Web-based Supplementary Materials for Inverse Probability Weighted Cox Regression for Doubly Truncated Data

Micha Mandel
Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel
email: msmic@huji.ac.il

Jacobo de Uña-Álvarez
Department of Statistics and OR and Center for Biomedical Research (CINBIO),
University of Vigo, Vigo 36310, Spain
email: jacobo@uvigo.es

David K. Simon
Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School,
Boston, MA 02115, USA
email: dsimon1@bidmc.harvard.edu

Rebecca A. Betensky
Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
email: betensky@hsph.harvard.edu

July 24, 2017

A Large sample properties

We consider independent realizations $(T_1, Z_1, L_1, U_1), \ldots, (T_n, Z_n, L_n, U_n)$ having the law of $(T, Z, L, U) \mid D$, where $D = \{L \leq T \leq U\}$. The following regularity conditions are assumed.

1. T is continuous and $t_{min} \leq T \leq t_{max}$ with probability one for some $t_{min} \geq 0$ and $t_{max} < \infty$.

2. The covariates Z are bounded.

3. $W(t) > \epsilon$ and $\widehat{W}(t) > \epsilon$ on $[t_{min}, t_{max}]$ for some $\epsilon > 0$.

4. The covariates are linearly independent.

5. The limiting second derivative of $PL(\beta)$ (defined below) is positive definite.

Under the conditions above and Conjectures 2.1 and 2.2, the following two conditions hold; these are used in our proofs below.

Condition A.1 (Uniform convergence)**.** $\max_{1 \leq i \leq n}\{W(T_i)/\widehat{W}(T_i) - 1\} \longrightarrow 0$ *in probability.*

Condition A.2 (IID representation)**.** $\sqrt{n}\left[\{W(t)\}^{-1} - \{\widehat{W}(t)\}^{-1}\right] = n^{-1/2}\sum_{i=1}^{n} \xi_n(\mathcal{D}_i, t) +$ $o_p(1)$ *uniformly on* $t \in [t_{min}, t_{max}]$*, where* $\mathcal{D}_i = (T_i, L_i, U_i)$ *is the data for subject* i*, and* $\xi_n(\mathcal{D}_i, t)$ *are independent and identically distributed zero mean random variables having finite variance.*

Theorem A.1. *Let* $\widehat{\beta}$ *be the solution of* $U(\beta) = 0$ *in (8) with* $\widehat{W}$ *substituted for* W*. If Conjectures 2.1 and 2.2 hold and under the regularity conditions 1-5 above,* $\widehat{\beta}$ *is consistent and* $n^{1/2}(\widehat{\beta} - \beta)$ *is asymptotically equivalent to a U-statistic with kernel given below;*

and hence, has an asymptotic normal distribution provided that the kernel's second order moment is finite.

A.1 Consistency

To prove consistency we follow the steps of Andersen and Gill (1982) and show:

1. The log-likelihood is concave in β.
2. The log-likelihood converges pointwise to a function that obtains its maximum at β_0.
3. Pointwise convergence of concave functions implies uniform convergence (Andersen and Gill, Theorem II.1).
4. Uniform convergence of concave functions implies convergence of the point of maximum (Andersen and Gill, Corollary II.2).

Concavity

We first show that the maximum is unique by proving concavity using standard arguments. Let $\widehat{W}$ be the estimator of W (satisfying $0 < \widehat{W}(T_j) < 1$ for all j), and consider the centered pseudo-log likelihood that vanishes at β_0:

$$PL(\beta) = n^{-1} \sum_{i=1}^{n} \left[(\beta^t - \beta_0^t) Z_i - \log \left\{ \frac{n^{-1} \sum_{j=1}^{n} e^{\beta^t Z_j} \{\widehat{W}(T_j)\}^{-1} I\{T_j \geq T_i\}}{n^{-1} \sum_{j=1}^{n} e^{\beta_0^t Z_j} \{\widehat{W}(T_j)\}^{-1} I\{T_j \geq T_i\}} \right\} \right].$$

The derivative of PL with respect to β,

$$\frac{\partial PL(\beta)}{\partial \beta} = \widetilde{U}_n(\beta) = n^{-1} \sum_{i=1}^{n} \left[Z_i - \sum_{j=1}^{n} Z_j \frac{e^{\beta^t Z_j} \{\widehat{W}(T_j)\}^{-1} I\{T_j \geq T_i\}}{\sum_{j'=1}^{n} e^{\beta^t Z_{j'}} \{\widehat{W}(T_{j'})\}^{-1} I\{T_{j'} \geq T_i\}} \right],$$

is the estimating equation (8) after plugging-in $\widehat{W}$ for W. Let

$$p_{j,i}(\beta) = e^{\beta^t Z_j} \{\widehat{W}(T_j)\}^{-1} I\{T_j \geq T_i\} \Big/ \sum_{j'=1}^{n} e^{\beta^t Z_{j'}} \{\widehat{W}(T_{j'})\}^{-1} I\{T_{j'} \geq T_i\},$$

so that $\widetilde{U}_n(\beta) = n^{-1}\sum_{i=1}^n\{Z_i - \sum_{j=1}^n Z_j p_{j,i}(\beta)\}$. Differentiating $\widetilde{U}_n(\beta)$ with respect to β gives

$$\frac{\partial^2 PL(\beta)}{\partial\beta^2} = -n^{-1}\sum_{i=1}^n\left\{\sum_{j=1}^n Z_j^{\otimes 2}p_{j,i}(\beta) - \Big(\sum_{j=1}^n Z_j p_{j,i}(\beta)\Big)^{\otimes 2}\right\}. \quad \text{(A.1)}$$

Thus, the second derivative of $PL(\beta)$ is represented as a sum of minus of variance matrices and hance is negative definite (assuming the design matrix is of full rank), and therefore PL is concave and has a unique maximum.

Pointwise consistency of PL

Let $Y_i(t) = I\{T_i \geq t\}$ and for $k = 0, 1, 2$ define

$$\begin{aligned}
\widetilde{S}_n^{(k)}(\beta, t) &= n^{-1}\textstyle\sum_{j=1}^n Y_j(t)\{\widehat{W}(T_j)\}^{-1}Z_j^{\otimes k}\exp(\beta Z_j),\\
S_n^{(k)}(\beta, t) &= n^{-1}\textstyle\sum_{j=1}^n Y_j(t)\{W(T_j)\}^{-1}Z_j^{\otimes k}\exp(\beta Z_j),\\
s_n^{(k)}(\beta, t) &= E\{S_n^{(k)}(\beta, t)\}.
\end{aligned}$$

Here $S_n^{(k)}$ is a scalar, a vector and a matrix for $k = 0, 1$ and 2, respectively. Then,

$$PL(\beta) = n^{-1}\sum_{i=1}^n\left[(\beta^t - \beta_0^t)Z_i - \log\left\{\frac{\widetilde{S}_n^{(0)}(\beta, T_i)}{\widetilde{S}_n^{(0)}(\beta_0, T_i)}\right\}\right].$$

It is shown below that

$$PL(\beta) = n^{-1}\sum_{i=1}^n\left[(\beta^t - \beta_0^t)Z_i - \log\left\{\frac{s_n^{(0)}(\beta, T_i)}{s_n^{(0)}(\beta_0, T_i)}\right\}\right] + o_p(1), \quad \text{(A.2)}$$

where the leading term is an average of independent and identically distributed random variables, which are bounded by the regularity conditions. Thus,

$$PL(\beta) \to E\left[(\beta^t - \beta_0^t)Z - \log\left\{\frac{s^{(0)}(\beta, T)}{s^{(0)}(\beta_0, T)}\right\} \;\Big|\; D\right] \quad \text{(A.3)}$$

in probability by the law of large numbers.

To show (A.2), write

$$\log\left\{\frac{\widetilde{S}_n^{(0)}(\beta,T_i)}{\widetilde{S}_n^{(0)}(\beta_0,T_i)}\right\}-\log\left\{\frac{S_n^{(0)}(\beta,T_i)}{S_n^{(0)}(\beta_0,T_i)}\right\}=\log\left\{\frac{\widetilde{S}_n^{(0)}(\beta,T_i)}{S_n^{(0)}(\beta,T_i)}\right\}-\log\left\{\frac{\widetilde{S}_n^{(0)}(\beta_0,T_i)}{S_n^{(0)}(\beta_0,T_i)}\right\}.$$

Now, by Assumption A.1

$$\widetilde{S}_n^{(0)}(\beta,T_i)=n^{-1}\sum_{j=1}^{n}Y_j(T_i)\{W(T_j)\}^{-1}\exp(\beta Z_j)W(T_j)/\widehat{W}(T_j)=S_n^{(0)}(\beta,T_i)\{1+o_p(1)\}$$

and similarly $\widetilde{S}_n^{(0)}(\beta_0,T_i)=S_n^{(0)}(\beta_0,T_i)\{1+o_p(1)\}$. It follows that

$$\log\left\{\frac{\widetilde{S}_n^{(0)}(\beta,T_i)}{\widetilde{S}_n^{(0)}(\beta_0,T_i)}\right\}-\log\left\{\frac{S_n^{(0)}(\beta,T_i)}{S_n^{(0)}(\beta_0,T_i)}\right\}=o_p(1). \tag{A.4}$$

Lemma A.1. *Assume that T is continuous and Z is bounded, then $S_n^{(k)}(\beta,t)\to s^{(k)}(\beta,t)$ $(k=0,1,2)$ in probability uniformly in $t\in[0,\tau]$ for any constant τ.*

This follows from Theorem 16(a) of Ferguson (1996) using convergence in probability rather than almost sure convergence, and replacing continuity with continuity with probability 1. Thus, $S_n^{(0)}(\beta,T_i)=s_n^{(0)}(\beta,T_i)\{1+o_p(1)\}$, and

$$\log\left\{\frac{S_n^{(0)}(\beta,T_i)}{S_n^{(0)}(\beta_0,T_i)}\right\}-\log\left\{\frac{s_n^{(0)}(\beta,T_i)}{s_n^{(0)}(\beta_0,T_i)}\right\}=o_p(1). \tag{A.5}$$

Equations (A.4) and (A.5) together establish (A.2), as we assume that $T<t_{max}<\infty$ for some t_{max}.

Convergence of concave functions

As PL is a sequence of concave functions, so is its limit, the right hand side of Equation (A.3) (by Theorem II.1 of Andersen and Gill 1982). It remains to show that its maximum is indeed β_0. Simple differentiation (similar to the proof of concavity of PL) shows that this is indeed the case. Here the standard assumption that the limiting second derivative is positive definite is required (see Andersen and Gill, 1982). Consistency is established by Corollary II.2 of Andersen and Gill 1982.

A.2 Asymptotic distribution

Define the column vectors

$$\widetilde{U}_n(\beta) = \sum_{i=1}^{n} \left\{ Z_i - \frac{\widetilde{S}_n^{(1)}(\beta, T_i)}{\widetilde{S}_n^{(0)}(\beta, T_i)} \right\}$$

$$U_n(\beta) = \sum_{i=1}^{n} \left\{ Z_i - \frac{S_n^{(1)}(\beta, T_i)}{S_n^{(0)}(\beta, T_i)} \right\}.$$

Using a first order Taylor expansion we have

$$\widetilde{U}_n(\beta) - \widetilde{U}_n(\beta_0) = \widetilde{\varphi}(\beta)(\beta - \beta_0),$$

where $\widetilde{\varphi}(\beta) = \nabla \widetilde{U}_n(\beta)$ and β is on the line segment between β and β_0. By definition, our estimator satisfies $\widetilde{U}_n(\widehat{\beta}) = 0$, and by plugging $\widehat{\beta}$ for β we have

$$n^{1/2}(\widehat{\beta} - \beta_0) = -\{n^{-1}\widetilde{\varphi}(\beta)\}^{-1} n^{-1/2} \widetilde{U}_n(\beta_0).$$

So there are two tasks. The first is to show that $n^{-1}\widetilde{\varphi}(\beta)$ converges in probability to a fix matrix. The second is to show that $n^{-1/2}\widetilde{U}_n(\beta_0)$ converges to a Gaussian vector.

Convergence of $n^{-1}\widetilde{\varphi}(\beta)$

$$\widetilde{\varphi}(\beta) = \nabla \widetilde{U}_n(\beta) = \sum_{i=1}^{n} \frac{\widetilde{S}^{(2)}(\beta, T_i)\widetilde{S}^{(0)}(\beta, T_i) - \widetilde{S}^{(1)}(\beta, T_i)\widetilde{S}^{(1)}(\beta, T_i)^T}{\{\widetilde{S}^{(0)}(\beta, T_i)\}^2}$$

By Assumption A.1, we have that $n^{-1}\nabla\widetilde{U}_n(\beta) = n^{-1}\nabla U_n(\beta) + o_p(1)$, and by Lemma A.1

$$n^{-1}\nabla U_n(\beta) = n^{-1} \sum_{i=1}^{n} \frac{s^{(2)}(\beta, T_i)s^{(0)}(\beta, T_i) - s^{(1)}(\beta, T_i)s^{(1)}(\beta, T_i)^T}{\{s^{(0)}(\beta, T_i)\}^2} + o_p(1).$$

It follows, using arguments similar to those establishing Lemma A.1, that

$$n^{-1} \sum_{i=1}^{n} \frac{s^{(2)}(\beta, T_i)s^{(0)}(\beta, T_i) - s^{(1)}(\beta, T_i)s^{(1)}(\beta, T_i)^T}{\{s^{(0)}(\beta, T_i)\}^2}$$

converges in probability uniformly in a neighborhood of β_0. Consistency of $\widehat{\beta}$ and continuity (in β) of the above sequence establish the convergence of $n^{-1}\widetilde{\varphi}(\beta)$.

Asymptotic of $n^{-1/2}\widetilde{U}_n(\beta_0)$

We have

$$n^{-1/2}\widetilde{U}_n(\beta_0) = n^{-1/2}U_n(\beta_0) + n^{-1/2}\{\widetilde{U}_n(\beta_0) - U_n(\beta_0)\} \tag{A.6}$$

The first term gives the contribution of the weighted estimating equation and the second shows the variability due to estimation of the weight function. Focusing on the first term,

$$U_n(\beta_0) = \sum_{i=1}^{n}\left\{Z_i - \frac{S_n^{(1)}(\beta_0, T_i)}{s^{(0)}(\beta_0, T_i)}\right\} - \sum_{i=1}^{n}\frac{S_n^{(1)}(\beta_0, T_i)}{S_n^{(0)}(\beta_0, T_i)s^{(0)}(\beta_0, T_i)}\left\{s^{(0)}(\beta_0, T_i) - S_n^{(0)}(\beta_0, T_i)\right\}. \tag{A.7}$$

The first expression of (A.7) can be represented as a sum of

$$n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left\{Z_i - \frac{Y_j(T_i)\{W(T_j)\}^{-1}Z_j\exp(\beta_0 Z_j)}{s^{(0)}(\beta_0, T_i)}\right\} \tag{A.8}$$

and

$$n^{-1}\sum_{i=1}^{n}\left\{Z_i - \frac{\{W(T_i)\}^{-1}Z_i\exp(\beta_0 Z_i)}{s^{(0)}(\beta_0, T_i)}\right\}.$$

The first of the two expressions resembles a U-statistic whose kernel has zero expectation by the iid assumption and (7) (since we have $E(Z|T = t, D) = s^{(1)}(\beta_0, t)/s^{(0)}(\beta_0, t)$). The second expression converges to its expectation by the low of large numbers.

For the second term of (A.7), we have by Lemma A.1 $S_n^{(k)}(\beta_0, t) \to s^{(k)}(\beta_0, t)$ in probability, uniformly in t, so the second term is asymptotically equivalent to

$$n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{s^{(1)}(\beta_0, T_i)}{\{s^{(0)}(\beta_0, T_i)\}^2}\left\{s^{(0)}(\beta_0, T_i) - Y_j(T_i)\{W(T_j)\}^{-1}\exp(\beta_0 Z_j)\right\}.$$

This again can be represented as a sum of a U-type statistic

$$n^{-1}\sum_{i=1}^{n}\sum_{j\neq i}\frac{s^{(1)}(\beta_0, T_i)}{\{s^{(0)}(\beta_0, T_i)\}^2}\left\{s^{(0)}(\beta_0, T_i) - Y_j(T_i)\{W(T_j)\}^{-1}\exp(\beta_0 Z_j)\right\} \tag{A.9}$$

and an average

$$n^{-1} \sum_{i=1}^{n} \frac{s^{(1)}(\beta_0, T_i)}{\{s^{(0)}(\beta_0, T_i)\}^2} \left\{ s^{(0)}(\beta_0, T_i) - \{W(T_i)\}^{-1} \exp(\beta_0 Z_i) \right\}.$$

Turning now to the second term of (A.6), $n^{-1/2}\{\widetilde{U}_n(\beta_0) - U_n(\beta_0)\}$, we express it in a similar way as

$$\begin{aligned} \widetilde{U}_n(\beta_0) - U_n(\beta_0) = \quad & \sum_{i=1}^{n} \frac{\widetilde{S}_n^{(1)}(\beta_0, T_i) - S_n^{(1)}(\beta_0, T_i)}{S_n^{(0)}(\beta_0, T_i)} \\ & + \sum_{i=1}^{n} \frac{\widetilde{S}_n^{(1)}(\beta_0, T_i)}{\widetilde{S}_n^{(0)}(\beta_0, T_i) S_n^{(0)}(\beta_0, T_i)} \{S_n^{(0)}(\beta_0, T_i) - \widetilde{S}_n^{(0)}(\beta_0, T_i)\}. \end{aligned} \tag{A.10}$$

Next,

$$\widetilde{S}_n^{(k)}(\beta_0, T_i) - S_n^{(k)}(\beta_0, T_i) = n^{-1} \sum_{j=1}^{n} Y_j(T_i) Z_j^{\otimes k} \exp(\beta_0 Z_j) [\{W(T_j)\}^{-1} - \{\widehat{W}(T_j)\}^{-1}],$$

which by Assumption A.2 is asymptotically equivalent to

$$n^{-2} \sum_{\ell=1}^{n} \sum_{j=1}^{n} Y_j(T_i) Z_j^{\otimes k} \exp(\beta_0 Z_j) \{\xi_n(\mathcal{D}_\ell, T_j) + R_n(T_j)\},$$

where $R_n(T_j) = o_p(1)$. Using this and Lemma A.1, we can express the first term of (A.10) as

$$n^{-2} \sum_{i=1}^{n} \sum_{\ell=1}^{n} \sum_{j=1}^{n} \frac{Y_j(T_i) Z_j \exp(\beta_0 Z_j) \{\xi_n(\mathcal{D}_\ell, T_j) + R_n(T_j)\}}{s^{(0)}(\beta_0, T_i) + o_p(1)}, \tag{A.11}$$

and, since the ratio $\widetilde{S}/S$ equals $o_p(1)$, the second term can be expressed as

$$n^{-2} \sum_{i=1}^{n} \sum_{\ell=1}^{n} \sum_{j=1}^{n} \frac{s^{(1)}(\beta_0, T_i) + o_p(1)}{\{s^{(0)}(\beta_0, T_i) + o_p(1)\}\{s^{(0)}(\beta_0, T_i) + o_p(1)\}} Y_j(T_i) \exp(\beta_0 Z_j) \{\xi_n(\mathcal{D}_\ell, T_j) + R_n(T_j)\}. \tag{A.12}$$

Combining the terms in (A.8), (A.9), (A.11), (A.12) we have

$$n^{-1/2} \widetilde{U}_n(\beta_0) = n^{1/2} \binom{n}{3}^{-1} \sum_{i<\ell<j} (\eta_{i\ell j}^1 + \eta_{i\ell j}^2 + \eta_{i\ell j}^3 + \eta_{i\ell j}^4) + o_p(1),$$

where

$$\eta^1_{i\ell j} = \left\{ Z_i - \frac{Y_j(T_i)\{W(T_j)\}^{-1} Z_j \exp(\beta_0 Z_j)}{s^{(0)}(\beta_0, T_i)} \right\}$$

$$\eta^2_{i\ell j} = \frac{s^{(1)}(\beta_0, T_i)}{\{s^{(0)}(\beta_0, T_i)\}^2} \left\{ s^{(0)}(\beta_0, T_i) - Y_j(T_i)\{W(T_j)\}^{-1} \exp(\beta_0 Z_j) \right\}$$

$$\eta^3_{i\ell j} = \frac{Y_j(T_i) Z_j \exp(\beta_0 Z_j) \xi_n(\mathcal{D}_\ell, T_j)}{s^{(0)}(\beta_0, T_i)}$$

$$\eta^4_{i\ell j} = \frac{s^{(1)}(\beta_0, T_i)}{\{s^{(0)}(\beta_0, T_i)\}\{s^{(0)}(\beta_0, T_i)\}} Y_j(T_i) \exp(\beta_0 Z_j) \xi_n(\mathcal{D}_\ell, T_j).$$

Thus, $n^{-1/2}\widetilde{U}_n(\beta_0)$ is asymptotically equivalent to a U-statistic with kernel $\eta^1 + \eta^2 + \eta^3 + \eta^4$, and hence has an asymptotic normal distribution provided that the kernel's second order moment is finite.

B Simulation - Left Truncated Data

		Estimation approach			
		Delayed Entry		Inverse Weighting	
n	Parameter	Bias	MSE	Bias	MSE
100	β_1	-0.015	0.091	-0.012	0.087
	β_2	-0.043	0.101	-0.038	0.100
200	β_1	-0.011	0.042	-0.009	0.041
	β_2	-0.022	0.039	-0.019	0.039
400	β_1	-0.011	0.019	-0.010	0.018
	β_2	-0.003	0.019	-0.002	0.019

Table 1: Bias and MSE of $\widehat{\beta}$ for the model $h(t \mid z_1, z_2; \beta) = \exp(t-2z_1-3z_2)$, $z_1 \sim \text{Ber}(0.5)$, $z_2 \sim \text{U}\{1, 2, 3, 4\}$, and $L \sim \text{Exp}(1/4)$.

C Distribution of Weights

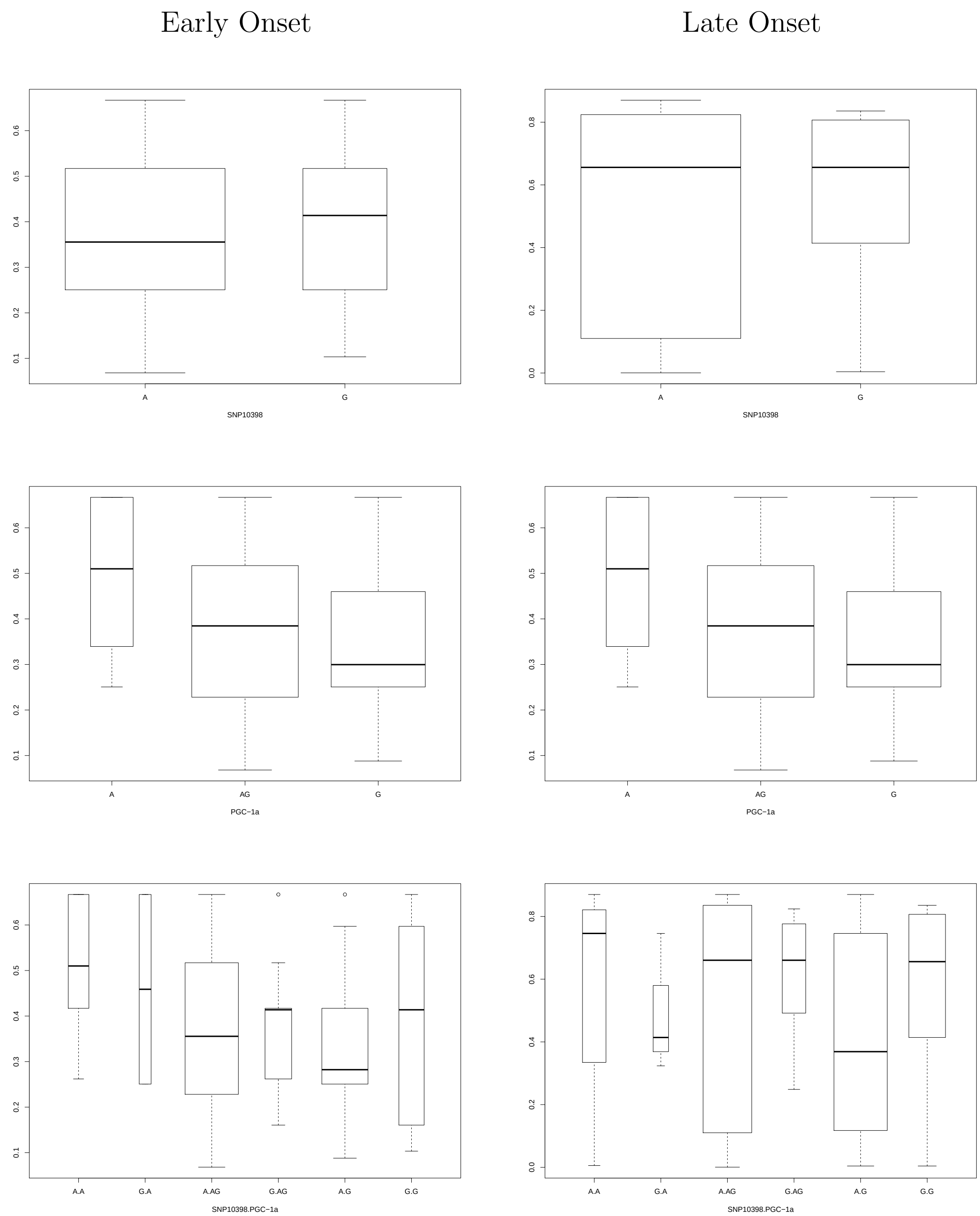

Figure 1: Box-plots of the weights for the different genotype groups.