Notes on Method of Moments and Monte Carlo Methods for Calculating Beta Product Confidence Procedure and Its Mid-p Version

Michael P. Fay and Erica H. Brittain
National Institute of Allergy and Infectious Diseases,
Bethesda, MD, 20892-9820, USA
correspondence to M.P. Fay: mfay@niaid.nih.gov

June 4, 2015

Summary

These notes describe some additional details of the calculation methods for the beta product confidence procedure (BPCP) for survival distributions with right censored data developed in Fay *and others* (2013) and the mid-p version developed in Fay and Brittain (2015). The notes are primarily of interest to software developers or researchers developing new calculation methods. Both Fay *and others* (2013) and Fay and Brittain (2015) suggested two different methods of calculating the confidence intervals, a Monte Carlo estimator and a method of moments estimator, with Fay and Brittain (2015) additionally recommending enforcing monotonicity on the method of moments estimator. The notes cover two issues. First, although in general there is good agreement between the Monte Carlo estimator and the method of moments estimators (with or without enforced monotonicity), in some cases they may give slightly different results. Second, we explore the non-monotonicity with the method of moments, which occurs most often in the mid-p upper BPCP limit.[1]

1 Overview

Because pseudo-random number generators are very good, in general the Monte Carlo estimator allows one to get as precise an estimator as needed by choosing a large enough Monte Carlo sample size. The disadvantages of the Monte Carlo estimator are that it generally takes longer, and there is simulation error so that two researchers may get slightly different answers even with the same number of Monte Carlo replications if they use different random number generators or different seeds.

The idea of the method of moments estimator is to approximate the beta product distribution with a beta distribution with the same two first moments. The method of moments estimator is generally faster than the Monte Carlo one and has no simulation error. However, in some cases the beta approximation is noticeably different from the Monte Carlo one (see Section 2).

Another issue with the method of moments estimator is that in some cases the method of moments BPCP limits can increase over time. This phenomenon occurs mostly with the mid-p BPCP for the upper limit. Fay and Brittain (2015) suggested improving the method of moments estimators by enforcing monotonicity of the confidence limits over time (for all method of moments estimators, lower and upper), and ran extensive simulations using that suggestion. The simulations show that using the method of moments with enforced monotoncity on the standard BPCP guaranteed coverage, while using method of moments with enforced monotoncity on the mid-p BPCP had generally closer to nominal coverage (but not guaranteed coverage). Section 3.2 shows a case where the upper mid-p BPCP (Section 3.2) and a much rarer case where the upper standard BPCP (Section 3.3) is increasing. In almost all cases studied so far, the lower limits (both versions) are non-increasing over time (see Section 3.1).

2 When the Method of Moments BPCP Estimators Differ from the Monte Carlo Ones

Figure 1 is a small data set with 34 subjects, and the two method of moments estimators are nearly indistinguishable. We use $10,000$ replications for the Monte Carlo method so that the limits will be very close to the ideal BPCP limits. The method of moments estimators are similar to the Monte Carlo one, except they are slightly different at the later time points when there are very few subjects left. To explore this further, we consider the upper 95% limit at the $11th$, and $12th$ failure times.

Let $T_1 < \cdots < T_k$ be the k observed failure times, and let $Y(T_i)$ be the number at risk just before T_i. Let $B(a, b)$ represent a beta random variable with parameters a and b. Then the upper $100(1-\alpha)\%$ BPCP at T_j is the $(1-\alpha/2)^{th}$ percentile of the random variable:

$$\prod_{i=1}^{j} B(Y(T_i), 1). \quad (1)$$

[1] R script for producing the figures in these notes is in the demo folder.

First we show that the method of moments estimator for the upper limit of the BPCP does fairly well up until T_{11}. The method of moments estimator for the upper BPCP distribution at T_{10} is $B(22.5475, 9.7069)$. So at the $11th$ failure (at day 1801=year 4.93) the BPCP method of moments estimator approximates $B(22.5475, 9.7069)B(13, 1)$, where the first B comes from the method of moments distribution at T_{10} and the second beta distribution is $B(Y(T_{11}), 1)$. For simple illustrative purposes we calculate the Monte Carlo estimator just using these two distributions (unlike the way the bpcp function calculates the Monte Carlo; it takes a Monte Carlo sample from each of the beta distributions up to that point (see equation 1), takes the product, then repeats the process many times). That simplified Monte Carlo estimate that uses only two beta distributions it is nearly equal to the method of moments one (see Figure 2). If we repeat this process at T_{12}, we see that the two estimators do not match as well (see Figure 3). So the lack-of-fit for the method of moments in this case is caused primarily by the last beta distribution in the product. This lack-of-fit is why the upper limits do not match after T_{12} in Figure 1.

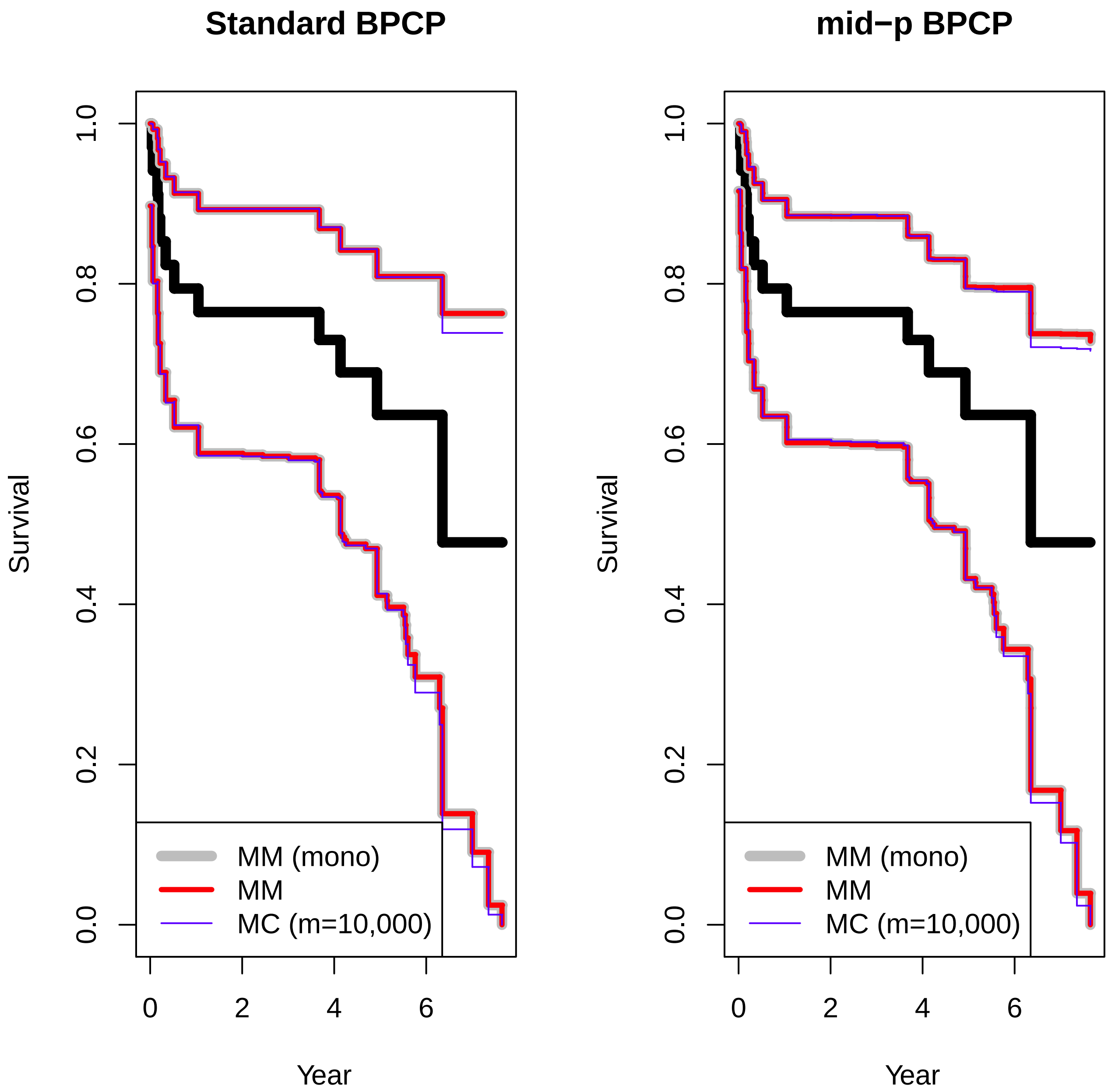

Figure 1: Sclerosis data from Nash *and others* (2007). Kaplan-Meier survival estimate (black), and 95% BPCP methods estimated by method of moments (red), method of moments with forced monotonicity (gray), or Monte Carlo (blue). Monte Carlo simulations based on $10,000$ replications.

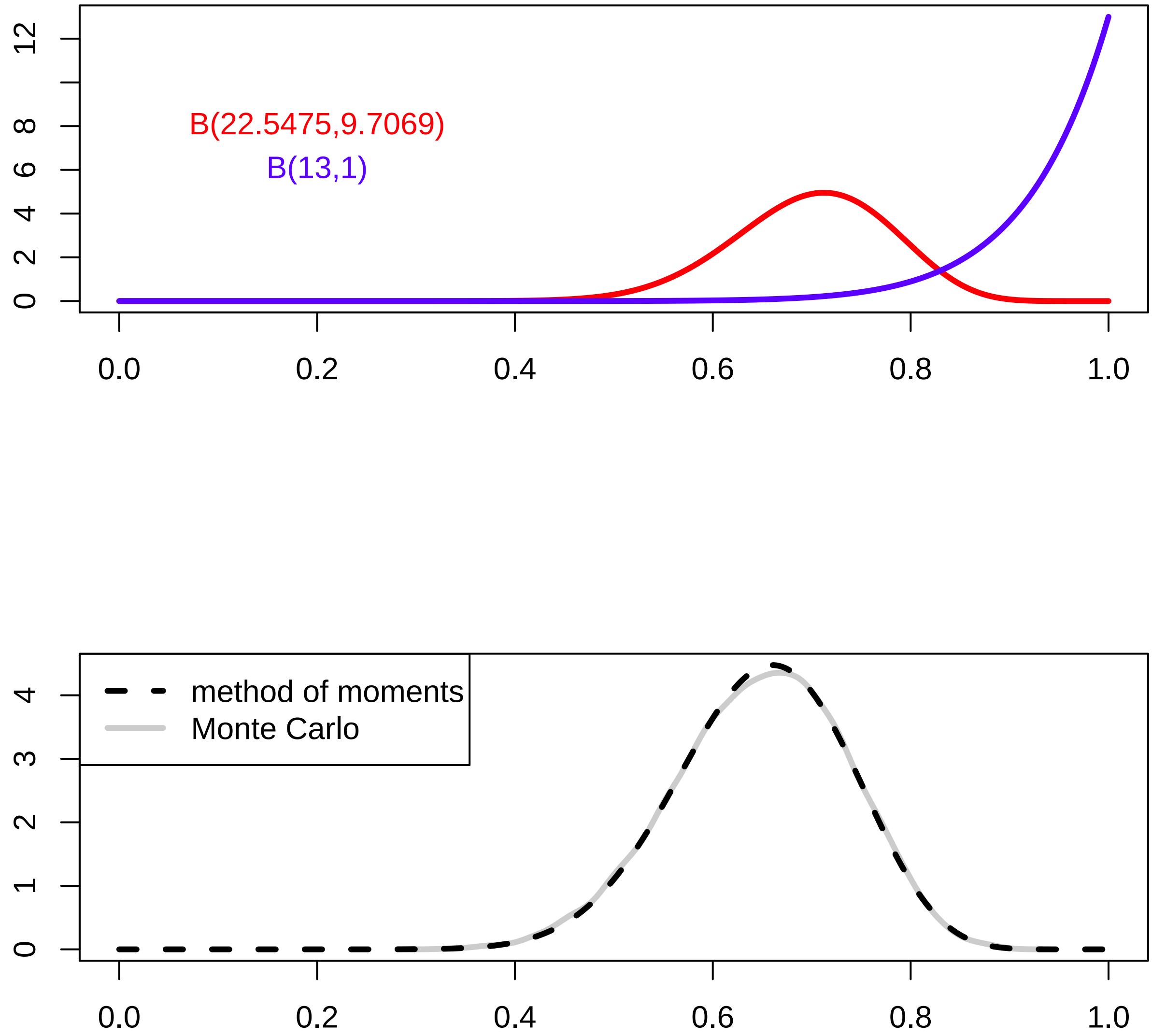

Figure 2: Top panel is the two beta distributions for which we want the distribution of the product. The red distribution is the method of moments beta estimate at T_{10}, and the blue distribution is $B(Y(T_{11}), 1)$, where $Y(T_{11}) = 13$. The bottom panel are the two estimated distributions for the product, the method of moments and the Monte Carlo. The Monte Carlo is calculated using $10,000$ replications and the density function (kernel density estimator) in the R stats package.

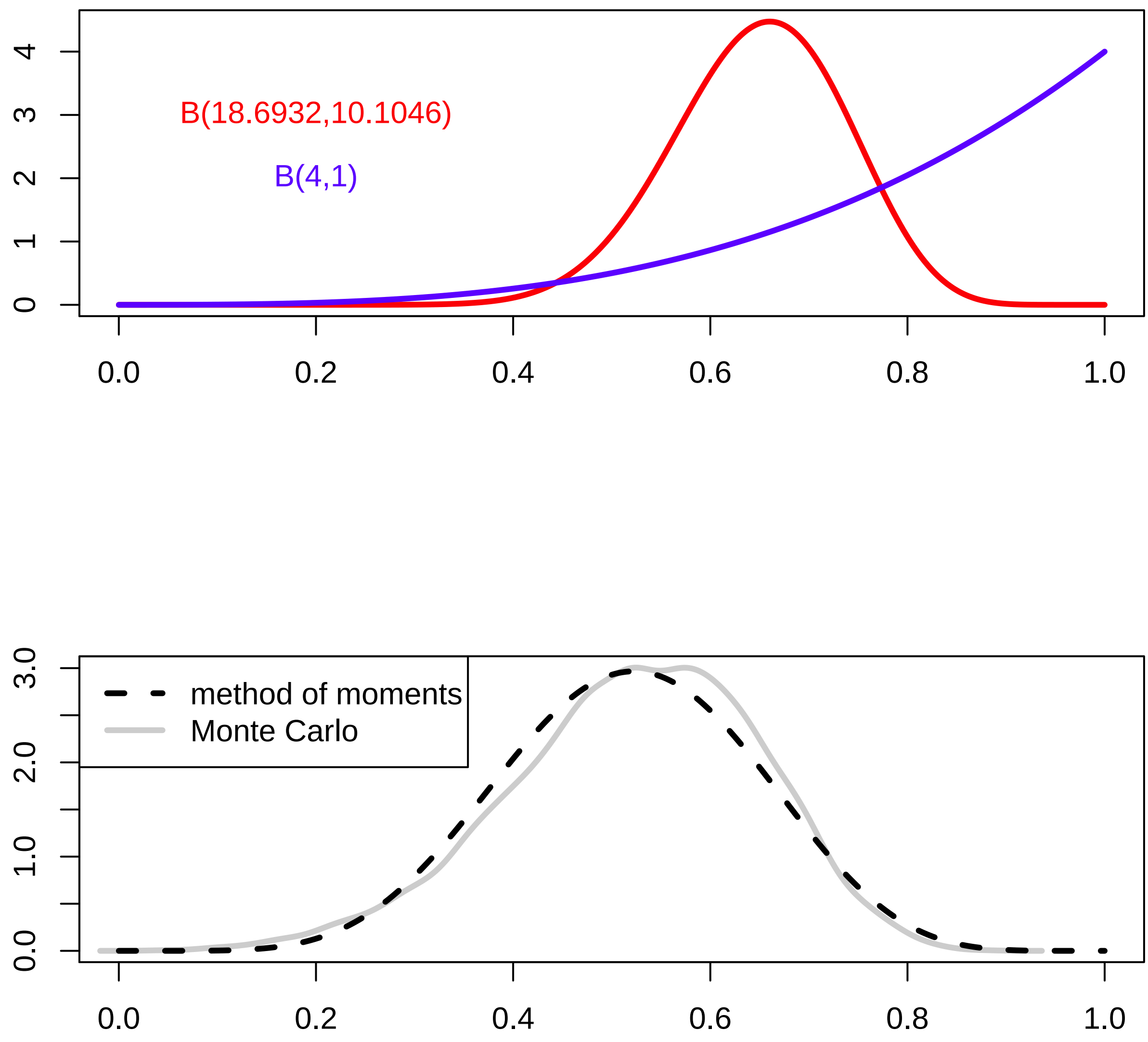

Figure 3: Top panel is the two beta distributions for which we want the distribution of the product. The red distribution is the method of moments beta estimate at T_{11}, and the blue distribution is $B(Y(T_{12}),1)$, where $Y(T_{12})=4$. The bottom panel are the two estimated distributions for the product, the method of moments and the Monte Carlo. The Monte Carlo is calculated using $10,000$ replications and the density function (kernel density estimator) in the R stats package.

3 The Method of Moments Estimators Can Be Non-Monotonic

3.1 Experience from Simulations

We examine another phenomenon in this section: the method of moments BPCP estimators may be non-monotonic, that is under certain circumstances they increase over time. Although it is more complicated than the standard BPCP, we discuss the mid-p BPCP first, since the method of moments upper limit increasing over time is more common for the mid-p BPCP than for the standard BPCP. In addition to the simulations in Fay and Brittain (2015), we also did the simulations using the same simulated data as Fay and Brittain (2015) but without enforcing monotonicity. The type I error rates without enforcing monotonicity were very nearly (if not exactly) identical to those of the simulations in Fay and Brittain (2015) that enforced monotonicity. Even in the simulations that did not enforce monotonicity, we did not see non-monotonic behaviour for the 95% standard BPCP in any of the 11 scenarios for the main simulations when $n = 10, 30, or 100$, but we did see it in 1 out of the 11 scenarios when $n = 1000$ for the upper limit only. For the 95% mid-p BPCP, we see this non-monotoic behaviour with the upper limit for most scenarios at all 4 sample sizes, but almost never in the lower limit (we saw it only in 1 scenario at $n = 1000$, but in that case the increase was imperceivable and on the order of 10^{-7}).

Theoretically, we know that the true BPCPs (both the standard and mid-p versions) do not increase over time (since multiplying a random variable by a beta random variable can only decrease any quantile of the product, since the maximum beta value is 1). It is in the method of moments implementation that this non-monotonicity issue arises. **That is why we recommend enforcing monotonicity for the method of moments estimator for both versions of the BPCP.**

3.2 Mid-p BPCP Upper Limit Example

Figure 4 is a larger data set with 4028 subjects. In this case the method of moments with forced monotonicity and the Monte Carlo methods are very close. In contrast, for the mid-p version only, the method of moments without forced monotonicity has a noticeable increase in the upper limit compared to the other two methods at some later time points. For these data there are 571 observed events, but the last one occurs at 11.42 years (4173 days) when there are still $n = 701$ at individuals risk of the event. To illustrate, we first pick a time point when all three methods are very similar. At time 13.70 years (5003 days) there are only $n = 330$ left at risk. In Figure 5 we plot the lower and upper confidence distributions for the BPCP at 5003 days. The two-sided upper 95% mid-p BPCP (which is also the one-sided upper 97.5% mid-p BPCP) is 0.8602, and is between the 97.5 percentile of the lower distribution (0.8593) and the upper distribution (0.8609). In contrast, Figure 6 gives an analogous plot at 5843 days (16.00 years) when there are $n = 38$ subjects still at risk. Notice how the lower confidence distribution of the BPCP generally has lower values, but its 97.5 percentile (0.8686) is higher than the same quantile from the upper distribution (0.8609, notice the upper distribution does not change since no events were observed between day 5003 and 5843). If we could easily calculate the beta product distributions exactly then the lower beta product distribution (the distribution used in calculating the lower BPCP limit) would be stochastically less than the upper beta product distribution, but the method of moments is only an approximation and does not always retain this stochastic ordering. This moves the 95% upper mid-p BPCP limit (red tick) to be higher than the 97.5% percentile of the upper confidence distibution (black tick). That is why the upper mid-p BPCP increases at the end (see Figure 4).

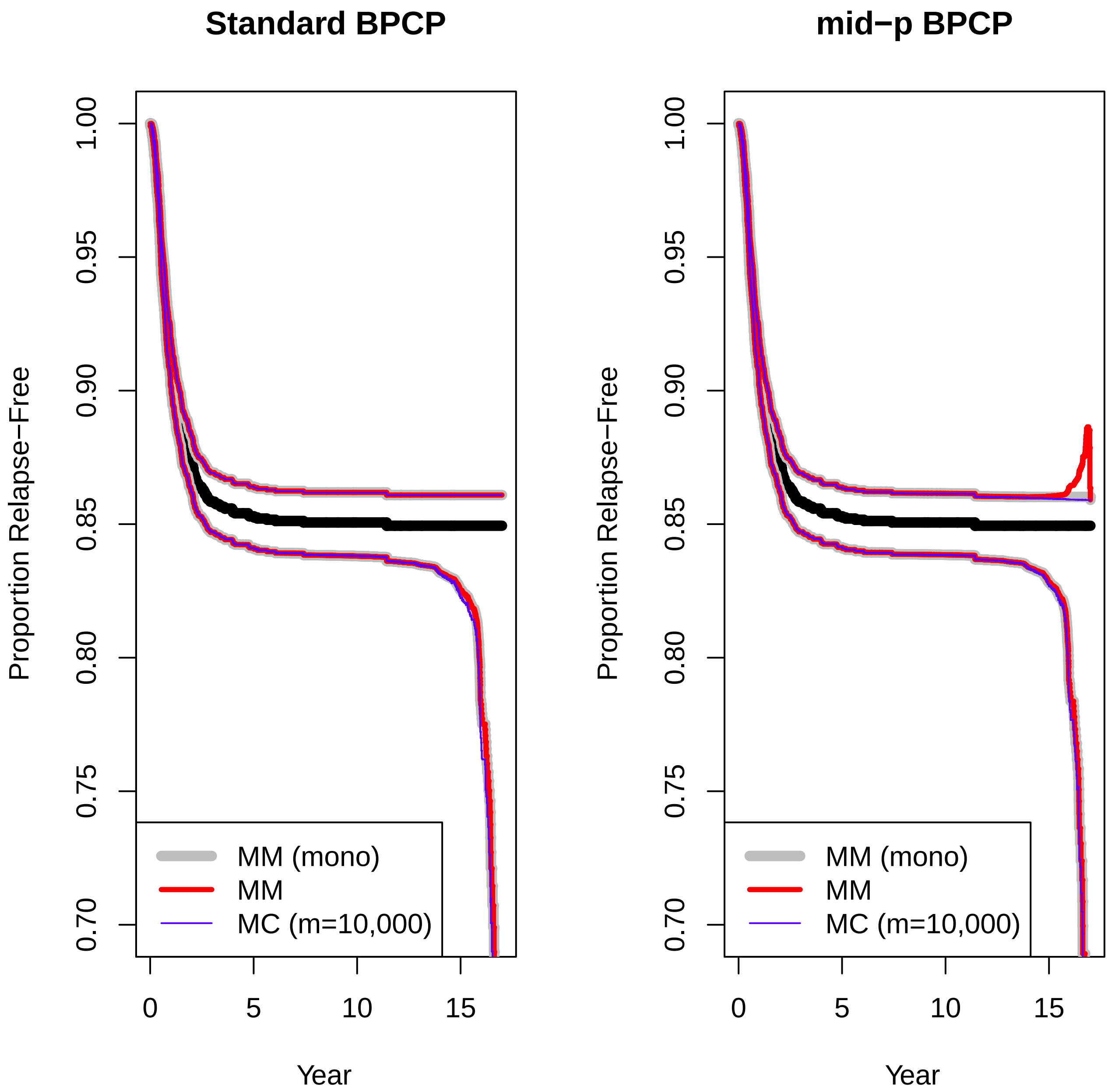

Figure 4: National Wilms Tumor Study Group data (Breslow and Chatterjee, 1999); data in the survival R package (Therneau, 2015). Kaplan-Meier survival estimate (black), and 95% BPCP methods estimated by method of moments (red), method of moments with forced monotonicity (gray), or Monte Carlo (blue). Monte Carlo simulations based on $10,000$ replications.

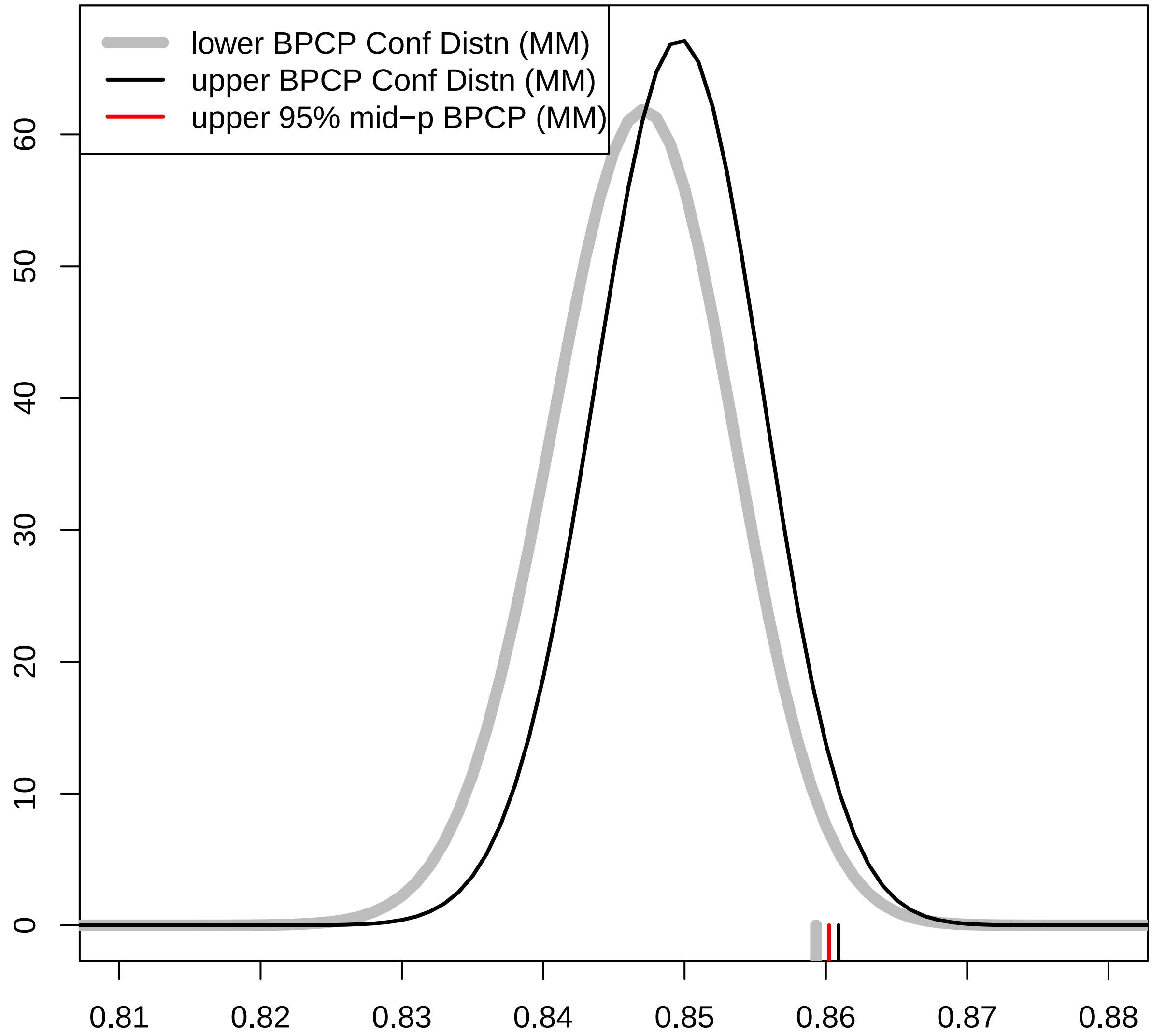

Figure 5: Beta distributions from method of moments BPCP (without enforced monotonicity) at time=5003 days from Wilms Tumor data. Gray (black) distribution is the beta distribution associated with the lower (upper) limits of the standard BPCP. The upper mid-p BPCP is represented by the red tick. The gray and black ticks are the 97.5 percentile of the lower and upper beta distributions, respectively. The red is between the lower and upper tick marks.

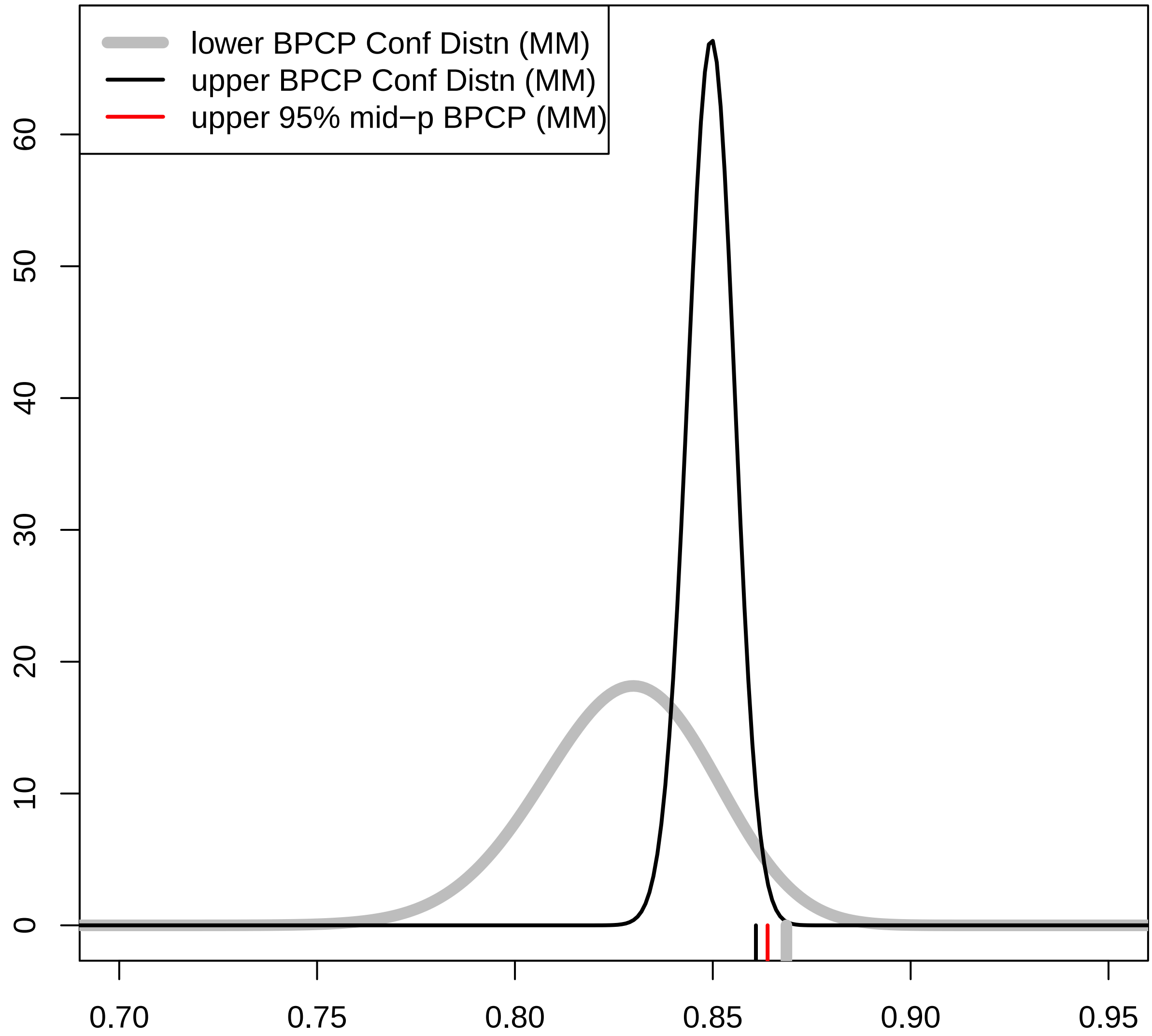

Figure 6: Beta distributions from method of moments BPCP (without enforced monotonicity) at time=6132 days from Wilms Tumor data. Gray (black) distribution is the beta distribution associated with the lower (upper) limits of the BPCP. The upper mid-p BPCP is represented by the red tick. The gray and black ticks are the 97.5 percentile of the lower and upper beta distributions, respectively. The red is between the lower and upper tick marks.

3.3 Standard BPCP Upper Limit Example

Although the issue with the upper limit of the method of moments estimator increasing over time is much more common in the mid-p BPCP, that issue can occur in the standard BPCP. Consider a simulated data example. This is simulated from the 11^{th} scenario from Fay and Brittain (2015) with sample size$= 1000$. In 11.62% of the simulations done, the upper confidence limit for the method of moments BPCP was non-monotonic at some point. In Figure 7 we give one simulated data example. In this example, there are 516 failures. Just before T_{515} (the 515^{th} failure) there are there are 21 individuals at risk, while just before T_{516} (the 516^{th} failure) there are only 2 at risk. In Figure 8 we plot three beta distributions from the method of moments calculations. The upper beta distribution from the method of moments at T_{515} is $B(112.4, 184.8)$ (density is gray curve), and the 97.5% of that distribution is 0.4340. The red beta distribution is $B(2, 1)$, and the blue beta distribution is the method of moments estimator of $B(112.4, 184.8)B(2, 1)$, which is $B(5.448, 16.155)$ and the 97.5% of that distribution is 0.4501. The beta product distribution for $B(112.4, 184.8)B(2, 1)$ (density estimated as green curve, calculated by Monte Carlo with 10^5 replications and using the density function) will be stochastically smaller than the beta distribution $B(112.4, 184.8)$ (the gray curve), but the method of moments estimator of that beta product is slightly larger at the upper tails. The 97.5% quantile of the Monte Carlo estimator is 0.3930 and is smaller than both the other estimators, and more likely closer to the 97.5% of the beta product distribution.

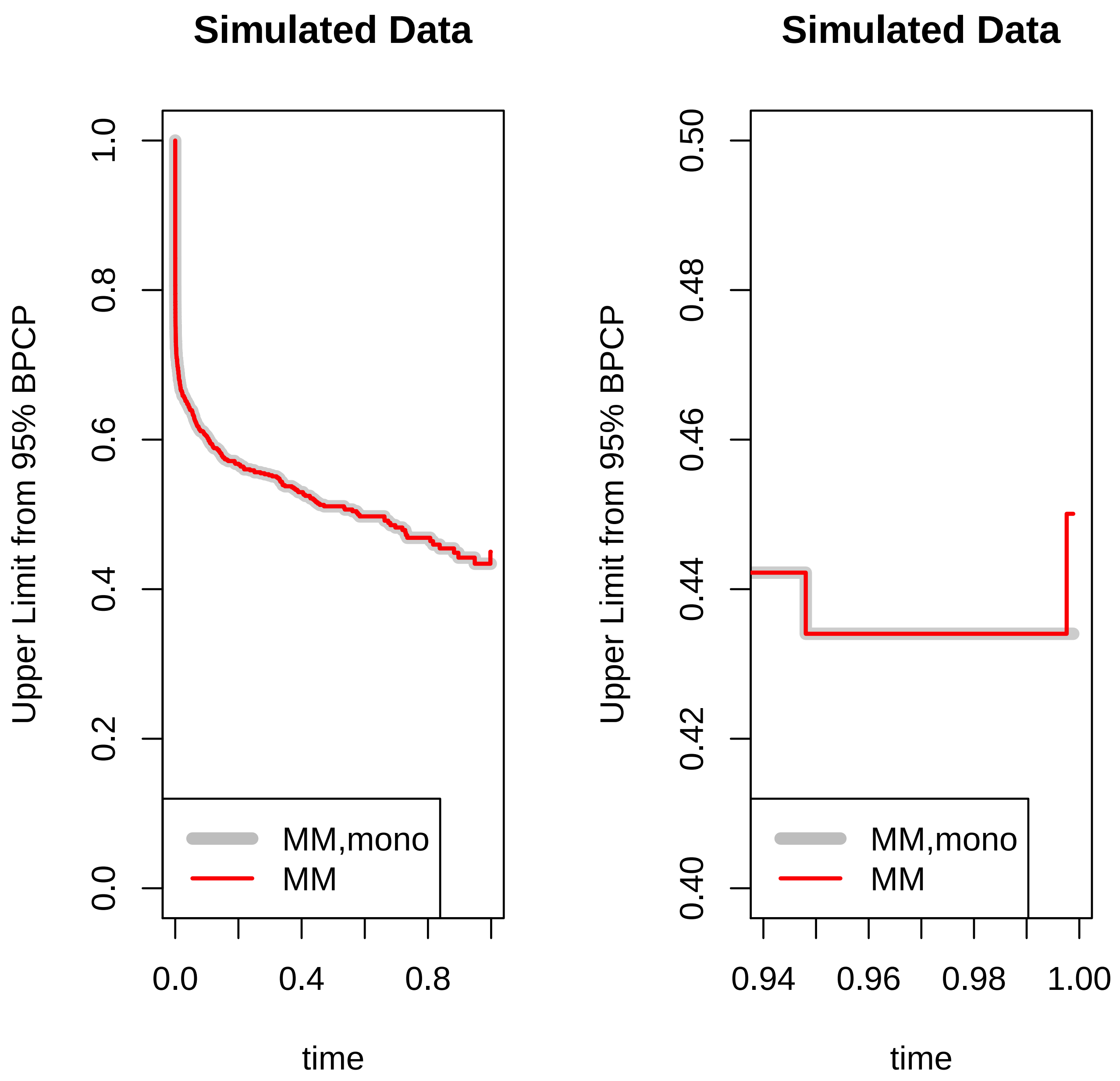

Figure 7: Simulated data, where the censoring distribution is uniform on $(0, 1)$, and the failure distribution is $Beta(0.10, 0.10)$ and $n = 1000$. Left panel is full range, and right panel focuses onto the non-monotonic part.

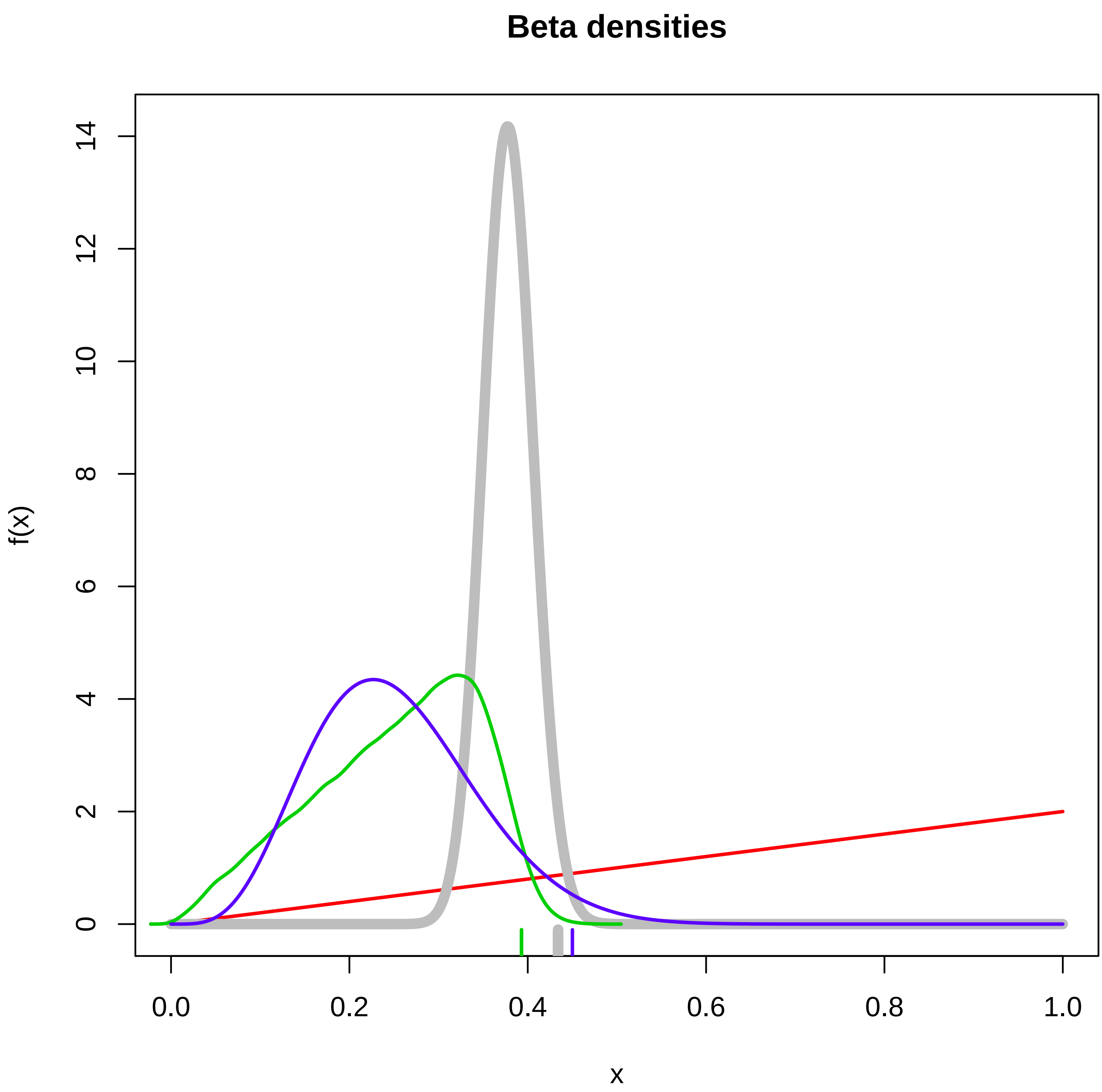

Figure 8: Beta distributions from the upper BPCP method of moments (without enforced monotonicity). The upper beta distribution from the method of moments at T_{515} is $B(112.4, 184.8)$ (density is gray curve), and the 97.5% of that distribution is 0.4340. The red beta distribution is $B(2, 1)$, and the blue beta distribution is the method of moments estimator of $B(112.4, 184.8)B(2, 1)$, which is $B(5.448, 16.155)$. The green distribution is calculated by Monte Carlo with 10^5 replications and using the density function. The 97.5% of the gray (0.4340) and blue (0.4501) and green (0.3930) distributions are given as ticks.

References

BRESLOW, NE AND CHATTERJEE, N. (1999). Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **48**(4), 457–468.

FAY, MICHAEL P AND BRITTAIN, ERICA H. (2015). Finite sample pointwise confidence intervals for a survival distribution with right-censored data. *unpublished manuscript*.

FAY, MICHAEL P, BRITTAIN, ERICA H AND PROSCHAN, MICHAEL A. (2013). Pointwise confidence intervals for a survival distribution with small samples or heavy censoring. *Biostatistics* **14**(4), 723–736.

NASH, R.A., MCSWEENEY, P.A., CROFFORD, L.J., ABIDI, M., CHEN, C.S., GODWIN, J.D., GOOLEY, T.A., HOLMBERG, L., HENSTORF, G., LEMAISTRE, C.F. *and others*. (2007). High-dose immunosuppressive therapy and autologous hematopoietic cell transplantation for severe systemic sclerosis: long-term follow-up of the us multicenter pilot study. *Blood* **110**(4), 1388.

THERNEAU, TERRY. (2015). *A Package for Survival Analysis in S*. R package version 2.38.