

1 To better mimic the sampling process, a negative binomial error was added to
 2 the abundance counts every time a sample was randomly drawn from the simulated assemblage
 3 in the *BiogTest* randomization algorithm. The negative binomial distribution was used to generate
 4 realistic heterogeneity that often results from spatial clustering of individuals and other small-scale
 5 processes. In practice this implies that each sample is drawn from a slightly different vector of
 6 abundance counts. An example for abundance-based data better illustrate this process. For better
 7 reproducibility of this example, we have included the R code that generates it, but note that due to
 8 the random nature of the process itself, application of the code will generate each time different
 9 results. Let's have two samples with the following counts of observed species abundances.

```
> x <- data.frame(sample1 = c(5, 7, 1, 0, 1), sample2 = c(0, 10, 1, 2, 1))
> x

  sample1 sample2
1        5        0
2        7       10
3        1        1
4        0        2
5        1        1
>
```

10 The total number of individuals in each sample is 14 and 15 respectively. Now let's create a
 11 simulated assemblage with a log-normal distribution of species abundances from which to draw
 12 random samples:

```
> library(rareNMtests)
> SR <- do.call("rbind", apply(x, 2, chao1))
> mr <- match(max(SR[,2]), SR[,2])
> range.SR <- c(SR[mr, 4], SR[mr, 5])
> richness <- round(runif(1, range.SR[1], range.SR[2]))
> sdlog <- runif(1, 0.1, 3.5)
> com <- rlnorm(richness, 0, sdlog)
> com <- ceiling(com/min(com[com>0]))
> com

 [1] 478  36  12   9  27  51  47  25  54  88  11  22  10 164 1194
[16]  16  19   1  43  67   3  18
>
```

13 The algorithm will now extract two random samples (with replacement) of 14 and 15 individ-
 14 uals from this assemblage, respectively, with probability weights proportional to the abundance
 15 counts in the assemblage. The vector of probabilities is thus the result of adding a negative bino-
 16 mial error to the abundance count of each species in the assemblage (i.e. the expectation μ of the
 17 negative binomial). The variance of the negative binomial is:

$$var = \mu + \frac{\mu^2}{k}$$

18 where k is the dispersion parameter. For every species, k was randomly drawn from a uniform
 19 distribution between 0.01 and 25 each time a sample was drawn from the assemblage, as shown in
 20 the example below:

```
> prob1 <- rbinom(n = length(com), mu=com, size=runif(length(com), 0.01, 25))
> prob1

[1] 670 22 6 14 25 0 37 36 77 61 11 19 9 186 394 11 21 2 44
[20] 88 4 28

> prob2 <- rbinom(n = length(com), mu=com, size=runif(length(com), 0.01, 25))
> prob2

[1] 351 51 12 17 31 43 41 24 41 76 13 38 9 228 1622
[16] 14 20 0 95 49 0 23

>
```

21 And now the assemblage is sampled using these vector of probabilities for each of the samples.

```
> sample1 <- sample(1:length(com), sum(x[,1]), replace=TRUE, prob=prob1)
> sample1 <- data.frame(table(sample1))
> colnames(sample1) <- c("species", "sample1")
> sample2 <- sample(1:length(com), sum(x[,2]), replace=TRUE, prob=prob2)
> sample2 <- data.frame(table(sample2))
> colnames(sample2) <- c("species", "sample2")
> df <- merge(sample1, sample2, by="species", all=TRUE)
> df[is.na(df)] <- 0
> df

  species sample1 sample2
1      1         5       0
2      8         1       0
3      9         3       0
4     15         4      11
5     20         1       0
6     14         0       1
7     17         0       1
8     19         0       1

>
```