Time Series

Jonathan M. Lees
University of North Carolina, Chapel Hill
Department of Geological Sciences
CB #3315, Mitchell Hall
Chapel Hill, NC 27599-3315
email: jonathan.lees@unc.edu
ph: (919) 962-0695

November 4, 2008

1 Time Series

A time series is sequence of observations which are ordered in time or space. The term time series applies to that have units of time, but they can also have other kinds of units, as in spatial time series where location along a line is important. Time series are incredibly important for geological sciences since geology deals with events occurring in space and time.

1.1 Series of Events

The first kinds of events we will deal with are events that are recorded as discrete points in time (or space). The classic examples of such data sets are catastrophic events, like volcanic explosions or earthquake ruptures. But any specific event can be analyzed using the following approach. We take here an exploratory approach, where the methods are intended to reveal patterns rather than test specific hypotheses.

```
> aso <- c(1229, 1239, 1240, 1265, 1269, 1270, 1272, 1273, 1274,
+     1281, 1286, 1305, 1324, 1331, 1335, 1340, 1346, 1369, 1375,
```

```
+     1376, 1377, 1387, 1388, 1434, 1438, 1473, 1485, 1505, 1506,
+     1522, 1533, 1542, 1558, 1562, 1563, 1564, 1576, 1582, 1583,
+     1584, 1587, 1598, 1611, 1612, 1613, 1620, 1631, 1637, 1649,
+     1668, 1675, 1683, 1691, 1708, 1709, 1765, 1772, 1780, 1804,
+     1806, 1814, 1815, 1826, 1827, 1828, 1829, 1830, 1854, 1872,
+     1874, 1884, 1894, 1897, 1906, 1916, 1920, 1927, 1928, 1929,
+     1931, 1932, 1933, 1934, 1935, 1938, 1949, 1950, 1951, 1953,
+     1954, 1955, 1956, 1957, 1958, 1962)
```

```
> postscript(file = "barcode.eps", width = 10, height = 4, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(aso, rep(1, length(aso)), type = "n", xlab = "Time, year",
+     ylab = "", axes = FALSE)
> axis(1)
> abline(v = aso)
> box()
> dev.off()
```

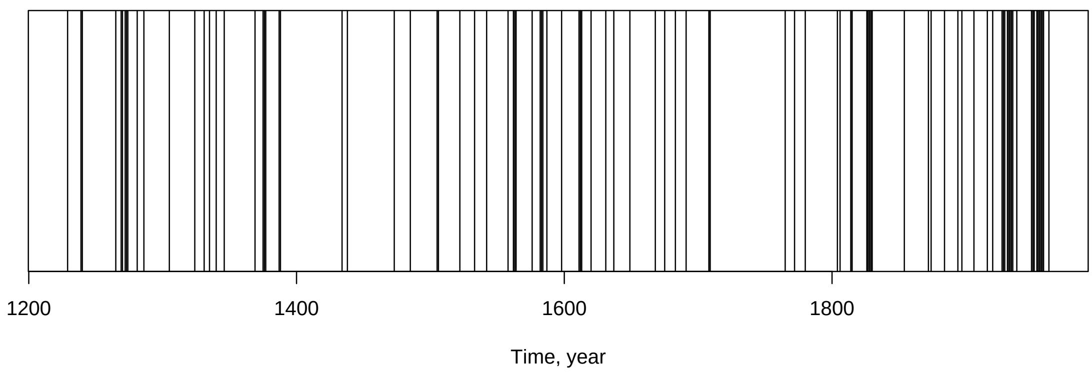


Figure 1: Barcode plot of series of event

1.1.1 Histograms

Occasionally histograms of the time intervals or spacings of the events show some intrinsic structure.

```
> a = diff(aso)
> length(a)
> postscript(file = "asohist.eps", width = 10, height = 4, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> hist(a, main = "")
> dev.off()
```

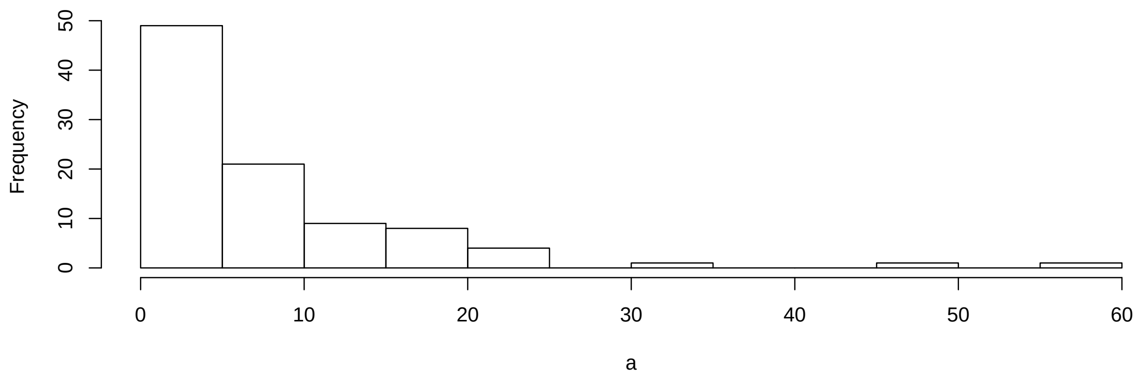


Figure 2: Histogram of series of event

We can combine the two plots,

```
> ma = 100 * floor(min(aso)/100)
> Ma = 100 * ceiling(max(aso)/100)
> intv = 100
> K = seq(from = ma, by = intv, to = Ma)
> K1 = K[1:(length(K) - 1)]
> K2 = K[2:length(K)]
> KD = rep(1, length(K1))
> for (i in 1:length(KD)) {
+     r = sum(aso >= K1[i] & aso < K2[i])
+     KD[i] = r
+ }
> postscript(file = "asohistbar1.eps", width = 10, height = 4,
+     paper = "special", horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(range(c((K1 + K2)/2 - intv/2, (K1 + K2)/2 + intv/2)), range(0,
+     KD), type = "n", ylab = "Number of events", xlab = paste(sep = "",
+     intv, "-year intervals"), main = "Aso Eruptions")
> rect((K1 + K2)/2 - intv/2, rep(0, length(KD)), (K1 + K2)/2 +
+     intv/2, KD, col = gray(0.9))
> points(aso, rep(max(KD), length(aso)))
> dev.off()
```

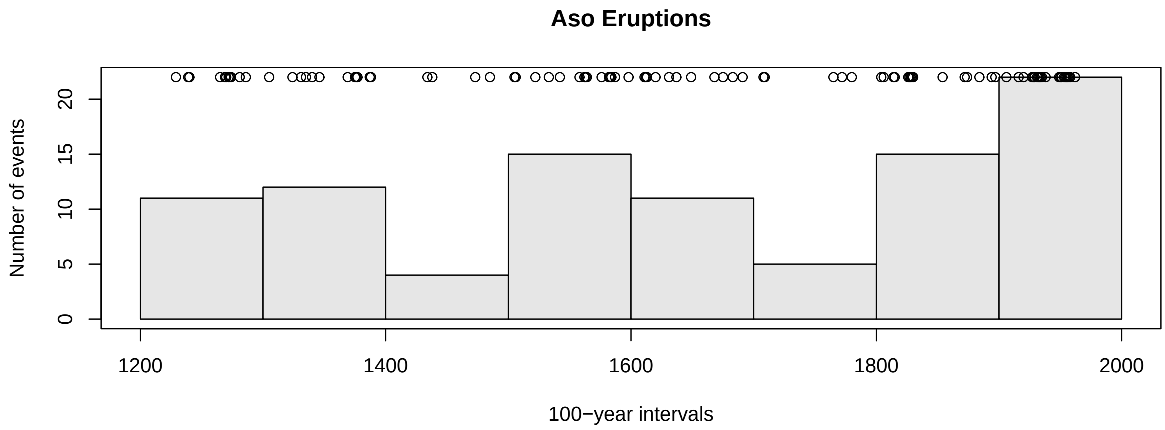


Figure 3: Histogram Barcode of series of event

We change the interval,

```
> postscript(file = "asohistbar2.eps", width = 10, height = 4,
+     paper = "special", horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(range(K1 - intv/2, K1 + intv/2), range(KD), type = "n",
+     ylab = "Number of events", xlab = paste(sep = "", intv, "-year intervals"),
+     main = "Aso Eruptions")
> rect((K1 + K2)/2 - intv/2, rep(0, length(KD)), (K1 + K2)/2 +
+     intv/2, KD, col = gray(0.9))
> points(aso, rep(max(KD), length(aso)))
> box()
> dev.off()
```

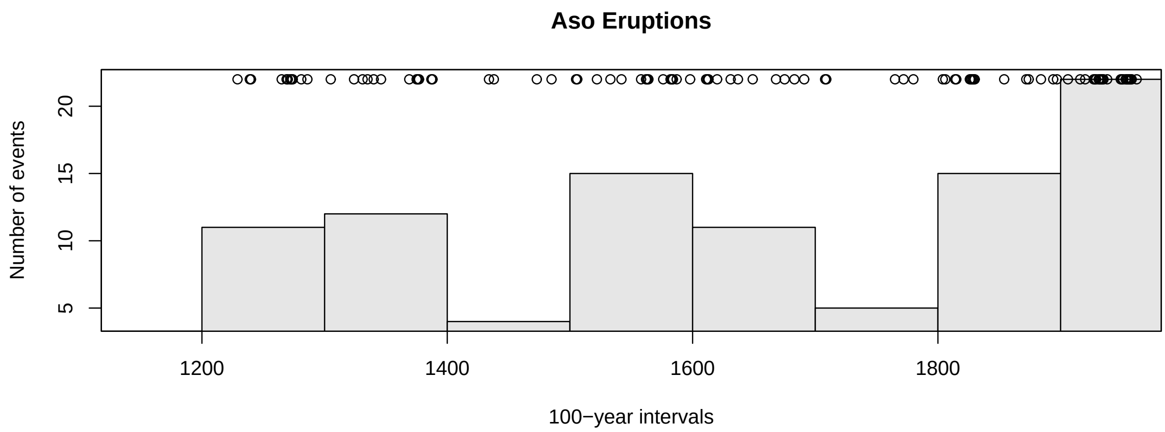


Figure 4: Histogram Barcode of series of event, 100 year

1.1.2 Cumulative Plots

A common way to visualize a sequence of related events is to plot their cumulative number versus the time of occurance. This is often used for illustrating the rate of earthquake occurance. When the slope of this plot is steep, events are occuring at a high rate, as during swarms.

In R this can be easily achieved by,

```
> CSK = rep(0, length(aso))
> for (i in 1:length(aso)) {
+     r = sum(aso <= aso[i])
+     CSK[i] = r
+ }
```

```
> postscript(file = "cumsum.eps", width = 10, height = 4, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(aso, CSK, pch = 1, col = 2, ylab = "Number of Events", xlab = "Year of event")
> abline(h = seq(from = 10, to = max(CSK), by = 10), v = pretty(aso),
+     lty = 2, col = gray(0.8))
> dev.off()
```

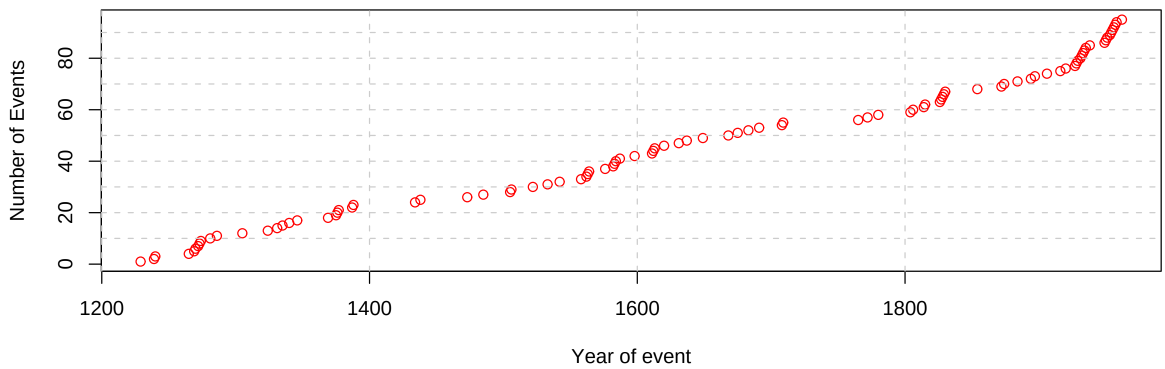


Figure 5: Cumulative Sum

1.1.3 Empirical Survivor

Empirical survivor plots show the percent of survivors versus that occur in time intervals longer than a specified time. if x_i is a specific time interval, we plot the propotion of time intervals Y

longer than x_i versus x_i.

```
> b = rep(0, length(a))
> sa = sort(a)
> for (i in 1:length(sa)) {
+     r = sum(sa >= sa[i])
+     b[i] = r
+ }
```

```
> postscript(file = "empsurv.eps", width = 10, height = 6, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(sa, b, type = "p", col = 2, ylab = "Percent survivors",
+     xlab = "Length of interval, years")
> dev.off()
```

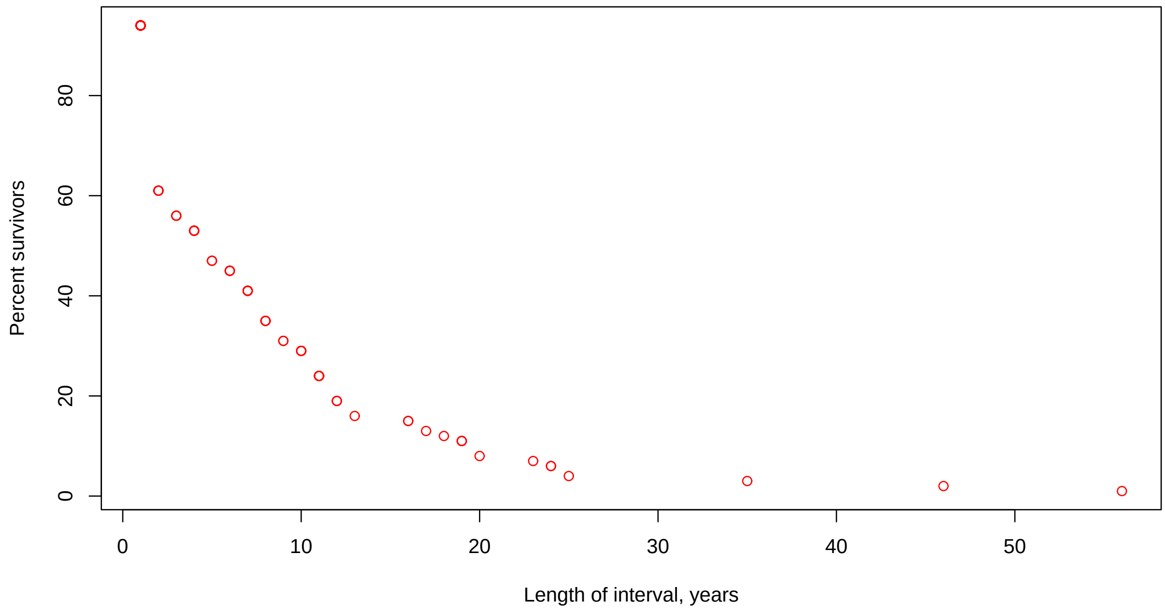


Figure 6: Cumulative Sum

It is common to plot the Empirical survivor plot with logarithmic axis - deviations from astraight line indicate

```
> sa = min(a):max(a)
> b = rep(0, length(sa))
> for (i in 1:length(sa)) {
+     r = sum(a >= sa[i])
+     b[i] = r
+ }
> postscript(file = "logempsurv.eps", width = 10, height = 6, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(sa, b, log = "y", type = "p")
> dev.off()
```

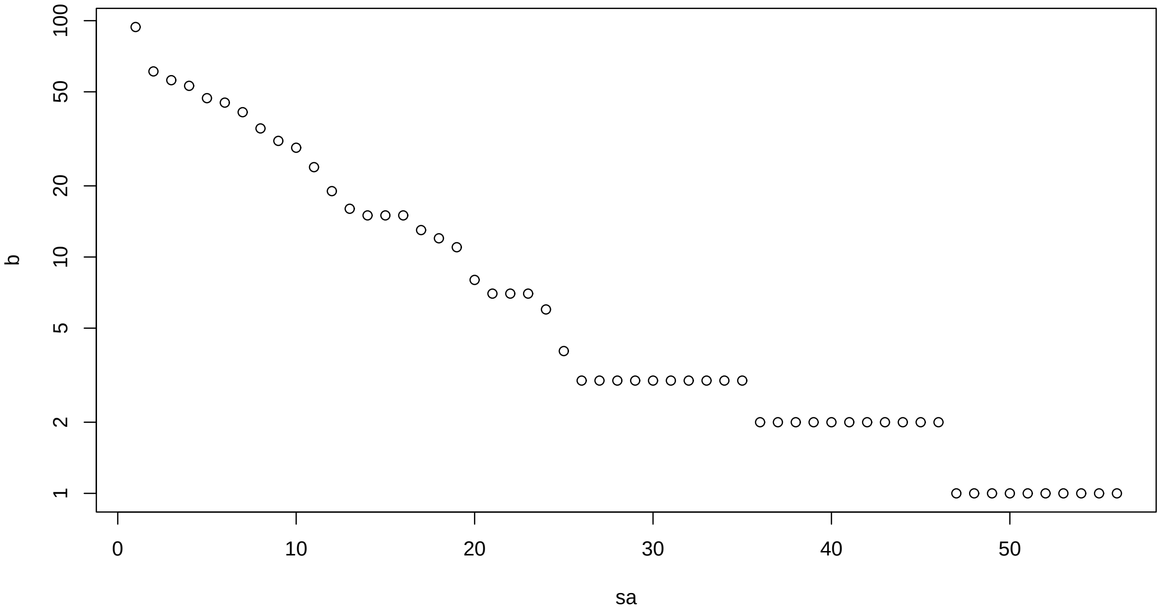


Figure 7: Cumulative Sum

1.1.4 Serial Correlation

Serial correlation of series of events indicates a non-random relation of one event to near by neighbors.

```
> a1 = aso[1:(length(aso) - 2)]
> a2 = aso[2:(length(aso) - 1)]
> a3 = aso[3:(length(aso))]
```

```
> postscript(file = "sericor.eps", width = 10, height = 6, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> plot(a3 - a2, a2 - a1, type = "p", main = "Serial Correlation, Aso",
+     xlab = "t[i+1]-t[i]", ylab = "t[i]-t[i-1]")
> dev.off()
```

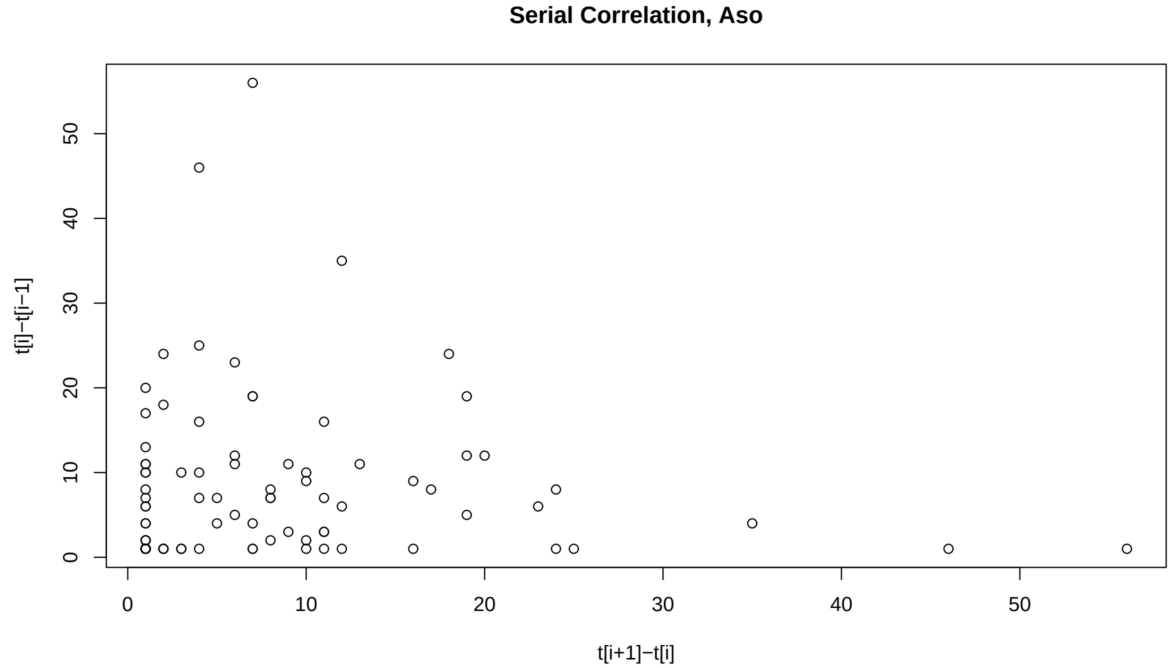


Figure 8: Cumulative Sum

1.2 Regular Time Series

When time series have a predetermined spacing or time interval the we refer to them as properly time series.

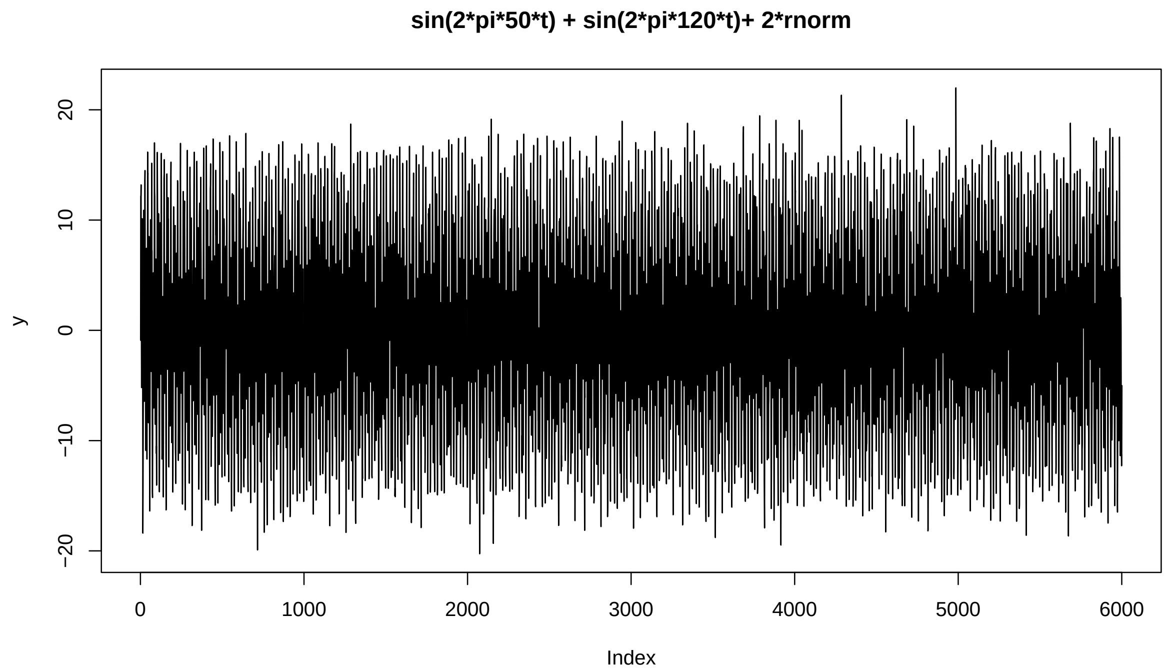


Figure 9: Example Time Series

1.3 Correlation

1.4 Fourier Analysis

1.4.1 Meaning of the DFT

We can create a time series with one sinusoidal frequency and see how the fourier transform provides information about the underlying power spectrum of the signal.

The signal can be represented by,

$$x_k = A \sin\left(\frac{2\pi k}{T}\right) \tag{1.1}$$

Where $k = 0, 1, 2, \ldots, (MT - 1)$ is the index of the samples and $\frac{2\pi}{T}$ is the frequency of the sinusoid. There are T samples per cycle and by the Nyquist theorem we expect that T 2. We will use M cycles such that the total number of samples is $N = MT$.

Given this signal we can compute the DFT following the definition,

$$
\begin{aligned}
X_m &= \sum_{k=0}^{N-1} x_k e^{-i2\pi km/N} && (1.2)\\
&= \sum_{k=0}^{N-1} A\sin\left(\frac{2\pi k}{T}\right) e^{-i2\pi km/N} && (1.3)\\
&= \sum_{k=0}^{N-1} \frac{A}{2i}\left[e^{\frac{2\pi ik}{T}} - e^{\frac{-2\pi ik}{T}}\right] e^{-i2\pi km/N} && (1.4)\\
&= \sum_{k=0}^{N-1} \frac{A}{2i}\left[e^{\frac{2\pi ik}{T}-i2\pi km/N} - e^{\frac{-2\pi ik}{T}-i2\pi km/N}\right] && (1.5)
\end{aligned}
$$

Next we use the fact that the sum of the sinusoids sum to zero if the arguments in the exponents is non-zero,

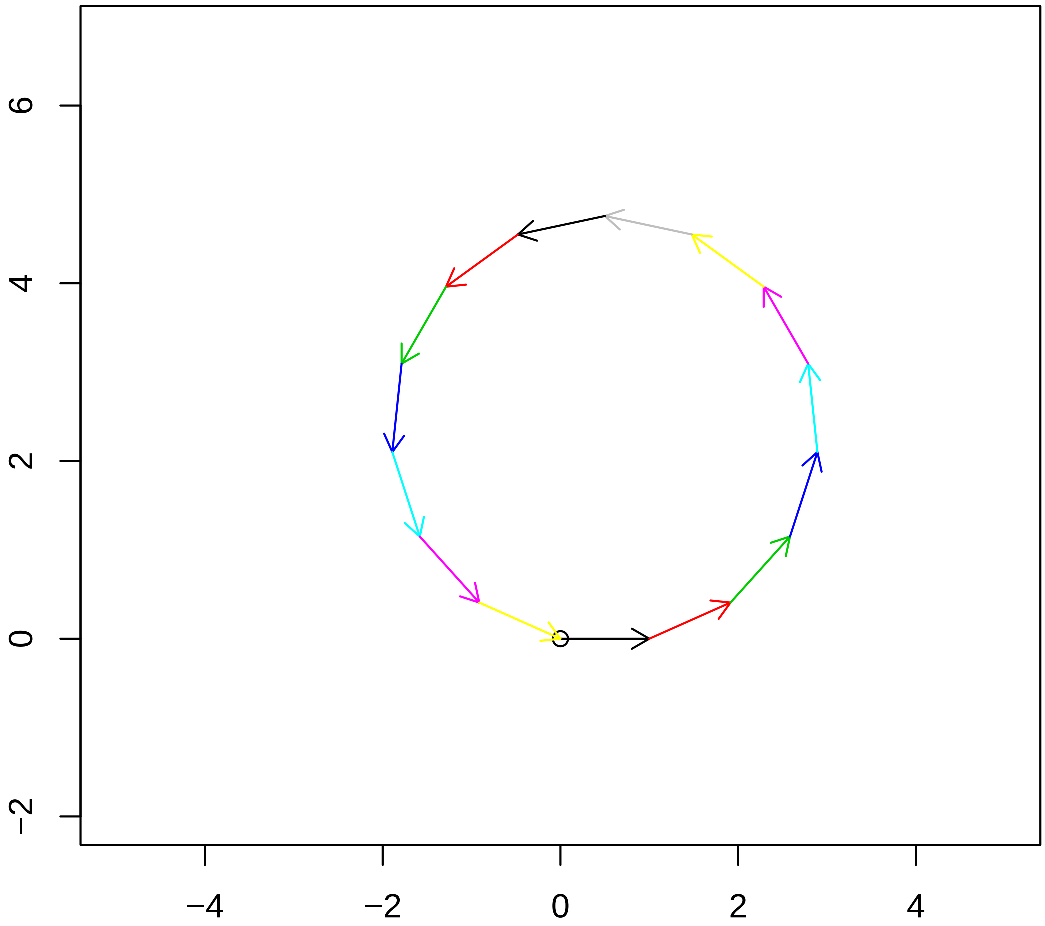
$$
\sum_{\nu=0}^{N-1} e^{i2\pi\nu(k'-k)/N} = N\delta_{k'k} \qquad (1.6)
$$

and the Kronecker delta is defined as,

$$
\delta_{k'k} = \begin{cases} 0 & \text{if } k' \neq k \\ 1 & \text{if } k' = k \end{cases} \qquad (1.7)
$$

Since complex exponentials can be represented in the complex plane as vectors the sum can be illustrated by a vector diagram for the case where $k' \neq k$,

```
> N = 15
> nu = 0:(N - 1)
> R = cos(2 * pi * nu/N)
> I = sin(2 * pi * nu/N)
> plot(c(-N/3, N/3), c(-0.2, N/3), type = "n", asp = 1, xlab = "",
+     ylab = "")
> x1 = 0
> y1 = 0
> points(x1, y1)
> for (i in 1:N) {
+     x2 = x1 + R[i]
+     y2 = y1 + I[i]
+     arrows(x1, y1, x2, y2, length = 0.1, col = i)
+     x1 = x2
+     y1 = y2
+ }
```



In the case where Since, $k' = k$ the vectors sum along the real axis and add up to N.

Using this mathematical fact and noting that $\left(\frac{2\pi k}{T} + 2\pi km/N\right)$ is never zero,

$$\begin{aligned}
X_m &= \sum_{k=0}^{N-1} \frac{A}{2i}\left[e^{\frac{2\pi ik}{T}-i2\pi km/N} - e^{-i\frac{2\pi k}{T}-i2\pi km/N}\right] && (1.8)\\
&= \sum_{k=0}^{N-1} \frac{A}{2i}\left[e^{i\left(\frac{2\pi k}{T}-2\pi km/N\right)} - e^{-i\left(\frac{2\pi k}{T}+2\pi km/N\right)}\right] && (1.9)\\
&= \sum_{k=0}^{N-1} \frac{A}{2i}e^{i\left(\frac{2\pi k}{T}-2\pi km/N\right)} - 0 && (1.10)\\
&= \sum_{k=0}^{N-1} \frac{A}{2i}e^{i2\pi k\left(\frac{1}{T}-\frac{m}{N}\right)} && (1.11)\\
&= \sum_{k=0}^{N-1} \frac{A}{2i}e^{i2\pi k\frac{N-mT}{TN}} && (1.12)
\end{aligned}$$

So when $\frac{1}{T} = \frac{m}{N}$ the exponential sum is non-zero and thus

$$X_m = \frac{A}{2i}N \tag{1.13}$$

or using $N = MT$,

$$X_m = \begin{cases} \frac{AMT}{2i} & \text{when } m = M \\ 0 & \text{otherwise} \end{cases} \tag{1.14}$$

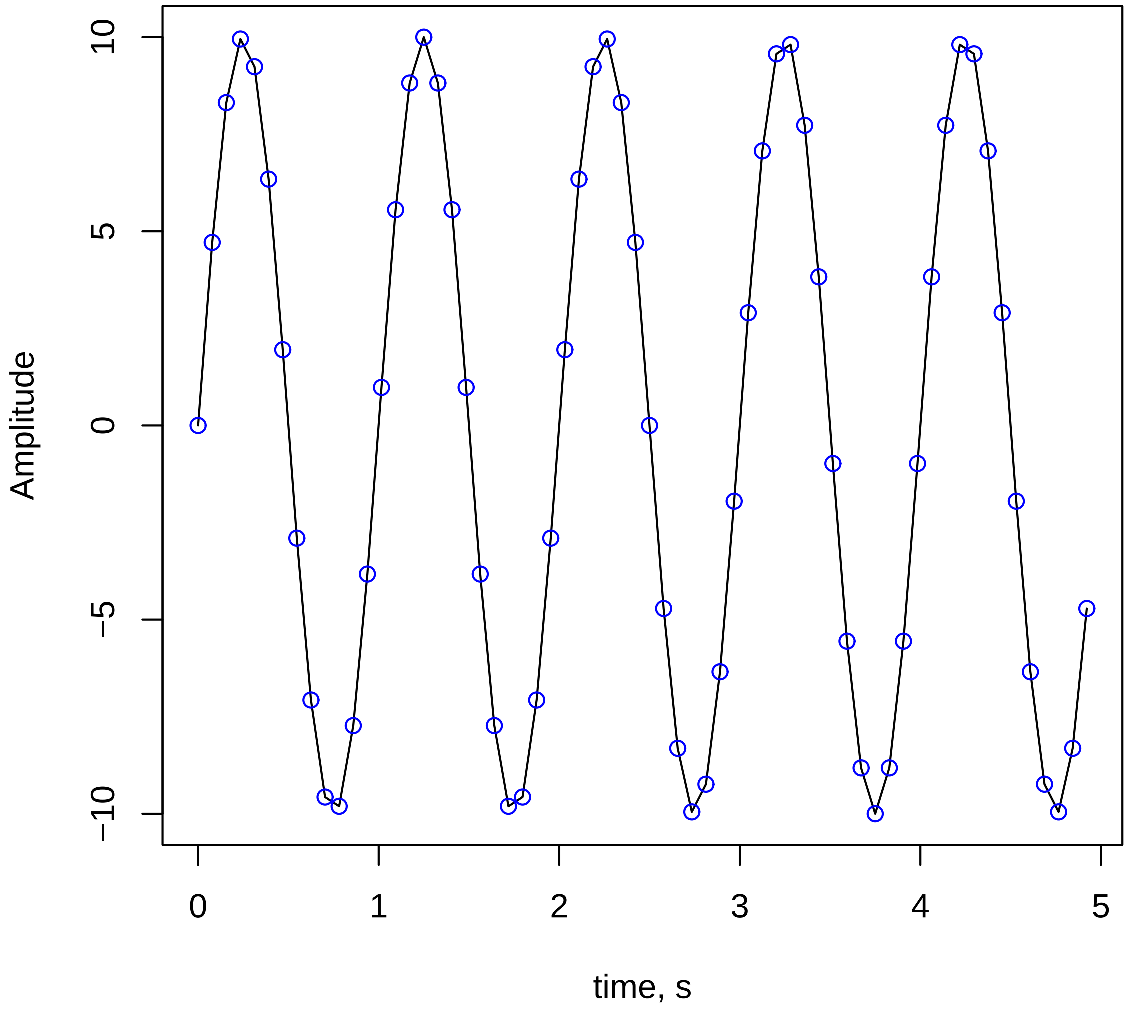
The Discrete Fourier Transform is zero everywhere except at one frequency when $m = M$ and then it has an amplitude $AMT/2i$. If the power spectrum is defined as,

$$P_m \equiv \frac{1}{N}X_m X_m^* = \frac{1}{N}\left|X_m\right|^2 \tag{1.15}$$

the power density is,

$$P_m = \frac{A^2 N}{4} \tag{1.16}$$

As an example, consider a signal of 64 points. The signal will have 5 cycles and an amplitude of 10 . Plotting this sinusoid



Since there are 5 cycles in 64 samples the sample rate is 12.8 and frequencies are calculated at $m = 0, 1, 2, \ldots, N/2$

```
> Xm = fft(xk)
> Pm = (Mod(Xm)^2)/N
> Pmax = max(Pm[0:32])
> f = seq(from = 0, to = N/2, by = 1)
```

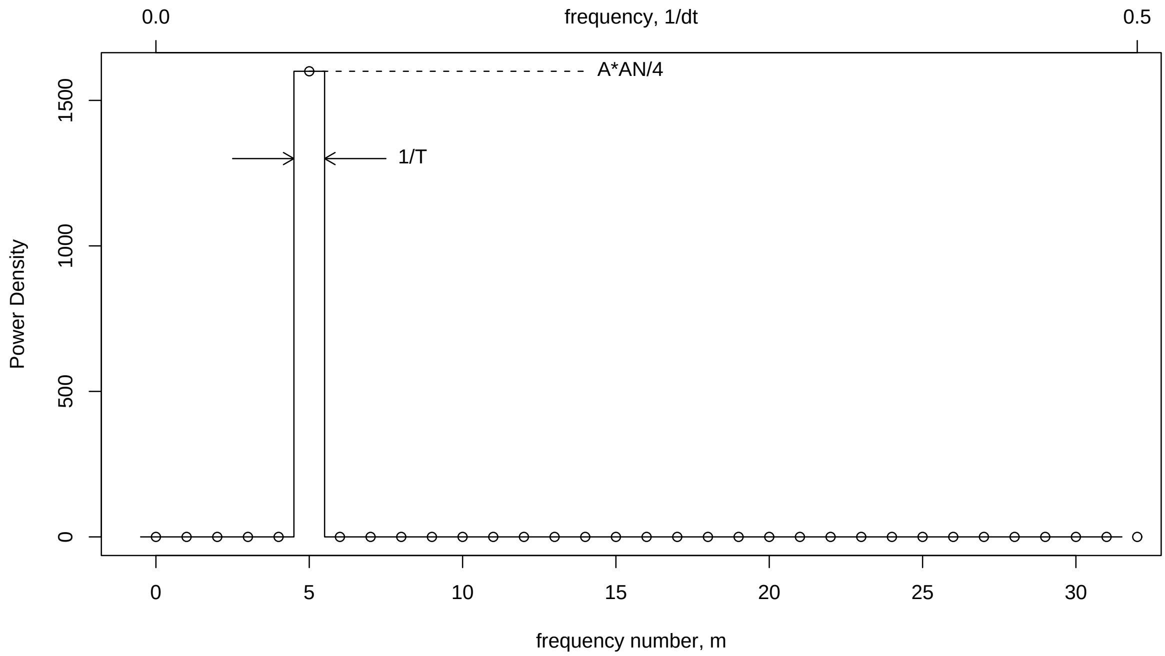


Figure 10: Understanding the DFT: Power Spectrum Density

1.4.2 Periodogram

The above computations can be used to estimate the power spectrum of any signal.

1.4.3 Improvements

1.4.4 Fourier Transform

Suppose we have a function, $y = f(x)$. The formal definition of the Fourier Transform ($\mathcal{FT}$) for continuous signals is

$$Y(\omega) = \mathcal{FT}[y] = \int_{-\infty}^{\infty} f(x) e^{-i2\pi x\omega} dx \tag{1.17}$$

The information in the time series in one domain is exactly the same as in the other domain. Taking the fourier transofrm only transforms the data, nothing else. It will be useful to be able to "jump" from one domain to the other at least mentally and calculationally. In different circumstances

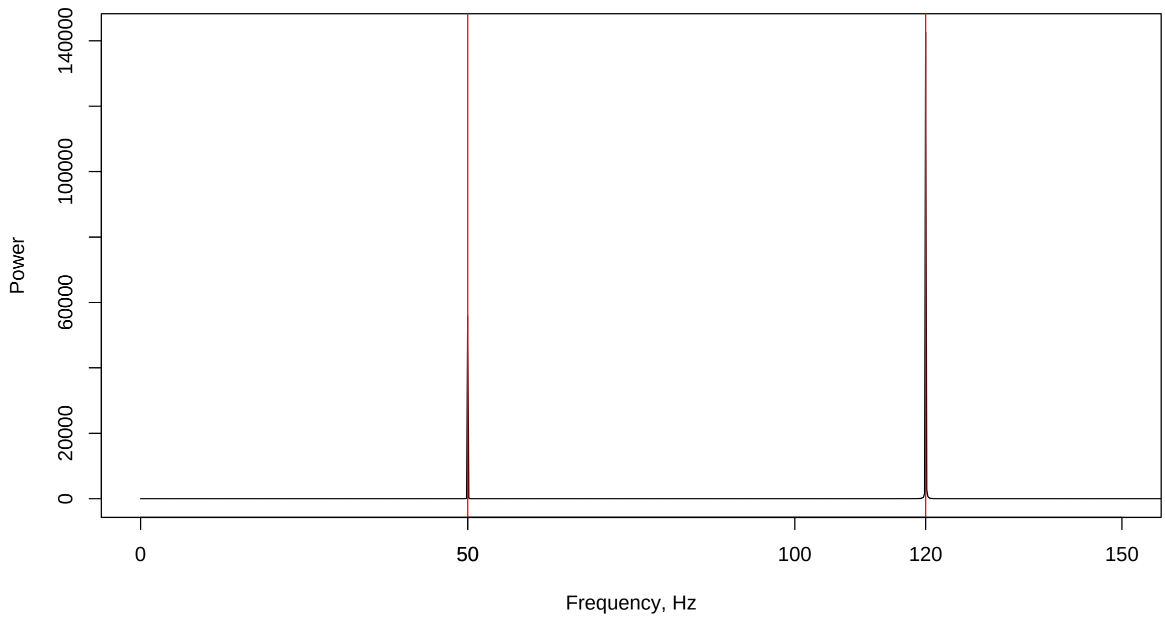


Figure 11: Periodogram

it is convenient to think ofthe signal in one domain or the other, depending on the situation.

What happens if the time series is shifted by a (time) units? Then,

$$Y(\omega) = \mathcal{FT}\left[f(x-a)\right] = \int_{-\infty}^{\infty} f(x-a)e^{-i2\pi x\omega}dx \tag{1.18}$$

sustituing $u = x - a$ and $du = dx$ gives

$$\begin{aligned} \int_{-\infty}^{\infty} f(u)e^{-i2\pi(u+a)\omega}du &= \int_{-\infty}^{\infty} f(u)e^{-i2\pi u\omega}e^{-i2\pi a\omega}du & (1.19) \\ &= e^{-i2\pi a\omega}Y(\omega) & (1.20) \end{aligned}$$

Shifting the time series by a constant is equivalent to multiplying the fourier transofrm by an exponential, $e^{-i2\pi a\omega}$. The shift theorem will be used in the derivation of other results, such as the convolution theorem and the derivative theorem.

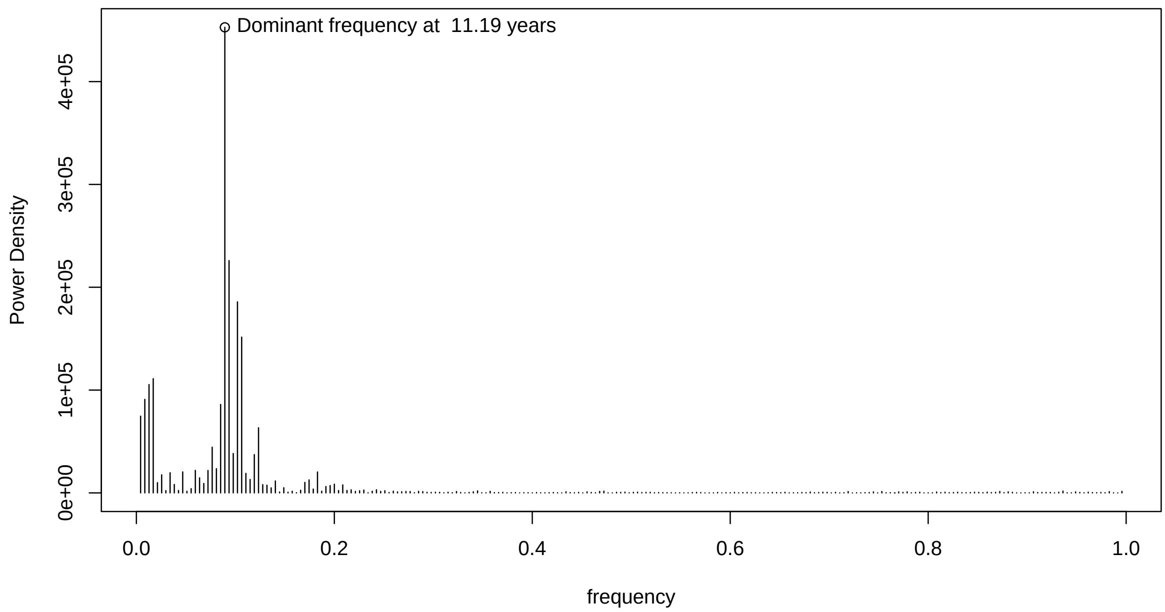


Figure 12: Sun Spots Power Spectrum/Periodogram

1.4.5 Convolution

Convolution is the correlation of one time series with the time reversed version of another. In symbols this can be represented as,

$$C(t) = A \otimes B = \int A(t)B(-t)dt \tag{1.21}$$

where C is the convolution of the two time series A and B. Convolutioin is everywhere in the earth sciences, and perhaps in all science. Consider a seismogram, $SIGNAL$. It consists of a rupture(source, S), the waveform through the earth (path, E) and the recording device (instrument, I). This can be represented symbolucally as,

$$SIGNAL = S \otimes E \otimes I \tag{1.22}$$

Seismologists are interested in all three of these components - sometimes specialists concentrate on just one of these and consider the other parts noise. There work often involves spending considerable time removing the noise to extract the part ofthe seismogram they consider signal. This operation is called deconvolution.

Consider a thermometer reading that is conducted once per hour. If a measurement is recorded during a sudden drop in temperature, it takes a finite amount of time for the mercury in the thermaometer to respond to the sudden change in temperature. Furthermore, there is a a lag in operation of actually recording the temperature in a notebook, transcribing the information as accurately as one can, and there is some error in actually estimating the correct position of the mercury height. These processes, even in the best circumstances, distort the underlying, or 'true', temperature variations introducing noise. If, on the other hand we could estimate the exact nature of the distortion introduced, it may be possible to remove the effect by deconvolution.

The convolution theorem states that convolution in the time domain is equivalent to multiplication in the frequency domain. Formally this is written,

$$\mathcal{FT}\left[f \otimes g\right] = \mathcal{FT}\left[f\right]\mathcal{FT}\left[g\right] = F(\omega)G(\omega) \tag{1.23}$$

Using the symbol $\Leftrightarrow$ to denote a Foutier Transform pair,

$$\mathcal{FT}\left[f \otimes g\right] \Leftrightarrow F(\omega)G(\omega) \tag{1.24}$$

The shift theorem and the definition of the Foutier Transform provides a proof of the convolution theorem,

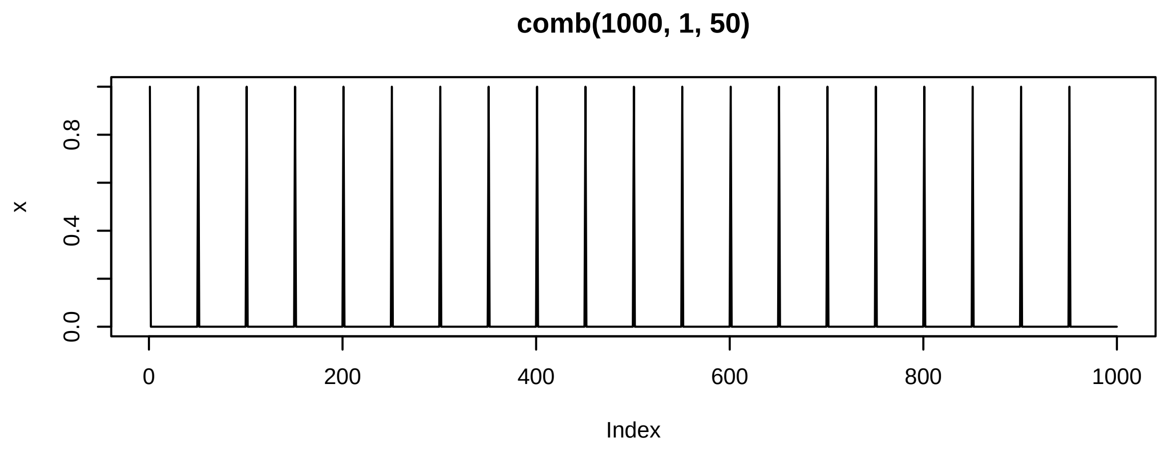
$$\begin{aligned}
\mathcal{FT}\left[f \otimes g(t-\tau)\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tau)g(t-\tau)e^{-i2\pi t\omega}d\tau dt \\
&= \int_{-\infty}^{\infty} f(\tau)\int_{-\infty}^{\infty} g(t-\tau)e^{-i2\pi t\omega}dt d\tau \\
&= \int_{-\infty}^{\infty} f(\tau)e^{-i2\pi\tau\omega}G(\omega)d\tau \\
&= G(\omega)\int_{-\infty}^{\infty} f(\tau)e^{-i2\pi\tau\omega}d\tau \\
&= F(\omega)G(\omega)
\end{aligned}$$

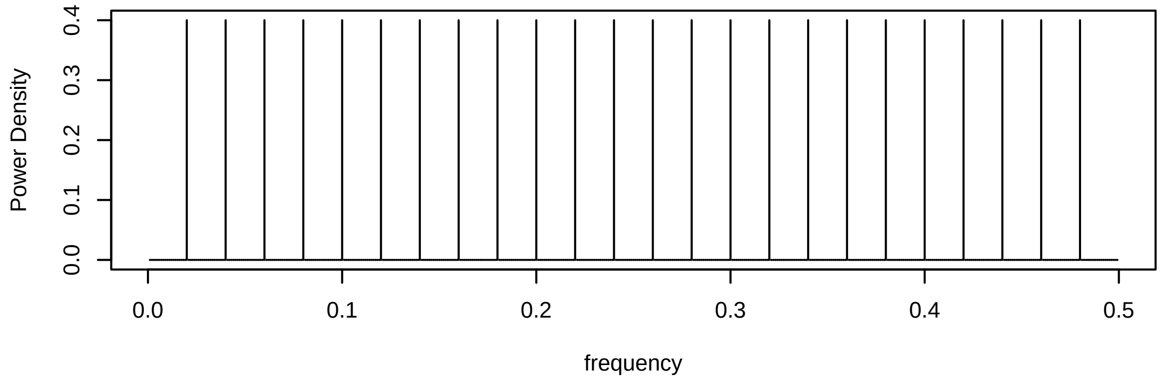
Noting that τ is just a dummy variable for integration.

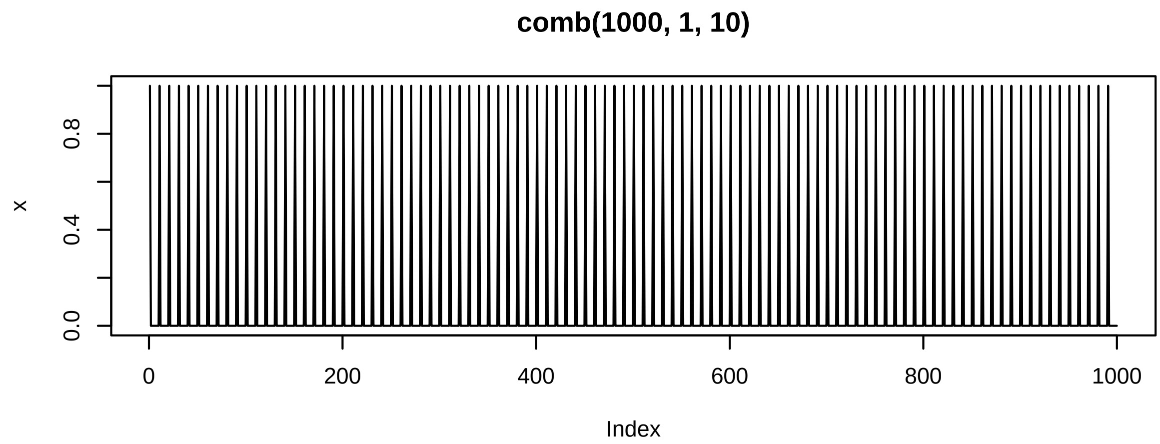
The sampling function can be considered like a comb. Every prong is a sample. If the prongs are close together sampling is dense. The Fourier Transform of a comb function is a comb function. As the sampling density in the time domain increases the spacing in the frequency domain decreases, and vice versa. We can see this by experimenting with the comb function and its DFT.

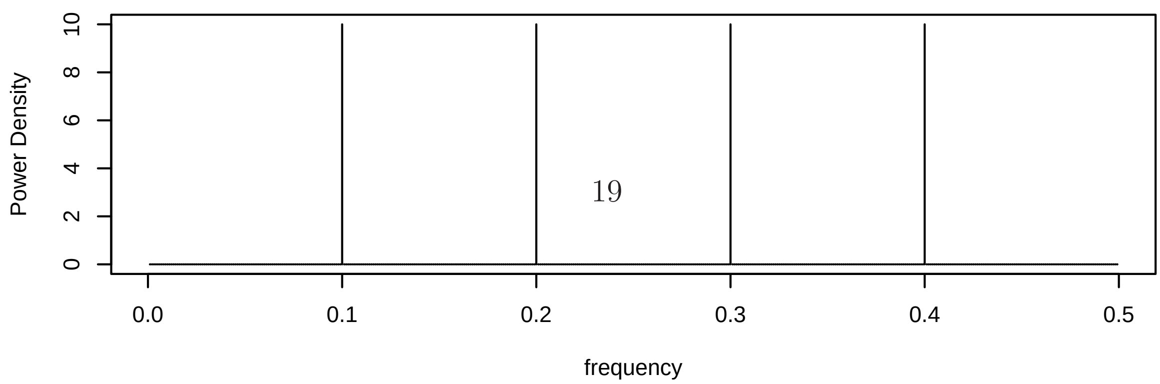
```
> postscript(file = "combs.eps", width = 6, height = 10, paper = "special",
+     horizontal = FALSE, onefile = TRUE, print.it = FALSE)
> par(mfrow = c(4, 1))
> x = comb(1000, 1, 50)
> title(main = "comb(1000, 1, 50)")
> BetterPowerspec(x, deltat = 1)
> x = comb(1000, 1, 10)
> title(main = "comb(1000, 1, 10)")
> BetterPowerspec(x, deltat = 1)
> dev.off()
```

The convolution theorem can be used in a variety of contexts to understand many, much more complex interactions and processes. As an example, consider some process that is assumed to be stationary in time. This could be a long running climate fluctuation recorded in ice cores or sediment records.


comb(1000, 1, 50)
x
0.0
0.4
0.8
0
200
400
600
800
1000
Index


Power Density
0.0
0.1
0.2
0.3
0.4
0.0
0.1
0.2
0.3
0.4
0.5
frequency


comb(1000, 1, 10)
x
0.0
0.4
0.8
0
200
400
600
800
1000
Index


Power Density
0
2
4
6
8
10
0.0
0.1
0.2
0.3
0.4
0.5
frequency