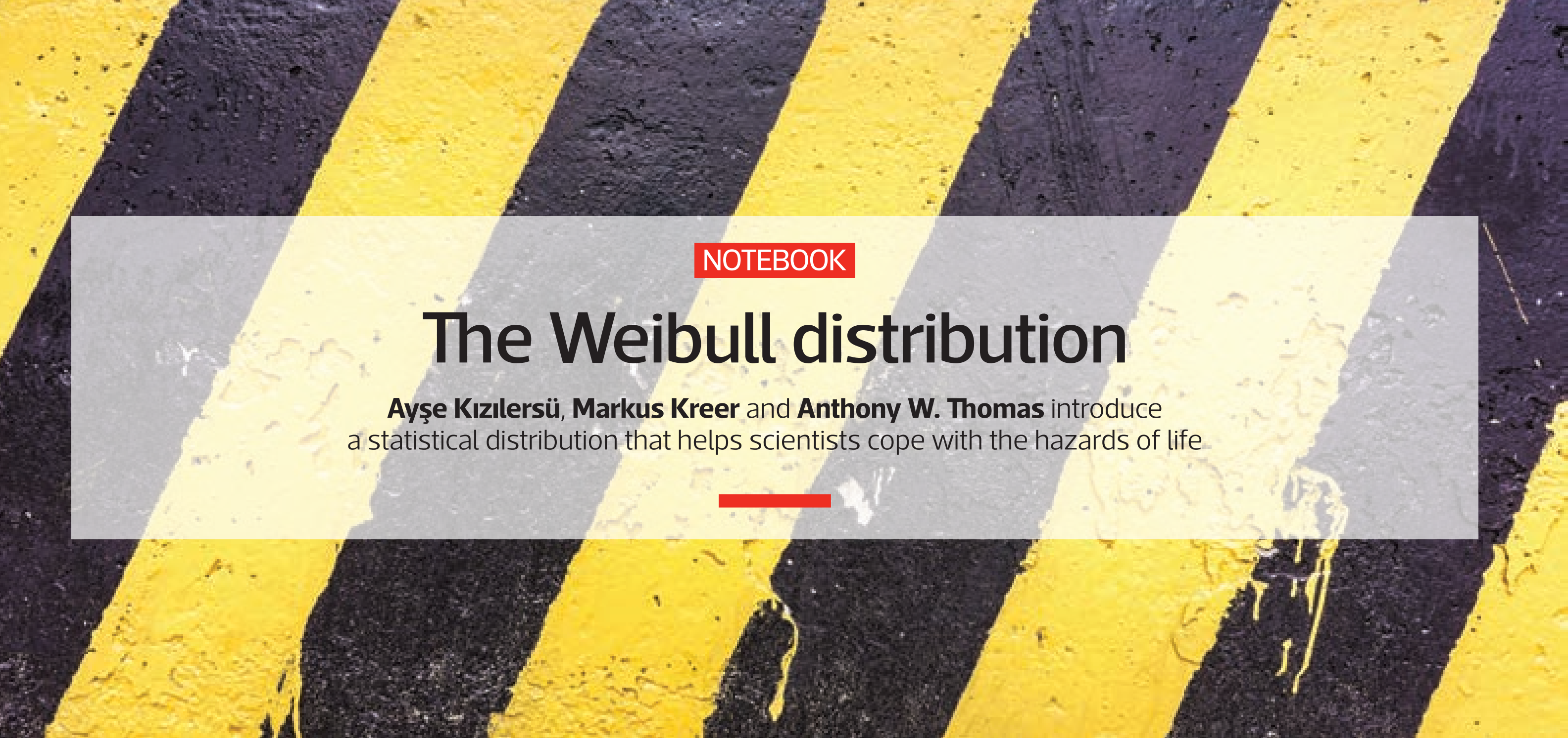

NOTEBOOK

The Weibull distribution

Ayşe Kızılersü, **Markus Kreer** and **Anthony W. Thomas** introduce a statistical distribution that helps scientists cope with the hazards of life

What is the Weibull distribution?

Life is hazardous. There are earthquakes, divorce, cancer and stock-market crashes to contend with. If you want to understand the statistics of events like these then you need to understand the Weibull distribution.[1]

The properties of the Weibull distribution are best described in terms of the hazard function (see box). This tells us how likely something is to fail given that it has survived so far. A simple power law is used for the hazard function, which accommodates three distinct behaviours: (1) if something is going to fail it will most likely fail at the start; (2) the rate of failure is fairly constant; (3) failure becomes more likely as time goes on.

Imagine, for example, that your car has been running for 10 years without any problems. Should you sell it, because it is likely to fail any minute? Or should you keep it, because if it has not yet failed it will probably outlive you? To make this decision you need to know the hazard function of your car.

Weibull distributions are characterised by a scale parameter, α, and a shape parameter, β (see Figure 1). There are also three-parameter Weibull distributions, which allow for different starting points; however, these are not elaborated on here.

The key to understanding the behaviour of Weibull distributions is the shape parameter, and the hazard function behaviours discussed above are also characterised by the shape parameter. There are three different regimes distinguished by their shapes:

1. For $\beta < 1$, the corresponding probability density tends to infinity as time (t) approaches zero. For example, this behaviour describes the waiting time between two subsequent stock exchange transactions of the same stock ($\beta \approx 0.54$).
2. When $\beta = 1$, we recover the well-known exponential distribution, which is finite at the starting point. One of its most recognised applications is to the radioactive decay of unstable atoms.
3. Finally, for $\beta > 1$, we find a distribution with a hump, like the bell-shaped curve of the normal distribution, except that it is asymmetric. These distributions describe phenomena as diverse as wind speeds ($\beta \approx 1.67$), the duration of ethnically mixed marriages ($\beta \approx 1.2$), and the distribution of the size of water droplets ($\beta \approx 4.5$).

It is helpful to visualise the differences between values of β using the "bathtub" diagram shown in Figure 2.

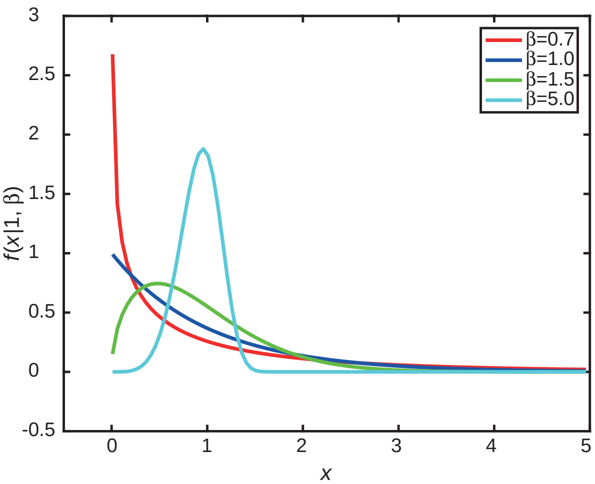

FIGURE 1 Weibull distributions for various choices of shape parameter β (= 0.7, 1.0, 1.5, 5.0) for α = 1. There are three basic shape types: β < 1, β = 1 and β > 1. When β ≤ 1, the distribution has no turning point. When β < 1, the function is unbounded. Two cases are shown where β > 1; as β increases and develops a single peak, the function looks more like a normal distribution.

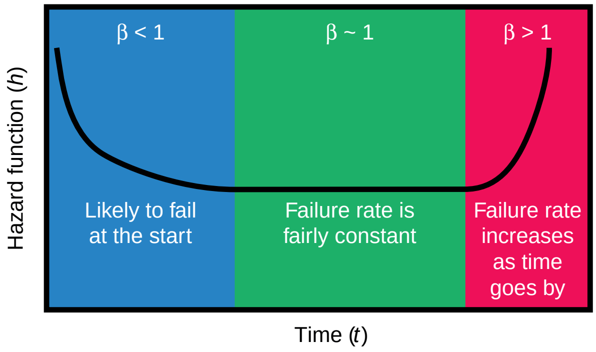

FIGURE 2 "Bathtub" diagram showing time evolution of the Weibull hazard function for different values of the shape parameter β.

Who discovered it?

The Weibull distribution is named after the Swedish engineer and scientist Ernst Hjalmar Waloddi Weibull (1887–1979). He did not discover it but instead popularised the distribution in his 1951 paper to the American Society of Mechanical Engineers.[2]

The distribution had previously been studied by French mathematician Maurice Fréchet in the context of extreme value distributions,[3] and was used many years earlier by Rosin and Rammler to model the grain size distribution of ground coal.[4] For this reason, the Weibull distribution is sometimes called the Rosin–Rammler distribution.

When should it be used?

The Weibull distribution is widely used in engineering, medicine, energy, the social sciences, finance, insurance, and elsewhere. With $\beta < 1$, it is particularly well suited to time series data with "heavy tails", where values far from the maximum probability are still fairly common.

As an extreme value distribution, the Weibull distribution has proven quite successful in predicting the occurrence of extreme phenomena like floods, earthquakes, high wind speeds and torrential rains.

Also, because the Weibull distribution is derived from the assumption of a monomial hazard function, it is very good at describing survival statistics, such as survival times after a diagnosis of cancer, light bulb failure times and divorce rates, among other things.

More recently, the distribution has been shown to accurately describe the behaviour of high-frequency trading.[5]

When should it not be used?

Random variables that can have negative values cannot be described by the Weibull distribution, therefore it cannot be used to describe the variation around a mean. Furthermore, in this distribution the mean and median are different, meaning it cannot be used to describe processes where they are the same.

The Weibull distribution is asymmetric, so that the probability of events occurring before the mode is not the same as after. Phenomena with "super-heavy tails", such as random variables described by Pareto or log-normal probability distributions, cannot be suitably described by the Weibull distribution. Examples include life times affected by chemical corrosion, and the distribution of personal income.

Keep in mind...

To use the Weibull distribution requires knowledge of the scale and shape parameters, α and β, and there is a substantial literature on how to estimate these parameters from a given data set. A rigorous mathematical approach uses maximum likelihood estimation to estimate the parameters and then performs a goodness-of-fit test, such as the Kolmogorov–Smirnov (KS) test, to assess validity.[6] In practice, most real-life problems involve truncated data for which the standard KS critical values are not appropriate. To address this issue, recent studies have explored the behaviour of critical values of the KS statistic as a function of truncation when the data is incomplete.[6] ■

References

1. Rinne, H. (2009) *The Weibull Distribution: A Handbook*. Boca Raton, FL: CRC Press.

2. Weibull, W. (1951) A statistical distribution function of wide applicability. *Journal of Applied Mechanics, Transactions of the American Society of Mechanical Engineers*, **18**(3), 293–297.

3. Fréchet, M. (1927) Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathematique, Cracovie*, **6**, 93–116.

4. Rosin, P. and Rammler, E. (1933) The laws governing the fineness of powdered coal. *Journal of the Institute of Fuel*, **7**, 29–36.

5. Kızılersü, A., Kreer, M., Thomas, A. W. and Feindt, M. (2016) Universal behaviour in the stock market: Time dynamics of the electronic orderbook. *Physics Letters A*, **380**, 2501–2512.

6. Kızılersü, A., Kreer, M. and Thomas, A. W. (2016) Goodess-of-fit testing for left-truncated two-parameter Weibull distributions with known truncation point. *Austrian Journal of Statistics*, **45**, 15–42.

The Weibull hazard function

At its core, the Weibull distribution is defined by a simple hazard function. The hazard function, $h(\cdot)$, is the conditional density given that the event we are concerned about has not yet occurred. Consider the probability that a light bulb will fail at some time between t and $t + dt$ hours of operation. This is the probability that it has survived until time t and it will fail in the next dt hours, and may be expressed as

$$f(t)dt = S(t)h(t)dt$$

where $f(t)$ is the probability density function for failure at time t, $S(t)$ is the survival function describing the probability that the light bulb has not failed up to time t, and $h(t)$ is the hazard function.

The cumulative distribution function (cdf) is the integral of the probability density function (pdf) and in this case represents the probability that the light bulb failed before some time t,

$$F(t) = \int_1^t f(s)ds$$

The survival function is the probability that the light bulb has survived until time t, which is therefore

$$S(t) = 1 - F(t)$$

From these three equations, we determine that the hazard function is the negative rate of change of the log of the survival function,

$$h(t) = -\frac{d \log S(t)}{dt}$$

Given the hazard function, we can integrate it to find the survival function, from which we can obtain the cdf, whose derivative is the pdf.

It turns out that the hazard function for light bulbs, earthquakes, etc. can be described by the monomial function

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$$

This defines the Weibull distribution with corresponding cdf and pdf given by

$$F(t) = 1 - e^{-(t/\alpha)^{\beta}}$$

$$f(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-(t/\alpha)^{\beta}}$$



Ayşe Kızılersü is a senior research scientist in the Centre for Subatomic Structure of Matter in the physics department at the University of Adelaide, Australia. **Markus Kreer** has an MSc in theoretical physics and a PhD in mathematics. He teaches and undertakes research at Campuservice of Frankfurt University. **Anthony W. Thomas** is the Elder Professor of Physics at the University of Adelaide and an Australian Research Council Australian Laureate Fellow (2009-2014).