STAT 661: Project

LS v.s. EM

Jacob, Robin, Ryan & Shen

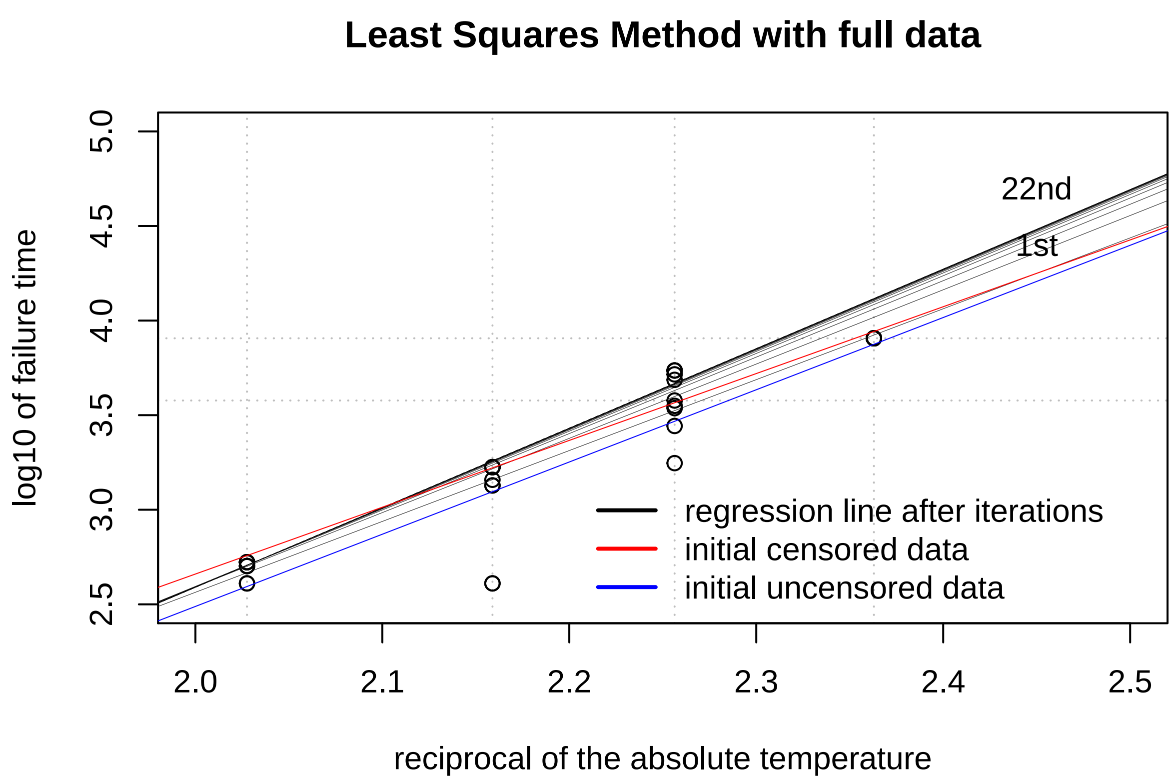
Dec, 2019

1 Appendix

1.1 Least Square Method with Full data

Using Schmee & Hahn's Least Square method, we reproduced the algoritm and got the exact same results after 22 iterations. The results of $\hat{\beta}_0$,$\hat{\beta}_1$, and $\hat{\sigma}$ are -5.8181182, 4.2041845, 0.2043471.

In page 422, The authors say "If one ignores the 150c data, the iterative least squares estimates and the maximum likelihood estimates were even closer to each other than before". I will try to confirm this concultion.



Iteration	Intercept	Slope	Sigma	mu150	mu170	mu190	mu220
1	-4.930507	3.747043	0.1572178	4.038379	3.808887	3.329470	2.829844
2	-5.260094	3.926205	0.1798671	4.098530	3.839728	3.365551	2.858422
3	-5.485695	4.040237	0.1910656	4.130648	3.855994	3.381991	2.869259
4	-5.622963	4.108341	0.1968923	4.148614	3.864902	3.390365	2.874186
5	-5.704072	4.148284	0.2000799	4.158863	3.869919	3.394903	2.876685
6	-5.751540	4.171581	0.2018782	4.164772	3.872793	3.397450	2.878038
7	-5.779242	4.185155	0.2029105	4.168199	3.874456	3.398907	2.878799
8	-5.795405	4.193069	0.2035086	4.170195	3.875424	3.399750	2.879235
9	-5.804842	4.197688	0.2038570	4.171359	3.875988	3.400240	2.879488
10	-5.810356	4.200386	0.2040605	4.172040	3.876318	3.400526	2.879636
11	-5.813580	4.201964	0.2041795	4.172438	3.876511	3.400694	2.879722
12	-5.815466	4.202887	0.2042491	4.172671	3.876624	3.400791	2.879772
13	-5.816571	4.203427	0.2042899	4.172807	3.876690	3.400849	2.879802
14	-5.817217	4.203743	0.2043138	4.172887	3.876729	3.400882	2.879819
15	-5.817596	4.203929	0.2043278	4.172934	3.876752	3.400902	2.879829
16	-5.817817	4.204037	0.2043359	4.172961	3.876765	3.400913	2.879835
17	-5.817947	4.204101	0.2043407	4.172977	3.876773	3.400920	2.879838
18	-5.818023	4.204138	0.2043436	4.172987	3.876777	3.400924	2.879840
19	-5.818068	4.204160	0.2043452	4.172992	3.876780	3.400927	2.879842
20	-5.818094	4.204173	0.2043462	4.172995	3.876782	3.400928	2.879842
21	-5.818109	4.204180	0.2043467	4.172997	3.876783	3.400929	2.879843
22	-5.818118	4.204185	0.2043471	4.172998	3.876783	3.400929	2.879843
NA	NA	NA	NA	NA	NA	NA	NA

1.2 Removing 150c

After removing the 150c censored data and 16 iterations, the results of $\hat{\beta}_0$,$\hat{\beta}_1$,$\hat{\sigma}$ are -4.6042301, 3.6366424, 0.2410175. From the figure and table we can know the estimats of $\hat{\beta}_1$ and log time to failure are smaller.

However, the censored data are underestimated values. After iterations, we anticipate the estimate should be larger than the before.

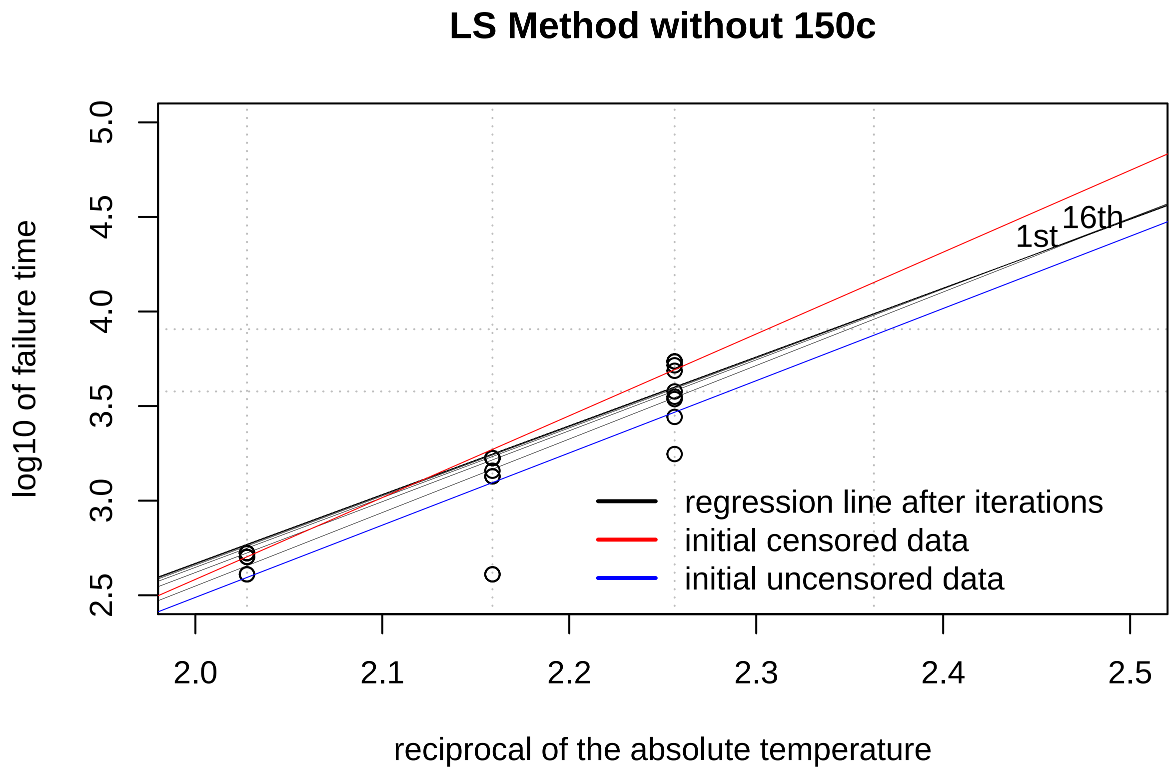
The reason might be below:

For the full data, the initial total fitted $\hat{\beta}_1$ (3.747043) is lager than the $\hat{\beta}_1$ (3.532411) fitted merely by censored data. The positive difference will be accumulated during the iterations and make the $\hat{\beta}_1$ and $\hat{\mu}$ larger and larger until convergency.

Without the 150c data, the initial total fitted $\hat{\beta}_1$ (3.886073) is smaller than the $\hat{\beta}_1$ (4.324944) fitted merely by censored data. Thus, the negative difference will be accumulated during the iterations and make the $\hat{\beta}_1$ and $\hat{\mu}$ smaller and smaller.

	fit0.coef	fit_c.coef	fit0_no150.coef	fit_c_no150.coef
(Intercept)	-4.930507	-4.404674	-5.223416	-6.066011
x	3.747043	3.532411	3.886073	4.324944

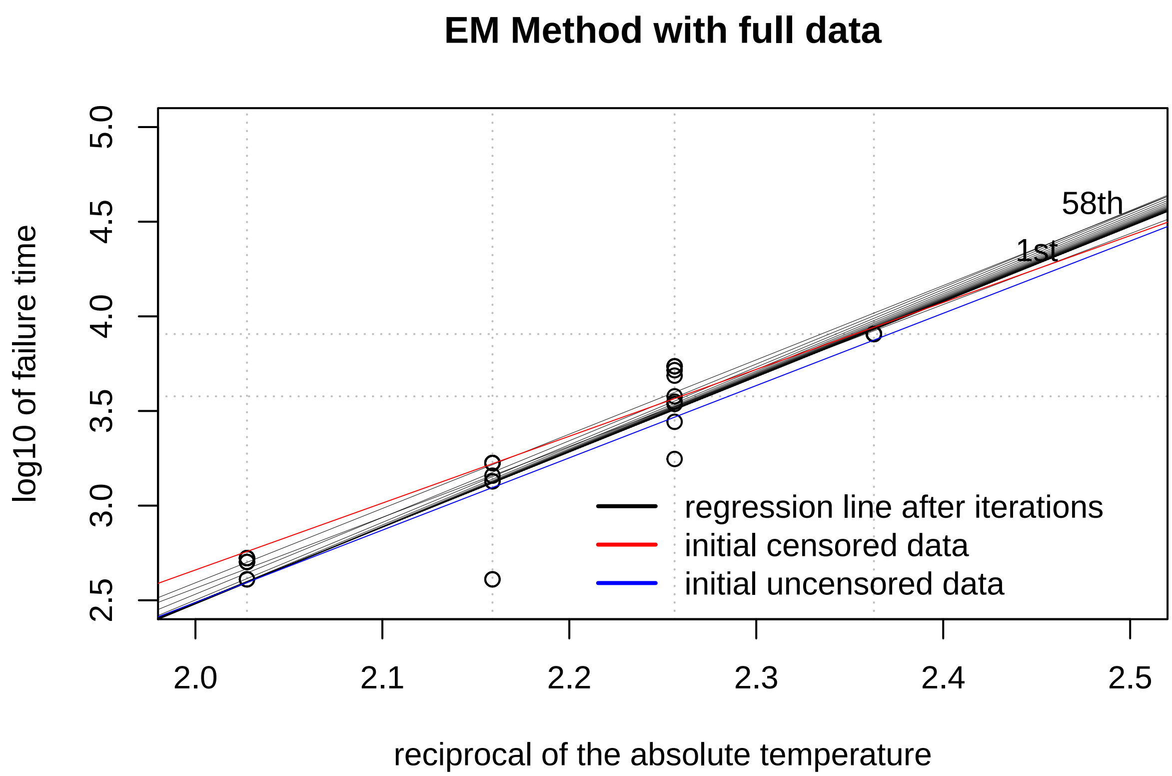
Note:
The initial regression coefficients



Iteration	Intercept	Slope	Sigma	mu170	mu190	mu220
1	-5.223416	3.886073	0.1822782	3.830161	3.351240	2.846310
2	-4.875219	3.747684	0.2150043	3.862018	3.393390	2.894494
3	-4.725422	3.686426	0.2292010	3.875601	3.411077	2.915461
4	-4.659330	3.659284	0.2355862	3.881710	3.418975	2.924904
5	-4.629525	3.647036	0.2385086	3.884509	3.422581	2.929223
6	-4.615901	3.641438	0.2398561	3.885801	3.424242	2.931214
7	-4.609628	3.638860	0.2404795	3.886399	3.425010	2.932134
8	-4.606728	3.637669	0.2407683	3.886676	3.425365	2.932560
9	-4.605385	3.637117	0.2409022	3.886805	3.425530	2.932758
10	-4.604763	3.636861	0.2409643	3.886864	3.425607	2.932850
11	-4.604475	3.636743	0.2409931	3.886892	3.425642	2.932892
12	-4.604341	3.636688	0.2410065	3.886905	3.425659	2.932912
13	-4.604279	3.636662	0.2410127	3.886911	3.425666	2.932921
14	-4.604250	3.636650	0.2410156	3.886913	3.425670	2.932925
15	-4.604236	3.636645	0.2410169	3.886915	3.425672	2.932927
16	-4.604230	3.636642	0.2410175	3.886915	3.425672	2.932928
NA	NA	NA	NA	NA	NA	NA

1.3 EM Method with full data

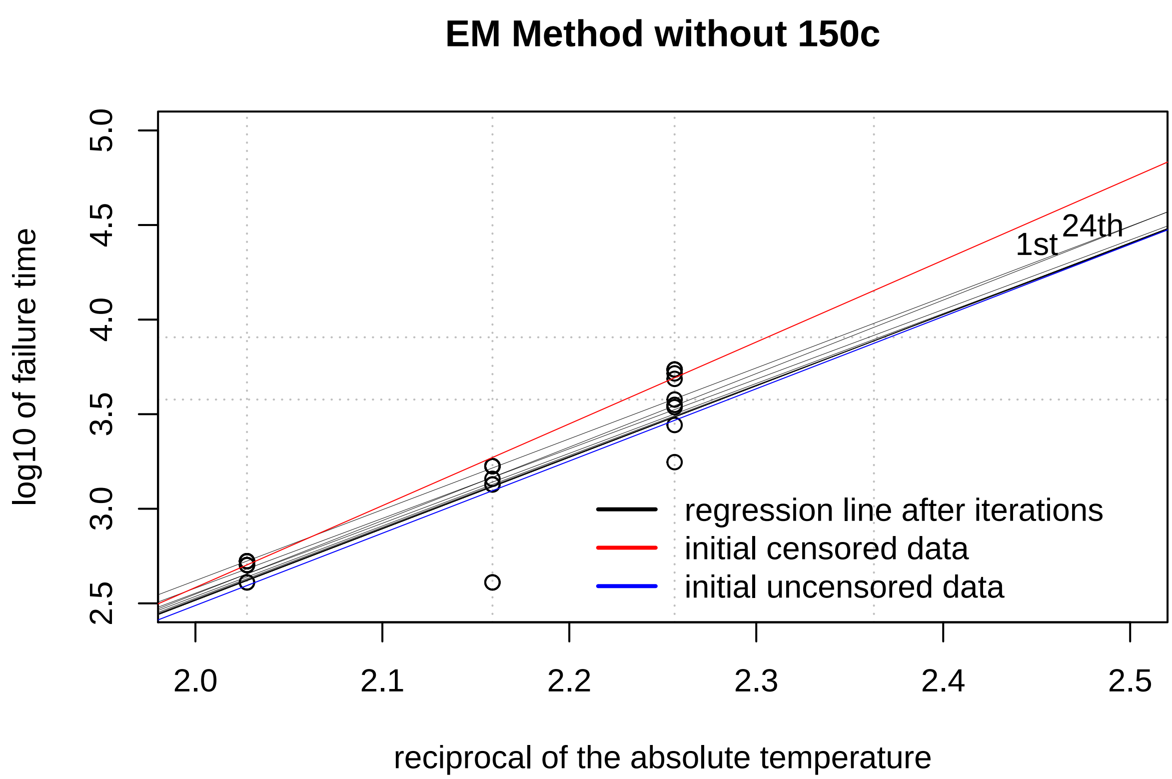
Using the EM method, the results of $\hat{\beta}_0$,$\hat{\beta}_1$,$\hat{\sigma}$ are -5.4538875, 3.9713585, 0.017898 after 58th iterations. Although the estimates are smaller than SL method, Thay shows a same trend: $\hat{\beta}_1$ and $\hat{\mu}$ grow larger and converges.



Iteration	Intercept	Slope	Sigma	mu150	mu170	mu190	mu220
1	-4.930507	3.747043	0.1572178	4.038379	3.808887	3.329470	2.829844
2	-5.260094	3.926205	0.0286122	4.038224	3.650517	3.247200	2.730235
3	-5.574756	4.053400	0.0226390	4.019754	3.612023	3.200660	2.667279
4	-5.698418	4.099947	0.0201760	4.004282	3.588923	3.174817	2.635442
5	-5.726997	4.107110	0.0191617	3.991887	3.574668	3.160601	2.620335
6	-5.712469	4.096952	0.0186849	3.982066	3.565414	3.152684	2.613774
7	-5.681420	4.080596	0.0184281	3.974277	3.559091	3.148141	2.611392
8	-5.646594	4.063283	0.0182740	3.968081	3.554575	3.145424	2.610956
9	-5.613612	4.047264	0.0181739	3.963139	3.551232	3.143715	2.611355
10	-5.584595	4.033334	0.0181054	3.959189	3.548694	3.142583	2.612056
48	-5.453914	3.971371	0.0178981	3.943304	3.539191	3.139269	2.616888
49	-5.453908	3.971368	0.0178980	3.943303	3.539191	3.139268	2.616888
50	-5.453903	3.971366	0.0178980	3.943303	3.539191	3.139268	2.616889
51	-5.453899	3.971364	0.0178980	3.943302	3.539191	3.139268	2.616889
52	-5.453896	3.971363	0.0178980	3.943302	3.539190	3.139268	2.616889
53	-5.453894	3.971362	0.0178980	3.943302	3.539190	3.139268	2.616889
54	-5.453892	3.971361	0.0178980	3.943301	3.539190	3.139268	2.616889
55	-5.453890	3.971360	0.0178980	3.943301	3.539190	3.139268	2.616889
56	-5.453889	3.971359	0.0178980	3.943301	3.539190	3.139268	2.616889
57	-5.453888	3.971359	0.0178980	3.943301	3.539190	3.139268	2.616889
58	-5.453888	3.971359	0.0178980	3.943301	3.539190	3.139268	2.616889

Note:
The first and last 10 rows

1.4 EM Method removing 150c data



Iteration	Intercept	Slope	Sigma	mu170	mu190	mu220
1	-5.223416	3.886073	0.1822782	3.830161	3.351240	2.846310
2	-4.875219	3.747684	0.0376498	3.639688	3.253855	2.762823
3	-4.781585	3.681179	0.0295009	3.570508	3.195638	2.713099
4	-4.838978	3.696305	0.0257825	3.541423	3.167127	2.682491
5	-4.910991	3.723894	0.0245124	3.529672	3.153387	2.665091
6	-4.964891	3.746028	0.0240298	3.524958	3.146782	2.655566
7	-4.999065	3.760469	0.0238283	3.523053	3.143581	2.650462
8	-5.019060	3.769047	0.0237380	3.522270	3.142012	2.647764
9	-5.030241	3.773885	0.0236955	3.521939	3.141234	2.646349
10	-5.036323	3.776532	0.0236748	3.521796	3.140844	2.645611
11	-5.039573	3.777951	0.0236645	3.521732	3.140648	2.645228
12	-5.041290	3.778702	0.0236593	3.521703	3.140548	2.645029
13	-5.042190	3.779097	0.0236567	3.521689	3.140497	2.644926
14	-5.042659	3.779303	0.0236554	3.521682	3.140471	2.644873
15	-5.042903	3.779410	0.0236547	3.521679	3.140457	2.644846
16	-5.043029	3.779465	0.0236543	3.521677	3.140451	2.644832
17	-5.043094	3.779494	0.0236541	3.521676	3.140447	2.644824
18	-5.043128	3.779509	0.0236540	3.521676	3.140445	2.644821
19	-5.043146	3.779517	0.0236540	3.521676	3.140444	2.644819
20	-5.043155	3.779521	0.0236540	3.521676	3.140444	2.644818
21	-5.043159	3.779523	0.0236540	3.521676	3.140443	2.644817
22	-5.043162	3.779524	0.0236540	3.521676	3.140443	2.644817
23	-5.043163	3.779524	0.0236540	3.521676	3.140443	2.644817
24	-5.043164	3.779524	0.0236540	3.521675	3.140443	2.644817
NA	NA	NA	NA	NA	NA	NA

1.5 Summary

Using Maximum Likelihood Mehtod, both Schmee & Hahn (1979), and Aitkin (1981) get a smaller $\hat{\beta}_0 = -6.019$, a larger $\hat{\beta}_1 = 4.311$, and larger estimate of the expected log time to failure times. I cannot reproduce these results. The estimates of removing 150c were not closer to each other too. In my attempts, Least Square Method and EM Method give similar results that ignoring the 150c data make $\hat{\beta}_1$ and $\hat{\mu}$ smaller.

	Iteration	Intercept	Slope	Sigma	mu150	mu170	mu190	mu220
LS_full	22	-5.818118	4.204185	0.2043471	4.172998	3.876783	3.400929	2.879843
EM_full	58	-5.453888	3.971359	0.0178980	3.943301	3.539190	3.139268	2.616889
LS_no150	16	-4.604230	3.636642	0.2410175	NA	3.886915	3.425672	2.932928
EM_no150	24	-5.043164	3.779524	0.0236540	NA	3.521675	3.140443	2.644817

- E-step

$$Q(\vec{\theta},\vec{\theta}^\star) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(t_j-\beta_0-\beta_1\nu_j)^2 - \frac{1}{2\sigma^2}\sum_{i=m+1}^{n} E[(T_i-\beta_0-\beta_1\nu_i)^2|T_i>w_i,\vec{\theta}^\star]$$

$$E[T_i|T_i>w_i,\vec{\theta}^\star] = \mu_i^\star + \sigma^\star H(\frac{w_i-\mu_i^\star}{\sigma^\star})$$

$$E[T_i^2|T_i>w_i,\vec{\theta}^\star] = \mu_i^{\star 2} + \sigma^{\star 2} + \sigma^\star(w_i+\mu_i^\star)H(\frac{w_i-\mu_i^\star}{\sigma^\star})$$

$$E[(T_i-\beta_0-\beta_1\nu_i)^2|T_i>w_i,\vec{\theta}^\star] = \mu_i^{\star 2}+\sigma^{\star 2}+\sigma^\star(w_i+\mu_i^\star)H(\frac{w_i-\mu_i^\star}{\sigma^\star})-2(\beta_0+\beta_1\nu_i)[\mu_i^\star+\sigma^\star H(\frac{w_i-\mu_i^\star}{\sigma^\star})]+(\beta_0+\beta_1\nu_i)^2$$

- M-step

$$\frac{\partial Q}{\partial\beta_0} = -\frac{1}{\sigma^2}\left\{\sum_{j=1}^{m}[t_j-\beta_0-\beta_1\nu_j] + \sum_{i=m+1}^{n}[\mu_i^\star+\sigma^\star H(\frac{w_i-\mu_i^\star}{\sigma^\star})-\beta_0-\beta_1\nu_i]\right\} = 0$$

$$\frac{\partial Q}{\partial\beta_1} = -\frac{1}{\sigma^2}\left\{\sum_{j=1}^{m}[t_j-\beta_0-\beta_1\nu_j]\nu_j + \sum_{i=m+1}^{n}[\mu_i^\star+\sigma^\star H(\frac{w_i-\mu_i^\star}{\sigma^\star})-\beta_0-\beta_1\nu_i]\nu_i\right\} = 0$$

$$\frac{\partial Q}{\partial\sigma^2} = \frac{1}{2\sigma^2}\left\{-n + \frac{1}{\sigma^2}\sum_{j=1}^{m}[t_j-\beta_0-\beta_1\nu_j]^2 + \frac{1}{\sigma^2}\sum_{i=m+1}^{n}E[(T_i-\beta_0-\beta_1\nu_i)^2|T_i>w_i,\vec{\theta}^\star]\right\} = 0$$

$$\sum_{j=1}^{m}[t_j-\beta_0-\beta_1\nu_j]^2 + \sum_{i=m+1}^{n}\left\{\mu_i^{\star 2}+\sigma^{\star 2}+\sigma^\star(w_i+\mu_i^\star-2\mu_i)H(\frac{w_i-\mu_i^\star}{\sigma^\star})-2\mu_i\mu_i^\star+\mu_i^2\right\} = n\sigma^2$$

1.6 the EM algorithm's pseudo code

Algorithm 1: EM algorithm

input : observed data $\mathcal{D}=\{\vec{x}_1,\vec{x}_2,\cdots,\vec{x}_N\}$,joint distribution $P(\vec{x},\vec{z}|\vec{\theta})$
output: model's parameters $\vec{\theta}$
// 1. identify hidden variables $\vec{z}$, write out the log likelihood function $\ell(\vec{x},\vec{z}|\vec{\theta})$
$\vec{\theta}^{(0)}=...$ // initialize
while *(!convergency)* **do**
  // 2. E-step: plug in $P(\vec{x},\vec{z}|\vec{\theta})$, derive the formula of $Q(\vec{\theta},\vec{\theta}^{t-1})$
  $Q(\vec{\theta},\vec{\theta}^{t-1})=\mathbb{E}\left[\ell_c(\vec{\theta})|\mathcal{D},\theta^{t-1}\right]$
  // 3. M-step: find $\vec{\theta}$ that maximizes the value of $Q(\vec{\theta},\vec{\theta}^{t-1})$
  $\vec{\theta}^t=\arg\max_{\vec{\theta}} Q(\vec{\theta},\vec{\theta}^{t-1})$

2 Reference

Schmee, J., & Hahn, G. (1979). A Simple Method for Regression Analysis with Censored Data. Technometrics, 21(4), 417-432. doi:10.2307/1268280

Aitkin, M. (1981). A Note on the Regression Analysis of Censored Data. Technometrics, 23(2), 161-163. doi:10.2307/1268032