Exponential and Piecewise Exponential Distributions

Mai Zhou

Summary

For the purpose of learning *Survival Analysis*, we review some properties of exponential distributions. Then we discuss some **extensions** of the exponential distribution. These extensions include the Weibull distribution/extreme value distributions; and the piecewise exponential distributions.

Since lifetimes are almost always non-negative, the normal model/distribution may not be appropriate. An easy to use, positive distribution is the exponential distribution. To a less extend, exponential distributions to the survival analysis is like normal distributions to the linear model/ANOVA.

Piecewise exponential distribution is also used to bridge/connect the parametric and nonparametric method/model, with the view that when the number of pieces grows to infinite (along with the sample size) the parametric model becomes the non-parametric model.

1 Exponential distribution, Weibull and Extreme Value Distribution

1. (Definition) Let X be a random variable. We say $X \sim \exp(\lambda)$, we mean $P(X > t) = P(X \geq t) = e^{-\lambda t}$ for $t > 0$, where $\lambda > 0$ is a parameter (called hazard parameter).

[Some other books use a different parameter. The best way to identify which parameter a particular book is using is to ask what is the mean value of the r.v. in terms of the parameter. In our definition, mean $= 1/\lambda$. Some books use the parameter $\beta =$ mean. We shall NOT use the mean parameter.]

2. How can you get [generate] a sample of iid (standard) exponential random variables from a sample of iid uniform(0,1) random variables?

2.5. If X is $\exp(\lambda)$ then $c \times X, (c > 0)$ is exp(?).

3. (Definition) If X is exp(1) then $X^{1/\beta}/\lambda$ is called a Weibull(λ, β) random variable. (notice this definition is same as Miller, Kalbfleish & Prentice, Cox & Oakes, ABGK but different from software R, Allison, and Klein & Moeschberger and some other text books.)

In the above, $\beta > 0$ is called the shape parameter, $\lambda > 0$ is called the scale parameter (of the Weibull).

Therefore, if X is exp(1) then $(X)^p$ is Weibull(?,?) for $(p > 0)$.

Find the survival function of an Weibull r.v.

4. (Definition) If X is exp(1) then $\log(X)$ is called a (standard) extreme value random variable. Its distribution is called an extreme value distribution.

$E\log(X) =$?? (not 0). $Var\log(X) =$

5. Define a two-parameter family of general extreme value random variable/distribution as: $\mu + \sigma\log(X)$. Here μ and $\sigma > 0$ are two parameters. (notice μ is not the mean since mean of $\log X$ is not zero. Similarly σ is not the standard deviation. They are called location and scale parameters.)

Therefore, if X is exp(1) then $\log[(cX)^p]$ is ..?...

If Y is Weibull(λ, β) then $\log Y$ is ..?....

2 Hazard and Cumulative Hazard function

Definition: The hazard function $h(t)$ (or sometimes denote by $\lambda(t)$), of a general r.v. X (that have a density function), is

$$h(t) = \frac{f(t)}{1 - F(t-)} \ .$$

The minus in the denominator F is superfluous here but will be useful later when we talk about discrete r.v.s.

For X that is exp(λ), the hazard function $h(t)$ is a constant: $h(t) \equiv \lambda$, (memoryless, forever young).

Prove that a r.v. is memoryless if and only if its hazard function is constant.

Definition: Cumulative hazard function $H(t)$ or $\Lambda(t)$ is:

$$H(t) = \int_{-\infty}^{t} h(s)ds = \int_{-\infty}^{t} \frac{dF(s)}{1 - F(s-)} \ .$$

Notice the last expression is also valid even for discrete CDF $F(t)$. So, we can still have a cumulative hazard function when the CDF of a random variable is discrete.

The intuitive meaning of the hazard $h(t)$. We notice

$$h(t) = \lim_{\epsilon \downarrow 0} P(t \le X < t + \epsilon | X \ge t)/\epsilon$$

So $h(x)dx$ is a conditional probability: $P(x \le X < x + dx | X \ge x)$.

5. The relation between density $f(t)$, cdf $F(t)$ and hazard $h(t)$, cumulative hazard $H(t)$.

For continuous cdf $F(t)$, we have $H(t) = -\log[1 - F(t)]$. Thus $1 - F(t) = e^{-H(t)}$.

For (purely) discrete cdf $F(t)$ we can also **define** $H(t)$:

$$H(t) = \sum_{s \le t} \frac{\Delta F(s)}{1 - F(s-)} \ ;$$

where $\Delta F(t) = F(t+) - F(t-)$, then we can show the following relation holds (exercise)

$$1 - F(t) = \prod_{s \le t} (1 - \Delta H(s)) \ .$$

For partly discrete, partly continuous distribution $F(t)$, there is also a pair of formula (which are applicable to all cases) but let us not go there. See Gill 1980

6. If X_1 and X_2 are two independent exponential r.v.s (with parameters λ_1 and λ_2) then $\min(X_1, X_2)$ is also an exp r.v. (with parameter $\lambda_1 + \lambda_2$).

Generalizations. (If Y_1 and Y_2 are independent general r.v.s with hazard functions $h_1(t)$ and $h_2(t)$, then $\min(Y_1, Y_2)$ has hazard function .....)

7. Example of using memoryless: The expectation of order statistics from an independent exponential sample. (hint: Find the expectation of the minimum first, and then use memoryless).

8. Given an i.i.d. exp(λ) sample the MLE of the hazard, is .....

Do you know the distribution of the MLEs?

9. The Fisher information for λ in the sample is .....

9.5 The MLE of two parameters in the weibull distribution is implicit

10. In general, for any distribution, we have, based on an iid sample $x_1, \cdots x_n$

$$\log lik = \sum_{i=1}^{n} [\log h(x_i) - H(x_i)] \ .$$

Generalize this to right censored case.

If (Y_i, δ_i) $i = 1, \cdots, n$ is a sample of right censored observations from a general distribution, ($Y_i = \min(X_i, C_i)$, with X_i iid) the log likelihood is

$$\log lik(Y\delta) = \sum_{i=1}^{n} [\delta_i \log h(Y_i) - H(Y_i)] \ .$$

In the above, if X_i is exponential, then h and H simplify to ...

The exponential rv view of any positive rv: any positive random variable can be thought of as the result of a (monotone increasing) transformation of an exponential random variable. Or an exponential rv under a crazy clock. ....And that is via the hazard function clock.....

3 Piecewise Exponential random variable

12. **Piecewise exponential distribution**: its definition and how to estimate the parameters from a sample.

Definition: If a random variable Y's hazard function, $h_Y(t)$, is a piecewise constant function, then Y is called a piecewise exponential random variable. We suppose the boundary or the cut points of the pieces are given (non-random).

As an example a three piece exponential r.v. with cut points $0 < T_1 < T_2 < T_3 = \infty$ has hazard function

$$h(t) = \lambda_1 I[t < T_1] + \lambda_2 I[T_1 \le t < T_2] + \lambda_3 I[T_2 \le t]$$

Its cumulative hazard is $H(t) =$??

The log likelihood function based on an iid sample of X_i from a piecewise exponential distribution can be written down, using $h(t)$ and $H(t)$.

Theorem 1 Based on a sample of n iid observations from a piecewise exponential distribution, the MLE of hazard of the i^{th} piece (for the interval $[T_{i-1}, T_i)$) is

$$\hat{\lambda}_i = \frac{\#\{x_j \in [T_{i-1}, T_i)\}}{\sum_{j=1}^n [\min(T_i, x_j) - T_{i-1}]^+} \quad \text{where} \quad [t]^+ = \max(0, t) \ .$$

The denominator in the above is identical to $\sum_j (x_j - T_{i-1}) I[T_{i-1} \le x_j \le T_i] + \sum_j (T_i - T_{i-1}) I[x_j \ge T_i]$.

9.5. Re-work the above with right censored observations.

Corollary If the sample in the above theorem is subject to right censoring, then the MLE $\hat{\lambda}_i$ only needs modification in the numerator: replace the numerator in the above by $\#\{$uncensored $x_j \in [T_{i-1}, T_i)\}$.

Define $R(t) = \sum I[x_i \ge t]$ then the denominator above is approximately equal to $R(T_{i-1})(T_i - T_{i-1})$.

You see the connection to Nelson-Aalen estimator here.

Theorem 2 The (approximate) variance of the MLE in the above theorem $\hat{\lambda}_i$ is (by using Fisher information for MLE theory)

$$\frac{\hat{\lambda}_i^2}{\sum_j \delta_j I[T_{i-1} \le x_j < T_i]} \ .$$

and $\hat{\lambda}_i$ is asymptotically independent of $\hat{\lambda}_j$ for $i \ne j$.

Notice the estimator $\hat{\lambda}_i$ for different i can be considered independent, at least asymptotically. It is easy to verify that the information matrix for $(\hat{\lambda}_1, \cdots, \hat{\lambda}_k)$ is diagonal, therefore the (approximate)

variance-covariance matrix is also diagonal.

Although this (the variance-covariance calculation) is obtained under the assumption of piecewise exponential population, it is in fact (approx.) true for a random sample from any distribution. [since a piecewise exponential can approximate any distribution] And it is in fact un-correlated for finite samples.

From here you get the Variance estimator for the Nelson-Aalen estimator.

Because the Nelson-Aalen estimator, $\hat{H}(t) = \hat{\Lambda}(t)$, is (approx.) equal to $\sum_{T \le t} \hat{\lambda}_i (T_i - T_{i-1})$. The variance of a sum is the sum of the variances (uncorrelated terms) and (please verify)

$$Var(\hat{\lambda}_i(T_i - T_{i-1})) \approx \frac{\#\{x_j \in [T_{i-1}, T_i); \delta_j = 1\}}{R(T_{i-1})^2}$$

Therefore

$$Var(\hat{\Lambda}(t)) \approx \sum_{T \le t} (T_i - T_{i-1})^2 Var(\hat{\lambda}) = ??$$

10a. Given any continuous r.v. X with cdf $F(t)$, what transformation convert it into exp(1)?

10b. Any continuous r.v. can be thought of as obtained by a transformation of an exponential r.v.

10c. Any positive (continuous?) r.v. can be viewed as an exp() r.v. but with time-changing hazard rate. (instead of constant hazard) Crazy Clock!

11. The best rank test for testing the equality of 2 samples of exp data ..... Savage test. Since the rank do not change under monotone increasing transformation....

Parametric Regression Models

There are two approaches to a regression model with exponential distribution.

12. Exponential regression model, MLE method.

The first approach is the generalized linear model approach. Model Assumption (postulates) X_i are independent and

$$X_i \sim \exp(\lambda_i) \qquad \text{where} \qquad \log(\lambda_i) = \alpha + \beta z_i \quad i = 1, 2, \cdots n.$$

We observe a sample of (X_i, z_i) for $i = 1, \cdots, n$ and need to estimate α, β. X_i is the survival times, the response, z_i the covariate(s).

Let $Z_i \sim \exp(1)$, then $X_i \sim \frac{Z_i}{\lambda_i}$.

In terms of hazard, This approach try to model the (log) hazard as a linear function.

The second approach is the log-linear model approach. Equivalent/alternative modeling with log transformation and extreme value distribution.

$$\log X_i = -\log \lambda_i + \log Z_i = -(\alpha + \beta z_i) + \log Z_i$$

where $\log Z_i = \epsilon_i$ is standard extreme value r.v.

So in this approach, the covariate z_i affects the location parameter of $\log X_i$ which is assumed to be an extreme valued distribution.

13. Generalization: Weibull regression i.e. $X_i \sim Weibull(b, \lambda_i)$. This can be achieved by adding a scale parameter in the above extreme value regression.

$$\log X_i = -\log \lambda_i + \sigma \log Z_i = -(\alpha + \beta z_i) + \sigma \log Z_i$$

We note $\sigma = 1/b$.

14. The estimation of α, β and σ can be obtained by the MLE method. The MLE do not have an easy expression. But are easily obtained using numerical maximization. This can all be done in SAS or R. (code examples?)

15. Generalization (project): replace Z_i above by a piecewise exponential random variable. (problem: there is no 'standard' piecewise exponential; and we might as well take/replace (the exp of) $-\alpha + \sigma \log Z_i$ as a piecewise exponential) (This is a project topic) How MLE/LR test of β works out here?

If we generalize the log linear model so that the error is a general, unknown distribution, we get the AFT model.

If we generalize the regression the hazard way, then the responses are considered as the result of a unknown transformation of the X_i from the exponential regression model.

Exponential random variables and Poisson process (this section can be omitted at first reading)

16. The relation between exp(λ) distributions and Poisson processes. (should be covered in Sta624, the stochastic process)

17. Notations of the Poisson process: $N_\lambda(t)$.

Let

$$N_\lambda(t) - \lambda \times t = N_\lambda(t) - \int_0^t \lambda ds = M(t) \ .$$

The expectation of $M(t)$ is zero for any t, in fact, $M(t)$ is a so called *martingale*.

Problems

1) Plot the hazard function $h(t)$ for

a. $N(\mu, \sigma)$ distributions with several different μ's and σ's.

b. log-normal distributions with several location/scale parameters.

c. Gamma distributions.

2). Use a discrete distribution (with 5-point mass) to verify the discrete formula connecting the CDF (F) and cumulative hazard function (H).

3). Work out the "?"s and "$\cdots$"s in the handout (on first page).

4). In computing the sum of the Savage scores, we can do a few change:

(a). change the scores to centered version:

$$a_{ni} = 1 - (\frac{1}{N} + \cdots + \frac{1}{N-i+1}).$$

(b). Originally we need to sum those scores correspond to sample 1. We could sum this differently as follows: when the score a_{ni} correspond to a sample 1 obs. we sum $1 - R_1 \times (\frac{1}{N-i+1})$; when the score a_{nj} correspond to a sample 2 obs. we sum $-R_1 \times (\frac{1}{N-j+1})$. Convince yourself that the resulting sum will be the same.

5). Write the sum as $\int (dN_1(t) - R_1(t) \cdot (\frac{d[N_1(t)+N_2(t)]}{R_1(t)+R_2(t)})$

6). I have written an R code for generating random numbers that follow the 2 piece, piecewise exponential distribution. Generalize it to k pieces.

Why proportional (cumulative) hazard is a point-wise property and shift model in probability is an interval property.

To compare $H_1(t_0)$ to $H_2(t_0)$, we can immediately find, for example, $H_1(t_0) = \eta * H_2(t_0)$. Here η is a parameter. This only involve the H_1 and H_2 value at one point: t_0.

To model $F1(t0)$ shift to $F2(t0)$, if we use the model $F_1(t_0) = F_2(t_0 - \eta)$, then this not only involve the F2(t0), it involves all F2 values from $F_2(t_0)$ to $F_2(t_0 - \eta)$.

For the right censored data: $(X_1, \delta_1), (X_2, \delta_2), \cdots, (X_n, \delta_n)$ define

$$N(t) = \sum_{i=1}^{n} \delta_i I_{[x_i \leq t]}$$

This counts the number of observed failures up to time t, whereas $dN(t) = \Delta N(t) =$ number of observed failures at time t.

Define also

$$R(t) = \sum_{i=1}^{n} I_{[x_i \geq t]}$$

This counts the number of subjects alive at time t. In medical term, those 'at risk' at time t.

Nelson-Aalen estimator is

$$\hat{\Lambda}(t) = \sum_{s \leq t} \frac{dN(s)}{R(s)}$$

We also have (from the computation of information with piecewise exponential random variables)

$$Var(\hat{\Lambda}(t)) \approx \sum_{s \leq t} \frac{dN(s)}{R(s)^2}.$$

The Kaplan-Meier estimator is

$$1 - \hat{F}(t) = \prod_{s \leq t} \left(1 - \frac{dN(s)}{R(s)}\right).$$

The variance of Kaplan-Meier estimator (Greenwood formula) can be derived as follows.

$$Var(\hat{F}(t)) = Var(1 - \hat{F}(t)) \approx Var(e^{-\hat{\Lambda}(t)})$$

since for continuous CDF we have $1 - F(t) = e^{-\Lambda(t)}$. Using the delta method, we have

$$Var(e^{-\hat{\Lambda}(t)}) \approx [e^{-\hat{\Lambda}(t)}]^2 Var(\hat{\Lambda}(t))$$

Finally, we get the Greenwood formula by replace $e^{-\hat{\Lambda}(t)}$ back with $1 - \hat{F}(t)$, and variance estimator of Nelson-Aalen

$$Var(\hat{F}(t)) \approx [1 - \hat{F}(t)]^2 \sum_{s \leq t} \frac{dN(s)}{R(s)[R(s) - dN(s)]},$$

The only point is that in the denominator we use $R(R - dN)$ instead of R^2. This can be explained with the special case of binomial variance, with no censoring. (this is only a finite sample adjustment).

Next is a discussion of the 'weighted average' in Cox model.

Given k independent exponential random variables X_i with hazards $\lambda_1, \cdots, \lambda_k$.

We have

$$P(the\ failed\ one\ is\ X_i | one\ failed\ out\ of\ k\ X's) = \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j} \tag{1}$$

i.e.

$$P(X_i = t | \min_{1 \le j \le k} X_j = t) = \frac{\lambda_i}{\sum_{j=1}^{k} \lambda_j}$$

Now imaging the random variables X_i carry with them a value z_i, the covariate.

Then the conditional expectation of z value for the failed one, given there is one fail among the k is

$$\sum_{i=1}^{k} \frac{z_i \lambda_i}{\sum_{j=1}^{k} \lambda_j}$$

Or

$$\frac{\sum_{i=1}^{k} z_i \lambda_i}{\sum_{j=1}^{k} \lambda_j}.$$

If we identify λ_i with $\exp(\beta z_i)$, this is the term in the Partial likelihood.

$$\frac{\sum_{i=1}^{k} z_i \exp(\beta z_i)}{\sum_{j=1}^{k} \exp(\beta z_j)}.$$

So, at least the score function of the Cox partial likelihood has mean zero. (each term as mean zero, at true β.) Because it is in a form of 'observed' - 'expected'.

Also, using this we see the Cox partial likelihood is the product of many conditional probabilities, as in (1). So it is not a 'real' likelihood.

Notes on piecewise exponential regression: (This is a topic not well developed.)

There is an example in Allison's book. In it, he cut the time interval for each obs. T_i into pieces defined by intervals $a0 < a1 < a2 < ...$

For a genuine piecewise exponential regression, we should cut the time interval for the error variable, Z_i or $\log Z_i$, not the responses T_i.

AFT model: $T_i = exp(-\beta x_i)Z_i$

where Z_i is a standard exp, or exp(1) r.v. After taking log, we have

$$\log(T_i) = -\beta x_i + \log Z_i$$

To model $\log Z_i$ or Z_i as iid piecewise exponential we should divide the time interval for the variable Z_i to $0 = a0 < a1 < a2....$

These cut intervals in terms of the variable T_i is different:

So, for T_i we need to cut (denote $\phi_i = exp(-\beta x_i)$)

$$\phi_i a0 < \phi_i a1 < \phi_i a2 < \phi_i a3 < ...$$

If $a0 = 0$ then the first is always zero.

This (β dependent cut) could be done iteratively. [first estimate β, then refine the cut, ie. it lead to new cut, then update the β estimate with this new cut,....]

This way every subject has its own set of intervals.

This should lead to the AFT model with general error distribution, at least the estimator of beta should be comparable.

The hazard of AFT model: the hazard for T_i is seen given as

$$h_i(t) = h_0(t\phi_i)\phi_i$$

Question:

(1) understand what SAS is doing in the example.

(2) How do you implement the piecewise exp regression in R?

(3) use a cut that is different for each subject and is dependent on βx_i.

Can R function survreg() take (start, stop, event) type input? No.