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

```
MC_error
```

```
##         stat theta           30           90
## 1   log zeta     0 1.336298e-02 6.628455e-04
## 2   log zeta    10 5.147109e-05 3.620481e-05
## 3   log zeta    20 8.239899e-05 5.094377e-05
## 4  logit phi     0 6.725974e-02 1.392724e-02
## 5  logit phi    10 1.465628e-03 4.601789e-04
## 6  logit phi    20 1.495869e-03 4.387971e-04
## 7        phi     0 2.038560e-05 5.685029e-06
## 8        phi    10 1.290800e-05 4.663213e-06
## 9        phi    20 1.435253e-05 5.239551e-06
## 10      zeta     0 1.946935e+20 6.967989e+10
## 11      zeta    10 6.436653e-06 5.630551e-06
## 12      zeta    20 6.258580e-06 5.359242e-06
```

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

```r
median_bias[c(2,6)]
```

```
## $`30.logit phi`
##   k_priors         0         10        20
## 1     0.00 0.2052003        NA        NA
## 2     1.01        NA 0.1565762 0.1935027
## 3     1.05        NA 0.1435088 0.1818983
## 4     1.10        NA 0.1333815 0.1687110
##
## $`90.logit phi`
##   k_priors          0          10         20
## 1     0.00 0.02724142         NA         NA
## 2     1.01         NA 0.03613800 0.04936004
## 3     1.05         NA 0.03624013 0.05028867
## 4     1.10         NA 0.03387578 0.04938390
```

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.

```r
max_bias[c(2,6)]
```

```
## $`30.logit phi`
##   k_priors      stat      0        10        20
## 1     0.00 logit phi 13.409        NA        NA
## 2     1.01 logit phi     NA 1.6328619 1.6919673
## 3     1.05 logit phi     NA 1.0332414 1.1314231
## 4     1.10 logit phi     NA 0.7955099 0.8798068
##
## $`90.logit phi`
##   k_priors      stat        0        10        20
## 1     0.00 logit phi 2.028474        NA        NA
## 2     1.01 logit phi       NA 0.3639644 0.4345758
## 3     1.05 logit phi       NA 0.2790620 0.3553346
## 4     1.10 logit phi       NA 0.2443701 0.3185634
```

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.

```r
median_rmse[c(2,6)]
```

```
## $`30.logit phi`
##   k_priors        0         10        20
## 1     0.00 2.604194         NA        NA
## 2     1.01       NA 0.8295573 0.8604821
## 3     1.05       NA 0.7762014 0.8168703
## 4     1.10       NA 0.7551695 0.7998022
##
## $`90.logit phi`
##   k_priors         0        10        20
## 1     0.00 0.3572772        NA        NA
## 2     1.01        NA 0.3623736 0.3662804
## 3     1.05        NA 0.3608815 0.3668847
## 4     1.10        NA 0.3622393 0.3758862
```

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.

```r
max_rmse[c(2,6)]
```

```
## $`30.logit phi`
##   k_priors        0       10       20
## 1     0.00 22.71783       NA       NA
## 2     1.01       NA 3.116967 3.092707
## 3     1.05       NA 2.344413 2.306007
## 4     1.10       NA 2.002932 2.039530
##
## $`90.logit phi`
##   k_priors        0       10       20
## 1     0.00 8.587835       NA       NA
## 2     1.01       NA 1.559925 1.543646
## 3     1.05       NA 1.249048 1.240259
## 4     1.10       NA 1.121849 1.115212
```

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.

Logit zeta resuts

```
median_bias[c(1,5)]
```

```
## $`30.log zeta`
##   k_priors         0        10        20
## 1     0.00 0.5623677        NA        NA
## 2     1.01        NA 0.1842935 0.2222192
## 3     1.05        NA 0.2100204 0.2449534
## 4     1.10        NA 0.2375740 0.2653742
##
## $`90.log zeta`
##   k_priors         0         10         20
## 1     0.00 0.1380762         NA         NA
## 2     1.01        NA 0.05325754 0.06747348
## 3     1.05        NA 0.06176446 0.07205399
## 4     1.10        NA 0.07275810 0.08844880
```

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.

```
max_bias[c(1,5)]
```

```
## $`30.log zeta`
##   k_priors     stat       0       10       20
## 1     0.00 log zeta 3.18601       NA       NA
## 2     1.01 log zeta      NA 1.333485 1.744768
## 3     1.05 log zeta      NA 1.383020 1.788898
## 4     1.10 log zeta      NA 1.444989 1.862074
##
## $`90.log zeta`
##   k_priors     stat         0       10       20
## 1     0.00 log zeta 0.9565459       NA       NA
## 2     1.01 log zeta        NA 1.203948 1.458549
## 3     1.05 log zeta        NA 1.249768 1.482248
## 4     1.10 log zeta        NA 1.297090 1.531017
```

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.

```r
median_rmse[c(1,5)]
```

```
## $`30.log zeta`
##   k_priors        0        10        20
## 1     0.00 1.532037        NA        NA
## 2     1.01       NA 0.4868519 0.5396810
## 3     1.05       NA 0.4887670 0.5515686
## 4     1.10       NA 0.4946791 0.5628773
##
## $`90.log zeta`
##   k_priors         0        10        20
## 1     0.00 0.7755638        NA        NA
## 2     1.01        NA 0.3333505 0.3440413
## 3     1.05        NA 0.3316138 0.3435764
## 4     1.10        NA 0.3289654 0.3448164
```

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.

```r
max_rmse[c(1,5)]
```

```
## $`30.log zeta`
##   k_priors        0       10       20
## 1     0.00 8.706765       NA       NA
## 2     1.01       NA 1.378095 1.823819
## 3     1.05       NA 1.428852 1.867555
## 4     1.10       NA 1.492315 1.941375
##
## $`90.log zeta`
##   k_priors        0       10       20
## 1     0.00 2.022429       NA       NA
## 2     1.01       NA 1.258320 1.535601
## 3     1.05       NA 1.304116 1.562397
## 4     1.10       NA 1.349635 1.610334
```

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.