Model exemplification

Alessio Crippa

2019-04-26

Aim

The aim of this document is to clarify the modelling strategy that will be used to analyze the accumulated data in the ProBio platform study.
Similar analyses have been also implemented in the simulations for defining the operational characteristics of ProBio. I first describe the assumptions behind the chosen parametric model and then use fictitious data to exemplify the presented strategy.

Modelling strategy: the Weibull parametric model

We are going to implement Bayesian methods for survival analysis. In a Bayesian framework, a parametric distribution is oftentime selected for modelling a time to event variable, in our case progression free survival (PFS).

The Weibull distribution is typically adopted in many bio-medical contexts, given its flexibility in describing different shapes and phenomena. A Weibull distribution can be parameterized in terms of a scale (λ) and shape (k) parameters, is such a way that it density function assumes the following form:

$$T \sim \text{Weibull}(\lambda, k)$$

$$f(t; \lambda, k) = \lambda k t^{k-1} \exp(-\lambda t^k)$$

Under the previous parametrization, the mean PFS is defined as $\lambda^{-\frac{1}{k}}\Gamma(1 + \frac{1}{k})$.

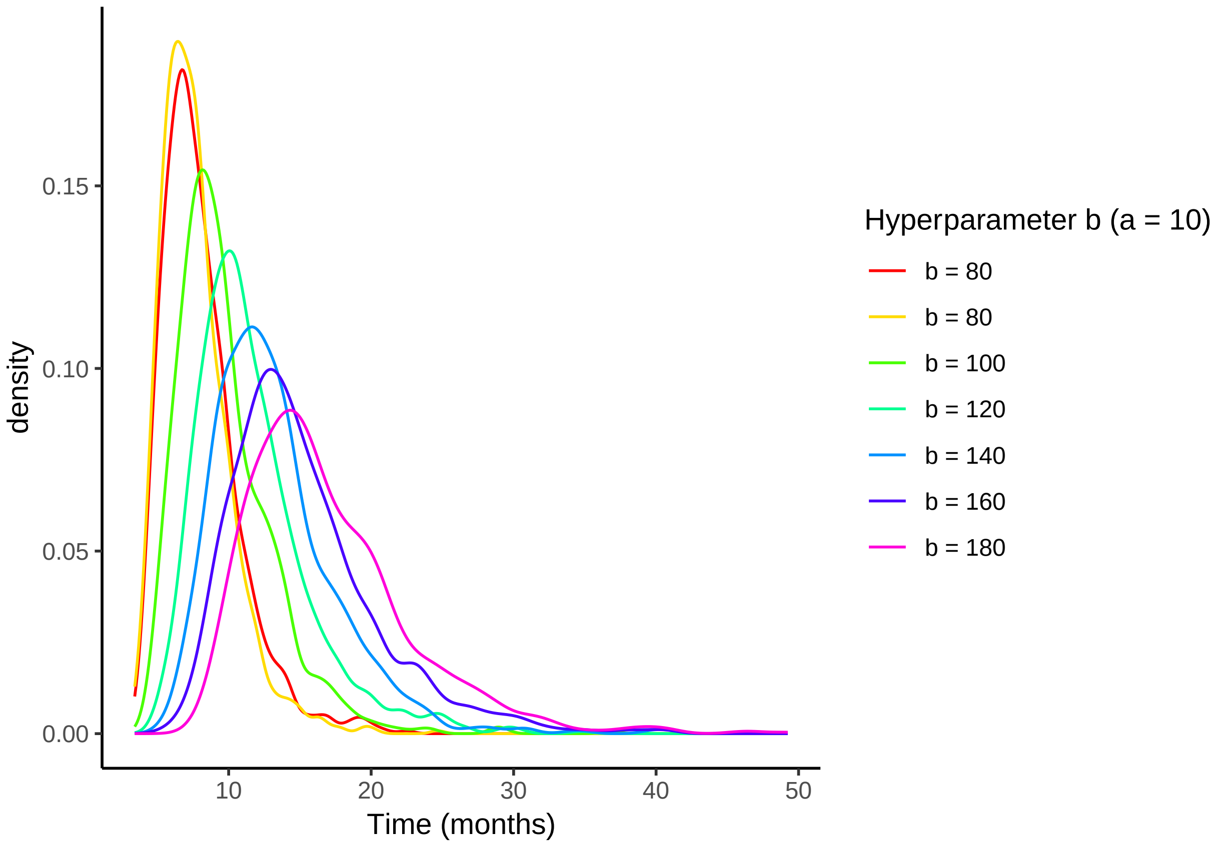
In a Bayesian perspective, our inference will be on the belief on the parameters of interest rather than the parameters themselves. Our belief (or historical data) in the parameters is represented by the definition of a distribution function.
In particular, we are going to assume a distribution for the scale parameter while fixing the shape parameter at 1.05, based on previous data from the BESENE study. Given that the scale parameter is strictly positive, a gamma distribution is typically used for this parameter, as it is also a conjugate model for the Weibull distribution:

$$\lambda \sim \text{Gamma}(a, b)$$

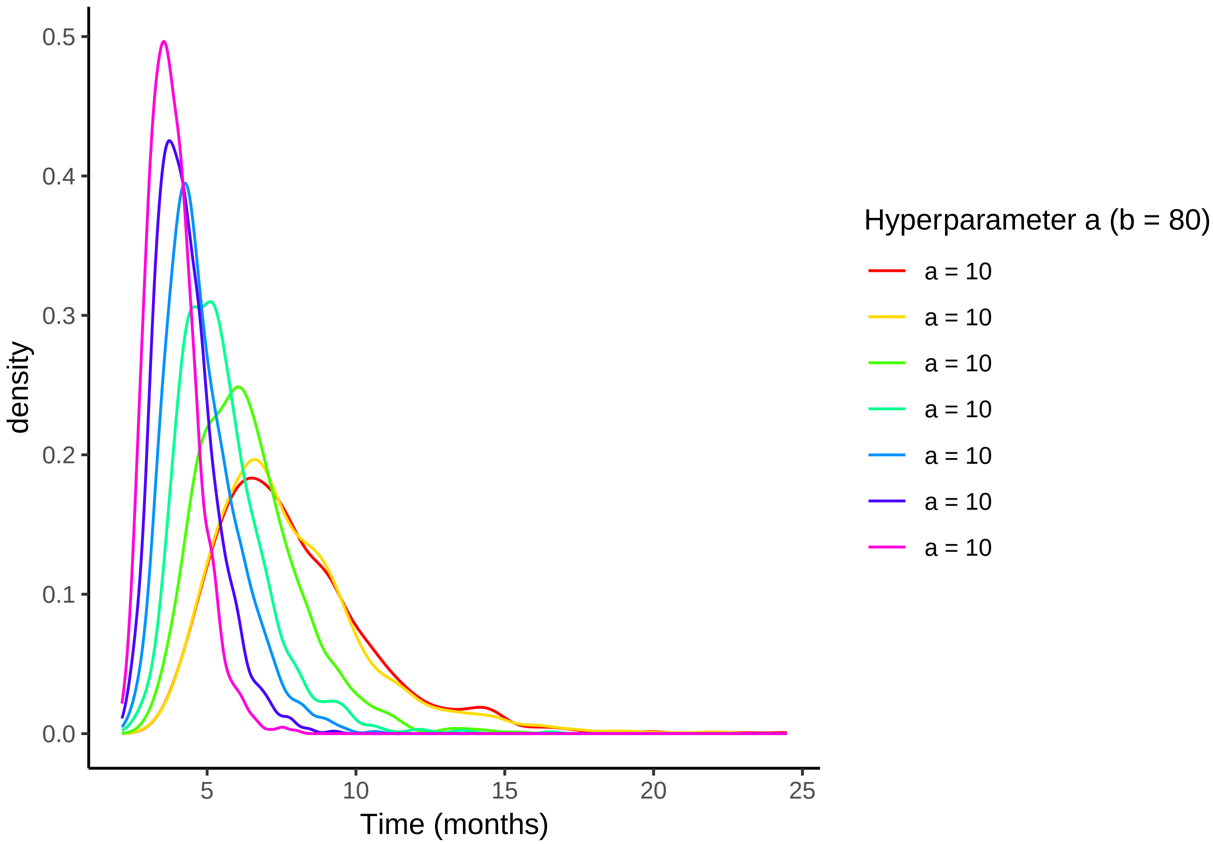
$$f(\lambda; a, b) = \frac{b^a \lambda^{a-1} \exp(-b\lambda)}{\Gamma(a)}$$

We are going to use $a = 10$ and $b = 80$ as apriori hyperparameters, that corresponds approximately to the information of 10 patient with a mean (rate) $E[\lambda] = \frac{a}{b} = 1/8 = 0.125$, which then gives a mean PFS time equal to $E[T] = 0.125^{-1/1.05}\Gamma(1 + \frac{1}{1.05}) = 7.1$.

Let's compare how the distribution of PFS times changes as the b hyperparameter increases from 80 to 180:



As the b hyperparameter increases, the disitribution shifts towards the right, with grater PFS times. Alternatively, we can let the other hyperparameter, a, to change while fixing $b = 80$:



The behaviour is opposite, as a increases the distribution of PFS times shift towards smaller values. We can compare the distributions by comparing the respective PFS means (mean time in the table below) in the two settings (OBS mean gamma is the mean of the scale paramter of the Weibull distribution).

a	b	mean gamma	mean time
10	80	0.1250000	7.106618
10	100	0.1000000	8.789380
10	120	0.0833333	10.456081
10	140	0.0714286	12.109545
10	160	0.0625000	13.751758
10	180	0.0555556	15.384200

a	b	mean gamma	mean time
10	80	0.125	7.106618
12	80	0.150	5.973822
14	80	0.175	5.158144
16	80	0.200	4.542166
18	80	0.225	4.060190
20	80	0.250	3.672550

In addition, we can quantify the extent by which two distribution with different values for the hyperparameters differ from each other. For example, what is the probability that then mean PFS modelled with a Weibull distribution where the λ parameter has a gamma distribution with $a = 10$ and $b = 140$ is greater than the mean PFS in a similar distribution but with $b = 80$? This can be computed using Monte Carlo simulations

a	b	mean gamma	mean time	prob of superiority
10	80	0.1250000	7.106618	0.459
10	100	0.1000000	8.789380	0.659
10	120	0.0833333	10.456081	0.794
10	140	0.0714286	12.109545	0.863
10	160	0.0625000	13.751758	0.927
10	180	0.0555556	15.384200	0.954

a	b	mean gamma	mean time	prob of superiority
10	80	0.125	7.106618	0.481
12	80	0.150	5.973822	0.322
14	80	0.175	5.158144	0.196
16	80	0.200	4.542166	0.119
18	80	0.225	4.060190	0.072
20	80	0.250	3.672550	0.028

Exemplification of a fictitious clinical trial

Let's use a fictitious example data set to illustrate how the hyperparameters changes throughout the trial based on the accumulated data, and how we can compute the quantities which let us to decide to ealier stop the trial or continue patients' enrollment.

For sake of clarity, we consider one active treatment being compared to a control group. Each group consists of 25 patients, whose PFS time has been recorded in the first 20 months. The PFS times for those men still alive at the end of the follow-up are marked with a "+" in the table below

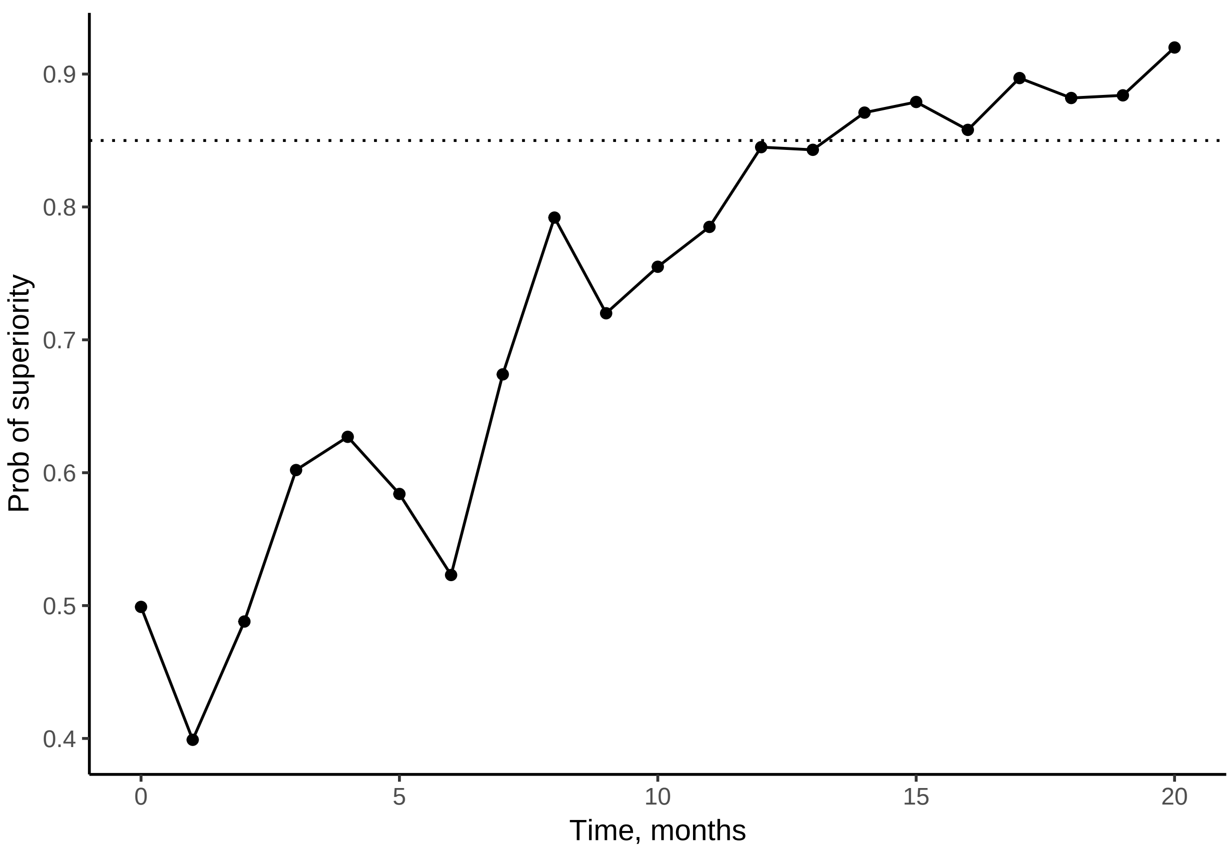
Control	Treatment
2.55, 6.43, 2.87, 6.68, 11.91, 6.95, 3.08, 7.43, 10.29, 6.34, 7.99, 19.93, 1.15, 20.00+, 7.43, 5.49, 8.69, 0.93, 2.63, 10.88, 16.88, 5.81, 1.42, 3.97, 20.00+	8.49, 20.00+, 3.18, 19.61, 20.00+, 15.35, 17.06, 20.00+, 10.51, 3.89, 20.00+, 8.12, 6.09, 20.00+, 2.66, 8.46, 0.48, 4.81, 5.26, 6.78, 5.62, 1.19, 20.00+, 0.31, 5.23

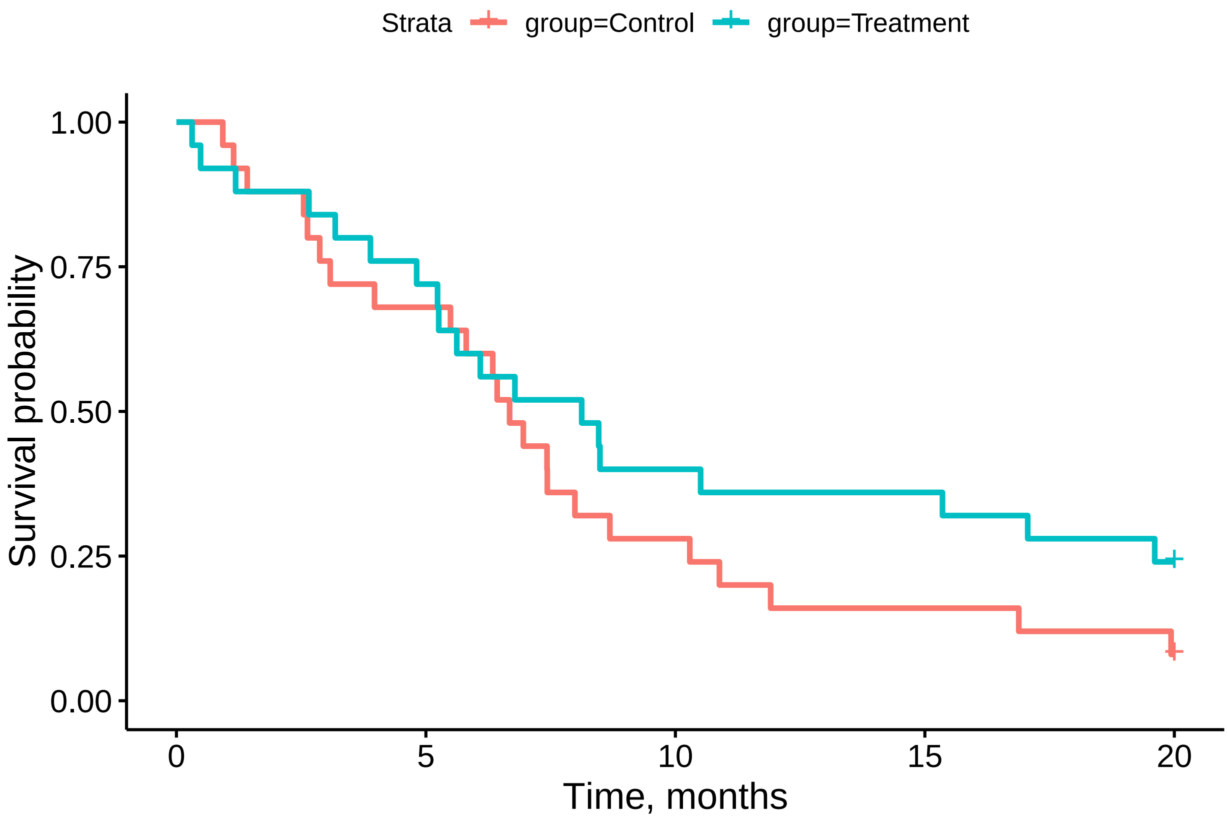
The hyperparameters of the Gamma distibution are updated monthly. In particular, the hyperparameter a is updated from month $t-1$ to the next month t with the number of progressions that have been observed during the month ($d_{(t)}$): $a_{(t)} = a_{(t-1)} + d_{(t)}$. Intuitevely, the distribution of PFS in the treatment arm shifts towards smaller times as the number of progressions increases (in the active arm). The other hyperparameter, instead, is updated with the amount of time the patients stayed in the trial during that month ($\sum_{i=1}^{n_{(t)}} PT^k_{i_{(t)}}$): $b_{(t)} = b_{(t-1)} + \sum_{i=1}^{n_{(t)}} PT^k_{i_{(t)}}$

For example, in the first month there have been 1 and 2 progressions in the control and treatment groups. After the first month $a = 10 + 1$ in the control group, while $a = 10 + 2$ in the treatment group. Similarly, the sum of the observed person times (elevated to the power of 1.05) in the first month were 24.93 and 23.76, so that $b = 80 + 24.93$ and $b = 80 + 23.76$ in the control and treatment group.

Given the hyperparameters it is possible to compare if the treatment is superior to control group using the quantities described in the previous sectin, i.e. the probabilities of superiority. This can be done monthly in the fictitious trial: ´

	Control						Treatment						
month	a	b	d	PT	mu gam	mu time	a	b	d	PT	mu gam	mu time	p
0	10	80.0	1	24.925	0.1250	7.107	10	80.0	2	23.763	0.1250	7.107	0.499
1	11	104.9	2	22.534	0.1048	8.403	12	103.8	1	22.171	0.1156	7.653	0.399
2	13	129.1	3	21.011	0.1007	8.730	13	127.5	1	21.642	0.1020	8.629	0.488
3	16	152.2	2	18.041	0.1051	8.381	14	151.3	2	20.051	0.0925	9.464	0.602
4	18	172.4	0	17.000	0.1044	8.434	16	173.7	1	18.806	0.0921	9.507	0.627
5	18	191.6	2	16.275	0.0939	9.329	17	195.0	3	16.061	0.0872	10.018	0.584
6	20	210.3	4	13.346	0.0951	9.218	20	213.4	2	13.854	0.0937	9.351	0.523
7	24	225.7	3	9.810	0.1063	8.290	22	229.4	0	13.000	0.0959	9.148	0.674
8	27	237.1	1	7.674	0.1139	7.767	22	244.5	3	11.027	0.0900	9.721	0.792
9	28	246.1	0	7.000	0.1138	7.773	25	257.5	0	10.000	0.0971	9.040	0.720
10	28	254.3	2	6.149	0.1101	8.020	25	269.2	1	9.489	0.0929	9.432	0.755
11	30	261.6	1	4.905	0.1147	7.715	26	280.5	0	9.000	0.0927	9.447	0.785
12	31	267.4	0	4.000	0.1159	7.636	26	291.1	0	9.000	0.0893	9.789	0.845
13	31	272.2	0	4.000	0.1139	7.766	26	301.9	0	9.000	0.0861	10.132	0.843
14	31	277.0	0	4.000	0.1119	7.896	26	312.6	0	9.000	0.0832	10.476	0.871
15	31	281.8	0	4.000	0.1100	8.026	26	323.4	1	8.335	0.0804	10.820	0.879
16	31	286.6	1	3.877	0.1082	8.157	27	333.5	0	8.000	0.0810	10.747	0.858
17	32	291.3	0	3.000	0.1099	8.037	27	343.1	1	7.053	0.0787	11.043	0.897
18	32	294.9	0	3.000	0.1085	8.133	28	351.7	0	7.000	0.0796	10.921	0.882
19	32	298.6	1	2.931	0.1072	8.228	28	360.2	1	6.592	0.0777	11.172	0.884
20	33	302.2	0	0.000	0.1092	8.082	29	368.2	0	0.000	0.0788	11.035	0.920





Strata
group=Control
group=Treatment
Survival probability
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
Time, months