] + (1 - α) * fCf[u, v, θ2f],
  {u, v}, {α, 0.5}, RegressionReport → {BestFitParameters, EstimatedVariance, ParameterCITable}]
```

```
{BestFitParameters → {α → 0.318455}, EstimatedVariance → 0.0000444997,
                          Estimate     Asymptotic SE     CI
 ParameterCITable →  α    0.318455     0.00110118        {0.316297, 0.320613} }
```

Frank-Gumbel

```
regfg = NonlinearRegress[CeXYZ, α * fCf[u, v, θ2f] + (1 - α) * fCg[u, v, θ2g],
  {u, v}, {α, 0.5}, RegressionReport → {BestFitParameters, EstimatedVariance, ParameterCITable}]
```

```
{BestFitParameters → {α → 0.634488}, EstimatedVariance → 0.0000475463,
                          Estimate     Asymptotic SE     CI
 ParameterCITable →  α    0.634488     0.00173447        {0.631089, 0.637888} }
```

parameters summary

```
{αcg, αcf, αfg} = (θ /. (BestFitParameters /. #)) & /@ {regcg, regcf, regfg}
```

```
{0.443701, 0.318455, 0.634488}
```

comparing to empirical copula (L2-norm distance)

```
√((Length[CeXYZ] - 1) (EstimatedVariance /. #)) & /@ {regg, regc, regf}
```

```
{3.95317, 4.84968, 5.01294}
```

Standard error estimation

This is an extra topic - not included in PhD thesis. Procedures are summarized in Genest & Favre (2006).

■ Non-parametric approach

mediating variables and standard error (as function)

```
Z = Table[1/n Sum[If[x[[j]] ≤ x[[i]] && y[[j]] ≤ y[[i]], 1., 0], {j, n}], {i, n}];
Z̃ = Table[1/n Sum[If[x[[i]] ≤ x[[j]] && y[[i]] ≤ y[[j]], 1., 0], {j, n}], {i, n}];
Z̄ = Mean[Z];
S = Sqrt[1/n Sum[(Z[[i]] + Z̃[[i]] - 2 Z̄)^2, {i, n}]];
fSE[g_] := 1/√n 4 S * Abs[g]
```

gumbel

```
D[θ /. Solve[t == (θ - 1) / θ, θ][[1]], t] /. t → τ;
fSE[%]
```

```
0.0454617
```

clayton

```
D[θ /. Solve[t == θ / (θ + 2), θ][[1]], t] /. t → τ;
fSE[%]
```

```
0.0909233
```

frank

```
Normal[Simplify[Series[1 + 4 / θ (fD1[θ] - 1), {θ, 2, 1}], θ > 0]];
Solve[t == %, θ][[1]];
Re[D[θ /. %, t] // N];
fSE[%]
```

```
0.268614
```

(kendall's tau)

```
4 n/(n - 1) Z̄ - (n + 3)/(n - 1)
```

```
0.237088
```

■ Semi-parametric approach

```
fDLg[uu_, vv_, θθ_, dv_] = D[Log[fcg[u, v, θ]], θ] /. {u → uu, v → vv, θ → θθ};
fDLc[uu_, vv_, θθ_, dv_] = D[Log[fcc[u, v, θ]], θ] /. {u → uu, v → vv, θ → θθ};
fDLf[uu_, vv_, θθ_, dv_] = D[Log[fcf[u, v, θ]], θ] /. {u → uu, v → vv, θ → θθ};
```

```
{tmp1g, tmp1c, tmp1f} = MapThread[Table[#1[xR⟦i⟧, i / (n + 1), #2, θ], {i, n}] &, {{fDLg, fDLc, fDLf}, {θ1g, θ1c, θ1f}}];
{tmp2g, tmp2c, tmp2f} = MapThread[Table[
    #1⟦i⟧ -
     1/n Sum[#2[xR⟦j⟧, j / (n + 1), #3, θ] * #2[xR⟦j⟧, j / (n + 1), #3, u], {j, i, n}] -
     1/n Sum[If[xR⟦j⟧ ≥ xR⟦i⟧, #2[xR⟦j⟧, j / (n + 1), #3, θ] * #2[xR⟦j⟧, j / (n + 1), #3, v], 0], {j, 1, n}],
    {i, n}] &, {{tmp1g, tmp1c, tmp1f}, {fDLg, fDLc, fDLf}, {θ1g, θ1c, θ1f}}];
```

standard error for gumbel, clayton and frank copula parameter estimate (respectively)

```
(MapThread[Variance[#2] / Variance[#1] &, {{tmp1g, tmp1c, tmp1f}, {tmp2g, tmp2c, tmp2f}}] / n) // Sqrt
```

```
{0.0372861, 0.0378833, 0.0370627}
```