Weibull Distribution and Extensions

2024-05-17

Contents

1 Introduction **1**

2 Lifetime Distributions **2**
- 2.1 Fundamental Functions . . . 2
- 2.2 Relationship Between Functions . . . 3
- 2.3 Shape Analyses of Hazard Rates . . . 3

3 Weibull Distribution **3**
- 3.1 Density Function and Reparametrization . . . 4
- 3.2 Survivor and Hazard Rate Functions . . . 5

4 **7**

1 Introduction

Let X be random variable with density function $f(x,\theta)$ (sometimes is denoted by $f_\theta(x)$), where θ is a vector of parameters. For example, a normal random variable has a vector of two parameters, denoted by $\theta = (\mu, \sigma)$, the corresponding density function is denoted by

$$f(x,\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ where } \quad -\infty < x < \infty.$$

The parameters $\sigma > 0$ and $-\infty < \mu < \infty$.

In general, for any given random variable X with probabilty density function (pdf) $f(x,\theta)$ and domain $x \in D$, the following properties are satisfied (the ranges of parameters in θ must also be specified!).

1. $f(x,\theta) \geq 0$.
2. $\int_D f(x,\theta)dx = 1$.

The cumulative distribution function (CDF) of X is defined to be

$$F(x,\theta) = \int_{-\infty}^{x} f(t,\theta)dt$$

The relationship between pdf and CDF is

$$F'(x,\theta) = f(x,\theta).$$

A distribution can be characterized its momonets. The momoent generation function of a random variable X (if exists) is given by

$$MGF_X(t,\theta) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x,\theta)dx$$

To find the k-th moment of X, we simply take the k-th derivative of the moment generating function and set $t = 0$.

$$E[X^k] = \left. \frac{d^k(MGF_X(t, \theta))}{dt^k} \right|_{t=0}$$

The k-th moment is a function of parameters of the distribution only.

2 Lifetime Distributions

A large family of distributions involve positive random variables. Because these positive random variables are commonly used in modeling lifetime, this family of distributions are called lifetime distributions. Lifetime distributions are widely used in clinical (e.g., in cancer research and related studies focusing patient survival time, the time a patient stays with a treatment, time to recovery, etc.), reliability modeling (time to failure, time to maintenance, etc.), and many other fields.

2.1 Fundamental Functions

Let X be a positive random variable with density function $f(x, \theta)$ that satisfies the following conditions

1. $f(x, \theta) \geq 0$, where θ is a vector of model parameters.
2. $\int_0^\infty f(x, \theta)dx = 1$. The actual range of X could be finite.

The CDF associated with X is defined to be

$$F(x) = \int_0^x f(t, \theta)dt.$$

Clearly, $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to 0} F(x) = 0$, and $F(x)$ is increasing.

The mean survival time (also called mean time between failure, MTBF, or mean time to faliure, MTTF) is defined to be

$$E[X] = \int_0^\infty x f(x)dx.$$

The shape of the density function has practical implications in modeling. Most of procedures and tools in statistical and process control are defined/formulated based on the shape of the density distribution.

There are well-defined and well-studied measures such as skewness and kurtosis coefficients (all based on moments) running in practice. The parametric and some non-parametric control chart and process capability in statistical process control are either explicitly or implicitly formulated based on the shape of the density function.

The survivor and hazard (also called failure rate) functions are the two primary functions used in the survival and reliability studies.

1. The survivor function is the probability of surviving beyond time x which is defined as

$$S(x) = \int_x^\infty f(t, \theta)dt.$$

2. The hazard (failure rate or hazard rate) function is defined to be

$$h(x) = \frac{f(x)}{S(x)}.$$

The notation of the vector of parameters (θ) was dropped from the above definition. $h(x)$ is an instantaneous hazard rate because the mathematical definition of the hazard is the following limit

$$h(x) = \frac{f(x)}{S(x)} = \frac{\lim_{\Delta x \to 0} \frac{P(X<x+\Delta x)-P(X<x)}{\Delta x}}{P(X>x)} = \lim_{\Delta x \to 0} \frac{P(x \le X \le X+\Delta x)/\Delta x}{P(X>x)}.$$

The following cumulative hazard rate function is also an important function used to build survival regression modeling.

$$H(x) = \int_0^x h(t)dt$$

2.2 Relationship Between Functions

Several functions have been introduced in the above sections, the following formulas explains the relationship between these functions.

1. $S(x) = 1 - F(x)$.
2. $f(x) = h(x)S(x)$.
3. $S(x) = \exp[-H(x)]$, hence, $F(x) = 1 - \exp[-H(x)]$

2.3 Shape Analyses of Hazard Rates

The shape of the hazard rate is of practical importance since it tells survival behavior. Although hazard rate function $h(x)$ could be a constant (e.g., the exponential model with density function $f(x) = \lambda \exp(-\lambda x)$ has a constant hazard rate). In most applications, hazard (failure) rates are not constant (i.e., dependent on the time). The following are common shapes that are used in real-world applications.

1. **Increasing Hazard Rates (IHR)** describe processes which are more likely to fail with age, such as machines whose parts wear out.
2. **Decreasing Hazard Rates (DHR)** describe processes that are less likely to wear out with time: a business that has lasted two centuries is less likely to go bankrupt than one that has lasted two years.
3. **Increasing and Then Decreasing Hazard Rates (IDHR)** describe processes which are more likely to fail initially and then decreasing after some time point: recovery from a surgery in clinical studies.
4. **Bathtub-Shaped Failure Rates** describe things that have relatively high failure rates when very young or very old, but flat rates in middle age (such as human beings and some machines).

Example 1. The one-parameter *Rayleigh distribution* with CDF

$$F(x;\sigma) = 1 - \exp[-x^2/(2\sigma^2]$$

has hazard rate function $h(x) = x/\sigma^2$. Clearly, the hazard rate of the above one-parameter Rayleigh distribution is an increasing function of time x.

Examples of other types of failure rate function will be introduced in the subsequent sections.

3 Weibull Distribution

The standard two-parameter Weibull distribution is one of the most studied and used distribution in practice, particularly, in the reliability analysis. There two commonly used formulations for Weibull density.

3.1 Density Function and Reparametrization

Let X be a positive random variable representing lifetime. The two formulations of the density are

Formulation 1:

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k},$$

where both k and λ are positive parameters. k is called shape parameter and λ the scale parameter.

Formulation 2:

$$f(x) = bkx^{k-1}e^{-bx^k}.$$

The CDF can be derived (show this!) as

$$F(x) = 1 - e^{-bx^k}, \quad b > 0 \text{ and } k > 0.$$

The difference between **Formulation 1** and **Formulation 2** is the form of the parameters. Clearly, replacing $b = \lambda^{-k}$. The process of changing the parameter form to get different formulations of base density function is called **reparametrization**. We can say that **formulation 1** is a **reparametrization 2**.

Reparametrization is occasionally used in the practice for two reasons:

1. Working with a simple form of density function;
2. Making the parameter practically meaningful;

There is no mathematical difference between parameters and the corresponding reparameterized parameters. However, but the statistical inferences on the two form of parameters significantly different. This will be discussed later.

Example: Consider exponential distribution

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

The λ is called the **rate parameter**. However, the above distribution can reparametrized to get the following density

$$f(x) = \frac{1}{\gamma}e^{-\frac{x}{\gamma}}.$$

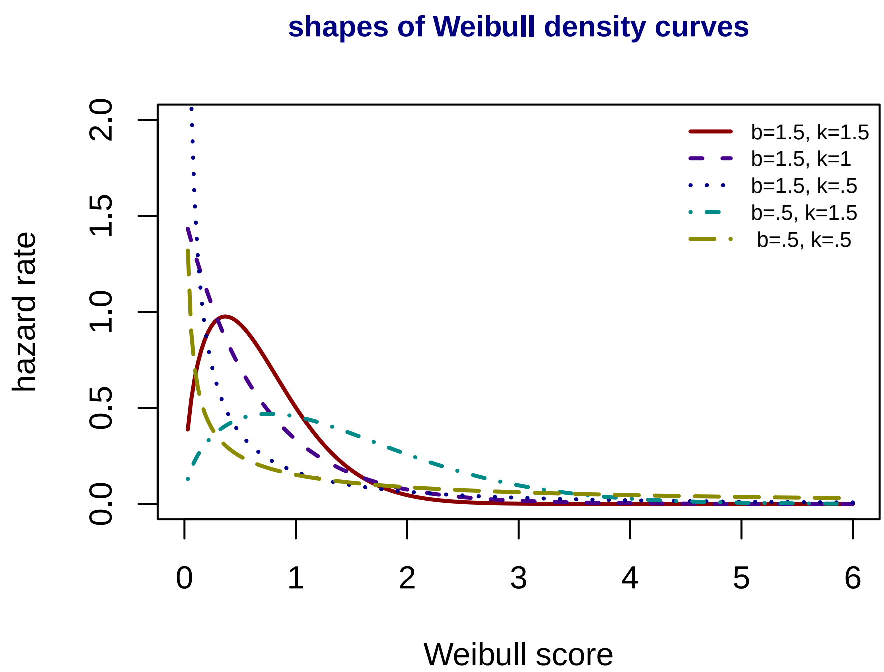
γ is the scale parameter in the above density function.

As mentioned earlier, we are interested in the shape of the density curves in many practical applications.

```
## write a function to evaluate h(x) for given x and values of parameters
fval=function(b, k, x){
   b*k*x^(k-1)*exp(-b*x^k)
}

## pre-selected x-values
x = seq(0, 6, length = 200)[-1]
##
## Evaluate y with different conbinations of parameters
y1 = fval(b=1.5, k=1.5, x)
y2 = fval(b=1.5, k=1, x)
```

```r
y3 = fval(b=1.5, k=.5, x)
##
y4 = fval(b=.5, k=1.5, x)
y5 = fval(b=.5, k=.5, x)
## range of y
ymax=max(c(y1, y2, y3, y4, y5))
##
plot(x, y1, type="l", ylim=c(0,2), xlim=c(0,6),
     xlab = "Weibull score",
     ylab = "hazard rate",
     col= "#8B0000",
     lwd=2,
     main = "shapes of Weibull density curves",
     cex.main = 0.9,     # font size of color
     col.main = "navy")
lines(x, y2, col="#46008B", lwd=2,lty=2)
lines(x, y3, col="#00008B", lwd=2,lty=3)
lines(x, y4, col="#008B8B", lwd=2,lty=4)
lines(x, y5, col="#8B8B00", lwd=2,lty=5)
## add legend to the plot
legend("topright", c("b=1.5, k=1.5", "b=1.5, k=1", "b=1.5, k=.5", "b=.5, k=1.5"," b=.5, k=.5"),
        col=c("#8B0000","#46008B","#00008B","#008B8B","#8B8B00"),
        lwd=rep(2,5), lty=1:5, cex = 0.7, bty="n")
```



3.2 Survivor and Hazard Rate Functions

For convenience, we use **Formulation 2** to define survivor and hazard rate functions in the following.

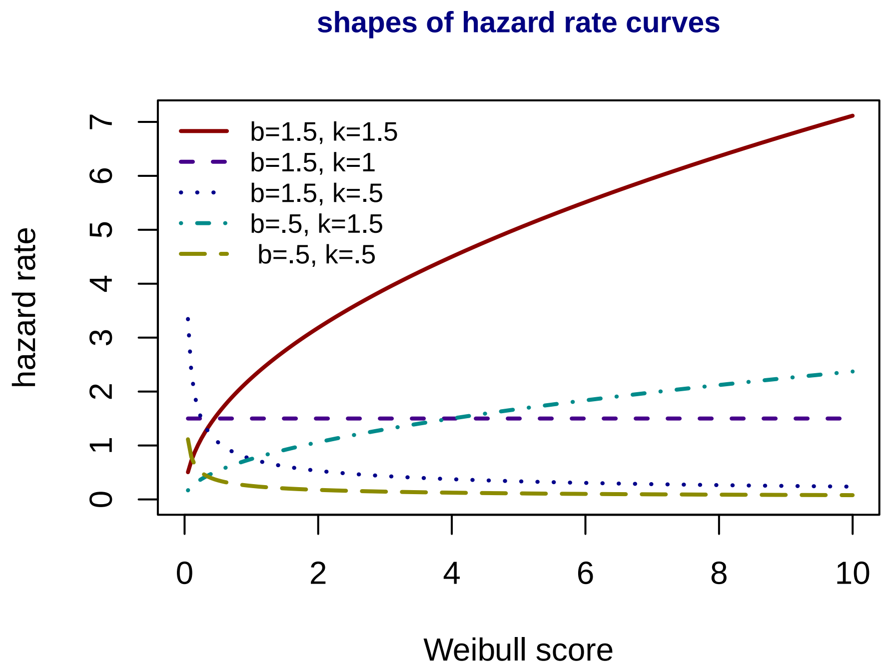
$$S(x) = 1 - F(x) = e^{-bx^k}$$

$$h(x) = \frac{f(x)}{S(x)} = \frac{bkx^{k-1}e^{-bx^k}}{e^{-bx^k}} = bkx^{k-1}, \text{ for } \; b > 0 \text{ and } k > 0.$$

Next, we use R to draw curve the hazard rate function with different values of the parameters.

```
## write a function to evaluate h(x) for given x and values of parameters
hval=function(b, k, x){
   b*k*x^(k-1)
}

## pre-selected x-values
x = seq(0, 10, length = 200)[-1]
##
## Evaluate y with different conbinations of parameters
y1 = hval(b=1.5, k=1.5, x)
y2 = hval(b=1.5, k=1, x)
y3 = hval(b=1.5, k=.5, x)
##
y4 = hval(b=.5, k=1.5, x)
y5 = hval(b=.5, k=.5, x)
## range of y
ymax=max(c(y1, y2, y3, y4, y5))
##
plot(x, y1, type="l", ylim=c(0,ymax), xlim=c(0,10),
     xlab = "Weibull score",
     ylab = "hazard rate",
     col= "#8B0000",
     lwd=2,
     main = "shapes of hazard rate curves",
     cex.main = 0.9,    # font size of color
     col.main = "navy")
lines(x, y2, col="#46008B", lwd=2,lty=2)
lines(x, y3, col="#00008B", lwd=2,lty=3)
lines(x, y4, col="#008B8B", lwd=2,lty=4)
lines(x, y5, col="#8B8B00", lwd=2,lty=5)
## add legend to the plot
legend("topleft", c("b=1.5, k=1.5", "b=1.5, k=1", "b=1.5, k=.5", "b=.5, k=1.5"," b=.5, k=.5"),
       col=c("#8B0000","#46008B","#00008B","#008B8B","#8B8B00"),
       lwd=rep(2,5), lty=1:5, cex = 0.8, bty="n")
```



The above shows that the shape of hazard curves are controlled by the shape parameter. when $k = 1$, the Weibull distribution is reduced to the one parameter exponential distribution which has a constant hazard rate. If $k > 1$, the hazard rate function is increasing. The hazard rate is decreasing if $k < 1$. The above pattern can be sunnarized in the following theorem.

Theorem 1. The hazard rate function of Weibull distribution $f(x) = bkx^{k-1}e^{-bx^k}$ is monotonic.

Proof. *Note that the hazard rate function of Weibull distribution and its derivative have the following explicit forms*

$$h(x) = bkx^{k-1} \quad \text{and} \quad h'(x) = bk(k-1)x^{k-2}.$$

Therefore,

$$\begin{cases} h'(x) > 0 & \text{if} \quad k > 1; \\ h'(x) = 0 & \text{if} \quad k = 1; \\ h'(x) < 0 & \text{if} \quad 0 < k < 1. \end{cases}$$

This proves the monotone hazard rate of Weibull distribution.

The above theorem show that the Weibull distribution has only monotonic hazard rate. In the real world applications, non-monotonic hazard rates are also common. The standard 2-parameter Weibull cannot be used to model distributions with non-monotonic hazard rates.

We need to define new distribution, or transform the random variables, or expand the existing family of distributions to capture pattern of the non-monotonic hazard rates.

4