MTS Preliminary Simulation Results

Daniel M. Swan

Tuesday, March 03, 2015

This document provides a summary of the preliminary simulation of the momentary time sampling (MTS) estimator using both the regular and the penalized log likelihood function. I will summarize the results of the simulation by characterizing:

- The Monte Carlo error for phi and zeta in both their natural parameterizations and the logit of phi and log of zeta
- The error in the estimates of logit phi in terms of the median and maximum bias and RMSE
- The error in the estimates of log zeta in terms of the median and maximum bisa and RMSE

Monte Carlo error

```
##            stat  MLE       30       60       90
## 1    log zeta  MLE 1.33e-02 2.94e-03 8.13e-04
## 2    log zeta MPLE 2.34e+01 1.24e+01 5.02e+00
## 3    log zeta  PLE 9.00e-04 9.61e-04 1.04e-03
## 4   logit phi  MLE 6.72e-02 4.81e-02 1.43e-02
## 5   logit phi MPLE 2.35e+01 1.16e+01 4.99e+00
## 6   logit phi  PLE 1.57e-03 9.32e-04 4.92e-04
## 7         phi  MLE 2.08e-05 8.74e-06 5.78e-06
## 8         phi MPLE 2.05e-05 9.00e-06 5.68e-06
## 9         phi  PLE 1.25e-05 6.74e-06 4.73e-06
## 10       zeta  MLE 2.07e+20 4.38e+25 1.25e+11
## 11       zeta MPLE 1.30e+40 2.01e+49 2.80e+54
## 12       zeta  PLE 4.81e-03 4.81e-03 4.80e-03
```

The maximum Monte Carlo error is quite small in all cases except for zeta in the unpenalized estimates for the natural parameterization of zeta are highly unstable. Given these results, the sample sizes chosen are sufficient in all cases but zeta. In the case of zeta the magnitude of the bias and the fact that unpenalized MTS estimator sometimes produces very large or infinite estimates of zeta, it may be that there is no reasonable sample size large enough to provide adequately small Monte Carlo error.

Logit phi results

Median bias

```
##    K_intervals k_priors        5      10      20      Inf
## 13          30     1.00 0.273138 0.27417 0.31849 0.20002
## 14          30     1.01 0.066473 0.06828 0.08258 3.75042
## 15          30     1.05 0.059851 0.06383 0.07520 3.74041
## 16          30     1.10 0.040400 0.05435 0.07516 3.62445
## 17          60     1.00 0.018139 0.02560 0.03420 0.04647
## 18          60     1.01 0.015157 0.02149 0.02940 0.04565
## 19          60     1.05 0.012528 0.02187 0.03127 0.04353
## 20          60     1.10 0.013782 0.02132 0.03124 0.04575
## 21          90     1.00 0.009597 0.01502 0.02041 0.02710
## 22          90     1.01 0.009604 0.01189 0.01982 0.02899
## 23          90     1.05 0.009670 0.01231 0.01762 0.02959
## 24          90     1.10 0.010311 0.01227 0.01950 0.02973
```

The median absolute bias for logit phi suggests that, on average, the estimator performs worse when the penalized log likelihood is used. For both values of K, larger values for the scale parameter increase the median absolute bias. When K = 30, increasing the value of the shape parameter slightly decreases the median absolute bias of the estimates. However, the median bias when K = 30 is simply too large for the estimates to be trusted. The bias changes considerably when K = 90. The median absolute bias for the unpenalized estimates is 2.7% and the bias of the penalized estimates range from approximately 3.4% to 5%, with the best estimates at shape = 1.10 and scale = 10. Increasing values for shape may have an impact on the estimates, but the impact does not appear to be uniformly better or worse. Given the very small magnitude, the differences might simply be attributed to Monte Carlo error.

Maximum bias

```
##    K_intervals k_priors        5       10       20      Inf
## 13          30     1.00 23.2052 23.4448 23.4289   13.520
## 14          30     1.01  1.5855  1.6997  1.5812 264.382
## 15          30     1.05  0.9393  0.9828  0.9826 244.710
## 16          30     1.10  0.6723  0.6988  0.7231 249.936
## 17          60     1.00 10.1402 10.6782 11.2586    6.067
## 18          60     1.01  0.7092  0.6871  0.7419   90.034
## 19          60     1.05  0.4282  0.4599  0.5077   98.967
## 20          60     1.10  0.3332  0.3372  0.3784   89.103
## 21          90     1.00  4.5738  4.2575  4.3779    2.099
## 22          90     1.01  0.3484  0.2890  0.3424   37.366
## 23          90     1.05  0.1904  0.2063  0.2252   32.905
## 24          90     1.10  0.1292  0.1618  0.1982   31.862
```

The maximum absolute bias is slightly different. At the absolute worst, the unpenalized MTS estimates are considerably more biased than the penalized estimates. In all cases, the absolute maximum bias is quite large. Increasing the value of K decreases the maximum absolute bias. Unlike with the median bias, at both levels of K increasing the shape parameter reduces the maximum absolute bias. As with the median absolute bias, increasing the values of the scale parameter increases the maximum absolute bias at both levels of K.

Median RMSE

```
##    K_intervals k_priors      5      10      20      Inf
## 13          30     1.00 3.8829 4.1212 4.2869   2.7253
## 14          30     1.01 0.8129 0.7963 0.7950  51.5634
## 15          30     1.05 0.7105 0.7229 0.7472  50.0295
## 16          30     1.10 0.7023 0.6840 0.7011  49.6645
## 17          60     1.00 0.4380 0.4524 0.4429   0.4566
## 18          60     1.01 0.4324 0.4309 0.4419   0.4735
## 19          60     1.05 0.4297 0.4343 0.4355   0.4551
## 20          60     1.10 0.4289 0.4348 0.4373   0.4510
## 21          90     1.00 0.3434 0.3450 0.3562   0.3535
## 22          90     1.01 0.3451 0.3499 0.3528   0.3580
## 23          90     1.05 0.3431 0.3486 0.3506   0.3562
## 24          90     1.10 0.3465 0.3487 0.3541   0.3561
```

In terms of the median RMSE, when $K = 30$ the unpenalized estimates perform worse than the penalized estimates. Increasing values of the shape parameter slightly decrease the median RMSE, and increasing values of the scale parameter slightly increase the median RMSE. However when $K = 90$, the median RMSE is fairly stable across all both the unpenalized estimates and the penalized estimates across all levels of the priors.

Maximum RMSE

```
##    K_intervals k_priors       5      10      20     Inf
## 13          30     1.00 38.989 39.181 39.149   22.77
## 14          30     1.01  3.150  3.210  3.116  433.01
## 15          30     1.05  2.303  2.306  2.285  406.15
## 16          30     1.10  2.000  1.971  1.989  410.47
## 17          60     1.00 28.361 29.090 29.801   16.65
## 18          60     1.01  2.272  2.220  2.264  245.56
## 19          60     1.05  1.680  1.687  1.694  260.35
## 20          60     1.10  1.467  1.443  1.412  244.69
## 21          90     1.00 19.872 19.218 19.449    8.71
## 22          90     1.01  1.607  1.554  1.562  162.25
## 23          90     1.05  1.266  1.268  1.211  148.73
## 24          90     1.10  1.144  1.147  1.092  146.19
```

In terms of the maximum RMSE, the unpenalized estimates perform worse than the penalized estimates at both levels of K. Much like with the median RMSE, increasing values of the shape parameter slightly decrease the median RMSE, and increasing values of the scale parameter slightly increase the median RMSE. The differences in magnitude may be indicative of a practical difference.

Taken together, this suggests that the estimator may produce approximately unbiased estimates in some cases, the estimates of prevalence are not approximately unbiased across the entire parameter space whether the unpenalized log likelihood or the penalized log likelihood with any of the tested combinations of priors are used.

Log zeta resuts

Median bias

```
##    K_intervals k_priors      5      10      20      Inf
## 1           30     1.00 7.8107 6.1751 5.05396 0.54164
## 2           30     1.01 1.5465 1.0206 0.63498 3.72099
## 3           30     1.05 0.8046 0.3755 0.18759 3.76970
## 4           30     1.10 0.4663 0.1894 0.11318 3.74139
## 5           60     1.00 3.9263 3.0323 2.32354 0.29415
## 6           60     1.01 0.9185 0.5520 0.32369 0.18526
## 7           60     1.05 0.4863 0.2364 0.13739 0.10416
## 8           60     1.10 0.2878 0.1258 0.08226 0.05741
## 9           90     1.00 2.3364 1.6181 1.21218 0.14699
## 10          90     1.01 0.5720 0.3299 0.20024 0.11517
## 11          90     1.05 0.3447 0.1733 0.10539 0.08055
## 12          90     1.10 0.2059 0.1032 0.06968 0.04733
```

The median absolute bias for log zeta suggests that, on average, the estimator performs considerably better when the penalized log likelihood is used. When K = 30, values for the median absolute bias are once again too large for the estimates to be trusted. As with logit phi, increasing the value of the scale parameter increases the median absolute bias at both levels of K. However, unlike logit phi, increase the value of the shape parameter increases the median absolute bias across both levels of K. When K = 90, the absolute bias might be considered adequate at approximately 5.3% when the shape parameter = 1.01 and the scale parameter = 10, but otherwise the median absolute bias is probably too large for the estimates to be considered approximately unbiased.

Maximum bias

```
##    K_intervals k_priors       5      10      20       Inf
## 1           30     1.00 27.105 26.073 25.6125   3.1491
## 2           30     1.01  4.093  3.518  2.9268 260.8511
## 3           30     1.05  2.931  2.332  1.7947 258.7264
## 4           30     1.10  2.416  1.849  1.3319 264.7653
## 5           60     1.00 17.128 15.777 15.7323   1.8223
## 6           60     1.01  3.807  2.850  2.2153  94.9347
## 7           60     1.05  2.342  1.760  1.3346 102.5774
## 8           60     1.10  1.937  1.428  1.0018  93.9782
## 9           90     1.00 16.119 13.159 12.4140   0.9597
## 10          90     1.01  3.883  2.620  1.9828  37.2859
## 11          90     1.05  2.216  1.357  0.9872  34.3951
## 12          90     1.10  1.603  1.099  0.7619  33.5268
```

The maximum absolute bias is somewhat different. When K = 30, the the maximum absolute bias of the unpenalized estimates is considerably larger than the penalized estimates. However, when K = 90, the maximum absolute bias for the unpenalized estimates are better than the penalized ones, although only to a moderate degree. As with the median absolute bias, at both levels of K, increasing the shape parameter and the scale parameter increases the maximum absolute bias.

Median RMSE

```
##    K_intervals k_priors       5      10     20      Inf
## 1           30     1.00 12.0389 10.1879 9.0364  1.5394
## 2           30     1.01  2.2705  1.7248 1.2422 51.0976
## 3           30     1.05  1.2148  0.7753 0.6575 51.9778
## 4           30     1.10  0.7504  0.5969 0.5772 51.0811
## 5           60     1.00  8.1520  6.7239 6.2807  0.9988
## 6           60     1.01  1.7072  1.2191 0.8559  0.6820
## 7           60     1.05  0.8469  0.6056 0.5357  0.5766
## 8           60     1.10  0.6203  0.5041 0.4805  0.5100
## 9           90     1.00  5.8416  4.7819 4.2055  0.7666
## 10          90     1.01  1.3485  0.9350 0.6477  0.5471
## 11          90     1.05  0.6809  0.4960 0.4522  0.4678
## 12          90     1.10  0.5027  0.4334 0.4113  0.4377
```

The median RMSE for log zeta is somewhat analogous to the median RMSE for logit phi. When K = 30, the unpenalized estimates perform worse than the penalized estimates. Increasing the shape parameter and the scale parameter increases the median RMSE. When K = 90, the unpenalized estimates perform worse than the penalized estimates, however the median RMSE is relatively stable across all the levels of the priors.

Maximum RMSE

```
##    K_intervals k_priors      5      10     20      Inf
## 1           30     1.00 34.606 34.426 34.170    8.716
## 2           30     1.01  4.503  3.904  3.289  425.903
## 3           30     1.05  3.155  2.536  1.963  427.494
## 4           30     1.10  2.569  1.982  1.439  432.281
## 5           60     1.00 26.136 26.046 26.819    4.246
## 6           60     1.01  4.124  3.236  2.747  258.698
## 7           60     1.05  2.689  2.135  1.653  269.349
## 8           60     1.10  2.218  1.677  1.218  257.783
## 9           90     1.00 20.444 18.333 18.510    2.198
## 10          90     1.01  4.187  3.056  2.409  162.734
## 11          90     1.05  2.517  1.804  1.390  156.153
## 12          90     1.10  1.914  1.414  1.039  153.901
```

The maximum RMSE for log zeta is approximately the same as median absolute bias. Increasing the value K decreases the maximum RMSE The unpenalized estimates perform worse than the penalized estimates at both levels of K. Increasing the shape parameter and the scale parameter both increase the maximum RMSE.

Taken together, this suggests that approximately unbiased estimates of zeta might be produced when the penalized log likelihood is used with shape = 1.01 and scale = 10, however unbiased estimates are not produced across the entire parameter space of phi and zeta.

Further investigation

These results suggest two possible directions for more investigation via simulation. It definitely seems as though it would be valuable to investage smaller values for the scale parameter, given that the smaller of the two tested scale values produced better estimates of both phi and zeta. It also might be worth examining the odd relationship between logit phi's median bias the varying levels of the shape parameter when K = 90. However, given the RMSE when K = 90, it's possible that the error for phi is on average relatively stable at larger values of K.