📏

What is the Weibull distribution?

▼ You can also see this file by checking out the link below.

👉 What is the Weibull distribution?

What is the weibull distribution?

- a kind of parametric estimation methods for analysis of life data
    - There are the exponential distribution, the weibull distribution, etc.
- When the failure rate is ambigous(constant/increase/decline), f(t) is affected by the Weibull distribution

cf) What is the exponential distribution?

- failure rate is constant regardless of change of time beacuse f(t) is a constant.
- mean time : expected time
    - The mean time is devided by repairable and non-repairable cases.
        - The mean time of repairable case : MTBF
        - The mean time of non-repairable case : MTTF

Why do you use the Weibull distribution?

- In the exponential distribution, failure rate is constant regardless of the change in time beacuse f(t) is a constant.

  → Therefore, You can check the change of failure rate by the change in time using the weibull distribution.

- The weibull distribution is used more universally, and It is known as a reasonable method.

Parameters of the Weibull distribution

- The 2-parameter Weibull distribution has a scale and shape parameter. The 3-parameter Weibull includes a location parameter.
- β(beta)

  is the shape parameter, also known as the Weibull slope

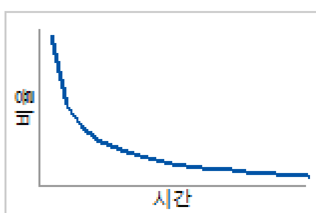
  - A shape of the Weibull distribution is totally decided by β
- η(eta or alpha)

  is the scale parameter
- γ

  is the location parameter

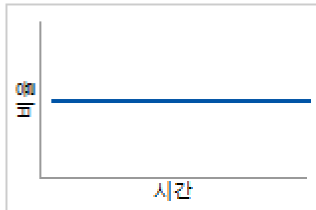
Probability density function according to β of the Weibull distribution

- **0 < ß < 1 : Early failure**
  - High failure rate in the early stages decreases as time goes on.



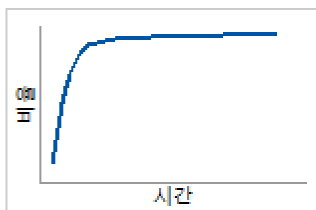
  - Modeling the early life of the product. : "burn-in" period
  - Early failures occur in initial period of product life. These failures may necessitate a product “burn-in” period to reduce risk of initial failure.

- **β = 1 : Random Failure**
    - The failure rate remains constant.

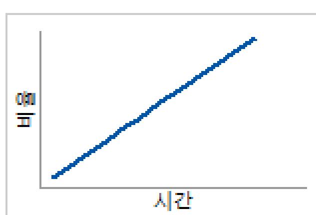


-
    - Random failures, multiple-cause failures. Models “useful life” of product.
    - There may be several causes of failure.

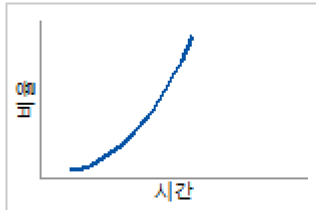
- **ß = 1.5 : Early wear-out failure**
    - Increasing failure rate, with largest increase initially



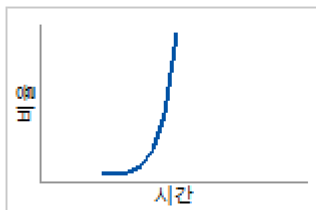
- **ß = 2**
    - Risk of wear-out failure increases steadily during the life of the product.
    - Linearly increasing failure rate.



- **3 ≤ ß ≤4 : Fast wear-out failures**
  - Models the final period of product life, when most failures occur.



- **ß > 10 : Very fast wear-out failures**
  - Models the final period of product life, when most failures occur.



  - refer to : https://support.minitab.com/ko-kr/minitab/20/help-and-how-to/statistical-modeling/reliability/supporting-topics/distribution-models/weibull-distribution/

Formulas related to the Weibull distribution

- **β(beta) / η(eta or alpha)**
  - Those can be derived from computer programs.
  - The Python code for estimation of alpha, beta using Reliability Library.(2-parameter)

    ▼ code

    ```
    ## https://github.com/MatthewReid854/reliability

    x, y = plotting_positions(failures=list_fail, right_censored=list_censored)
    ```

```
x = np.array(x)
y = np.array(y)

def linear_regression(...) # skip -> you can see this code at github.

xlin = np.log(x)
ylin = np.log(-np.log(1 - y))
slope, intercept = linear_regression(
    xlin, ylin
)
LS_beta = slope
LS_alpha = np.exp(-intercept / LS_beta)
guess = [LS_alpha, LS_beta]

print(guess)
```

- **MTTF**
    - $MTTF = \eta \cdot \Gamma(1/\beta + 1)$
- **Reliability / Probability of Failure**
    - Probability of Failure
        - F(t) is the cumulative probability of failure from time zero till time t.

$$F(t) = 1 - \eta^{-(t/n)^{\beta}}$$

    - Reliability
        - R(t) is the chance of survival from from time zero till time t.

$$R(t) = \eta^{-(t/n)^{\beta}}$$

- **Bearing Life**
    - The “BX” or “Bearing Life” nomenclature, which refers to the time at which X% of items in a population will fail.

$$B(1-P)\,Life = \eta[-\ln(1-\mathrm{P})]^{1/\beta}$$

Estimation of life data parameters

- Least Square Estimator: LSE
  - If the data format is complete, there isn't bias in theory. However, There is can be Bias by cumulative failure rate.
  - It is not possible to estimate variance directly. Thus, Variance is estimated highly.
  - It is more accurate than MLE when the shape of distribution is determined.
  - It is possible to estimate parameters without using computer programs
- Maximum Likelihood Estimator: MLE
  - As the sample size increases, the bias decreases.
  - The variance is estimated to be small.
  - In the case of estimation of parameters, it has a better estimates.
  - If you do not use computer programs, the estmation is difficult.
- etc.

Goodness of Fit

- **Evaluate the Anderson-Darling goodness-of-fit statistic and the Pearson correlation coefficient**
  - Substantially lower values of Anderson-Darling generally indicate a better fitting distribution. The Anderson-Darling statistic is calculated for both the maximum likelihood estimation method (MLE) and the least squares estimation method (LSE).

- Substantially higher values of the Pearson correlation coefficient identify a better fitting distribution. The correlation coefficient is available for the LSE method.

- **Log-likelihood**
  - The **log-likelihood value** of a regression model is a way to measure the goodness of fit for a model.
  - The higher the value of the log-likelihood, the better a model fits a dataset.

```
#calculate log-likelihood value of each model
logLik(model1)

'log Lik.' -91.04219 (df=3)

logLik(model2)

'log Lik.' -111.7511 (df=3)
```

→ The first model has a higher log-likelihood value (**-91.04**) than the second model (**-111.75**), which means the first model offers a better fit to the data.

- **AIC**
  - The Akaike information criterion (AIC)

$$AIC = -2\ LogLikelihood + 2p$$

  - AIC is calcuated from the maximum likelihood estimate of the model (how well the model reproduces the data).
  - **lower is better.**
  - The more variables and unnecessary paramters there are, AIC is increases.
  - AIC increases as 2k increases → It is bad model.
- **BIC**
  - **Bayesian information criterion** (**BIC**)

$$BIC = -2\ LogLikelihood + \log(n)p$$

- **lower is better.**
- BIC is similar to AIC. it improves AIC by modifying the last variable.
- In the case of BIC, the more variables there are, the more penalties will be added than AIC.
- Therefore, the meaning of minimizing AIC and BIC means the optimal model (parsimonious & explainable) with the largest likelihood and the smallest number of variables.
    - refer to : https://rk1993.tistory.com/entry/AIC-BIC-Mallows-Cp-쉽게-이해하기

Estimation of Parameters using R

```
library("fitdistrplus")

data <- read.csv("C:\\~~~")

parameters <- fitdist(data$uptime, 'weibull', 'mle')$estimate

parameters
```

Estimation of Parameters using Python

👉 Link