Chapter 6
G-computation: Parametric Estimation of Optimal DTRs

In this chapter, we will focus on fully parametric estimation of optimal DTRs by modeling the full, longitudinal distribution of the trajectories data. As noted in Chap. 3, optimal dynamic treatment regimes may be determined using dynamic programming. In a classic, likelihood-based framework, this requires modeling the complete longitudinal distribution of the data. If a joint distribution for the longitudinal data can be decomposed into its conditional components, it can be used to predict the expected outcome under a variety of treatment regimes using a Monte Carlo approach. The link between the models used in dynamic programming to those used in G-computation has been made previously (Lavori and Dawson 2004), and forms the basis of the likelihood-based methods of estimation.

6.1 Frequentist G-computation

The semi-parametric approaches to estimating DTRs of the previous two chapters relied on correct modeling of the propensity score, $\pi_j(A_j|H_j)$ (at least with respect to confounding variables). This is trivially satisfied in RCT settings, but could potentially be problematic in an observational study, where $\pi_j(A_j|H_j)$ is unknown and potentially difficult to model. In this case, one can proceed by expressing the value function V^d alternatively as

$$V^d = E\left\{ \sum_{\{(h_j,a_j):1\le j\le K\}} \mathbb{I}[d_1(h_1)=a_1,\ldots,d_K(h_K)=a_K] \times \quad E\Big[\sum_{j=1}^{K} Y_j \Big| H_j=h_j, A_j=a_j\Big]\right\},$$

and then fitting a parametric model, say $\phi_j(h_j,a_j;\theta_j)$, for the inside conditional expectation. Note that in a single-stage setting, the above expression simply gives

B. Chakraborty and E.E.M. Moodie, *Statistical Methods for Dynamic Treatment Regimes*, Statistics for Biology and Health 76, DOI 10.1007/978-1-4614-7428-9_6,

$V^d = E\Big\{E[Y|H=h,A=d(h)]\Big\}$, which is estimated by $\mathbb{P}_n[Y|A=d(h),H=h] = \mathbb{P}_n\phi(h,d(h);\hat{\theta})$ where $\hat{\theta}$ is an estimator of θ. The resulting estimator is known as the *G-computation formula* (Robins 1986), and is given by

$$\hat{V}_G^d = \mathbb{P}_n\left[\sum_{\{(h_j,a_j):1\le j\le K\}} \mathbb{I}[d_1(h_1)=a_1,\ldots,d_K(h_K)=a_K]\ \phi_j(h_j,a_j;\hat{\theta}_j)\right]. \quad (6.1)$$

This estimator is consistent if the models $\phi_j(h_j,a_j;\hat{\theta}_j)$ are correctly specified. Like inverse probability weighting and G-estimation, G-computation can be highly variable due to the presence of non-smooth indicator functions; however in comparison to the (semi-parametric) IPTW estimator, the variability is reduced to some extent by employing the parametric model ϕ.

Note that unlike Q-learning or G-estimation, which are performed recursively in time (backwards induction methods), G-computation models and then simulates data forward in time. That is, in G-computation, a sequence of models for responses at stage j, $j=1,\ldots,K$ are posited and estimated from the observed data. Then, beginning at the first interval and using these models and their estimated parameters, given observed baseline data, data is simulated under a particular regime of interest, d, to generate a distribution of potential outcomes $O_2(d_1)$; next, the stage 2 model is used to simulate (or generate) potential outcomes in response to following the regime of interest, d, in the second stage, generating the distribution of potential outcomes $Y(d)$ in a two-stage setting; more generally, in the K-stage setting, the second stage of simulation would produce $O_3(d_1,d_2)$, and simulation would continue forward until the final stage outcome was simulated under regime $d=d_1(H_1),d_2(H_2),\ldots,d_K(H_K)$.

As noted earlier, when d is not one of the embedded DTRs in the study, estimating V^d is really a problem of counterfactual estimation. In his original work, Robins introduced the above method from a purely causal inference perspective as an approach to estimating counterfactual means, and more generally, entire counterfactual distributions. When the interest lies in estimating counterfactual distributions rather than means, instead of modeling the conditional expectation $E\Big(\sum_{j=1}^K Y_j\Big|H_j=h_j,A_j=a_j\Big)$, one must model the corresponding conditional likelihood, e.g. $\prod_{j=1}^{K+1} f_j(o_j|h_{j-1},a_{j-1})$ in Eq. (3.1). Thus the key idea underlying G-computation is to estimate the marginal mean (or distribution) of the outcome by first fitting models for conditional means (or conditional likelihoods) of stage-specific, time-varying outcomes given history and action, and then to substitute values corresponding to specific treatment patterns into Eq. (6.1) (or corresponding expression of the data likelihood). Note that in G-computation, a potentially greater part of the likelihood of the data is modeled (the states and responses), in contrast to some of the semi-parametric approaches of the previous two chapters, where efforts are focused on modeling the treatment allocation probabilities and the final outcome model. G-computation requires the assumption of *no unmeasured confounding* introduced in Chap. 2. See Robins and Hernán (2009) or Dawid and Didelez (2010) for a detailed exposition of G-computation.

6.1.1 Applications and Implementation of G-computation

G-computation has seen considerable use in the last decade. Thall et al. (2002) considered G-computation to evaluate a phase II clinical trial of prostate cancer treatment. Lavori and Dawson (2004) demonstrated (with R pseudocode) how to evaluate two-stage data, motivated by the treatment for major depressive disorder in the sequentially randomized STAR*D trial; see Chap. 2 for a brief description of this trial. Bembom and Van der Laan (2007) demonstrated the use of G-computation and compared results with marginal structural models (see Sect. 5.2) to examine the optimal chemotherapy for the treatment of prostate cancer, choosing from among four first-line treatments and the same four treatments offered as salvage therapy (Thall et al. 2007b).

One of the most complex and realistic implementations of G-computation using epidemiological data was performed by Taubman et al. (2009), who used more than 20 years of data from the Nurses' Health Study to examine the effect of composite lifestyle interventions on the risk of coronary heart disease. Similarly, Young et al. (2011) analyzed data from a large multi-country cohort study of HIV+ individuals to determine when to initiate antiretroviral therapy as a function of CD4 cell count, and Westreich et al. (2012) used G-computation to evaluate the impact of antiretroviral therapy on time to AIDS or death. The question was the same as that investigated by Cain et al. (2010) using a marginal structural modeling approach (albeit with different data). G-computation has also been adopted in the econometric literature (e.g. Abbring and Heckman 2007), where it has been used to explore the effects of television-watching on performance in math and reading (Huang and Lee 2010), and of spanking on behavior (Lee and Huang 2012).

Diggle et al. (2002) provided a simple expositional example on two stages where all variables are binary; in such a case, it is simple to implement G-computation non-parametrically (i.e. without using a parametric model for the conditional mean or distribution). More recently, Daniel et al. (2013) demonstrated the use of G-computation, as well as two semi-parametric approaches to estimating time-varying treatment effects, using simulated data. In the tutorial, a small by-hand example of a non-parametric implementation of G-computation is given as is a more complex scenario which requires parametric modeling. The supplementary material in the tutorial include a worked example in which there is loss to follow-up, so that the treatment of interest is redefined to be not simply treatment pattern $\bar{a}$, but rather receipt of treatment pattern $\bar{a}$ *and* continued observation. G-computation has been implemented as a SAS macro

`http://www.hsph.harvard.edu/causal/software/`

and as a Stata command (Daniel et al. 2011), facilitating dissemination and use of the method.

There are two potentially serious drawbacks to G-computation. The first is that in complex settings (many stages, or high dimensional intermediate observations), G-computation typically requires an estimate of the distribution of each intermediate outcome O_j, given each possible history up to that time point. Using a Monte Carlo

approach to simulate the counterfactual distribution for complex longitudinal data with many stages can be computationally intensive, however algorithms have been proposed to reduce the computational burden (Neugebauer and Van der Laan 2006). The second, and more worrisome, limitation of G-estimation is that incorrect model specification can lead to biased results and incorrect conclusions in longitudinal settings with time-varying treatments, even when the "sharp" null hypothesis holds, i.e. when there is no treatment effect for any individual so that $Y(\bar{a}_K) - Y(\bar{a}'_K) = 0$ with probability 1 for all regimes $\bar{a}_K$ and $\bar{a}'_K$. This is known as the null paradox, and can occur even in sequentially randomized trials where randomization probabilities are known (Robins and Wasserman 1997).

6.1.2 Breastfeeding and Vocabulary: An Illustration

Background and Study Details

The PROmotion of Breastfeeding Intervention Trial (PROBIT) (Kramer et al. 2001), briefly introduced in Chap. 2, randomized hospitals and affiliated polyclinics in the Republic of Belarus to a breastfeeding encouragement intervention modeled on the WHO/UNICEF Baby-Friendly Hospital Initiative or to standard care. All study infants were born from June 17, 1996, to December 31, 1997, at term in one of 31 Belarusian maternity hospitals, weighing at least 2,500 g, initiated breastfeeding, and were recruited during their postpartum stay. This resulted in the enrollment of 17,046 mother-infants pairs who were followed regularly for the first year of life. In a later wave of PROBIT, follow-up interviews and examinations were performed on 13,889 (81.5 %) children at 6.5 years of age. One of the components of these visits was the administration of the Wechsler Abbreviated Scales of Intelligence (WASI), which consists of four subtests: vocabulary, similarities, block designs, and matrices. We focus our analysis on the vocabulary subtest.

Many studies from developed countries have observed higher cognitive scores on IQ and other tests among both children and adults who were breastfed compared with those who were formula-fed (Anderson et al. 1999). Based on an intention-to-treat analysis, PROBIT demonstrates that prolonged and exclusive breastfeeding improves children's cognitive development (Kramer et al. 2008). We consider 13,739 children (excluding 159 children from the follow-up due to missing data in the variables of interest from the first year of life). A simple random effects model controlling for within-hospital correlation reveals a statistically significant effect of the intervention on the vocabulary subset score of 7.5 points (95 % CI: 0.7 to 14.4 points) in these children and, as noted by Kramer et al. (2008), the intervention also served to significantly and meaningfully increase the duration and intensity of breastfeeding; for instance, 43.3 % of infants in the intervention group were exclusively breastfed at 3 months of age, as compared to 6.4 % of the infants in the control group. Here, we provide a demonstration of G-computation to examine the evidence

that actual breastfeeding (rather than exposure to breastfeeding encouragement) increases verbal cognitive ability, and consider whether tailoring breastfeeding habits to infant growth can improve this outcome.

Analysis and Results

In this example, consider two key stages (intervals): birth to 3 months, and 3–6 months of age. The "treatment" or exposure of interest for our analysis is any breastfeeding measured in each of the stages. That is, A_1 takes the value 1 if the child was breastfed up to 3 months of age (and is set to -1 otherwise), and A_2 is the corresponding quantity for any breastfeeding from 3 to 6 months of age. Note that any breastfeeding allows for exclusive breastfeeding or breastfeeding with supplementation with formula or solid foods. The outcome, Y, is the vocabulary subtest score on the WASI measured at age 6.5 years. A single tailoring variable is considered at each stage: the birthweight of the infant at the first stage, and the infant's 3-month weight at the second stage.

Implementing G-computation to address the question of whether breastfeeding itself produces higher vocabulary subtest scores requires models for both the vocabulary subtest score, as well as for the 3-month weight. A linear model was used to fit the vocabulary subtest score on the log-scale (Y) as a function of baseline covariates (intervention group status, geographical location (eastern/rural, eatern/urban, western/rural, or western/urban), mother's education, mother's smoking status, family history of allergy, mother's age, mother's breastfeeding of any previous children, whether birth was by cesarean section, gender) as well as birthweight, 3 month weight, breastfeeding from 0 to 3 months (A_1), breastfeeding from 3 to 6 months (A_2), and the first-order interactions (i) $A_1 \times A_2$, (ii) A_1 by birthweight, and (iii) A_2 by 3-month weight. Note that O_1 includes all baseline covariates and the tailoring variable birthweight, while O_2 includes all variables in O_1 in additional to 3-month weight. Three-month weight was also fit on the log scale using a linear model that conditioned on the baseline covariates and birthweight, breastfeeding from 0 to 3 months (A_1), and the interaction between A_1 and birthweight.

The G-computation procedure used can be described by the following steps, for any regime of interest, $d = (d_1(h_1), d_2(h_2))$:

1. Fit an appropriate joint distribution model for the baseline variables O_1. For PROBIT, a non-parametric approach is adopted, and the empirical distribution was used.
2. Fit an appropriate model to the intermediate variable, O_2, as a function of O_1 and A_1. For PROBIT, a linear model on the log-transformed 3-month weight is used.
3. Fit an appropriate model to the response, Y, as a function of O_1, A_1, O_2, and A_2. For PROBIT, a linear model on the log-transformed subtest score is used.
4. Create a hypothetical population by drawing a random sample with replacement from the distribution of baseline covariates fit in Step (1).

5. Using coefficient estimates and randomly sampled residuals from the model fit in Step (2), determine the (counterfactual) intermediate variable $o_2(d_1)$ under treatment regime d_1 with history $h_1 = o_1$.
6. Using coefficient estimates and randomly sampled residuals from the model fit in Step (3), determine the response under treatment regime d with history $h_1 = o_1$, $h_2 = (o_1, d_1(h_1), o_2(d_1))$.

Using this approach, we can compare distributions under different treatment regimes, such as the static regimes "never breastfeed" or "breastfeed until at least 6 months of age", or the dynamic regime "breastfeed until three months of age, then continue only if 3-month weight exceeds 6.5 kg". Note that in steps 5 and 6, one could assume a likelihood for the potential outcomes, e.g. a normal distribution, rather than the less parametric approach of selecting a random residual.

Table 6.1 Parameter coefficients from a linear regression model for the log-transformed vocabulary subtest score of the WASI and log-transformed 3-month weight

	Vocab. score		Weight at 3 months	
	Est.	SD	Est.	SD
Intercept	4.315	0.047	1.312	0.020
Intervention	0.071	0.035	0.012	0.006
East Belarus (rural)	0.034	0.048	−0.015	0.008
West Belarus (urban)	0.008	0.053	−0.011	0.008
West Belarus (rural)	−0.002	0.044	−0.016	0.007
Attended some university	0.047	0.003	0.008	0.002
Completed university	0.099	0.004	0.009	0.003
Smoker	−0.008	0.008	0.002	0.005
Allergy	0.023	0.006	−0.003	0.004
Age	0.009	0.002	−0.001	0.001
Age^2	0.000	0.000	0.000	0.000
BF previously	−0.049	0.003	0.001	0.002
Did not BF	−0.042	0.003	−0.002	0.002
Cesarean	0.000	0.004	−0.005	0.002
Gender	−0.011	0.002	0.045	0.001
Birthweight	0.010	0.004	0.139	0.002
A_1: Breastfed 0–3 months	0.048	0.019	0.008	0.012
Weight at 3 months	0.017	0.002		
A_2: Breastfed 3–6 months	−0.121	0.057		
$A_1 \times$Birthweight	−0.012	0.005	0.001	0.004
$A_2 \times$Weight at 3 months	0.016	0.007		
$A_1 \times A_2$	0.019	0.037		

Results from regression models which account for within-hospital clustering are presented in Table 6.1; coefficient estimates from models which ignored clustering are very similar. Statistically significant effects of breastfeeding and its interaction with weight are found in the model for the log vocabulary score. However, when these models are subsequently used to produce samples from the counterfactual distribution of outcomes, it is evident that the impact of breastfeeding itself on the

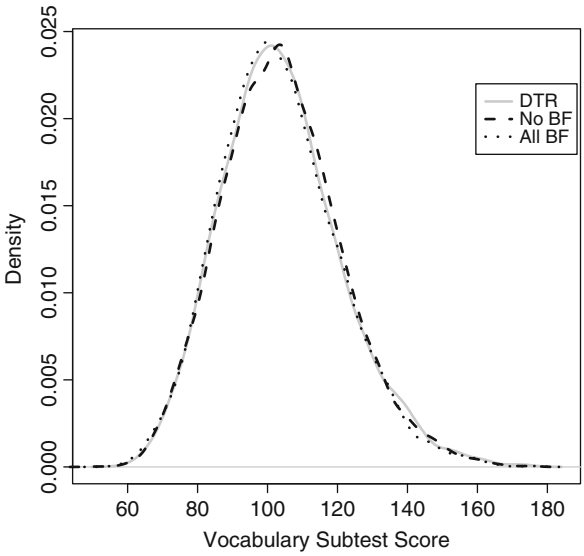

Fig. 6.1 Counterfactual vocabulary subtest score under three different breastfeeding regimes estimated by G-computation: a DTR (*gray*, *solid line*), no breastfeeding (*dashed line*) and breastfeeding until at least 6 months (*dotted line*)

vocabulary subtest score is minimal (see Fig. 6.1), with mean test scores varying by less than one point under the three regimes considered. These results are broadly in line with the findings of Moodie et al. (2012).

6.2 Bayesian Estimation of DTRs

A detailed presentation of the many modeling choices required for any particular application of a Bayesian estimation of a dynamic treatment regime is beyond the scope of this text, however a great number of resources are available to the interested reader (see, e.g. Chen et al. 2010).

6.2.1 Approach and Applications

Parametric frequentist methods of estimating optimal DTRs typically rely on the assumption that all models are correctly specified, while semi-parametric approaches are often able to relax some modeling assumptions. In the Bayesian setting, a number of different approaches have been proposed. Wathen and Thall (2008) considered at the outset a number of candidate models, and choose from among them using an approximate Bayes factor. Arjas and Saarela (2010) used model

averaging over random draws from the space of possible models, so that inference is based on results from the averaged model. They argued that this is a distinct advantage of a Bayesian approach over frequentist methods (semi-parametric or otherwise), as it allows the analyst to incorporate uncertainty regarding model specification into the estimation procedure. As in the frequentist approaches, Bayesian estimation of optimal dynamic treatment regimes may be computationally burdensome in complex settings with many covariates and/or stages, although some advances have been made. For example, Wathen and Thall (2008) adapted the forward-sampling approach of Carlin et al. (1998) so as to be able to sample from the predictive distribution of the outcome under each of several regimes, where the distribution is estimated from the observed data, however in this case the "regimes" of interest were stopping rules for group sequential clinical trials.

Arjas and Saarela (2010) considered data on HIV treatment from the Multi-Center AIDS Cohort (Kaslow et al. 1987), focusing on a two-stage setting in which there is a single (continuous) tailoring variable at each stage, treatment is binary, and the outcome is a continuous variable. They postulated appropriate prior distributions for each component of the joint likelihood, and thus obtained the joint posterior distribution. Following this, the posterior predictive distribution was used to see how the outcomes of individuals drawn from the same population as those who formed the sample data were distributed under different treatment patterns. This approach uses the principles set forth by Arjas and Parner (2004), who suggested using summaries of the posterior predictive distributions as the main criterion for comparing different treatment regimes, leading to what they refer to as an "integrated causal analysis" in which the same probabilistic framework is used for inference about model parameters as well as for treatment comparisons and hence the choice of an optimal regime.

Lavori and Dawson (2000) used multiple imputations performed by an approximate Bayesian bootstrap (Rubin and Shenker 1991) to generate draws from the counterfactual distributions, and thereby allow a non-parametric means of comparing mean outcomes under different treatment strategies. Zajonc (2012) proposed a similar approach, though from a more overtly Bayesian perspective, and considers data from the North Carolina Education Research Data Center, examining the impact of honors programs on tenth grade mathematics test scores. Two stages with a binary exposure were considered; several baseline confounders, and two time-dependent variables were used in the analysis. Tailoring of the decision rule was performed in a variety of ways including using the single, continuous mathematics score at the start of each stage as well as by creating an index score that was a composite of five variables including sex, race, and test score. The approach was the same in spirit as that of Arjas and Saarela (2010), however the motivation was somewhat different. Like Lavori and Dawson (2000) and Zajonc (2012) framed the estimation problem as one of missing data, where the missing information is on the potential outcomes, and undertakes estimation through what is effectively a multiple imputation approach. Thus, Bayesian machinery was used to estimate the posterior predictive distribution of the potential outcomes, and the optimal regime was selected as that which maximized the expected posterior utility,

where the utility was simply some analyst-defined function of the outcome and potentially other factors such as treatment cost.

The Bayesian posterior predictive approach to dynamic treatment regime estimation is in many ways similar to G-computation, but is more readily able to capture three primary sources of variability in estimators: (i) randomness in covariates and outcomes as specified by the predictive distribution for the outcome given data, (ii) potential randomness in the regime (if, for example, the DTR had a stochastic component such as "treat within three months of the occurrence of a particular health-related event"); and (iii) variability in the unknown model parameters (Arjas 2012). There have also been a number of applications of Bayesian predictive inference to examine questions of causation for non-continuous outcomes, many by Elja Arjas and colleagues. For example, Arjas and Andreev (2000) used a Bayesian nonparametric intensity model for recurrent events to study the impact of child-care setting on the number of ear infections.

6.2.2 Assumptions as Viewed in a Bayesian Framework

Although rarely discussed explicitly in Bayesian analyses, most implicitly require the assumption of no unmeasured confounding or exchangeability (Saarela et al. 2012b). Arjas and Saarela (2010) used this assumption so that in using the posterior predictive distribution to estimate mean response, the values of treatment at each stage could be set to specific values, i.e. they noted "we can switch from observed treatment values in the data to 'optional' or 'forced' in the predictions is again consequence of the no unmeasured confounders postulate" and that this "can be viewed as representing 'do'-conditioning (Pearl 2009)." Arjas and Saarela (2010) further related this condition to the assumption of missing at random (Little and Rubin 2002). Arjas (2012) elaborated on the mathematical conditions under which the forced 'do'-probabilities can be identified from observational data (what he terms a 'see'-study, following Lindley (2002)) and related these probabilistic statements to Rubin's causal model (1974) and to the sequential randomization design of Robins (1986), which rely on the potential outcomes framework. Zajonc (2012), too, linked the NUC assumption to the idea that the treatment mechanism may be considered ignorable, i.e. possibly dependent on observed data, but not on unobserved counterfactual quantities.

6.2.3 Breastfeeding and Vocabulary: An Illustration, Continued

We now return to the PROBIT trial, and re-analyze the data using a Bayesian predictive approach.

A Bayesian G-computation procedure is designed to complement and compare with the analysis performed in Sect. 6.1.2. A variety of summary measures of the

counterfactual outcome under different regimes may be computed. For example, to estimate the counterfactual distribution of outcomes (see Fig. 6.1 for a frequentist version) for a regime of interest, $d = (d_1(h_1), d_2(h_2))$, we may:

1. Fit an appropriate joint distribution model for the baseline variables O_1. For PROBIT, a non-parametric approach is adopted, and the empirical distribution was used.
2. Fit an appropriate model to the intermediate variable, O_2, as a function of O_1 and A_1. For PROBIT, a linear model on the log-transformed 3-month weight is used with a normal, mean-zero, variance 10 (covariance 0) prior is used for all regression coefficients, β_2, and the (proper) Inverse Gamma prior with parameters 5, 0.2 is used for the variance parameters.
3. Fit an appropriate model to the response, Y, as a function of O_1, A_1, O_2, and A_2. As in the previous step, a normal (0,diag(10)) prior is used for regression coefficients, β_1, and an Inverse Gamma(5,0.2) prior is used for the variance parameters in the PROBIT analysis.
4. Draw a sample from the posterior predictive distribution of the counterfactual mean outcome using the following steps:
   (a) Draw a random sample of size 10,000 with replacement from the distribution of baseline covariates fit in Step (1) to create a hypothetical population.
   (b) For each member of the hypothetical population, draw a sample value from the posterior distribution of β_1 found in Step (2) and use this to determine the mean (counterfactual) intermediate variable $o_2(d_1)$, and call this $\mu_{o_2}(\beta_1)$. Next, draw a value of $o_2(d_1)$ from the posterior predictive distribution with mean $\mu_{o_2}(\beta_1)$.
   (c) Then, for each member of the hypothetical population, draw a sample value from the posterior distribution of β_2 found in Step (3) and use this to determine the mean (counterfactual) outcome, and call this $\mu_y(\beta_2)$. Draw a value of $y(d_1, d_2)$ from its posterior predictive distribution with mean $\mu_y(\beta_2)$.

Note that in this approach, the uncertainty in the estimation of the regression parameters β_1 and β_2 is incorporated into the sampling from the counterfactual means. Figure 6.2 compares the counterfactual distribution of vocabulary subtest scores under two static regimes "never breastfeed" and "breastfeed until at least 6 months of age". Note that these are nearly identical to their frequentist analogues, which is reassuring though not surprising given the large sample size.

The approach described above may be altered to produce the distributions of the *mean* counterfactual outcomes, averaging over the covariate space, rather than the distribution of the counterfactual outcomes themselves. To do this, follow Steps (1)–(3) as above. Perform Step (4), and then in (5) take the average counterfactual outcome. Repeat Steps (4) and (5) a large number of times; boxplots of 1,000 means of counterfactual distributions can be found in Fig. 6.3.

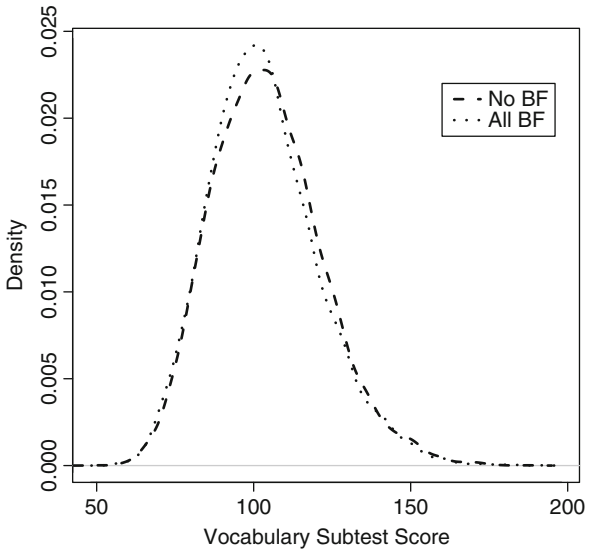

Fig. 6.2 Counterfactual vocabulary subtest score under two different breastfeeding regimes estimated by a Bayesian implementation of G-computation: no breastfeeding (*dashed line*) and breastfeeding until at least 6 months (*dotted line*)

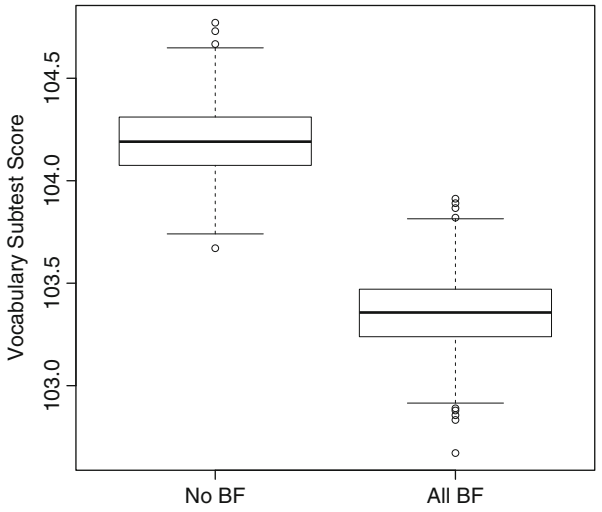

Fig. 6.3 Distribution of the mean counterfactual vocabulary subtest score under two breastfeeding regimes estimated by a Bayesian implementation of G-computation: no breastfeeding (*left*) and breastfeeding until at least 6 months (*right*)

6.3 Discussion

In this chapter, we have presented the fully parametric estimation approach of G-computation. We presented the approach in the frequentist setting, then turned our attention to the Bayesian DTR literature, describing the general approach which is effectively a G-computation based on posterior predictive distributions. We implemented G-computation in the PROBIT data, using both a frequentist and then a Bayesian approach; the conclusions of the two analyses were nearly identical, but the Bayesian calculations permitted the incorporation of the variability due to the estimation of the time-varying parameter distributions into the counterfactual response distributions. A Bayesian perspective on the variance of marginal structural models has recently been considered (Saarela et al. 2012a), but this has not yet been used in the context of DTRs.