Solution 1

(a) We have $\mathrm{P}(S \le s) = 0$ for $s < 0$ and $\mathrm{P}(S \le 0) = 1 - \theta$, so $\mathrm{P}(S = 0) = 1 - \theta$. Hence

$$\mathrm{E}(S) = 0 \times (1-\theta) + \int_{0+}^{\infty} s \times \theta\lambda\theta e^{-\lambda\theta s}\,\mathrm{d}s = \theta(\lambda\theta)^{-1} = 1/\lambda,$$

and

$$\mathrm{E}(S \mid S > 0) = \int_{0+}^{\infty} s \times \theta\lambda\theta e^{-\lambda\theta s}\,\mathrm{d}s/\mathrm{P}(S>0) = \theta(\lambda\theta)^{-1}/\theta = 1/(\lambda\theta),$$

so $\theta = \mathrm{E}(S)/\mathrm{E}(S \mid S > 0)$.

(b) If S^*/m has approximately the distribution of S, then $\mathrm{P}(S^* > s) = \mathrm{P}(S^*/m > s/m) \approx \mathrm{P}(S > s/m) = \theta\exp(-\lambda\theta s/m)$, for $s > 0$, and

$$\widehat{\theta} = \frac{n^{-1}\sum_{j=1}^{n} S_j^*/m}{n^{-1}\sum_{j=1}^{n} I(S_j^*/m > 1/m)S_j^*/m \div n^{-1}\sum_{j=1}^{n} I(S_j^*/m > 1/m)} \approx \mathrm{E}(S)/\mathrm{E}(S \mid S>0)$$

for large n and m, so we might hope that the ratio is a reasonable estimator of θ. Proving this mathematically would require the asymptotic dependence between the S_j^* to be weak enough for the sums in $\widehat{\theta}$ to converge in probability to their expectations.

(c) The conditional distribution of $S \mid S > 0$ is exponential with parameter $\lambda\theta$, so $S - u \mid S > u \sim \exp(\lambda\theta)$ using the lack of memory of the exponential distribution. If you doubt this, note that for $x, u > 0$,

$$\mathrm{P}(S-u > x \mid S > u) = \mathrm{P}(S > u + x \mid S > u) = \frac{\mathrm{P}(S > u+x)}{\mathrm{P}(S>u)} = \frac{\exp\{-\lambda\theta(u+x)\}}{\exp(-\lambda\theta u)} = \exp(-\lambda\theta x).$$

For any $u > 0$, therefore, the positive values of $S_1^* - u, \ldots, S_n^* - u$, are a random sample from the exponential distribution with mean $\lambda\theta/m$. If $X_1, \ldots, X_K \overset{\text{iid}}{\sim} \exp(\lambda\theta/m)$ for known λ/m, then the log likelihood is

$$\ell(\theta) = K\log(\lambda\theta/m) - \lambda\theta/m \times \sum_{k=1}^{K} X_k, \quad \theta > 0,$$

so the MLE is $\widehat{\theta} = Km/(\lambda\sum_k X_k)$. In the present case K is replaced by $n_u = \sum_{j=1}^{n} I(S_j^* > u)$ and $\sum_{k=1}^{K} X_k$ is replaced by $\sum_{j=1}^{n} I(S_j^* > u)(S_j^* - u)$, so

$$\widehat{\theta} = \frac{m}{\lambda}\frac{\sum_{j=1}^{n} I(S_j^* > u)}{\sum_{j=1}^{n} I(S_j^* > u)(S_j^* - u)},$$

and if we replace λ/m by its estimator $n/\sum_{j=1}^{n} S_j^*$, we (almost) get the estimator in (b), with $u = 1$.

Solution 2

(a) As F, F_X, F_Y are distributions, they are monotone increasing, and thus $C(u, v)$ is also monotone increasing in each of its arguments. Moreover $C(u, v) \ge 0$ since $F(x, y) \ge 0$, and it is straightforward to show that $\lim_{u\to 0, v\to 0} C(u, v) = 0$ and $\lim_{u\to 1, v\to 1} C(u, v) = 1$, so $C(u, v)$ is a distribution function. Now, let (U, V) be distributed according to $C(u, v)$. Since $F_X(x) = F(x, \infty)$ and $F_Y(y) = F(\infty, y)$, we have

$$\begin{aligned}\mathrm{P}(U \le u) &= \lim_{v\to 1}\mathrm{P}(U \le u, V \le v) = \lim_{v\to 1} C(u,v) = \lim_{v\to 1} F\left\{F_X^{-1}(u), F_Y^{-1}(v)\right\} \\ &= F\left\{F_X^{-1}(u), \lim_{v\to 1} F_Y^{-1}(v)\right\} = F\left\{F_X^{-1}(u), \infty\right\} = F_X\{F_X^{-1}(u)\} = u,\end{aligned}$$

and $\mathrm{P}(V \le v) = v$ by symmetry: the margins of $C(u, v)$ are uniform.

(b) Recall that $\log(1+a) \approx a$ as $a \to 0$, so $\log p = \log\{1+(p-1)\} \approx p-1$ as $p \to 1$. Hence

$$\begin{aligned} \mathrm{P}\left\{Y > F_Y^{-1}(u) \mid X > F_X^{-1}(u)\right\} &= \frac{\mathrm{P}\left\{X > F_X^{-1}(u), Y > F_Y^{-1}(u)\right\}}{\mathrm{P}\left\{X > F_X^{-1}(u)\right\}} \\ &= \frac{1-2u+C(u,u)}{1-u} = 2 - \frac{1-C(u,u)}{1-u} \\ &\approx 2 - \frac{\log C(u,u)}{\log u}, \quad u \to 1. \end{aligned}$$

The limit χ can be interpreted as a measure of extremal dependence. If $\chi = 0$, then the variables X and Y (and thus also U and V) are asymptotically independent. If $\chi > 0$, then the variables are asymptotically dependent. In practice, it often happens that dependence weakens at higher levels, casting doubt on the validity of asymptotically dependent models.

(c) Here $F_X(x) = \exp(-1/x)$ and $F_Y(y) = \exp(-1/y)$, so $F_X^{-1}(u) = -1/\log u$, $F_Y^{-1}(v) = -1/\log v$, and

$$C(u,v) = F\left\{F_X^{-1}(u), F_Y^{-1}(v)\right\} = \exp\left[-\left\{(-1/\log u)^{-1/\alpha} + (-1/\log v)^{-1/\alpha}\right\}^{\alpha}\right], \quad 0 < u, v < 1.$$

Thus,

$$\begin{aligned} \chi(u) = 2 - \frac{\log C(u,u)}{\log u} &= 2 - \frac{-\left\{(-1/\log u)^{-1/\alpha} + (-1/\log u)^{-1/\alpha}\right\}^{\alpha}}{\log u} \\ = 2 - \frac{-\left\{2(-1/\log u)^{-1/\alpha}\right\}^{\alpha}}{\log u} &= 2 - 2^{\alpha}\frac{-\left\{(-1/\log u)^{-1/\alpha}\right\}^{\alpha}}{\log u} = 2 - 2^{\alpha}, \end{aligned}$$

and therefore $\chi = 2 - 2^{\alpha}$. When $\alpha = 1$ the variables are asymptotically independent (in fact, exactly independent) and $\chi = 0$, whereas $\chi \to 1$ as $\alpha \to 0$.

(d) The model in (c) has $\chi = \chi(u) = 2 - 2^{\alpha}$ for all $u \in [0,1]$, so $\chi(u) = 2 - 2^{0.3} \approx 0.77$ when $\alpha = 0.3$ and $\chi(u) = 0$ when $\alpha = 1$. The left- and right-hand graphs correspond to these models, so the middle one must correspond to the bivariate normal distribution. For the latter we see that there is dependence for all u but that the dependence reduces towards zero when $u \to 1$. In fact the bivariate normal model is asymptotically independent, ie., $\chi = 0$.

Solution 3

(a) We first simulate the moving maximum process with Fréchet margins from the code below and inspect the plots in Figure 1.

```
n <- 10000; a <- 1; i <- c(1:n)   # we saw this before
z <- 1/rexp(n+1)    # independent Frechet variables
x <- pmax(a*z[i],z[i+1])/(a+1)  # moving maximum series
par(mfrow=c(1,2))  # two adjacent panels for figures

chi.lag <- function( x, lag=0)  chiplot(cbind(x[1:(n-lag)],x[(1+lag):n]), which=1)
chi.lag( x, 1)
chi.lag( x, 2)
```

The processes X_{t+1} and X_t are asymptotically dependent by construction and because of the standard Fréchet margins. Note that X_{t+h} and X_t have no noise variables Z in common for $h \geq 2$, so they are independent for such lags, as shown by the right plot of Figure 1 and the plots in Figure 2. Here we see that $\chi_h(u)$ behaves similarly for $h \geq 2$.

Figure 3 provides plots of $\chi(u)$ for the moving maximum for $a \in \{0.5, 0.1\}$, showing that the weakening dependence is reflected in a lower $\chi_h(u)$.

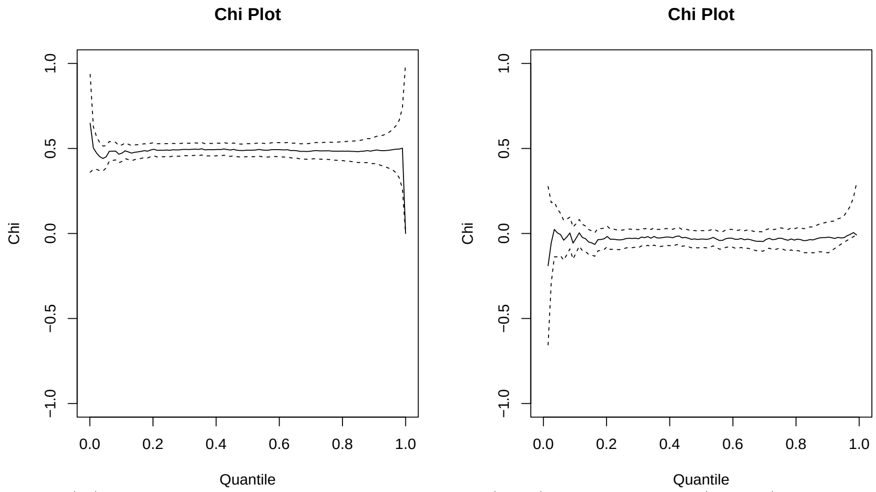

Figure 1: Plots of $\chi_h(u)$ for $a = 1$ and the lags $h = 1$ (left) and $h = 2$ (right) for the moving maximum process.

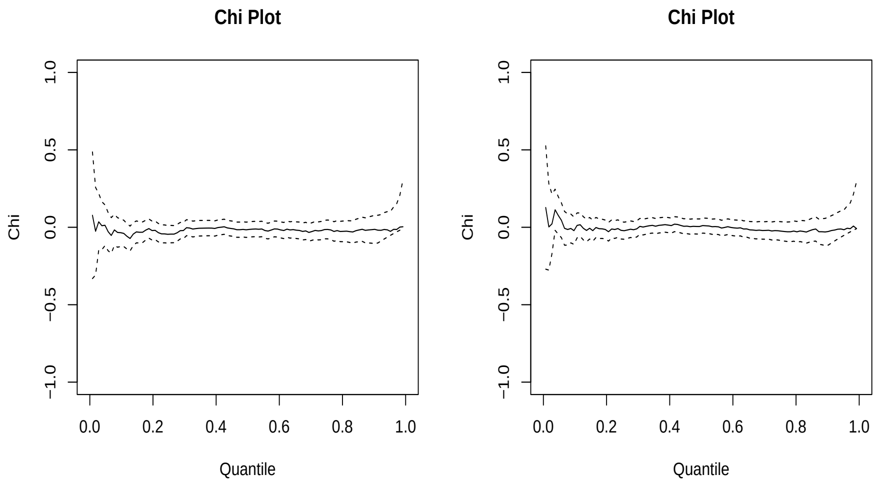

Figure 2: Plots of $\chi_h(u)$ for $a = 1$ and the lags $h = 3$ (left) and $h = 4$ (right) for the moving maximum process.

(b) For the Gaussian autoregressive process, we use bi-linearity of the covariance operator to compute

$$\begin{aligned}\operatorname{cov}(X_j, X_{j+1}) &= \operatorname{cov}(X_j, \rho X_j + (1-\rho^2)^{1/2}\varepsilon_{j+1}) \\ &= \rho \operatorname{cov}(X_j, X_j) + \operatorname{cov}(X_j, (1-\rho^2)^{1/2}\varepsilon_{j+1}) \\ &= \rho \operatorname{var}(X_j),\end{aligned}$$

since X_j is independent of ε_{j+1}. As $\operatorname{var}(X_j) = 1$, we find $\operatorname{corr}(X_j, X_{j+1}) = \rho$.

You were not asked to find the corresponding result for general h, but to do so write $\tilde{\varepsilon}_j = (1-\rho^2)^{1/2}\varepsilon_j$. Note that $\operatorname{cov}(X_{j+h}, X_j) = \operatorname{cov}(\rho X_{j+h-1} + \tilde{\varepsilon}_{j+h}, X_j) = \rho\operatorname{cov}(X_{j+h-1}, X_j)$ for $h \geq 1$, so one can start from $h = 1$ and use that $\operatorname{cov}(X_{j+h}, X_j) = \rho$ to show that $\operatorname{cov}(X_{j+2}, X_j) = \rho\operatorname{cov}(X_{j+1}, X_j)$. Proceeding similarly, and using that $\operatorname{var}(X_j) = 1$ gives $\operatorname{cov}(X_{j+h}, X_j) = \rho^h$. Applying a similar argument for $h < 0$ gives that $\operatorname{cov}(X_{j+h}, X_j) = \rho^{|h|}$.

In contrast to the moving maximum process with Fréchet margins, the Gaussian autoregressive process exhibits so-called asymptotic independence: while low levels of thresholds u show dependence, larger values of u lead to a decrease in χ, eventually giving $\lim_{u\to\infty}\chi(u) = 0$. Figure 6 shows plots of $\chi_h(u)$ for different lags h, indicating that convergence of $\chi(u)$ to zero is very slow.

(c) We inspect the moving average process $X_{j+1} = \varepsilon_{j+1} + \rho\varepsilon_j$, and compute

$$\operatorname{var}(\varepsilon_{j+1} + \rho\varepsilon_j) = \operatorname{var}(\varepsilon_{j+1}) + \rho^2\operatorname{var}(\varepsilon_j) = 1 + \rho^2.$$

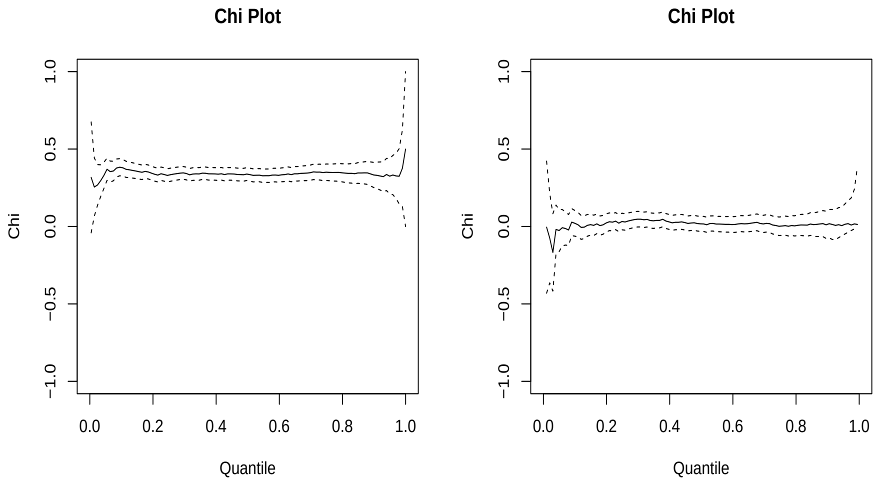

Figure 3: Plots of $\chi_h(u)$ for $a = 0.5$ and the lags $h = 1$ (left) and $h = 2$ (right) for the moving maximum process.

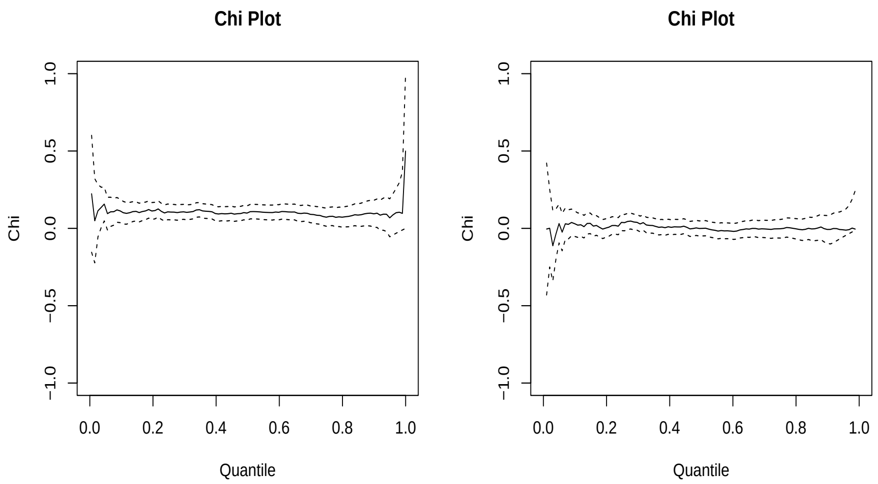

Figure 4: Plots of $\chi_h(u)$ for $a = 0.1$ and the lags $h = 1$ (left) and $h = 2$ (right) for the moving maximum process.

where we have used the independence of the ε_j's and that they have standard normal distributions. We then compute the covariance between X_{j+1} and X_j

$$\begin{aligned}\mathrm{cov}(\varepsilon_{j+1} + \rho\varepsilon_j, \varepsilon_j + \rho\varepsilon_{j-1}) &= \mathrm{cov}(\varepsilon_{j+1}, \varepsilon_j) + \mathrm{cov}(\varepsilon_{j+1}, \rho\varepsilon_{j-1}) + \mathrm{cov}(\rho\varepsilon_j, \varepsilon_j) + \mathrm{cov}(\rho\varepsilon_j, \rho\varepsilon_{j-1}) \\ &= \rho\mathrm{cov}(\varepsilon_j, \varepsilon_j) = \rho.\end{aligned}$$

Division by the variance leads to $\mathrm{corr}(X_{j+1}, X_j) = \rho/(1 + \rho^2)$.

For lags $h \geq 2$ the indices of the white noise ε involved in the computation of the covariance of X_{j+h} and X_j differ, and since the noise is independent, the processes X_{j+h} and X_j are independent for such h, implying that $\mathrm{cov}(X_{j+h}, X_j) = 0$. The dependence for the Gaussian process is illustrated by the left plot in Figure 7, exhibiting asymptotic independence similar to what we observed in Figure 6. The plot on the right on Figure 7 shows independence similar to what we observed for lags $h \geq 2$ in (a).

(d) To evaluate the impact of non-stationarity one must take into account both the availability of the data and the extent or frequency of the non-stationarity episodes relative to the observed data. For instance, if we observe daily data and non-stationary behaviour occurs on a yearly basis, we can expect its effect to be negligible. However, if we observe monthly or seasonal data, then we can expect the effect of such non-stationarity to be more significant.

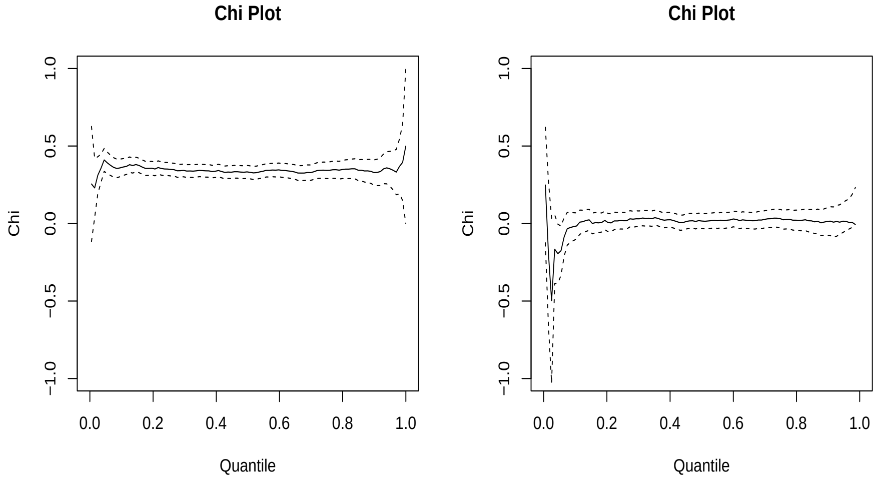

Figure 5: Plots of $\chi_h(u)$ for $a = 2$ and the lags $h = 1$ (left) and $h = 2$ (right) for the moving maximum process.

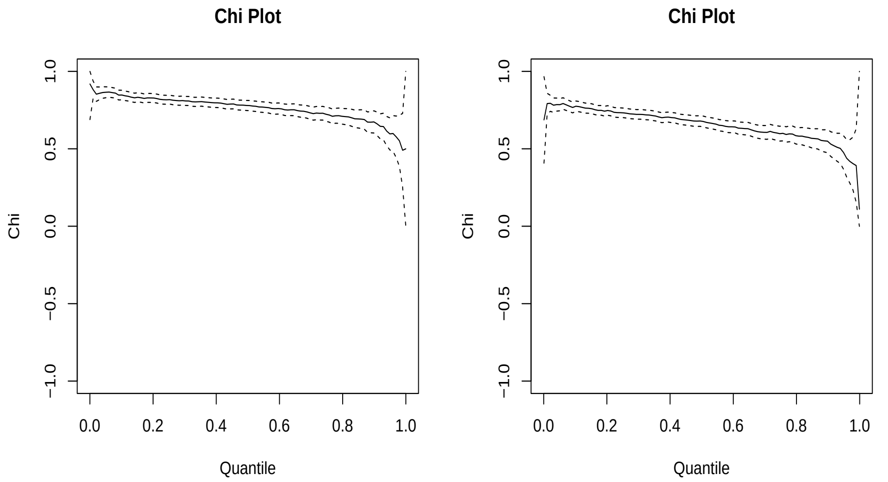

Figure 6: Plots of $\chi_h(u)$ for $a = 0.9$ and the lags $h = 1$ (left) and $h = 2$ (right) for the autoregressive process.

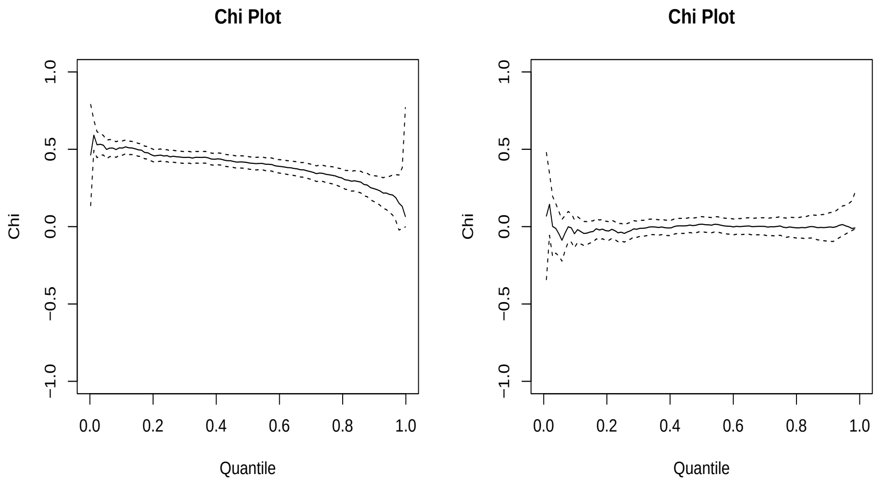

Figure 7: Plots of $\chi_h(u)$ for $a = 0.9$ and the lags $h = 1$ (left) and $h = 2$ (right) for the moving average process.