

Tail probabilities of extinction time in a large number of experimental populations

Supplementary Materials: Appendix A

John M. Drake

1 Data summary

This supplement describes the statistical analysis of extinction times of 1,074 experimental populations of *Daphnia pulicaria*. Code is contained in the file `pulicaria.R`.

The raw data are contained in table `data.csv` with the following columns:

1. *Tube.ID* – A unique identifier for each population
2. *Block* – Block number
3. *Startno* – The initial population size, either $x_0 = 3$ or $x_0 = 18$
4. *Start* – Date population was initialized
5. *End* – Date population was scored extinct
6. *Days* – Duration of population persistence (extinction time) in days, *i.e.*, $End - Start$
7. *Tetracycline* – Indicator for tetracycline addition (0: No, 1: Yes)

A total of $n = 1080$ populations were initialized, of which 6 were accidentally destroyed (indicated by *Days* = -9999) yielding a total experiment size of $n = 1,074$. Three additional populations were right-censored with the censoring time treated as the extinction time. This introduces a very small bias to the regression models and estimated extinction time. For right-censored populations, the extinction time (if it were observed) would be greater than the recorded time, placing the population more to the extreme of the distribution. Thus, censoring bias decreases the mass in the tail and cannot lead to spurious conclusion of power laws when distributions are truly exponential. Extinction time ranged from 1 day to 1239 days with a mean of 123.88 days (sd: 122.38) and median 87 days. The tail of the distribution is observed in the histogram of extinction times, shown on both log and original scales (Fig. A1). Variation among blocks is evident in histograms for each block (Fig. A2) and their empirical cumulative distributions (Fig. A3). Finally, the the mean, median, excess kurtosis (fourth standardized moment minus 3), skewness, variance and coefficient of variation were calculated for each block (Table 1, Fig. A4). Excess kurtosis and skewness were estimated using the `moments` package in R (R Development Core Team, 2011). Mean excess kurtosis was found to be 88.3 (sd: 41.5), *i.e.*, extinction times were highly peaked compared with a Gaussian distribution where the excess kurtosis is 0. Estimated mean skewness was 8.5 (sd: 8.5), greater than 0 as expected.

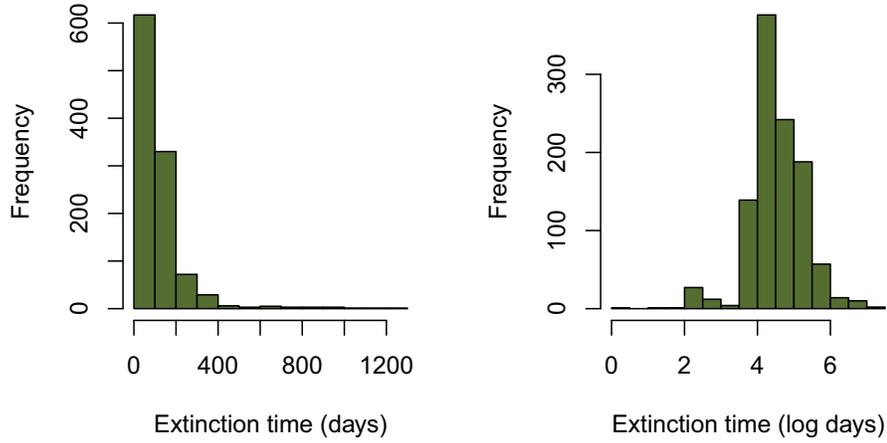


Figure A1: Histograms of extinction time (left) and log (extinction time) for all populations pooled.

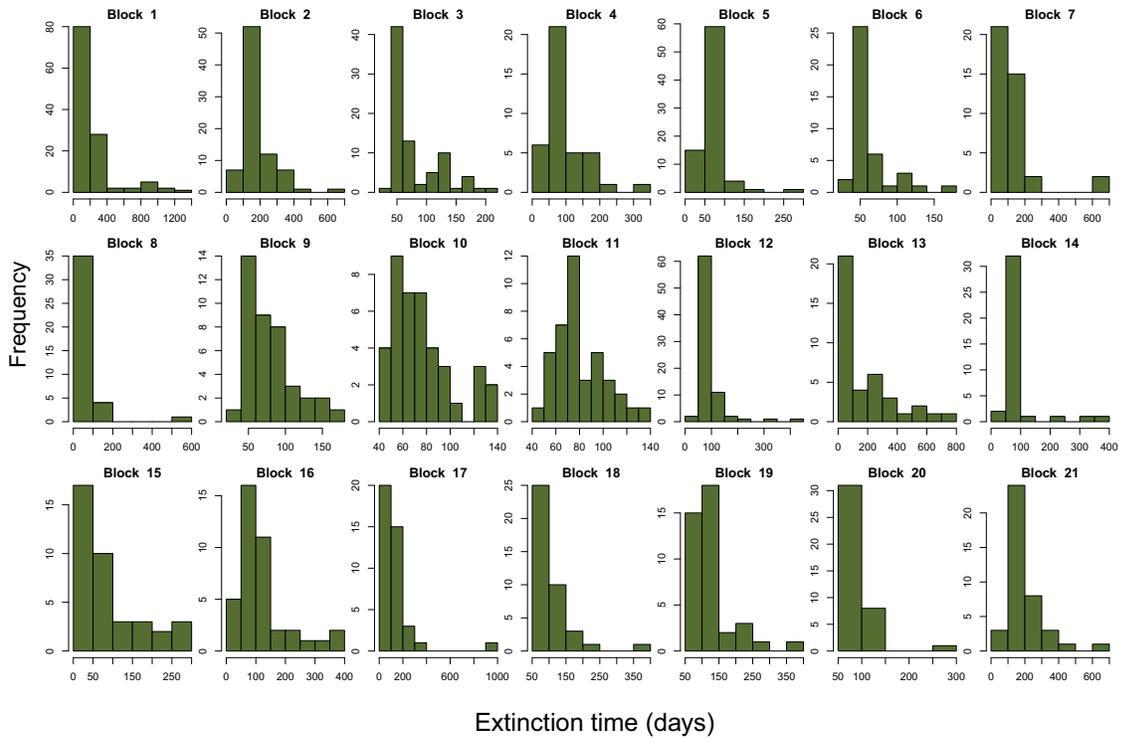


Figure A2: Histograms of extinction time by block.

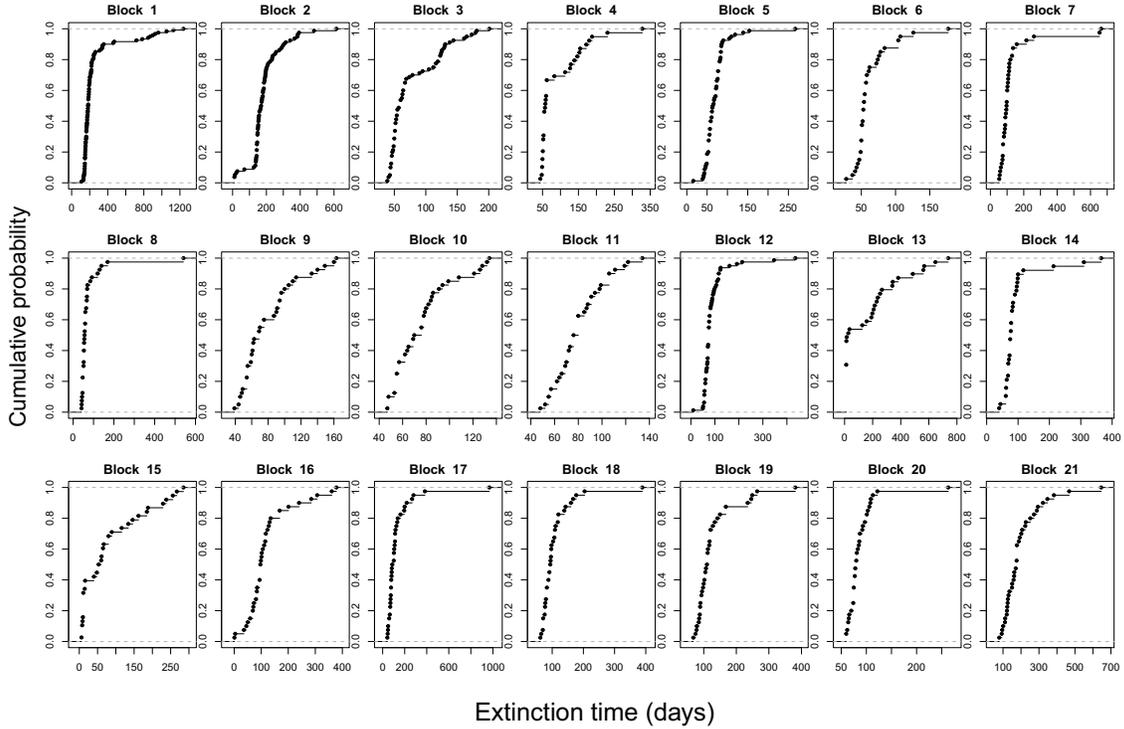


Figure A3: Empirical cumulative distribution of extinction time by block.

2 Effects of tetracycline and initial population size treatments

Some experimental blocks (Blocks 13 and 15) received an additional treatment (+tetracycline) and all blocks received treatments of initial population size (very nearly balanced in the final data set). Sample sizes are shown in Table 2.

I tested for effects of tetracycline and initial population size treatments on log extinction time with four mixed effects models, fit by maximum likelihood using the `lme4` package (Bates *et al.*, 2011).

- Model 1. Main effects of tetracycline and initial population size with random intercept for block
- Model 2. Effect of tetracycline with random intercept for block
- Model 3. Effect of initial population size with random intercept for block
- Model 4. Random intercept for block only

Analysis of variance and comparison of AIC values both show Model 1, the most complex model, to be best supported by the data (Table 3). This model has fixed effects for both treatments. Markov Chain Monte Carlo samples were generated to obtain empirical 95% confidence intervals (highest posterior density interval), which demonstrate statistical significance both for tetracycline (95% CI: -1.11, -0.17) and initial population size (95% CI: 0.01, 0.02). To determine if the random effect of block is warranted, I fit a simple fixed effects model (*i.e.* analysis of variance) which has a greater AIC by a difference of $\Delta_{AIC} = 296.72$. This indicates that the estimated random effects among blocks are indeed real. Analysis of residuals and comparison of predicted and observed values do not indicate any structural problem with this model. The dashed line in the top left panel is the one-to-one-line (Fig. A5).

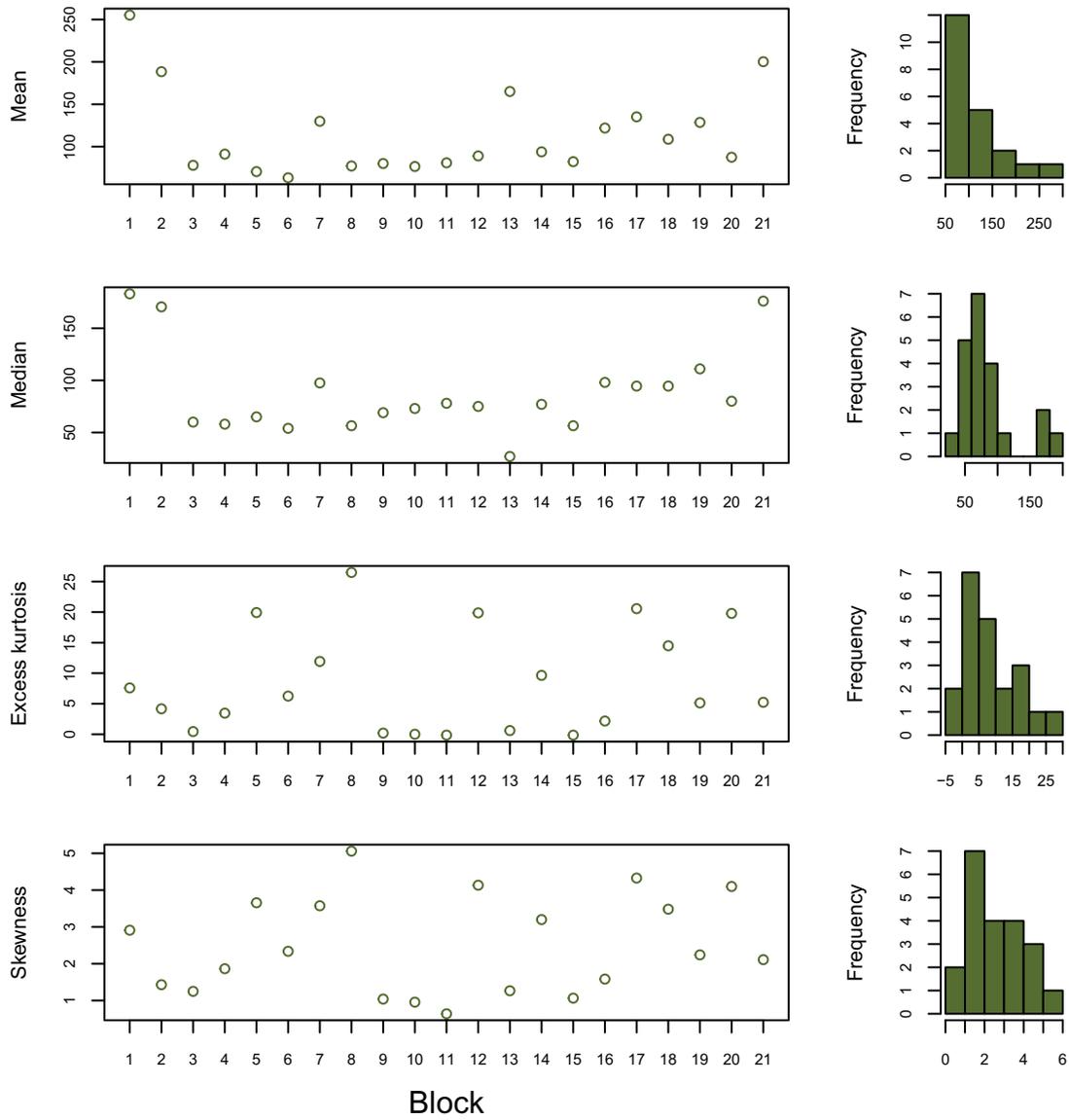


Figure A4: Summary statistics of extinction time by block.

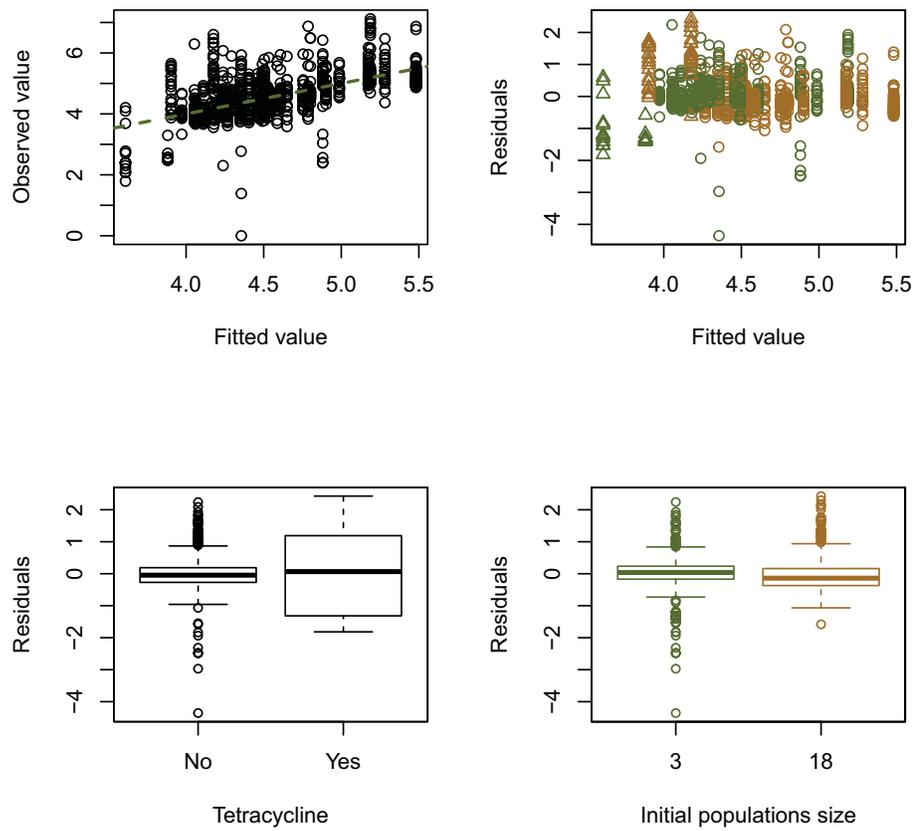


Figure A5: Regression diagnostics for mixed effects model of experimental treatments.

Table A1: Summary statistics of extinction time by block. Boxplots show interquartile range; whiskers extend to the most extreme points within a distance of 1.5 times the interquartile range. Note the logarithmic axis on extinction time plot.

Block	Mean	Median	Excess kurtosis	Skewness	Variance	Coefficient of variation
1	255.09	183.00	7.59	2.91	48605.83	0.86
2	188.45	170.50	4.17	1.43	9730.55	0.52
3	78.08	60.00	0.44	1.25	1673.56	0.52
4	91.23	58.00	3.46	1.86	3996.08	0.69
5	70.64	65.00	19.92	3.66	940.64	0.43
6	63.35	54.00	6.25	2.34	743.82	0.43
7	130.00	97.50	11.93	3.57	16646.21	0.99
8	77.28	56.50	26.49	5.06	6429.90	1.04
9	80.15	69.00	0.20	1.04	1097.98	0.41
10	76.65	73.00	0.01	0.95	613.67	0.32
11	81.00	78.00	-0.13	0.64	416.87	0.25
12	89.11	75.00	19.88	4.13	3278.81	0.64
13	165.13	27.00	0.61	1.26	42212.75	1.24
14	93.95	77.00	9.64	3.20	4138.54	0.68
15	82.34	56.50	-0.13	1.06	7308.18	1.04
16	122.00	98.00	2.17	1.58	7221.79	0.70
17	135.22	94.50	20.56	4.33	23290.79	1.13
18	108.80	94.50	14.49	3.48	3080.88	0.51
19	128.68	111.00	5.13	2.24	4013.05	0.49
20	87.60	80.00	19.79	4.10	1048.19	0.37
21	200.20	176.00	5.24	2.11	12317.39	0.55

In conclusion, these results provide evidence for effects of both initial population size and tetracycline addition on extinction time. These are not, however, expected to influence the tail probabilities (as will be statistically demonstrated below) for the following reasons. In the case of initial population size, any effects are quickly eroded as the majority of populations achieve quasi-stationarity (*i.e.*, stable fluctuations around the long term mean, at which point any memory of the initial condition is lost) (Drake *et al.*, 2011). Those populations that do not reach quasi-stationarity, *i.e.*, those populations that caused the detected treatment effect, reach extinction extremely quickly and therefore are not included in the set of populations making up the tail. In the case of tetracycline, the treatment was applied uniformly within blocks 13 and 15. Thus, since distributions were fit to each block separately (see below), any effect of the tetracycline treatments would apply to the block as a whole. If the addition of tetracycline would qualitatively change the shape of the extinction time distribution, this would be evident in the table of results by block.

3 Estimating tail probabilities

I used the methods of Clauset *et al.* (Clauset *et al.*, 2009) to estimate functions describing distribution tails for each block. Standard models for extinction in a stationary environment imply that the distribution of extinction times should be exponential in the tails. Alternative distributions could arise in a non-stationary environment or because extinction modes are mixed in heterogeneous populations. If the system is in the vicinity of a critical point, the extinction time might be a power law. My goal was to compare a discrete power law (zeta distribution), a discrete power law with exponential cutoff, a fat-tailed distribution that was not a power law (Weibull distribution), and “exponential” tails (geometric distribution). Models were fit using code from (Clauset *et al.*, 2009) in R v 3.02 (R Development

Table A2: Number of experimental populations by block and treatment.

	Tetracycline		0		1	
	Startno	3	18	3	18	
Block						
1		60	60	0	0	
2		40	40	0	0	
3		40	40	0	0	
4		19	20	0	0	
5		40	40	0	0	
6		20	20	0	0	
7		20	20	0	0	
8		20	20	0	0	
9		20	20	0	0	
10		20	20	0	0	
11		20	20	0	0	
12		40	40	0	0	
13		0	0	20	19	
14		20	18	0	0	
15		0	0	18	20	
16		20	20	0	0	
17		20	20	0	0	
18		20	20	0	0	
19		20	20	0	0	
20		20	20	0	0	
21		20	20	0	0	

Table A3: ANOVA table for estimated treatment effects in four linear mixed effects models

	Df	AIC	BIC	logLik	Chisq	Chi Df	Pr(>Chisq)
Model4	3	2099.61	2114.55	-1046.81			
Model2	4	2095.94	2115.85	-1043.97	5.68	1	0.0172
Model3	4	2040.04	2059.96	-1016.02	55.90	0	0.0000
Model1	5	2036.33	2061.22	-1013.16	5.71	1	0.0168

Core Team, 2011). Given the significance of the random effects in the preceding analysis and visually verifiable variation among blocks in Fig. A2, models were fit separately for each block and results were pooled for statistical inference. Following Clauset *et al.* (Clauset *et al.*, 2007), I first fit the threshold parameter and then fit the tails to all times greater than or equal to the fit threshold. I compared model fits using AIC_c . I found the optimization routine used by Clauset *et al.* occasionally to fail for the Weibull distribution (specifically for Blocks 8 and 14) due to numerical problems at small trial values of the shape parameter. To get around this problem, the shape and scale parameters (which both must be positive) were passed as the logarithm of their traditional values. As Fig. 3 shows, in the majority of blocks that can be fit both ways the results of the two approaches are equivalent. In 4 cases, the logged version actually returns a parameterization yielding a slightly higher likelihood (in one case higher by about 1.2 log-likelihoods). Using the version involving logged parameters, I therefore fit the four models, retaining the log-likelihood, and calculating AIC_c . The results, by block, are shown in Table 4. The total statistical support for each of the four distributions may be obtained by summing AIC_c over blocks for each model (Table 5).

This analysis leads to the conclusion that the power law is strongly supported overall and individually is best supported in 11 out of 21 blocks. The exponential distribution is most weakly supported overall, with a difference $\Delta_{AIC} = 98.95$, although it was (marginally) best supported in 8 out of 21 blocks. The

Table A4: Statistical fit of tail extinction times, by block, to four models. Third column is the number of observations in the tail (exceeding the cutoff). Column 4 is the fit exponent of the zeta (power law) distribution. Columns 5-8 are AIC values for the candidate distributions

Block	Threshold	n	Power	α	Power/cutoff	Weibull	Exponential
1	138	115	1289.37	3.2	1291.44	1291.70	1339.12
2	135	72	756.02	3.8	757.48	757.62	759.60
3	41	79	738.13	2.8	734.95	735.48	735.08
4	113	12	125.70	3.6	127.27	127.28	124.40
5	49	71	611.60	3.6	608.75	608.50	606.66
6	58	14	127.23	3.6	129.29	129.32	126.61
7	73	35	350.61	3.2	352.87	352.94	367.00
8	43	38	335.36	3.2	337.59	337.65	351.83
9	49	35	330.57	3.0	326.19	325.98	324.37
10	55	35	303.74	3.8	302.68	302.53	300.53
11	62	34	292.21	4.2	288.50	285.49	286.97
12	60	69	621.46	3.7	623.58	623.55	633.49
13	191	16	200.77	2.8	202.07	202.21	199.67
14	61	36	321.71	3.8	323.96	324.05	332.79
15	48	22	246.57	2.2	244.25	244.38	241.96
16	99	20	214.15	3.1	216.29	216.49	215.53
17	69	33	347.41	2.8	349.66	349.70	360.72
18	76	34	318.88	3.8	321.01	320.98	321.71
19	85	36	353.65	3.6	355.56	355.64	355.98
20	64	37	318.96	4.2	320.03	319.79	318.01
21	150	27	299.50	3.5	301.38	301.47	300.52

Table A5: Total statistical support (summed AIC) for four models of tail extinction probabilities.

AIC	
Power	8503.61
Power cutoff	8514.80
Weibull	8512.74
Exponential	8602.56

power law with exponential cutoff and Weibull distribution were best supported each for only one block.

4 Effect of experimental treatments on tail distribution

Above it was argued that the initial population size treatment, despite its statistical significance, should not affect the shape of the estimated extinction time distribution. I tested that claim by fitting the distributions separately for each $Block \times x_0$ combination and calculating AIC_c .

The $x_0 = 3$ treatment for Block 13 has only two records in the tail (and therefore cannot be used to estimate the two-parameter models and renders AIC_c for the remaining two models undefined). It is therefore excluded from further consideration. Tabulating the results separately for the $x_0 = 3$ and $x_0 = 18$ treatments and summing AIC_c values enables one to determine if the different treatments yielded qualitatively different patterns in the tail extinction probabilities. Results are shown by block in Table 6. Clearly, even when models are fit separately for the two treatments (initial condition of $x_0 = 3$ vs. $x_0 = 18$), the power-law is preferred (Table 7). These results therefore indicate that there is no effect

of initial population size on the shape of the extinction time tail and underscore the main conclusion that extinction times in this experiment followed a power law distribution.

References

- Bates, D., Maechler, M. & Bolker, B. 2011 *lme4: Linear mixed-effects models using eigen and classes*. R package version 0.999375-42.
- Clauset, A., Shalizi, C. R. & Newman, M. E. J. 2009 Power-law distributions in empirical data. *SIAM Review*, **51**(4), 661–703. (doi:10.1137/070710111)
- Clauset, A., Young, M. & Gleditsch, K. S. 2007 On the frequency of severe terrorist events. *Journal of Conflict Resolution*, **51**(1), 58–87. (doi:10.1177/0022002706296157)
- Drake, J. M., Shapiro, J. & Griffen, B. D. 2011 Experimental demonstration of a two-phase population extinction hazard. *Proceedings of the Royal Society Interface*, **8**(63), 1472–9. (doi:10.1098/rsif.2011.0024)
- R Development Core Team 2011 *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

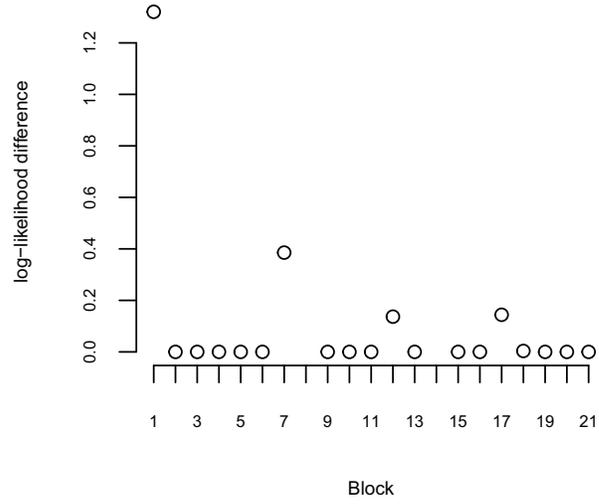


Figure A6: In the majority of blocks that can be fit both ways (all blocks except 8 and 14) the results of the two optimization approaches are equivalent. In 4 cases, the logged version actually returns a parameterization yielding a higher likelihood (in one case higher by about 1.2 log-likelihoods).

Table A6: Statistical fit of tail extinction times, by block-treatment combinations, to four models.

Block/x(0)	Threshold	x(0)	n	Power	Power/cutoff	Weibull	Exponential
1 3	145	3	59	684.88	687.03	687.22	716.52
1 18	138	18	56	590.07	592.22	592.32	605.38
2 3	199	3	9	106.21	109.30	109.33	106.02
2 18	151	18	29	314.29	315.53	315.62	313.56
3 3	43	3	38	316.08	317.67	317.80	317.01
3 18	41	18	40	403.89	399.19	399.26	397.04
4 3	48	3	19	128.46	130.98	130.98	129.38
4 18	129	18	10	102.49	105.34	105.35	102.17
5 3	47	3	33	274.33	268.73	262.43	267.57
5 18	54	18	36	314.89	316.77	316.70	316.63
6 3	62	3	4	38.76	50.06	49.46	38.14
6 18	52	18	12	108.13	110.90	110.98	108.69
7 3	68	3	16	146.17	148.43	148.45	146.11
7 18	76	18	20	213.05	215.53	215.57	223.35
8 3	41	3	20	182.93	185.41	185.44	194.93
8 18	42	18	20	172.26	174.60	174.66	174.13
9 3	49	3	17	151.59	153.68	153.77	151.60
9 18	46	18	20	197.57	193.15	192.01	191.58
10 3	47	3	20	160.74	163.06	163.08	161.34
10 18	69	18	17	145.84	146.11	145.19	143.77
11 3	55	3	19	164.33	163.28	161.64	161.21
11 18	62	18	19	165.26	164.96	164.02	162.75
12 3	68	3	34	317.79	320.04	320.05	323.61
12 18	51	18	39	337.62	339.58	339.03	340.27
13 3	15	3	2	Inf	3.47	3.48	Inf
13 18	191	18	16	200.77	202.07	202.21	199.67
14 3	61	3	20	181.23	183.72	183.77	188.35
14 18	61	18	16	142.84	145.48	145.50	146.65
15 3	8	3	17	112.45	114.98	115.09	119.30
15 18	48	18	20	229.19	225.79	225.69	223.31
16 3	81	3	17	179.81	182.23	182.38	182.24
16 18	97	18	11	107.90	110.96	110.98	110.13
17 3	72	3	17	174.04	175.64	175.74	173.39
17 18	59	18	18	195.24	197.78	197.82	206.53
18 3	70	3	20	185.28	186.16	186.16	183.68
18 18	80	18	15	146.86	149.54	149.55	149.37
19 3	85	3	17	166.07	168.65	168.65	168.40
19 18	86	18	19	188.77	190.90	191.02	189.19
20 3	64	3	19	159.92	159.55	157.81	157.37
20 18	64	18	18	161.04	163.45	163.34	161.87
21 3	121	3	16	162.39	163.85	163.82	161.22
21 18	150	18	17	197.93	199.88	200.01	198.02

Table A7: Total statistical support (summed AIC) for four models of tail extinction probabilities for two initial population sizes.

	Power	Power/cutoff	Weibull	Exponential
x(0)=3	3993.46	4032.42	4023.08	4047.42
x(0)=18	4635.88	4659.73	4656.83	4664.06