Semi-parameter model estimation, with quantiles (n = 200)

2019-11-25

Changes

- use inverse function in R package instead of using my own function
- calculated the estimated $\hat{S}(t)$ at time quantiles (e.g. $t_{0.1}, t_{0.25}, t_{0.5}, t_{0.75}, t_{0.9}$), instead of using order statistics of time.

Results highlight

- Four methods: KM: Kaplan Meier; Exp m(): estimate λ_F with $\exp(m(t)\lambda_H)$; Dikta 1: use Dikta's first formula; Dikta 2: use Dikta's updated formula.
- The 4 methods have similar mean absolute difference between $\hat{S}(t)$ and true $S(t)$
- The standard deviation of KM is larger than the other three methods.
- The other three methods have similar standard deviation over 500 iterations.
- When using the logistic regression estimated $m(t)$, the standard deviation get larger, but still smaller than the Kaplan Meier's sd

Simulation

All the simulations below have generated data with size 200. And 500 times iterations.

Example 3: exponential + extreme distribution

The survival time is denoted as T and censor time as C, the observed time is marked as Z

And

$$S_T(t) = P(T > t) = P(T > t, C > 0) = e^{-\theta t} e^{-(e^{\theta 0}-1)\left((t-0)^2+1\right)} = e^{-\theta t}$$

$$f_T(t) = \frac{\partial}{\partial t}(1 - S_T(t)) = \frac{\partial}{\partial t}(1 - e^{-\theta t}) = \theta e^{-\theta t}$$

$$S_Z(t) = P(T > t, C > t) = e^{-\theta t} e^{-(e^{\theta t}-1)} = e^{-e^{\theta t} - \theta t + 1}$$

$$f_Z(t) = \frac{\partial}{\partial t}(1 - S_x(t)) = 1 - e^{-e^{\theta t} - \theta t + 1} = \theta(1 + e^{\theta t}) e^{-e^{\theta t} - \theta t + 1}$$

$$\psi(t) = \int_t^\infty f(t, c)dc = \int_t^\infty \theta^2 e^{-e^{\theta c} + \theta c - \theta t + 1} dc = \theta e^{-e^{\theta t} - \theta t + 1}$$

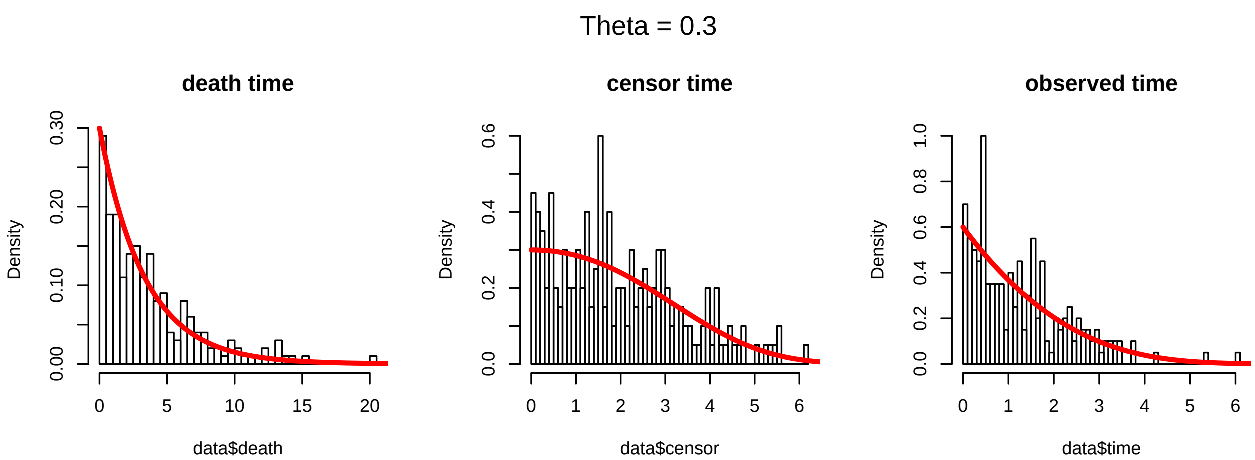
Therefore, the $m()$ function is:

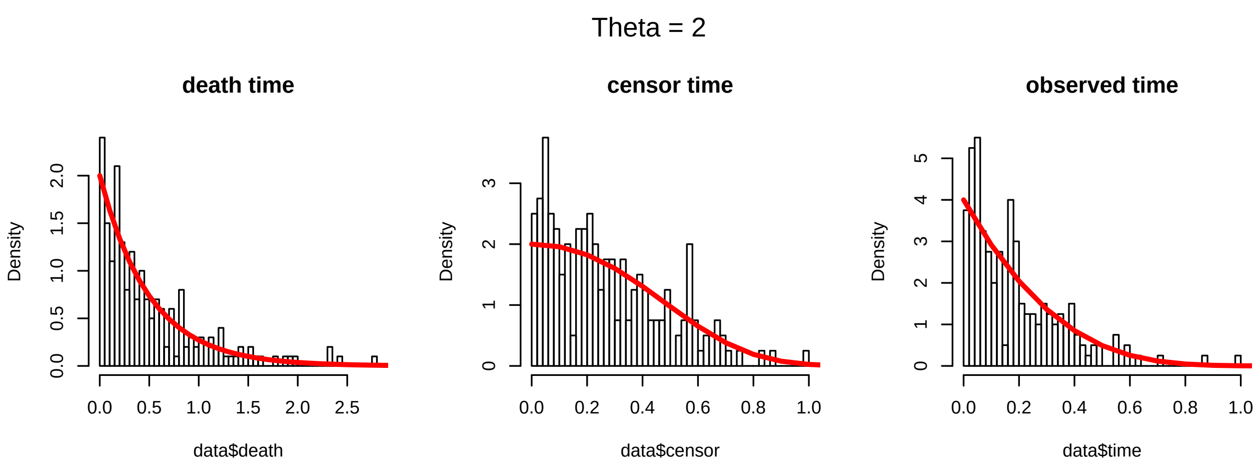
$$m(t) = \frac{\lambda_F(t)}{\lambda_H(t)} = \frac{f_T(t)}{S_T(t)} / \frac{f_Z(t)}{S_Z(t)} = \frac{\theta e^{-\theta t}}{e^{-\theta t}} / \frac{\theta(1+e^{\theta t})e^{-e^{\theta t}-\theta t+1}}{e^{-e^{\theta t}-\theta t+1}} = \frac{1}{1+e^{\theta t}}$$

And for the $\rho()$ function,

$$\begin{aligned}\rho =& \frac{f(t)/\psi(t)-1}{S(t)/S_Z(t)-1} \\ =& \frac{\theta e^{-\theta t}/(\theta e^{-e^{\theta t}-\theta t+1})-1}{e^{-\theta t}/e^{-e^{\theta t}-\theta t+1}-1} \\ =& 1\end{aligned}$$

After changing the inverse function, the data is generated more appropriately. For example:





Results

Table 1: Mean absolute difference between estimated and true S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
theta = 0.3								
t0.1	0.01662	0.01322	0.01333	0.01339	0.01662	0.01335	0.01347	0.01354
t0.25	0.02309	0.02454	0.02509	0.02539	0.02309	0.02634	0.02695	0.02729
t0.5	0.04416	0.06393	0.06671	0.06844	0.04416	0.07595	0.07880	0.08052
t0.75	0.11031	0.04176	0.08877	0.09731	0.11031	0.06177	0.10511	0.11322
t0.9	0.06725	0.07608	0.10000	0.10000	0.06725	0.05345	0.10000	0.10000
theta = 0.8								
t0.1	0.01723	0.01271	0.01281	0.01286	0.01723	0.01274	0.01284	0.01289
t0.25	0.02192	0.02369	0.02418	0.02445	0.02192	0.02462	0.02514	0.02542
t0.5	0.03803	0.05151	0.05410	0.05572	0.03803	0.05896	0.06160	0.06320
t0.75	0.09446	0.04072	0.08927	0.09814	0.09446	0.05589	0.10209	0.11020
t0.9	0.07936	0.07741	0.09997	0.09993	0.07936	0.06249	0.10000	0.09997
theta = 1								
t0.1	0.01599	0.01235	0.01244	0.01250	0.01599	0.01246	0.01256	0.01261
t0.25	0.02173	0.02319	0.02369	0.02397	0.02173	0.02463	0.02517	0.02547
t0.5	0.04158	0.05456	0.05715	0.05876	0.04158	0.06307	0.06572	0.06732
t0.75	0.10060	0.04363	0.10104	0.10926	0.10060	0.05994	0.11447	0.12218
t0.9	0.07493	0.07735	0.10000	0.10000	0.07493	0.06004	0.10000	0.10000
theta = 1.5								
t0.1	0.01536	0.01213	0.01220	0.01225	0.01536	0.01216	0.01223	0.01228
t0.25	0.02185	0.02198	0.02242	0.02266	0.02185	0.02288	0.02334	0.02361
t0.5	0.03697	0.04839	0.05086	0.05240	0.03697	0.05504	0.05762	0.05920
t0.75	0.06551	0.02553	0.03544	0.04023	0.06551	0.03332	0.04487	0.05018
t0.9	0.08145	0.08591	0.10004	0.10001	0.08145	0.07070	0.09999	0.09996
theta = 2								
t0.1	0.01620	0.01209	0.01215	0.01218	0.01620	0.01211	0.01217	0.01220
t0.25	0.02191	0.02112	0.02155	0.02178	0.02191	0.02162	0.02210	0.02236
t0.5	0.03738	0.04827	0.05079	0.05235	0.03738	0.05553	0.05811	0.05970
t0.75	0.05766	0.02827	0.02984	0.03140	0.05766	0.03218	0.03658	0.03936
t0.9	0.08913	0.10504	0.10003	0.10000	0.08913	0.08643	0.09997	0.09995
theta = 5								
t0.1	0.01691	0.01822	0.01846	0.01858	0.01691	0.01894	0.01918	0.01931
t0.25	0.02528	0.03892	0.03972	0.04016	0.02528	0.04447	0.04532	0.04577
t0.5	0.07636	0.09930	0.10262	0.10470	0.07636	0.12506	0.12846	0.13047
t0.75	0.17926	0.05657	0.10639	0.11643	0.17926	0.09625	0.13797	0.14676
t0.9	0.05939	0.08286	0.09032	0.08129	0.05939	0.04326	0.06463	0.05803

Table 2: MSE

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
theta = 1								

t0.1	0.00043	0.00028	0.00028	0.00029	0.00043	0.00028	0.00029	0.00029
t0.25	0.00082	0.00091	0.00095	0.00097	0.00082	0.00103	0.00107	0.00109
t0.5	0.00290	0.00496	0.00535	0.00560	0.00290	0.00675	0.00722	0.00750
t0.75	0.01683	0.00256	0.01605	0.01704	0.01683	0.00536	0.01834	0.01961
t0.9	0.00629	0.00592	0.01000	0.01000	0.00629	0.00396	0.01000	0.01000
theta = 2								
t0.1	0.00045	0.00026	0.00027	0.00027	0.00045	0.00026	0.00027	0.00027
t0.25	0.00075	0.00084	0.00087	0.00089	0.00075	0.00089	0.00093	0.00095
t0.5	0.00224	0.00346	0.00376	0.00395	0.00224	0.00435	0.00470	0.00491
t0.75	0.01327	0.00249	0.01644	0.01746	0.01327	0.00446	0.01809	0.01929
t0.9	0.00857	0.00612	0.00999	0.00999	0.00857	0.00513	0.01000	0.00999
t0.1	0.00041	0.00024	0.00025	0.00025	0.00041	0.00025	0.00025	0.00025
t0.25	0.00073	0.00084	0.00087	0.00089	0.00073	0.00093	0.00097	0.00099
t0.5	0.00259	0.00383	0.00415	0.00436	0.00259	0.00495	0.00532	0.00555
t0.75	0.01487	0.00279	0.01977	0.02069	0.01487	0.00492	0.02155	0.02269
t0.9	0.00801	0.00612	0.01000	0.01000	0.00801	0.00474	0.01000	0.01000
t0.1	0.00038	0.00024	0.00024	0.00024	0.00038	0.00024	0.00024	0.00024
t0.25	0.00075	0.00076	0.00078	0.00080	0.00075	0.00081	0.00083	0.00085
t0.5	0.00214	0.00312	0.00340	0.00358	0.00214	0.00390	0.00421	0.00441
t0.75	0.00674	0.00107	0.00357	0.00404	0.00674	0.00182	0.00447	0.00511
t0.9	0.00906	0.00751	0.01001	0.01000	0.00906	0.00614	0.01000	0.00999
t0.1	0.00039	0.00022	0.00022	0.00023	0.00039	0.00022	0.00023	0.00023
t0.25	0.00073	0.00068	0.00071	0.00072	0.00073	0.00071	0.00074	0.00076
t0.5	0.00211	0.00304	0.00332	0.00350	0.00211	0.00383	0.00415	0.00435
t0.75	0.00519	0.00120	0.00194	0.00210	0.00519	0.00166	0.00265	0.00299
t0.9	0.01072	0.01130	0.01001	0.01000	0.01072	0.00895	0.01000	0.00999
t0.1	0.00045	0.00049	0.00050	0.00051	0.00045	0.00053	0.00054	0.00054
t0.25	0.00098	0.00206	0.00213	0.00217	0.00098	0.00256	0.00265	0.00269
t0.5	0.00764	0.01085	0.01155	0.01200	0.00764	0.01666	0.01755	0.01809
t0.75	0.03483	0.00409	0.01859	0.02004	0.03483	0.01043	0.02385	0.02576
t0.9	0.00466	0.00750	0.00833	0.00699	0.00466	0.00272	0.00543	0.00492

Table 3: Standard deviations of the estimated S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
theta = 0.3								
t0.1	0.02066	0.01554	0.01559	0.01561	0.02066	0.01555	0.01559	0.01562
t0.25	0.02851	0.02267	0.02276	0.02281	0.02851	0.02261	0.02270	0.02274
t0.5	0.03751	0.03014	0.03050	0.03075	0.03751	0.03176	0.03208	0.03227
t0.75	0.07560	0.03568	0.09387	0.08993	0.07560	0.04535	0.08839	0.08466
t0.9	0.07913	0.01152	0.00000	0.00000	0.07913	0.03505	0.00000	0.00000
theta = 0.8								
t0.1	0.02135	0.01561	0.01566	0.01568	0.02135	0.01557	0.01562	0.01564
t0.25	0.02708	0.02333	0.02341	0.02346	0.02708	0.02324	0.02332	0.02337
t0.5	0.03561	0.02931	0.02966	0.02989	0.03561	0.03056	0.03087	0.03106

t0.75	0.07656	0.03473	0.09437	0.09002	0.07656	0.04445	0.09145	0.08724
t0.9	0.09197	0.01141	0.00820	0.00745	0.09197	0.03566	0.00898	0.00821
theta = 1								
t0.1	0.02015	0.01510	0.01514	0.01517	0.02015	0.01514	0.01518	0.01520
t0.25	0.02689	0.02321	0.02329	0.02334	0.02689	0.02355	0.02364	0.02368
t0.5	0.03773	0.03056	0.03092	0.03116	0.03773	0.03239	0.03272	0.03290
t0.75	0.08111	0.03559	0.10049	0.09558	0.08111	0.04320	0.09503	0.09030
t0.9	0.08932	0.01190	0.00000	0.00000	0.08932	0.03463	0.00000	0.00000
theta = 1.5								
t0.1	0.01946	0.01510	0.01513	0.01515	0.01946	0.01509	0.01512	0.01514
t0.25	0.02717	0.02346	0.02353	0.02358	0.02717	0.02346	0.02353	0.02358
t0.5	0.03716	0.02964	0.03000	0.03022	0.03716	0.03077	0.03110	0.03127
t0.75	0.06337	0.03276	0.05679	0.05730	0.06337	0.04042	0.05920	0.05939
t0.9	0.08697	0.01153	0.00987	0.00910	0.08697	0.03418	0.00882	0.00807
theta = 2								
t0.1	0.01958	0.01483	0.01488	0.01490	0.01958	0.01483	0.01487	0.01489
t0.25	0.02686	0.02226	0.02234	0.02239	0.02686	0.02189	0.02197	0.02201
t0.5	0.03590	0.02813	0.02845	0.02866	0.03590	0.02808	0.02836	0.02853
t0.75	0.05858	0.03163	0.04400	0.04552	0.05858	0.04080	0.04999	0.05102
t0.9	0.06311	0.01623	0.00965	0.00900	0.06311	0.03893	0.00837	0.00772
theta = 5								
t0.1	0.02131	0.01561	0.01565	0.01567	0.02131	0.01568	0.01572	0.01574
t0.25	0.02993	0.02550	0.02560	0.02566	0.02993	0.02572	0.02583	0.02588
t0.5	0.04504	0.03150	0.03201	0.03233	0.04504	0.03197	0.03241	0.03266
t0.75	0.05193	0.03115	0.08575	0.08085	0.05193	0.03412	0.06946	0.06503
t0.9	0.04473	0.02521	0.09116	0.08364	0.04473	0.03265	0.07112	0.06462

The estimate of $m()$

Table 4: mean absolute difference between hat m() and true m()

0.3	0.8	1	1.5	2	5
0.0182307	0.016181	0.016701	0.0145415	0.0156571	0.0268094

The row name shows the θ value

Table 5: standard deviation of estimated m()

0.3	0.8	1	1.5	2	5
0.013985	0.0122275	0.0123005	0.0113775	0.0120009	0.0155827

The row name shows the θ value

Table 6: estimated theta from logitic regression

0.3	0.8	1	1.5	2	5
0.2536484	0.721832	0.8876818	1.363414	1.80001	3.395434

The row name shows the true θ value

Example 4: exponential + weibull distribution

$$P(T \geq x, C \geq y) = S(x,y) = \begin{cases} e^{-\theta x} e^{-(\theta y)^k \left((\theta x - \theta y)^2 + 1\right)} & x \geq y \\ e^{-\theta x} e^{-(\theta y)^k} & x < y \end{cases}$$

Then

- $S_T(x) = P(T \geq x, C \geq 0) = S(x,0) = e^{-\theta x}$, $f_T(x) = \frac{1 - S_T(x)}{x} = \theta e^{-\theta x}$
- $S_C(x) = P(T \geq 0, C \geq x) = S(0,x) = e^{-\theta 0} e^{-(\theta x)^k} = e^{-(\theta x)^k}$, $f_C(x) = \frac{1 - S_C(x)}{x} = k\theta(\theta y)^{k-1} e^{-(\theta x)^k}$

The death time is from an exponential distribution with paramter θ, the censor time is from a Weibull distribution with shape parameter k and scale parameter $1/\theta$.

Beisdes,

- $S_Z(x) = P(T > x, C > x) = e^{-\theta x - (\theta x)^k}$, $f_Z(x) = (\theta + k\theta(\theta x)^{k-1}) e^{-\theta x - (\theta x)^k}$

Therefore the $m()$ function is

$$m(x) = \frac{f_T(x)/S_T(x)}{f_Z(X)/S_Z(x)} = \frac{\theta e^{-\theta x}/e^{-\theta x}}{(\theta + k\theta(\theta x)^{k-1}) e^{-\theta x - (\theta x)^k}/e^{-\theta x - (\theta x)^k}} = \frac{1}{1 + k(\theta x)^{k-1}}$$

We could also transform $m()$ function as:

$$m(x) = \frac{1}{1 + \exp(\log(k(\theta x)^{k-1}))} = \frac{1}{1 + \exp(\log(k) + (k-1)\log(\theta) + (k-1)\log(x))}$$

We can then estimate the k and θ by fitting logistic regression.

Results

Table 7: Mean absolute difference between estimated and true S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
theta = 1								
t0.1	0.01954	0.01711	0.01719	0.01720	0.01954	0.01773	0.01783	0.01784

t0.25	0.03197	0.02756	0.02806	0.02818	0.03197	0.02843	0.02893	0.02903
t0.5	0.05746	0.06405	0.06600	0.06680	0.05746	0.06425	0.06617	0.06697
t0.75	0.17775	0.11595	0.13769	0.14445	0.17775	0.11912	0.14461	0.15132
t0.9	0.09933	0.01935	0.09779	0.09784	0.09933	0.01744	0.09863	0.09863
theta = 2								
t0.1	0.01690	0.01571	0.01572	0.01572	0.01690	0.01651	0.01652	0.01652
t0.25	0.02054	0.01935	0.01943	0.01944	0.02054	0.02033	0.02043	0.02045
t0.5	0.02156	0.01744	0.01764	0.01777	0.02156	0.01793	0.01812	0.01826
t0.75	0.03437	0.02665	0.02556	0.02532	0.03437	0.02633	0.02552	0.02546
t0.9	0.08075	0.05828	0.09932	0.09901	0.08075	0.06000	0.09952	0.09929

Table 8: MSE

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
theta = 1								
t0.1	0.00063	0.00046	0.00047	0.00047	0.00063	0.00052	0.00053	0.00053
t0.25	0.00157	0.00116	0.00120	0.00121	0.00157	0.00128	0.00131	0.00132
t0.5	0.00424	0.00486	0.00512	0.00523	0.00424	0.00494	0.00521	0.00532
t0.75	0.03402	0.01443	0.02161	0.02337	0.03402	0.01503	0.02368	0.02546
t0.9	0.00989	0.00050	0.00975	0.00975	0.00989	0.00046	0.00984	0.00985
theta = 2								
t0.1	0.00045	0.00038	0.00038	0.00038	0.00045	0.00042	0.00042	0.00042
t0.25	0.00071	0.00060	0.00060	0.00060	0.00071	0.00068	0.00069	0.00069
t0.5	0.00071	0.00047	0.00048	0.00049	0.00071	0.00050	0.00051	0.00052
t0.75	0.00185	0.00106	0.00099	0.00099	0.00185	0.00106	0.00101	0.00101
t0.9	0.00872	0.00418	0.00989	0.00985	0.00872	0.00442	0.00993	0.00990

Table 9: Standard deviations of the estimated S()

	With true m()				With estimated m()			
Quantile	KM	Exp m()	Dikta 1	Dikta 2	KM	Exp m()	Dikta 1	Dikta 2
theta = 1								
t0.1	0.02335	0.02105	0.02111	0.02111	0.02335	0.02233	0.02239	0.02240
t0.25	0.02950	0.02634	0.02642	0.02643	0.02950	0.02830	0.02839	0.02840
t0.5	0.03209	0.02758	0.02776	0.02786	0.03209	0.02875	0.02894	0.02904
t0.75	0.04930	0.03141	0.05153	0.05016	0.04930	0.02904	0.05269	0.05070
t0.9	0.00541	0.01703	0.01588	0.01418	0.00541	0.01937	0.01270	0.01138
theta = 2								
t0.1	0.02125	0.01924	0.01929	0.01929	0.02125	0.02043	0.02048	0.02048
t0.25	0.02657	0.02445	0.02453	0.02454	0.02657	0.02609	0.02616	0.02618
t0.5	0.02667	0.02154	0.02162	0.02168	0.02667	0.02220	0.02226	0.02233
t0.75	0.04282	0.02884	0.02995	0.03086	0.04282	0.02926	0.03044	0.03136

t0.9	0.08340	0.02828	0.02515	0.02308	0.08340	0.02889	0.02157	0.01983

The estimate of $m()$

Table 10: mean absolute difference between hat m() and true m()

1	2
0.0416769	0.0270299

The colname shows the true θ value.

Table 11: standard deviation of estimated m()

1	2
0.0177885	0.0161354

The colname shows the true θ value.

Table 12: estimated k from logitic regression (true k = 2)

1	2
1.899808	2.164795

The colname shows the true θ value.

Table 13: estimated theta from logitic regression

1	2
1.438359e+22	1.674205

The colname shows the true θ value.

Table 14: estimated theta from logitic regression with true k

1	2
0.808364	2.622251

The colname shows the true θ value.