Numerically Stable Frank Copula Functions via Multiprecision: R Package Rmpfr

Martin Mächer
ETH Zurich
May 2011 – July 2012 (LaTeX’ed October 26, 2015)

Abstract

The package copula (formerly nacopula) has provided functionality for Archimedean copulas, one of them the “Frank copula”. Recently, explicit formulas for the density of those copulas have allowed for maximum likelihood estimation in high (e.g., $d = 150$) dimensions, (Hofert, Mächer, and McNeil (2012)).

However for non-small dimensions, the evaluation of these densities is numerically challenging, and in some cases difficult. Here, we use high precision arithmetic from MPFR, via R package Rmpfr in order to get accurate values for the diagonal density of the Frank copula.

Subsequently, judiciously analysing and circumventing the numerical infelicities of the “traditional” evaluation, we gain accurate values even with traditional (double precision) arithmetic.

Keywords: Archimedean copulas, Frank Copula, Multiple Precision, R, MPFR, Rmpfr.

1. The diagonal density of Frank’s copula

The “diagonal density” of a copula $C(\cdot)$ is the density of $\max(U_1, U_2, \ldots, U_d)$, where $U = (U_1, U_2, \ldots, U_d)^T \sim C$. The (cumulative) distribution function, by definition,

\[ F_D(u) := P[\max(U_1, U_2, \ldots, U_d) \leq u] = P[U_1 \leq u, U_2 \leq u, \ldots, U_d \leq u] = C(u, u, \ldots, u), \]

(1)
evaluates the copula only on the diagonal and is therefore called “the diagonal of $C$”. Its density $f_D(u) := \frac{d}{du} F_D(u)$, is therefore called the diagonal density of $C$. For Archimedean copulas, i.e., where

\[ C(u) = C(u; \psi) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)), \quad u \in [0,1]^d, \]

(2)
the diagonal density is

\[ f_D(u) = \frac{d}{du} F_D(u) = \frac{d}{du} \psi \left( \sum_{j=1}^{d} \psi^{-1}(u) \right) = \frac{d}{du} \psi(d \cdot \psi^{-1}(u)) = \psi'(d \cdot \psi^{-1}(u)) \cdot d \cdot \frac{d}{du} \psi^{-1}(u) = d \cdot \psi'(d \cdot \psi^{-1}(u)) \cdot [\psi^{-1}]'(u). \]

(3)
For this reason, the *copula* package’s dDiag() function for computing the diagonal density \( f^D(u) \) makes use of the following

```r
> copula:::dDiagA
```

```r
function (u, d, cop, log = FALSE) {
    stopifnot(is.finite(th <- cop@theta), d > 2)
    if (any(copAMH@name == c("AMH", "Frank", "Gumbel", "Joe")) &
        any(i0 <- u == 0)) {
        if (log)
            u[i0] <- -Inf
        u[!i0] <- dDiagA(u[!i0], d = d, cop = cop, log = log)
        return(u)
    }
    if (log) {
        log(d) + cop@absdPsi(d * cop@iPsi(u, th), th, log = TRUE) +
        cop@absdiPsi(u, th, log = TRUE)
    } else {
        d * cop@absdPsi(d * cop@iPsi(u, th), th) * cop@absdiPsi(u, th)
    }
}
<bytecode: 0x286306f8>
```

where the three functions

\[
\text{absdPsi}(t, \theta) = |\psi'_\theta(t)|, \quad (4) \\
\text{iPsi}(u, \theta) = \psi^{-1}_\theta(u), \text{ and} \quad (5) \\
\text{absdiPsi}(u, \theta) = |[\psi^{-1}_\theta]'(u)| \quad (6)
\]

are all provided by the slots of the corresponding Archimedean copula family.

For the following explorations, we need a definition of dDiagA which is more flexible as it does not work with the copula family object but gets the three functions as arguments,

```r
> dDiagA <- function(u, th, d, iPsi, absdPsi, absdiPsi, log = FALSE) {
    stopifnot(is.finite(th), d > 0, is.function(iPsi),
              is.function(absdPsi), is.function(absdiPsi))
    if (log) {
        log(d) + absdPsi(d * iPsi(u, th), th, log = TRUE) +
        absdiPsi(u, th, log = TRUE)
    } else {
        d * absdPsi(d * iPsi(u, th), th) * absdiPsi(u, th)
    }
}
```
Now, for the Frank copula (Hofert and Maechler (2011)),

\[
\psi_\theta(t) = -\frac{1}{\theta} \log \left( 1 - (1 - e^{-\theta})e^{-t} \right), \quad \theta > 0, \quad \text{hence,} \quad (7)
\]

\[
\psi_\theta^{-1}(u) = -\log \left( \frac{e^{-u\theta} - 1}{e^{-\theta} - 1} \right), \quad \text{and} \quad (8)
\]

\[
(-1)^k \psi_\theta^{(k)}(t) = \frac{1}{\theta} \text{Li}_{k-1}((1 - e^{-\theta})e^{-t}), \quad \text{and} \quad (9)
\]

\[
\left| \psi_\theta^{-1}'(u) \right| = \frac{\theta}{(e^{u\theta} - 1)}, \quad \text{and } (10)
\]

where \( \text{Li}_s(z) \) is the polylogarithm of order \( s \) at \( z \), defined as (analytic continuation of) \( \sum_{k=1}^{\infty} \frac{z^k}{k^s} \). When \( s = -n, n \in \mathbb{N} \), one has

\[
\text{Li}_{-n}(z) = \left( z \cdot \frac{\partial}{\partial z} \right)^n \frac{z}{1-z}, \quad (11)
\]

and we note that here, only the first derivative is needed, \(-\psi_\theta'(t) = |\psi_\theta'(t)|\), and hence only

\[
\text{polylog}(z, s = 0) = \text{Li}_0(z) = z/(1-z). \quad (12)
\]

First note that numerically, \( e^{-a} - 1 \) suffers from cancellation when \( 0 < a \ll 1 \), and the \( R \) (and \( C \)) function \( \expm1(-a) \) is advisedly used instead of \( \exp(-a) - 1 \). For this reason, in \texttt{copula},

I had replaced the original \texttt{iPsi.0(u, th)} for \( \psi_\theta^{-1}(u) \) by \texttt{iPsi.1()}, making use of \texttt{expm1()}. These and the derivative (10) originally were

\[
> \text{iPsi.0} <- \text{function}(u, \text{theta}) \ -\log( \ (\exp(-\text{theta}*u)-1) \ / \ (\exp(-\text{theta})-1) )
\]

\[
> \text{iPsi.1} <- \text{function}(u, \text{theta}) \ -\log(\expm1(-u*\text{theta}) \ / \ \expm1(-\text{theta}))
\]

\[
> \text{abedPsi.1} <- \text{function}(u, \text{theta}, \log=\text{FALSE}) \ 
\quad \text{if}(\log) \ \log(\text{theta})-\log(\expm1(u*\text{theta})) \ \text{else theta/\expm1(u*\text{theta})}
\]

and the general \( k \)-th derivative (9), simplified, for \( k = 1 \) (degree = 1),

\[
> \text{require("copula")# for polylog()}
\]

\[
> \text{abedPsi.1} <- \text{function}(t, \text{theta}, \log=\text{FALSE}) \{
\quad p <- \neg \expm1(-\text{theta})
\quad \text{Li.} <- \text{polylog}(\log(p) - t, s = 0, \log=\log, \ 
\quad \quad \text{method="negI-s-Eulerian", is.log.z=TRUE})
\quad \text{if}(\log) \ \text{Li.} - \log(\text{theta}) \ \text{else Li.} / \ \text{theta}
\}
\]

where we however now assume that the \texttt{polylog()} function for \( s=0 \) would basically use the direct formula (12), such that we use

\[
> \text{abedPsi.2} <- \text{function}(t, \text{theta}, \log=\text{FALSE}) \{
\quad w <- \log(\neg \expm1(-\text{theta})) - t
\quad \text{Li.} <- \text{if}(\log) \ w - \log(\neg \expm1(w)) \ \text{else} -\exp(w)/\expm1(w)
\quad \text{if}(\log) \ \text{Li.} - \log(\text{theta}) \ \text{else Li.} / \ \text{theta}
\}
\]

### 2. Computing the “diagonal MLE”

The most important use of the diagonal density is to compute the “diagonal maximum likelihood estimator”, \( \hat{\theta}^{D} \) which for a sample of observations \( u_1, u_2, \ldots, u_n, (u_i \in [0,1]^d) \) is
defined as minimizer of the negative log-likelihood,

\[ \hat{\theta}^D = \arg \min_{\theta} -l(\theta; u_1, \ldots, u_n), \]

where

\[ l(\theta; u_1, \ldots, u_n) = \sum_{i=1}^{n} \log f^D(\tilde{u}_i) \]

and

\[ \tilde{u}_i = \max_{j=1, \ldots, d} u_{i,j}. \]

In our exploration of the \texttt{dmle} estimator, we found cases with numerical problems, at first already in evaluating the logarithm of the diagonal density \( \log f^D = \log f^D_D(\theta, u) \) for non-small \( \theta \) and "large" \( u \), i.e., \( u \approx 1 \):

\begin{verbatim}
> curve(dDiagA(x, th = 38, d = 2, iPsi = iPsi.0, absdPsi=absdPsi.1, absdiPsi=absdiPsi.1, log = TRUE),
ylab = "dDiagA(x, theta= 38, *, log=TRUE)",
0, 1, col = 4, n = 1000)
> ## and using the slightly better \texttt{iPsi.1} does not help here:
> curve(dDiagA(x, th = 38, d = 2, iPsi = iPsi.1, absdPsi=absdPsi.2, absdiPsi=absdiPsi.1, log = TRUE),
add = TRUE, col = 2, n=1000)
> legend("bottom", c("iPsi.0()","iPsi.1()"),col=c(4,2), lty=1, bty="n")
\end{verbatim}

However, it’s not hard to see that indeed our initial computation of Frank’s \( \psi_{\theta}^{-1} \), i.e, (8), \( \psi_{\theta}^{-1}(u) = - \log \left( \frac{1-e^{-u} \theta}{1-e^{-\theta}} \right) \), suffers from “division cancellation” for “large” \( \theta \) (\( \theta = 38 \) in ex.) when computed directly with \texttt{iPsi.0()}, see above, and that the improvement of using \texttt{expm1(-t)}
instead of $\exp(-t) - 1$, as used in $\text{iPsi.1()}$, see above, helps only for $t \approx 0$. However, we can rewrite $\psi^{-1}$ as

$$\psi^{-1}(u) = -\log \left( 1 - \frac{e^{-u\theta} - e^{-\theta}}{1 - e^{-\theta}} \right),$$  

(16)

which when using $\log(1 + e)$ for $\log(1 + e)$ is much better numerically:

```r
> iPsi.2 <- function(u,theta) -log1p((exp(-u*theta)-exp(-theta)) / expm1(-theta))
> curve(dDiagA(x, th = 38, d = 2, iPsi = iPsi.2, 
    absdPsi=absdPsi.2, absdPsiI=absdPsiI.1, log = TRUE),
    ylab = "dDiagA(x, theta= 38, *, log=TRUE)",
    0, 1, col = 4, n = 1000)
> ## previously
> curve(dDiagA(x, th = 38, d = 2, iPsi = iPsi.1, 
    absdPsi=absdPsi.2, absdPsiI=absdPsiI.1, log = TRUE),
    add = TRUE, col = "darkgray", lwd=2, lty=3, n=1000)
```

Unfortunately, this is not enough to get numerically stable evaluations of the negative log-likelihood $l()$:

```r
> d <- 5
> (theta <- copFrank@iTau(tau = 0.75))
[1] 14.13852
> cop <- onacopulaL("Frank", list(theta, 1:d))
> set.seed(1); for(l in 1:4) U <- rnacopula(n = 100, cop)
> U. <- sort(apply(U, 1, max)) # build the max
```
> mlogL <- function(theta)
  -sum(dDiagA(U., theta, d=d, iPsi = iPsi.2,
          absdPsi=absdPsi.2, absdiPsi=absdiPsi.1,
          log = TRUE))

Now, plot this negative log likelihood function in an interval \( \theta \), close to what is proposed copula's initOpt() function, defining a utility function that we'll reuse later:

```r
> p.mlogL <- function(th, mlogL, col= "red2", lwd = 1, lty = 1,
  add= FALSE) {
  stopifnot(is.numeric(th), is.function(mlogL))
  nll <- vapply(th, mlogL, 0.)
  if(add) lines(nll ~ th, col=col, lwd=lwd, lty=lty)
  else plot(nll ~ th, xlab=expression(theta),
            ylab = expression(- logLik(theta, .)),
            type = "l", col=col, lwd=lwd, lty=lty)
  invisible(nll) # return invisibly
}
> thet <- seq(11, 99, by = 1/4)
> p.mlogL(thet, mlogL)
```

```r
> require("Rmpfr")## compute the same with *high* accuracy ...
> ## using three different precisions:
> MPrecBits <- c(160, 128, 96)
> mkNm <- function(bits) sprintf("%03d.bits", bits)
> ## As it takes a while, cache the result:
> fnam <- sprintf("mlogL_mfr_%s.rda", Sys.info()["machine"])
> if (!file.exists(fn <- file.path(copDDir,fnam))) {
  print(system.time(
    nllMP <- lapply(MPrecBits, function(pBit) {
      thM <- mpfr(thet, precBits = pBit)
      nllM <- sapply(thM, mlogL)
      nllM
    })))## 91.226 0.013 91.506 [nb-mm icore 5]
  names(nllMP) <- mkNm(MPrecBits)
  copSrcDDir <- if(Sys.getenv("USER") == "maechler")
    "/R/Pkgs/copula/inst/doc" else "
  if(file.exists(copSrcDDir))## only for certain users; not on CRAN etc
    save(nllMP, file = file.path(copSrcDDir, fnam))
  else load(fn)
}
```

```r
> colB <- c("blue3","violetred4","tan3")
> ltyB <- c(5:3)
> lwdB <- c(2,2,2)
> for(i in seq_along(nllMP)) {
  lines(thet, as.numeric(nllMP[[i]]),
        col=colB[i], lty = ltyB[i], lwd = lwdB[i])
}
```

```r
> leg <- c("double prec.", sprintf("mpfr(*, precBits = %d)", MPrecBits))
> legend("top", leg,
    col= c("red3",colB), lty=c(1, ltyB), lwd=c(1,lwdB), bty="n")
```
So, clearly, high-precision computations can solve the numerical problems, if the precision is high enough. E.g., for $\theta = 100$, it needs more than 128 bits precision.

Let us look at the phenomenon in more details now. The flesh in the \texttt{mlogL() computation} is (up to the constant $\log(d)$, $d = 5$), only the sum of the two terms

$$\text{absdPsi}(d\ast iPsi(u, th), \text{th, log} = \text{TRUE}) + \text{absdiPsi}(u, \text{th, log} = \text{TRUE})$$

currently, with the three functions

\begin{align*}
  \text{iPsi} &= \text{iPsi.2}; \text{absdPsi} = \text{absdPsi.2}; \text{absdiPsi} = \text{absdiPsi.1}
\end{align*}

where we have already tried to ensure that the \text{iPsi()} function is ok, but now can confirm it, e.g., for $\theta = 50$. Further note, that using high-precision arithmetic, we can also “partially afford” to use the simplistic \text{iPsi.0()} function instead of the more stable \text{iPsi.2()} one:

```r
> stopifnot(all.equal(iPsi.2(U., mpfr(50, 96)),
                        iPsi.2(U., mpfr(50, 96))),
          all.equal(iPsi.0(U., mpfr(50, 200)),
                    pI.U <- iPsi.2(U., mpfr(50, 200)), tol=1e-50 )
```

However, we can observe dramatic differences in \text{absdPsi.2()} ($= |\psi'(\cdot)|$):

```r
> psD.n <- absdPsi.2(as.numeric(pI.U), 40)
> psD.M <- as.numeric(absdPsi.2(pI.U, mpfr(40, 200)))
> matplot(U., cbind(psD.n, psD.M), type="l", log="y")
> legend("top", c("double prec.", "mpfr(*, precBits = 200)"),
          col=1:2, lty=1:2, bty="n")
```
where we see a very large difference (note the log scale!) for $u \approx 1$, i.e., for very small $pI.U = iPsi.2(U., \theta) = \psi^{-1}_\theta(u)$.

```r
> u0 <- 2^-(100:1)
> psD.n <- absdPsi.2(u0, 40)
> psD.M <- as.numeric(absdPsi.2(u0, mpfr(40, 200)))
> matplot(u0, cbind(psD.n, psD.M), type="l", log="xy")
> legend("top", c("double prec.", "mpfr(*, precBits = 200)"),
  col= 1:2, lty=1:2, bty="n")
```
Further investigation shows that the culprit is really the use of \( \log(-\exp(-\theta)) \) inside `absdPsi.2()` which underflows for large \( \theta \), and hence should be replaced by the generally accurate \( \log1mexp() \), see Mächler (2012).

```r
> log1mexp <- function(a) # accurately compute \( \log(1-\exp(-a)) \)
{ #
  stopifnot(a >= 0)
  r <- a
  tst <- a <= log(2)
  r[ tst] <- log(-expm1(-a[ tst]))
  r[!tst] <- log1p(-exp(-a[!tst]))
  r
}
```

so that we rather compute \( |\psi'(\cdot)| \) via

```r
> absdPsi.3 <- function(t, theta, log=FALSE) {
  w <- log1mexp(theta) - t
  Li. <- if(log) w - log1mexp(-w) else -exp(w)/expm1(w)
  if(log) Li. - log(theta) else Li. / theta
}
```

Does this already solve the “diagonal likelihood” problem? We investigate via the graphic

```r
> p.mlogL(th = seq(11, 99, by = 1/4),
          mlogL = (mlogL2 <- function(theta)
                    -sum(dDiagA(U., theta, d=d, iPsi = iPsi.2,
                                  absdPsi = absdPsi.3, absdiPsi=absdiPsi.1,
                                  log = TRUE)), lwd = 2))
```
Yes, indeed, using a numerically stable version for `absdPsi()` did solve the numerical problem of computing, the “diagonal likelihood”, and hence the `dmle()` ("diagonal MLE").

Well, if we really insist, there are more problems, but probably not really practical:

```r
> thet <- 9:1000
> nll <- p.mlogL(thet, mlogL = mlogL2, lwd=2)
> (th0 <- thet[i0 <- max(which(is.finite(nll))])]

[1] 710
```

```r
> abline(v = th0, col="red2", lty="15", lwd=2)
```
where we see that for $\theta > 710$, $\text{mlogL}()$ is not finite, e.g.,

```r
> dDiagA(0.999, 715, d = d, iPsi = iPsi.2, absdPsi = absdPsi.3,
        absdiPsi = absdiPsi.1, log = TRUE)
```

[1] -Inf

which after closer inspection is from the `absdiPsi(u, th, log = TRUE)` part of `dDiagA()` and that, see (6), uses

$$\log(\theta) - \log(\expm1(u*\theta)),$$

where clearly already $\expm1(u*\theta)$ numerically overflows to Inf:

```r
> absdiPsi.1(0.999, th = 715, log=TRUE)
```

[1] -Inf

However, as $\log(\expm1(y)) = \log(e^y - 1) = \log(e^y(1-e^{-y})) = y + \log(1-e^{-y})$, the “numerical stable” solution is to replace $\log(\expm1(y))$ by $y + \log1mexp(y)$, such that we will use

```r
> absdiPsi.2 <- function(u, theta, log = FALSE)
    if(log) log(theta)- {y <- u*theta; y + log1mexp(y)} else theta/expm1(u*theta)
```

Unfortunately, this improves the situation for large $\theta$ only slightly:

```r
> plot(nll ~ thet, xlab=expression(theta),
       ylab = expression(- logLik(theta, .)),
       type = "l", col="red2", lwd=2)
> abline(v = th0, col="red2", lty="15", lwd=2)
> nll3 <- p.mlogL(thet, mlogL = function(theta)
    -sum(dDiagA(U., theta, d=d, iPsi= iPsi.2, absdPsi= absdPsi.3,
                absdiPsi = absdiPsi.2, log = TRUE)),
    col = "blue2", lwd=3, lty=2, add = TRUE)
```
> nll3[thet == 800]
[1] -Inf

where we can see, that this time, the numerical overflow to \(-\infty\) happens in \texttt{absdPsi(d*iPsi(u, th), th, log = TRUE)}, as \texttt{iPsi(u, th) = iPsi(0.999, th=800)=0} underflows to zero — by necessity: the correct value is smaller than the smallest representable double precision number:

> pI <- iPsi.2(u=0.999, th= mpfr(800, 200))
> cat(sapply(list(pI, .Machine$double.xmin),
          format, digits = 7), "\n")

4.495130e-348 2.225074e-308

The only solution here is to allow passing the logarithm \(\log(t)\) instead of \(t\) to \texttt{absdPsi()}, i.e., we start computing \(\log(iPsi(.)\) directly (evading the underflow to 0 there).

......to be continued.

3. Session Information

> toLatex(sessionInfo())

- R version 3.2.2 Patched (2015-10-23 r69569), x86_64-pc-linux-gnu
- Locale: LC_CTYPE=de_CH.UTF-8, LC_NUMERIC=C, LC_TIME=en_US.UTF-8,
  LC_COLLATE=C, LC_MONETARY=en_US.UTF-8, LC_MESSAGES=C, LC_PAPER=de_CH.UTF-8,
4. Conclusion

References


Affiliation:

Martin Mächler
Seminar für Statistik, HG G 16
ETH Zurich
8092 Zurich, Switzerland
E-mail: maechler@stat.math.ethz.ch
URL: http://stat.ethz.ch/people/maechler