NOTES ON INPUT VARIABLES:
----

1) MAT

This must be a matrix. Rows denote subjects. In each row i, the event times for subject i are given. For example, suppose we have the following event time data:

```
subject 1:  9
subject 2:  11   17
subject 3:  11
subject 4:  17
subject 5:  10   14
subject 6:  16   17
subject 7:  11   12   16   17   19
subject 8:  2    9    11
subject 9:  10   16
subject 10: 5    7    9   16
```

Then MAT should be ....

```
> MAT
      [,1] [,2] [,3] [,4] [,5]
 [1,]    9   NA   NA   NA   NA
 [2,]   11   17   NA   NA   NA
 [3,]   11   NA   NA   NA   NA
 [4,]   17   NA   NA   NA   NA
 [5,]   10   14   NA   NA   NA
 [6,]   16   17   NA   NA   NA
 [7,]   11   12   16   17   19
 [8,]    2    9   11   NA   NA
 [9,]   10   16   NA   NA   NA
[10,]    5    7    9   16   NA
```

NOTE: For each row, after the last event time is given, the rest of the row should be filled in with NA's, as shown.

Event times must be non-negative integers (0's are allowed).

2) tau

This is a vector of the censoring times for each subject. tau[i] should be the censoring time for the subject corresponding to MAT[i,]. Censoring times must be non-negative.

3) Xc

This may be either a vector, matrix, or 3-dimensional array.

If Xc is a vector, its length must equal the number of rows of MAT. It defines the value of 1 time-independent covariate for each subject.

If Xc is a matrix, it defines the value of 1 or more time-independent covariates for each subject. Rows correspond to subjects, columns to covariates. Thus, MAT and Xc must have the same number of rows.

If Xc is a 3-dimensional array, it defines the value of 1 or more covariates for each subject at each distinct event time. Dimension 1 denotes time points; dimension 2, covariates; dimension 3, subjects. Suppose that there are k subjects, p covariates, and r distinct event times across all subjects. Then dim(Xc) = c(r,p,k) and Xc[j,l,i] is the value of covariate l for subject i at the jth distinct time point.

If Xc is not specified, then the common cumulative mean function (Nelson-Aalen estimate) will be calculated.

4) eps

Newton's Method is used to calculate the estimate of beta (B). In a given cycle of Newton's Method, let B be the current value of B and B.old its value at the previous iteration. Convergence is concluded when

```
sum(abs((B-B.old)/B)) <= eps
```

The default value of eps is 0.0001.

5) times

This vector is only used when no covariates are specified. It specifies at which time points the covariance matrix of the estimated cumulative mean function (CMF) should be calculated.

6) mtime

This is a single time value, used only when covariates are specified. In the covariate case, the covariance matrix of (B, Mo(t)) is calculated, where B is the estimator of the beta vector and Mo(t) is the estimator of the baseline cumulative mean function at time t. "mtime" specifies this time t.

7) cutpts

This vector of cut points defines the intervals for piecewise constant intensity functions. The vector should consist of all the cut points, including the lower and upper bounds. Thus, a vector of length r+1 will result in r intervals.

Suppose that we specify cutpts = c(0,100,200,300). This corresponds to specifying the intervals [0,100], (100,200], (200,300]. Note that for each interval the upper bound is included in the interval but the lower bound is not. The exception is the first interval, which includes the lower bound as well.

All the values in cutpts must be non-negative.

Notes concerning models with piecewise constant intensity functions:

1) Specifying a vector "cutpts" will result in a model with piecewise constant intensity functions. If "cutpts" is not specified in the function call, then the baseline cumulative mean function will be estimated at each distinct failure time.

2) This type of model is only implemented for the covariate case. Therefore, if cutpts is specified, Xc must be specified as well.

3) For this type of model, "mtime" must be the upper bound of one of the intervals defined by cutpts. Therefore, if cutpts = c(0,100,200,300) then mtime must be one of 100, 200, or 300.

4) In some situations it may be desirable to have an event time of 0 be an interval unto itself. (In other words, we allow probability mass at time 0.) Suppose we want [0,0] to be the first "interval" and then the succeeding intervals to be (0,100], (100,200], and (200,300]. Then specify cutpts = c(0,0,100,200,300).

5) All event times and censoring times must fall within the intervals defined by cutpts.

6) Each interval must contain at least 1 event.

```
********************** EXAMPLE WITH NO COVARIATES ************************

We have 10 subjects with the following event data:

subject 1:  9
subject 2:  11   17
subject 3:  11
subject 4:  17
subject 5:  10   14
subject 6:  16   17
subject 7:  11   12   16   17   19
subject 8:  2    9    11
subject 9:  10   16
subject 10: 5    7    9   16

Therefore, we create ...

> MAT
      [,1] [,2] [,3] [,4] [,5]
 [1,]    9   NA   NA   NA   NA
 [2,]   11   17   NA   NA   NA
 [3,]   11   NA   NA   NA   NA
 [4,]   17   NA   NA   NA   NA
 [5,]   10   14   NA   NA   NA
 [6,]   16   17   NA   NA   NA
 [7,]   11   12   16   17   19
 [8,]    2    9   11   NA   NA
 [9,]   10   16   NA   NA   NA
[10,]    5    7    9   16   NA

The 10 subjects were censored at the following times:

> tau
 [1] 17 17 19 18 15 17 20 19 17 19


We want the estimated covariance matrix of CMF at the following times:

> times
[1]  4 10 17
```

Run the function with the following command:

```
> recurrent(MAT,tau,times=times)

$dft:
 [1]  2  5  7  9 10 11 12 14 16 17 19

$CMF:
 [1] 0.100000 0.200000 0.300000 0.600000 0.800000 1.200000 1.300000 1.400000
 [9] 1.844444 2.288889 2.538889

$times:
[1]  4 10 17

$Vr:
            [,1]       [,2]        [,3]
[1,] 0.009000000 0.01200000 0.006123457
[2,] 0.012000000 0.09600000 0.053308642
[3,] 0.006123457 0.05330864 0.127319616

$Vp:
     [,1] [,2]      [,3]
[1,] 0.01 0.01 0.0100000
[2,] 0.01 0.08 0.0800000
[3,] 0.01 0.08 0.2387654
```

Output:
----

1) "dft" is the vector of distinct failure times across all subjects
2) "CMF" shows the value of the estimated cumulative mean function at each distinct failure time
3) "times" is the vector of time values defining the covariance matrix of the CMF
4) "Vr" is the robust estimated covariance matrix of the cumulative mean function at the time points specified
5) "Vp" is the estimated covariance matrix of the cumulative mean function under the Poisson process assumption

For example, at time 10, the value of the estimated cumulative mean function is 0.80 and its robust variance is 0.096. Its variance under the Poisson process assumption is 0.08. The robust estimate of the covariance of the CMF at time 10 and the CMF at time 17 is 0.053.

```
********************** EXAMPLE WITH COVARIATES ***********************

Suppose we have the same 10 subjects as above.

We now have 2 time-independent covariates.  We define the matrix ...

> Xc
      [,1] [,2]
 [1,]    0    4
 [2,]    0    8
 [3,]    0    2
 [4,]    0    5
 [5,]    0    3
 [6,]    1    8
 [7,]    1    1
 [8,]    1    4
 [9,]    1    7
[10,]    1    9


We choose eps = 0.0001 and mtime = 6.


> recurrent(MAT,tau,Xc,eps=0.0001,mtime=6)

$B:
[1]  0.76292475 -0.04105348

$dft:
 [1]  2  5  7  9 10 11 12 14 16 17 19

$m0:
 [1] 0.07873053 0.07873053 0.07873053 0.23619158 0.15746105 0.31492211
 [7] 0.07873053 0.07873053 0.33848302 0.33848302 0.15920268

$M0:
 [1] 0.07873053 0.15746105 0.23619158 0.47238316 0.62984421 0.94476632
 [7] 1.02349685 1.10222737 1.44071040 1.77919342 1.93839610

$mtime:
[1] 6

$Vr:
            [,1]         [,2]          [,3]
[1,]  0.066657724 0.0011661212 -0.0048248202
[2,]  0.001166121 0.0014606402  0.0009107667
[3,] -0.004824820 0.0009107667  0.0064716160
```

```
$Vp:
              [,1]         [,2]       [,3]
[1,]  0.214218679 -0.006421532 0.10553824
[2,] -0.006421532  0.005717559 0.79357322
[3,]  0.105538236  0.793573223 0.01730807

$count:
[1] 4
```

```
Output:
-------

1) "B" is the estimate of beta.
2) "dft" is the vector of distinct failure times across all subjects
3) "m0" is the value of the baseline mean function at each distinct
   failure time
4) "M0" is the value of the baseline cumulative mean function at
   each distinct failure time
5) "Vr" is the robust estimate of the covariance matrix of the estimators
   (B, M0(t)) where t = mtime
6) "Vp" is the estimate of the covariance matrix of the estimators
   (B, M0(t)) where t = mtime under the Poisson process assumption
7) "count" is the number of iterations of Newton's Method until convergence

For example, the robust estimate of the covariance of B[1] and B[2]
is 0.00117.  The robust variance estimate of M0(t) is 0.0065.
The robust covariance estimate of B[1] and M0(t) is -0.0048.

Under the Poisson process assumption, the estimate of the covariance
of B[1] and B[2] is -0.00642.  The variance estimate of M0(t) is 0.0173.
The covariance estimate of B[1] and M0(t) is 0.10554.
```

*************** EXAMPLE WITH TIME-DEPENDENT COVARIATES ****************

Suppose we have time-dependent covariates. Xc is now defined as a 3-dimensional array with dim(Xc) = c(r,p,k), where r is the number of distinct event times (across all subjects), p is the number of covariates, and k is the number of subjects.

Suppose we have two covariates, 1 time-independent, 1 time-dependent. There are 11 distinct event times (see "dft" in output above).

```
> Xc

, , 1
      [,1] [,2]
 [1,]    0    3
 [2,]    0    5
 [3,]    0    6
 [4,]    0    9
 [5,]    0    2
 [6,]    0    4
 [7,]    0    6
 [8,]    0    7
 [9,]    0    3
[10,]    0    5
[11,]    0    2

, , 2
      [,1] [,2]
 [1,]    0    2
 [2,]    0    4
 [3,]    0    4
 [4,]    0    1
 [5,]    0    6
 [6,]    0    6
 [7,]    0    3
 [8,]    0    5
 [9,]    0    7
[10,]    0    4
[11,]    0    2

.
.
.

, , 10
      [,1] [,2]
 [1,]    1    3
 [2,]    1    3
```

```
 [3,]    1    7
 [4,]    1    4
 [5,]    1    7
 [6,]    1    4
 [7,]    1    2
 [8,]    1    1
 [9,]    1    1
[10,]    1    8
[11,]    1    9
```

```
> recurrent(MAT,tau,Xc,eps=0.0001,mtime=6)

$B:
[1]  0.8005536 -0.1789978

$dft:
 [1]  2  5  7  9 10 11 12 14 16 17 19

$m0:
 [1] 0.1276202 0.1734543 0.1570566 0.4597648 0.2642011 0.4909697 0.1090065
 [8] 0.1283023 0.5856040 0.6379140 0.2668905

$M0:
 [1] 0.1276202 0.3010744 0.4581310 0.9178959 1.1820969 1.6730667 1.7820732
 [8] 1.9103755 2.4959795 3.1338935 3.4007839

$mtime:
[1] 6

$Vr:
             [,1]         [,2]        [,3]
[1,]  0.047953366 -0.001000681  0.00437368
[2,] -0.001000681  0.006881481 -0.01159470
[3,]  0.004373680 -0.011594696  0.04486017

$Vp:
             [,1]         [,2]      [,3]
[1,]  0.211619802 -0.006467119 0.1819169
[2,] -0.006467119  0.012291936 1.4174589
[3,]  0.181916853  1.417458914 0.0747382

$count:
[1] 4
```

********* EXAMPLE WITH PIECEWISE CONSTANT INTENSITY FUNCTIONS *********

Now suppose we want to fit a model where the baseline intensity function (mean function) is constant within intervals. Thus, instead of estimating the mean function at each distinct failure time, we will only estimate a baseline mean function for each interval. This is useful when the number of distinct failure times is very large, since it speeds up the computations.

Suppose we wish to define the intervals as [0,5], (5,10], (10,15], (15,20].

In creating these intervals, we can now think of each interval as a "distinct failure time". Thus, any time-dependent covariates will have a single value for each interval. (Note, of course, that for time-independent covariates, the Xc matrix will be a vector or matrix as before.)

Let's define Xc as follows:

```
> Xc

, , 1
     [,1] [,2]
[1,]    0    2
[2,]    0    5
[3,]    0    3
[4,]    0    4

, , 2
     [,1] [,2]
[1,]    0    7
[2,]    0    4
[3,]    0    6
[4,]    0    2

.
.
.

, , 10
     [,1] [,2]
[1,]    1    5
[2,]    1    6
[3,]    1    3
[4,]    1    5
```

Note that MAT and tau are still defined as before. The value of "mtime" must now be the value of the upper bound of one of the intervals.

```
> recurrent(MAT,tau,Xc,eps=0.0001,mtime=5,cutpts=c(0,5,10,15,20))

$B:
[1] 0.5283819 0.1629496

$dft:
[1]  5 10 15 20

$m0:
[1] 0.06638956 0.17969427 0.18392204 0.78855407

$M0:
[1] 0.06638956 0.24608383 0.43000587 1.21855994

$mtime:
[1] 5

$Vr:
              [,1]          [,2]          [,3]
[1,]  0.0996174796 -0.0006659604 -0.0029575901
[2,] -0.0006659604  0.0039620701 -0.0006069952
[3,] -0.0029575901 -0.0006069952  0.0017887175

$Vp:
            [,1]        [,2]        [,3]
[1,]  0.19015468 -0.01319892 0.042264433
[2,] -0.01319892  0.01031467 0.345931851
[3,]  0.04226443  0.34593185 0.003391849

$count:
[1] 9


Output:
-------

1) "B" is the estimate of beta.
2) "dft" is the upper bound of each interval
3) "m0" is the value of the (constant) baseline mean function for each
    interval multiplied by the length of the interval
4) "M0" is the value of the baseline cumulative mean function at
    the upper bound of each interval
5) "Vr" is the robust estimate of the covariance matrix of the estimators
   (B, M0(t)) where t = mtime
6) "Vp" is the estimate of the covariance matrix of the estimators
   (B, M0(t)) where t = mtime under the Poisson process assumption
7) "count" is the number of iterations of Newton's Method until convergence
```

Therefore, the constant baseline mean function for the first interval [0,5] is estimated as 0.06638956/(5-0) = 0.0133. The value of the baseline cumulative mean function is estimated as 0.06638956 at time 5 and 0.24608383 at time 10. To plot the baseline cumulative mean function, simply plot "dft" vs. "M0" and connect the dots with straight lines using the type="l" option.