Introduction to Probability and Statistics for Engineers and Scientists

Sheldon M Ross

Solutions

Anirudh Krishnan

October 11, 2021

Contents

Chapter 1

Life Testing

"Ch14 Quote here."

1.1 Introduction

The general problem being considered here is a population of independent items having some common underlying distribution of lifetimes, which is known but for a single parameter.

The concept of a hazard rate is used in engineering to analyze this problem and the most common choice of *a-priori* distribution assumed is an exponential RV.

1.2 Hazard rate functions

Consider a positive RV X with some CDF (F) and PDF (f). The *failure rate* (also called *hazard rate*) function of F is now defined as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} \tag{0.1}$$

$\lambda(t)$ represents the conditional probability that an item of age t will fail imminently.

$$P\{X \in (t,\ t + \mathrm{d}t) \mid X > t\} \approx \frac{f(t)}{1 - F(t)}\ \mathrm{d}t \tag{0.2}$$

For an underlying exponential distribution, which is memoryless, $\lambda(t) = \lambda$ constant equal to its rate. The rate function is uniquely able to determine an the underlying CDF for a positive continuous RV.

$$\lambda(s) = \frac{\mathrm{d}\ F(s)}{\mathrm{d}s}\ \frac{1}{1 - F(s)} = \frac{\mathrm{d}}{\mathrm{d}s}\ - \log\left[1 - F(s)\right] \tag{0.3}$$

$$1 - F(t) = \exp\left[-\int\limits_0^t\ \lambda(s)\ \mathrm{d}s\right] \tag{0.4}$$

A special case of $\lambda(t) = bt$ is called the *Rayleigh* density function. A linear relationship between death rates between two conditions leads to a power law relationship between the probability of both conditions surviving to the same age.

$$\begin{aligned} \lambda_y &= n\lambda_x \\ \implies\ P\{Y > b \mid Y > a\} &= P\{X > b \mid X > a\}^n \end{aligned} \tag{0.5}$$

Where $b > a$ are some ages and X, Y are two categories being compared.

1.3 Exponential distributions in life testing

Stopping at the r^{th} failure : Consider a population of n items which are all IID using an exponential distribution with unknown mean θ. A problem of great interest is to attempt to estimate θ using observations about the time taken for r out of n simultaneously initialized items to fail.

The observed data takes the form of an ascending ordered set of failure times $\{x_1,\ \ldots,\ x_r\}$ along with a set of indices for the failing items $\{i_1\ \ldots\ i_r\}$. If the lifetime of component i_j is denoted by X_{i_j}, then the independence of components gives

$$f_{X_{i_j}} = \frac{1}{\theta}\ \exp\left[-x_j/\theta\right] \tag{0.6}$$

$$f_{X_{i_1}\ldots X_{i_r}} = \prod_{j=1}^{r} \frac{1}{\theta}\ \exp\left[-x_j/\theta\right] \tag{0.7}$$

Additionally, the probability of the rest of the $(n-r)$ components having a lifetime larger than x_r is

$$P\{\text{rest of the components lasting longer}\} = \left[\exp(-x_r/\theta)\right]^{n-r} \tag{0.8}$$

Defining a likelihood function $\mathcal{L}$ as the product of the two expression above and finding an MLE $\widehat{\theta}$ of the unknown mean θ and defining $X_{(j)}$ as the time at which the j^{th} failure occurs,

$$\widehat{\theta} = \frac{1}{r}\ \left[\ \sum_{j=1}^{r} X_{(j)} + (n-r)X_{(r)}\ \right] = \frac{\tau}{r} \tag{0.9}$$

τ is called the *total time-on-test* statistic, because it represents the sum of lifetimes of all r failing items as well as all $n-r$ surviving items.

For an interval estimate of θ, consider a set of RVs $\{Y_j\}$ which measure the additional time on test contributed by the time interval between the $(i-1)^{\text{th}}$ and i^{th} failures

$$Y_j = (n-j+1)\ (X_{(j)} - X_{(j-1)}) \tag{0.10}$$

$$\tau = \sum_{j=1}^{r} Y_j \tag{0.11}$$

$$Y_j \sim \mathtt{Expon}(1/\theta) \qquad \text{and} \qquad \tau \sim \mathtt{Gamma}(r,\ 1/\theta) \tag{0.12}$$

Using the relation $\mathtt{Gamma}(n/2,\ 1/2) \equiv \chi^2_n$, the confidence interval for θ can be defined as

$$\frac{2\tau}{\theta} \sim \chi^2_{2r} \tag{0.13}$$

$$\theta \in \left(\frac{2\tau}{\chi^2_{\alpha/2,\ 2r}},\ \frac{2\tau}{\chi^2_{1-\alpha/2,\ 2r}} \right) \tag{0.14}$$

The length of the test $X_{(r)}$ is itself the sum of many exponential functions, with

$$\mathbb{E}[X_{(r)}] = \sum_{j=1}^{r} \frac{\theta}{n-j+1} \qquad \approxeq \theta\ \log\left[\frac{n}{n-r+1}\right] \tag{0.15}$$

$$\mathrm{Var}(X_{(r)}) = \sum_{j=1}^{r} \left(\frac{\theta}{n-j+1}\right)^2 \qquad \approxeq \theta^2\ \left[\frac{r-1}{n(n-r+1)}\right] \tag{0.16}$$

Sequential Testing : If an infinite supply of IID items (each exponentially distributed with mean θ) are put to test one by one instead of simultaneously, with the test ending at some fixed time T.

The observations at the end of such a test is a set of lifetimes $\{x_j\}$ with $j \in 1, \ldots, r$ for the total number of items r that failed within time T.

In order for the total number of failed items to be exactly r, the inequality

$$\sum_{i=1}^{r} x_i < T < \sum_{i=1}^{r} x_i + X_{r+1} \tag{0.17}$$

Once again defining a likelihood function $\mathcal{L}$ and finding the MLE $\widehat{\theta}$,

$$\mathcal{L}(r, x_1, \ldots, x_r \mid \theta) = \frac{1}{\theta^2}\ \exp[-T/\theta] \tag{0.18}$$

$$\widehat{\theta} = \frac{T}{r} \tag{0.19}$$

Once again, the interval estimate for θ involves relating `Gamma` and χ^2 RVs

$$\frac{2T}{\theta} \sim \chi^2_{2r} \tag{0.20}$$

$$\theta \in \left(\frac{2T}{\chi^2_{\alpha/2,\ 2r}},\ \frac{2T}{\chi^2_{1-\alpha/2,\ 2r}} \right) \tag{0.21}$$

Simultaneous testing with fixed time stop: Consider the earlier simultaneous test case except that the test stops either at a fixed time T or when all n items fail, whichever is sooner.

If r out of the n items, indexed by the set $\{i_j\}$ have failed by time T, with respective fail times $\{x_j\}$, then the likelihood of such a set of observations is

$$f(\{i_j\}, \{x_j\}) = \frac{1}{\theta^r}\ \exp\left[- \frac{\sum_{i=1}^{r} x_i}{\theta} - \frac{(n-r)T}{\theta}\right] \tag{0.22}$$

If R items fail by time T, with the ordered set of failure times $\{X_{(j)}\}$, the MLE of $\theta$is give by

$$\widehat{\theta} = \frac{1}{R}\ \left(\sum_{i=1}^{r} X_{(i)} + (n-R)T \right) = \frac{\tau}{R} \tag{0.23}$$

Note the resemblance of the above expression to the result from the simultaneous testing method. Since τ and R above are both random and dependent on each other, finding an interval estimate of θ is too complex to pursue here.

Bayesian approach : For a set of IID items being tested with common unknown mean $(1/\lambda)$, the likelihood function is given by

$$f(\text{data} \mid \lambda) = K\lambda^r e^{-\lambda t} \tag{0.24}$$

with t being the summed time on test of items that do and don't fail and r the number of failing items.

Assuming a prior distribution $g(\lambda)$ and applying Bayes' theorem,

$$f(\lambda \mid \text{data}) = \frac{f(\text{data} \mid \lambda)\ g(\lambda)}{\displaystyle\int f(\text{data} \mid \lambda)\ g(\lambda)\ \mathrm{d}\lambda} \tag{0.25}$$

$$g(\lambda) \sim \texttt{Gamma}(b, a) \implies f(\lambda \mid \text{data}) \sim \texttt{Gamma}(b + R, a + \tau) \tag{0.26}$$

The Bayes estimator of λ for the special case of a gamma-RV as prior distribution is simply

$$\mathbb{E}[\lambda \mid \text{data}] = \frac{b + R}{a + \tau} \tag{0.27}$$

1.4 Two-sample problem

Consider two sets of items produced using different processes, whose lifetimes are exponentially distributed with mean $\theta_1,\ \theta_2$ respectively. Let samples of size (n, m) be used to test lifetimes resulting in the set of observations $\{X_i\},\ \{Y_j\}$.

Testing the hypothesis $\theta_1 = \theta_2$, involves the sample mean lifetimes $\overline{X}, \overline{Y}$, and the relation between exponential and gamma RVs

$$\begin{aligned} \overline{X} &\sim \texttt{Gamma}(n, 1/\theta_1) \\ \frac{2}{\theta_1}\ \overline{X} &\sim \chi^2_{2n} \end{aligned} \tag{0.28}$$

Rearranging the χ^2-RVs into an F-RV yields the test statistic for the above hypothesis tests along with a p-value and rejection criterion

$$H_0 : \theta_1 = \theta_2 \qquad \text{vs} \qquad H_1 : \theta_1 \neq \theta_2 \tag{0.29}$$

$$F_{n,m} \sim \frac{2\overline{X}/\theta_1}{2n} \, \frac{2m}{2\overline{Y}/\theta_2} \tag{0.30}$$

$$\text{reject } H_0 \text{ if} \qquad P_{H_0}\{F \leq \nu\} \leq \frac{\alpha}{2} \qquad \text{or} \qquad P_{H_0}\{F \geq \nu\} \leq \frac{\alpha}{2}$$

$$\text{accept } H_0 \qquad \text{otherwise} \tag{0.31}$$

$$p \text{ value} = 2 \times \min(P\{F \leq \nu\}, 1 - P\{F \leq \nu\}) \tag{0.32}$$

1.5 Weibull distribution in life testing

Moving away from the simplistic constant value of hazard rate function, a commonly encountered real-world functional form for $\lambda(t)$ is

$$\lambda(t) = \alpha\beta \; t^{\beta-1} \qquad\qquad \alpha, \beta, t > 0 \tag{0.33}$$

$$F(t) = 1 - \exp\left[-\alpha t^{\beta}\right] \tag{0.34}$$

$$f(t) = \alpha\beta t^{\beta-1} \; \exp\left[-\alpha t^{\beta}\right] \tag{0.35}$$

Parameter estimation by maximum likelihood : Consider n IID Weibull RVs $\{X_1, \ldots, X_n\}$ being used to estimate the parameters α, β of the Weibull distribution. The MLE approach yields

$$\widehat{\alpha} = \frac{n}{\sum x_i^{\widehat{\beta}}} \qquad\qquad n + \widehat{\beta} \; \log\left(\prod x_i\right) = \frac{n\widehat{\beta} \sum x_i^{\widehat{\beta}} \; \log x_i}{\sum x_i^{\widehat{\beta}}} \tag{0.36}$$

The MLE estimators are unfortunately not closed-form and need to be solved for numerically.

Parameter estimation by least squares : Given n IID Weibull RVs, and an ordered set of sample values $X_{(i)}$,

$$F(x) = 1 - \exp\left[-\alpha x^{\beta}\right] \tag{0.37}$$

$$\log\log\left[\frac{1}{1 - F(x)}\right] = \beta \log x + \log\alpha \tag{0.38}$$

If the LHS above can be approximated using y_i, then a linear regression problem can be set up to yield approximate values $\widehat{\alpha}$, $\widehat{\beta}$. Multiple methods of estimating y_i exist, with no real winner in terms of computation time or accuracy.

Chapter 2

Life Testing

1 Weibull RV with parameters α, β

$$F(t) = 1 - \exp\left[-\alpha t^{\beta}\right] \qquad\qquad t \geq 0$$

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{1 - F(t)} = \frac{\alpha\beta\ t^{\beta-1} \exp\left[-\alpha t^{\beta}\right]}{\exp\left[-\alpha t^{\beta}\right]} \\ &= \alpha\beta\ t^{\beta-1} \qquad (1.1)\end{aligned}$$

2 Two RVs X, Y with respective failure rate functions $\lambda_x(t), \lambda_y(t)$. Their minimum is Z

$$\begin{aligned}Z &\equiv \min(X, Y) \\ P\{Z > t\} &= P\{X > t\} \times P\{Y > t\} \qquad (2.1) \\ [1 - F_z(t)] &= [1 - F_x(t)]\ \ [1 - F_y(t)]\end{aligned}$$

$$f_z(t) = \frac{\mathrm{d}}{\mathrm{d}t}\ F_z(t) = f_x(t)\,[1 - F_y(t)] + f_y(t)\,[1 - F_x(t)] \qquad (2.2)$$

$$\lambda_z(t) = \lambda_x(t) + \lambda_y(t) \qquad (2.3)$$

3 Given failure rate function $\lambda(t)$,

$$\lambda(t) = 0.027 + 0.025 \left(\frac{t-40}{10}\right)^4 \tag{3.1}$$

$$F(t) = 1 - \exp\left[-\int_0^t \lambda(s)\ \mathrm{d}s\right]$$

$$= 1 - \exp\left[-0.027t - 0.05\left(\frac{t-40}{10}\right)^5\right]$$

$$P\{t \geq 50 \mid t > 40\} = \frac{1-F(50)}{1-F(40)} = 72.6\% \tag{3.2}$$

$$P\{t \geq 60 \mid t > 40\} = 11.7\% \tag{3.3}$$

$$P\{t \geq 70 \mid t > 40\} = 0.0002\% \tag{3.4}$$

$$\tag{3.5}$$

4 Given $\lambda(t) = t^3$

$$1 - F(2) = \exp\left[-\int_0^2 \lambda(s)\ \mathrm{d}s\right] = \exp\left[\left(\frac{t^4}{4}\right)\Bigg|_2^0\right]$$

$$= 1.83\% \tag{4.1}$$

$$F(1.4) - F(0.4) = 99.36\% - 38.27\% = 61.09\% \tag{4.2}$$

$$\mathbb{E}[T] = \int_0^\infty [1 - F(t)]\ \mathrm{d}t = \frac{1}{4}\ (-0.25^{-0.25})\Gamma(1/4)$$

$$= 1.28 \tag{4.3}$$

$$P\{t \geq 2 \mid t > 1\} = 2.35\% \tag{4.4}$$

5 Failure rate is non-decreasing in t, called IFR

$$f(t) = \lambda e^{-\lambda t}(\lambda t)^{\alpha-1}$$

$$\lambda(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\int\limits_t^\infty \lambda e^{-\lambda s}(\lambda s)^{\alpha-1}\,\mathrm{d}s} \tag{5.1}$$

$$= \left[\int\limits_t^\infty e^{-\lambda(s-t)}\,(s/t)^{\alpha-1}\,\mathrm{d}s\right]^{-1} \tag{5.2}$$

$$= \left[\int\limits_0^\infty e^{-\lambda v}\,(1+v/t)^{\alpha-1}\,\mathrm{d}v\right]^{-1} \tag{5.3}$$

The above integrand is decreasing in t if $\alpha - 1 > 0$. This implies an IFR for $\alpha > 1$. The first expression is a special case of $\alpha = 2$ and is thus also IFR.

6 Uniform RV on $[a, b]$ is IFR

$$f(t) = \frac{1}{(b-a)} \qquad \forall\ x \in (a, b)$$

$$\lambda(t) = \frac{1}{(b-t)} \qquad \forall\ t \in (a, b) \tag{6.1}$$

$$\tag{6.2}$$

Clearly, since $t < b$, the failure rate is increasing in t

7 Consider the bar chart to be composed of horizontal bars first $\{h_j\}$ and then vertical bars $\{v_k\}$

$$h_1 = n \times \left[X_{(1)} - X_{(0)}\right]$$

$$h_r = (n - r + 1) \times \left[X_{(r)} - X_{(r-1)}\right]$$

$$\text{Area under curve} = \sum_j\ h_j \tag{7.1}$$

$$v_1 = X_{(1)}$$

$$v_k = X_{(r)} \qquad \forall\ k \geq r$$

$$\text{Area under curve} = \sum_k\ v_k \tag{7.2}$$

$$\tag{7.3}$$

The area under the curve is thus equal to both $\sum_{j=1}^{r} X_{(j)} + (n-r)X_{(r)}$ and also to τ.

8 Performing the simultaneous start test and stopping at r failures,

`Stopping at r failures`	
Total items (n)	30
Number failed (r)	10
Total time on test (τ)	1541.50
MLE of mean ($\widehat{\theta}$)	154.15
Stop time (T)	62.20
p value %	0.00

Confidence interval of 0.95 % for the mean lifetime is [90.23, 321.45]

Hypothesis test for $\theta_0 = 7.5$ is rejected

Lower confidence interval of 95 % for the mean lifetime is $(-\infty, 284.13]$

9 Placeholder P9

10 Performing the simultaneous start test and stopping at r failures,

Hypothesis test for $\theta_0 = 10$ is accepted

`Stopping at r failures`	
Total items (n)	20
Number failed (r)	8
Total time on test (τ)	44.31
MLE of mean ($\widehat{\theta}$)	5.54
Stop time (T)	2.72
p value %	16.20

11 Using the exact formula for the mean and variance of the testing period T

`Statistics of testing period`	
Total items (n)	20
Number failed (r)	10
Mean lifetime (θ)	10.00
Test Period mean (exact)	6.69
Test Period variance (exact)	4.64

12 Using the approximate formula for the mean testing period

$$\mathbb{E}[X_{(r)}] \approxeq \log\left(\frac{n}{n-r+1}\right) \tag{12.1}$$

$$3 = \theta \ \log\left(\frac{n}{n-9}\right)$$

$$9/n = 1 - \exp(-3/20)$$

$$n = 64.6 \tag{12.2}$$

A minimum of 65 items need to be tested simultaneously.

13 Performing the sequential test,

Confidence interval of 95 % for the mean lifetime is [12.13, 32.8]

Hypothesis test for $\theta_0 = 20$ is accepted

`Stopping at fixed time`	
Number failed (r)	16
Total time on test (T)	300.00
MLE of mean ($\widehat{\theta}$)	18.75
p value %	86.38

14 Inter arrival times are IID exponential RVs. The resulting process is a Poisson RV

Using the results from sequential testing with fixed stopping time T

$$P\{N(x/2) \geq n\} = P\{X_1 + \cdots + X_n \leq x/2\} \tag{14.1}$$

The sum of the first n inter-arrival times have to be smaller than $x/2$ for the Poisson process to yield an integer not smaller than n after time $x/2$ elapses.

$$\sum_{j=1}^{n} X_j \sim \texttt{Gamma}(n, 1/\theta)$$

$$\sim \frac{\theta}{2}\ \chi^2_{2n}$$

$$P\{N(x/2) \geq n\} = P\left\{\frac{\theta}{2}\ \chi^2_{2n} \leq x/2\right\} \tag{14.2}$$

Thus, the RHS can be rearranged to be the CDF of a χ^2 RV with $2n$ DOF.

15 First consider the case when time T elapses and the test stops before r failures. Let the number of failures be $k < r$. The observations $\{x_i\}$ indicate the failure times.

The differences between successive elements of $\{x_i\}$ are the lifetimes of each item.

$$\mathcal{L}(x_1, \ldots, x_r \mid \theta) = \prod_{i=1}^{k} \frac{1}{\theta} \exp\left[\frac{-(x_i - x_{i-1})}{\theta}\right] \times P\left\{x_{k+1} > T\right\} \tag{15.1}$$

$$= \frac{1}{\theta^k} \exp\left[\frac{-(x_k - x_0)}{\theta}\right] \times \exp\left[\frac{-(T - x_k)}{\theta}\right] \tag{15.2}$$

$$= \theta^{-k} \exp\left\{\frac{-T}{\theta}\right\} \tag{15.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\mathcal{L}) = 0 = \frac{-k}{\theta} + \frac{T}{\theta^2}$$

$$\widehat{\theta} = T/k = \frac{\text{total duration of test}}{\text{number of failures}} \tag{15.4}$$

Next consider the case where the test stops because all r items have failed before time T

$$\mathcal{L}(x_1, \ldots, x_r \mid \theta) = \prod_{i=1}^{r} \frac{1}{\theta} \exp\left[\frac{-(x_i - x_{i-1})}{\theta}\right] \tag{15.5}$$

$$= \frac{1}{\theta^r} \exp\left[\frac{-(x_r - x_0)}{\theta}\right] \tag{15.6}$$

$$= \theta^{-r} \exp\left\{\frac{-x_r}{\theta}\right\} \tag{15.7}$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log(\mathcal{L}) = 0 = \frac{-r}{\theta} + \frac{x_r}{\theta^2}$$

$$\widehat{\theta} = x_r/k = \frac{\text{total duration of test}}{\text{number of failures}} \tag{15.8}$$

In both cases the general formula does hold, as seen in the last step.

16 Find MLE given likelihood function

$$\mathcal{L} = \frac{K}{\theta^r}\ \exp\left[\frac{-1}{\theta}\ \left(\sum_{i=1}^{r} x_i + \sum_{j=1}^{s} y_j\right)\right] \tag{16.1}$$

$$= \frac{K}{\theta^r}\ \exp\left[\frac{-\tau}{\theta}\right] \tag{16.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ \log(\mathcal{L}) = 0 = \frac{-r}{\theta} + \frac{\tau}{\theta^2}$$

$$\widehat{\theta} = \frac{\tau}{r} = \frac{1}{r}\ \left(\sum_{i=1}^{r} x_i + \sum_{j=1}^{s} y_j\right) \tag{16.3}$$

17 Performing the sequential test with panel size 5

`Stopping at fixed time`	
Number failed (r)	9
Total time on test (T)	764.00
MLE of mean ($\widehat{\theta}$)	84.89

18 Using the general rule for the MLE of θ

$$\widehat{\theta} = \frac{\text{total time on test}}{\text{number of failures}} \tag{18.1}$$

$$= \frac{702}{12} = 58.57$$

19 Data from Problem 17, Prior distribution is `Gamma`(1, 100)

$$g(\lambda) = \texttt{Gamma}(a, b) \tag{19.1}$$

$$\mathbb{E}[\lambda \mid \text{data}] = \frac{b+R}{a+\tau}$$

$$= \frac{100+9}{1+764} = 0.1425 \tag{19.2}$$

20 Data from Problem 18, Prior distribution is $\texttt{Expon}(\lambda = 30)$

$$
\begin{aligned}
f(\lambda \mid \text{data}) &= \frac{\lambda^r e^{-\lambda t} g(\lambda)}{\int \lambda^r e^{-\lambda t} g(\lambda)\, \mathrm{d}\lambda} \\
&= \frac{\lambda^r e^{-(\lambda)t}}{\int \lambda^r e^{-(\lambda)t}\, \mathrm{d}\lambda} \\
&= \frac{(t)^{r+1}\ \lambda^r e^{-\lambda t}}{(r)!} \\
&= \texttt{Gamma}(r+1, t) && (20.1) \\
\mathbb{E}[f(\lambda \mid \text{data})] &= \frac{r+1}{t} = 0.0185 && (20.2) \\
\lambda_b &= 54.08 && (20.3)
\end{aligned}
$$

21 Performing the two sample test for equality of the exponential RV means, null hypothesis is accepted

`Two sample problem`	
Sample A size (n)	7
Sample B size (m)	7
Mean Lifetime A ($\overline{X}$)	123.17
Mean Lifetime B ($\overline{Y}$)	89.51
Test Statistic (F Test)	1.38
p vlaue %	68.43