Journal of Environmental Statistics
August 2014, Volume 6, Issue 4. http://www.jenvstat.org

# Characterization Theorems for Weibull Distribution with Applications

**Ratan Dasgupta**
Indian Statistical Institute

**Abstract**

We prove characterization theorems for Weibull distributions based on invariance of hazard rate under scale transformations in a countable dense set near origin. Similar characterizations are also obtained for two different types of discrete Weibull distributions, when logarithm of survival function / hazard rate are scale invariant on the set of non negative integers. Thus scale invariance of survival function / hazard rate of a variable on a small domain is equivalent to Weibull distribution. This assumption on reliability function (or hazard rate) when satisfied, leads to an appropriate model selection. Modified Weibull Distribution (MWD) with bathtub hazard rate are characterised and its discrete versions are also discussed. Survival function and hazard rate of yam plant lifetime, related to harvest scenario are examined under Weibull model in forecasting market supply of the crop and in analysing time lag in supply. The proposed analysis may be adopted for similar situations in production and marketing of other products. Observed growth curve of yam plant lifetime based on field experiment data has a good match with simulated growth curves.

Keywords: Weibull model, hazard rate, Cauchy equation, Elephant foot yam.

## 1. Introduction

Weibull distribution is extensively used in survival analysis, reliability, extreme value theory, weather forecasting, general insurance claims and inventory control among others. Weibull density function with appropriate choice of parameters gives rise to wide coverage of possibilities to explain many applied problems e.g., see Joh, Kim, and Malaiya (2008), Johnson, Kotz, and Balakrishnan (1994), Murthy, Xie, and Jiang (2004), Qin, she Zhang, and dong Yan (2012), Wang, Hong-we, and Xi-chao (2012), Weibull (1951), Zhu, Xia, Yu, Adnan, Liu, and Du (2011). Most reliability data are modelled using distributions such as exponential, Weibull, gamma, and lognormal. Weibull distribution is easy to

interpret and is extremely versatile. By adjusting the value of its shape parameter, one can model the characteristics of many different lifetime distributions. A modified Weibull model is used in Lai, Xie, and Murthy (2003) to explain bathtub-shaped hazard rate function. We characterise a class of such distributions in terms of instantaneous failure per cumulative hazard. Discrete versions of such variables are also discussed. With limited failure data Enkhmunkh, Kim, Hwang, and Hyun (2007) proposed Weibull parameter estimation. A well known property of Weibull distribution is that its hazard rate remains invariant under a change of scale. This property has applications in modelling lifetime distributions, especially in industrial context, where characteristics measured are nonnegative and a change of scale may cause proportionate change in hazard rate. Characterization of distribution in terms of invariance of hazard rate leads to an appropriate model selection as Weibull.

Modelling plant lifetime via Weibull distribution is of interest in analysing agricultural yield. As for example, plant lifetime of Elephant foot yam is seen to be approximately Weibull via probability plots. Under Weibull model, scale change in lifetime measured from sprouting to harvest, in terms of days / hours results in proportionate change in hazard rate i.e., proportionate change in rate of crop harvest. Such lifetime modelling of crops has applications in forecasting market-supply that regulates price level. Discrete versions of Weibull distribution are of interest while dealing with discrete time points. There are several extensions of Weibull distribution to discrete cases, see e.g., Englehardt and Li (2011), Nakagawa and Osaki (1975), Ali Khan, Khalique, and Abouammoh (1989), Stein and Dattero (1984); having applications in analysis of failure data mainly measured in discrete time. Measurement of variables are often made in discrete scale, conditions required for modelling discrete data is therefore of interest. Relevance of the characterisation results is apparent from the fact that survival function / hazard rate of appropriate form on a small domain is equivalent to Weibull model selection. Dasgupta (2013) proved a characterization of Pareto variable based on quantiles when conditional distribution above a threshold is considered.

In section 2 we prove characterization theorems for Weibull distribution based on properties of hazard rate in a countable dense set. The results are related to solution of Cauchy equation. Distributions with bathtub hazard are characterised in terms of instantaneous failure rate per cumulative hazard, discrete versions of such modified Weibull variables are discussed in section 3.

Interpretation of discrete Weibull variable as a power of a geometric variable is made in Nakagawa and Osaki (1975), see Remark 2 therein; however there is an error. To this end, a correction is suggested in Remark 2 of the present paper. Properties of discrete Weibull distribution are discussed in comparison with continuous Weibull distribution. In section 4 characterization results for discrete Weibull distributions of two types viz., (i) distribution based on invariance of logarithm of survival function, and (ii) distribution based on invariance of hazard rate; under change of scale are proved. These are by virtue of Erdös (1946) theorem on arithmetical functions. Survival function / hazard rate are examined for yam plant lifetime while forecasting market supply of the crop under Weibull model. Simulated growth curves of plant lifetime when compared with observed growth curve based on data obtained from field experiments on Elephant foot yam, show similarity. In section 5 experimental data on plant lifetime of Elephant foot yam is modelled by Weibull distribution. Terminologies used in industrial context are adopted for broader applications. Hazard rate of plant lifetime, which is equivalent to rate of crop-maturity, is estimated. This is directly related to the rate of crop harvest and market supply, when multiplied by the number of plantations. Section 6 discusses the relevance of discrete Weibull distribution for Yam lifetime and subsequent

crop arrival time to consumers. In the appendix we provide yam lifetime data from several conducted experiments in Giridih farm.
The overall idea of the paper is to focus for model selection in a particular class viz. Weibull class, under the assumption of invariance in a small region; discrete or connected. Mathematical techniques are suitably tailored, so as to obtain the results; with applications to market forecasting. The adopted technique provides insight into the intrinsic properties of Weibull and other variations of it; those are commonly used in practice.

## 2. Characterization of Weibull distribution

Let $X$ be a random variable with support $(0,\infty)$ having continuous density function $f$ and distribution function $F(x) = 1 - e^{-\int_0^x h(y)dy}$, where $h(y) = \frac{f(y)}{1-F(y)} = \frac{f(y)}{\bar{F}(y)}$ is the hazard rate. We prove the following.

**Theorem 1.** Let the hazard rate $h$ satisfies $h(cx) \propto h(x), x = c^m, m \in N^+$, the set of positive integers and $(0 <)c \in A_0$, a countable dense neighborhood of origin, whose upper point exceeds 1 (e.g., $c \in A_0 = (0,\delta) \cap Q, \delta > 1$, and $Q$ is the set of rational numbers). The variable $X$ is Weibull iff the above holds.

**Proof of the theorem.** One way implication is clear, we prove the 'only if' part. Write $h(c) = bh(1)$, where $b$ is the constant of proportionality, $h(c^2) = bh(c) = b^2h(1), h(c^3) = bh(c^2) = b^3h(1)$ and for integers $m$, $h(c^m) = b^m h(1)$. Write, $c^m = x$, i.e., $m = \log x/\log c$, then $h(x) = b^{\log x/\log c}h(1) = e^{\alpha \log x}h(1) = ax^\alpha,\ \alpha = \log b/\log c,\ a = h(1)$.
This specifies the distribution function $F$ to be Weibull in a dense set $x = c, c^2, \cdots, c^m, \cdots$ of $(0,\infty)$. For an arbitrary real number $z > 0$, there exist integer $m$ and $c \in A_0$, a dense set in $(0,\delta)$; e.g., $c \in Q \cap (0,\delta), \delta > 1$; such that $c^m$ is arbitrary close to the number $z$. Next from continuity of $f$, the form of $F$ is Weibull at $z$, where $z > 0$ is arbitrary.

Condition on invariance of the function $h$, i.e., $h(cx) \propto h(x)$ may be written in an alternative form. We have the following.

**Theorem 2.** For a positive random variable $X$ with continuous density $f$, let the hazard rate $h$ satisfies $h(cx) \propto h(x);\ c, x \in Q^+$, the set of positive rational numbers. Then $h(xy)h(1) = h(x)h(y);\ x, y \in Q^+$. Consequently $X$ is a Weibull variable.

**Proof.** From symmetry in product term, relation $h(cx) \propto h(x)$ is equivalent to $h(xy) = g(x)g(y)$, for some $g$; $x, y \in Q^+$. Without loss of generality assume that $h(1) \neq 0$, as otherwise $h \equiv 0$. The condition then reduces to $h(xy)h(1) = h(x)h(y)$. This under an appropriate scaling $h(1) = 1$ may be written as a multiplicative relation $h(xy) = h(x)h(y)$, where $h$ is continuous, as $f$ is so.
Note that $h(1/x) = 1/h(x)$. For a positive integer $m$, write $h(2^m) = h^2(2^{m-1}) = \cdots = h^m(2)$. Now $h(2) = h^n(2^{1/n})$, as $h$ is continuous. Thus for a rational number of the form $m/n$, one may write $h(2^{m/n}) = h^m(2^{1/n}) = h^{m/n}(2)$. Since the rational numbers are dense in $(0,\infty)$, hazard rate $h$ without the restriction $h(1) = 1$, is of the form $h(x) = ax^\alpha$, leading to Weibull

distribution.

**Remark 1.** Solution of $h(xy)h(1) = h(x)h(y)$ may also be obtained as follows. Write $g(x) = h(e^x)$, then $h(e^{\log x+\log y}) = h(e^{\log x})h(e^{\log y})$, for $h(1) = 1$. That is $g(\log x + \log y) = g(\log x)g(\log y)$. Now write $\psi(x) = \log g(x)$, then $e^{\psi(u+v)} = e^{\psi(u)+\psi(v)}$, i.e., $\psi(u + v) = \psi(u)+\psi(v)$, where $u = \log x, v = \log y$. Cauchy functional equation under continuity assumption admits linear solution, $\psi(u) = ku$. This states $h(e^x) = g(x) = e^{\psi(x)} = e^{kx}$, providing the desired result as $h(y) = y^k$, when $h(1) = 1$. Thus the general solution is $h(y) = h(1)y^k, y > 0$; as obtained earlier.

## 3. Characterisation of MWD and bathtub hazard rate

For Weibull distribution both the hazard rate $h(x)$ and survival function $\overline{F}(x)$ are monotone. Cox (1972) proposed scaling of $h$ by an unspecified baseline hazard in regression model. Scaling of $h$ may also be done by cumulative hazard. Observe that the *negative ratio of hazard rate and logarithm of survival function* i.e., $z(x) = -h(x)/\log \overline{F}(x)$ that is instantaneous failure rate with respect to logarithm of proportion of surviving elements till a time point, characterizes the form of distribution.
To see this, write

$$\frac{f(x)}{\overline{F}(x)\log \overline{F}(x)} = -z(x) \Rightarrow \log(-\log \overline{F}(x)) = \int z(x)dx \tag{1}$$

The above specifies $F$ in terms of $z$. The expression $-\log \overline{F}(x)$ is same as cumulative hazard upto $x$. Thus $z$ is a measure of instantaneous failure $h(x)$ per cumulative hazard. The function $h$, when scaled by logarithm of proportion of surviving elements, provides the index $z$; reflecting rate of future availability at current rate of loss. For Weibull density $f(x) = \alpha x^{\alpha-1}e^{-x^\alpha}$, with monotone hazard rate $h(x) = \alpha x^{\alpha-1}$, one has $z(x) = \alpha/x$; a decreasing function of time. The case of constant failure per log survival arises when $z \equiv c(> 0)$, and this refers to the distribution

$$F(x) = 1 - \exp(-e^{cx+a}), x \in (-\infty, \infty) \tag{2}$$

where $a$ appears from constant of integration in (1). This is extreme value distribution of Type I. Distribution (2) with $c = 1$ may be termed as reference baseline distribution with respect to which the indices $z(x)$ are evaluated.
A particular choice $z(x) = b/x + \lambda$ leads to the following modified Weibull distribution (MWD).

$$F(x) = 1 - \exp(-ax^b e^{\lambda x}), a > 0, \lambda > 0, b \geq 0 \tag{3}$$

This distribution has bathtub hazard rate for $0 < b < 1$, see Lai *et al.* (2003).
Although the hazard function $h(x) = a(b+\lambda x)x^{b-1}e^{\lambda x}$ for the above distribution may not be monotone, the function $z(x) = b/x + \lambda$, from which $F$ is derived; is monotone.
The unique minimum of the hazard rate $h$ of $F$ in (3) occurs at

$$x^* = b^{1/2}(1 - b^{1/2})/\lambda \ \leq \frac{1}{4\lambda}, b \in (0, 1), \lambda > 0 \tag{4}$$

The bound of $x^*$ is attained when $b = 1/4$. Model $F$ in (3) may be termed as flexible, since the possible range of time $x$ where minimum in hazard rate may occur is $(0, \infty)$.

**Remark 2.** Geometric distribution is a discrete version of exponential distribution. Similarly, if $U$ is Weibull with survival function $\overline{F}(u) = \exp(-\lambda u^{\alpha})$,
$\lambda > 0, \alpha > 0$; then the integer part $V = [U] \in \mathbf{N_0}$ is a discrete Weibull variable.

$$P(V \geq k) = P(U \geq k) = \exp(-\lambda k^{\alpha}) = q^{k^{\alpha}}, q = \exp(-\lambda); k \in \mathbf{N_0}. \tag{5}$$

In Remark 2 of Nakagawa and Osaki (1975), discrete Weibull variable $Y$ is interpreted as a (possibly fractional) power of a geometric variable $X$ with parameter $q$. However, there is an error, as $Y \equiv X^{1/\beta}$ may not be an integer, unlike a discrete Weibull variable $V$; see equation (4) of Nakagawa and Osaki (1975). The variable $Y = X^{1/\beta}$ is discrete, but not necessarily integer valued, distribution function of $Y$ considered at integers behaves like a discrete Weibull variable $V$. The jump points in cumulative distribution function of $Y$ where non zero probability mass are associated may not even be rational numbers. The problem cannot be resolved even by considering $M = \lfloor X^{1/\beta} \rfloor$, the floor i.e., integer part; or $N = \lceil X^{1/\beta} \rceil$, the ceiling of $X^{1/\beta}$. However, the following bound holds.

$$P(M \geq k) \leq P(X \geq k^{\beta}) = (q)^{k^{\beta}} \leq P(N \geq k). \tag{6}$$

Discrete Weibull variable $V$ of (5) preserves the form of the survival function as that of continuous type and also maintains pmf of same form as that of pdf $f(x) = \alpha x^{\alpha-1} e^{-x^{\alpha}}$ corresponding to continuous Weibull variable at some intermediate point between two successive integers. With an application of the mean value theorem, write

$$p_k = P(V = k) = q^{k^{\alpha}} - q^{(k+1)^{\alpha}} \propto \alpha \kappa^{\alpha-1} q^{\kappa^{\alpha}}, \kappa \in (k, k+1); k \in \mathbf{N_0}. \tag{7}$$

which is of the same form as that of the Weibull pdf $f(x)$ at some intermediate point $\kappa$ between two integers. Hazard rate of discrete Weibull variable $V$ for $k \in \mathbf{N_0}$ has the following representation.

$$r_k(q, \alpha) = p_k / \sum_{j=k}^{\infty} p_j = 1 - (q)^{(k+1)^{\alpha} - k^{\alpha}} = 1 - (q)^{\alpha \kappa_1^{\alpha-1}}, \kappa_1 \in (k, k+1). \tag{8}$$

The ratio in (8) lies in $(0, 1)$ unlike hazard rate of continuous Weibull variable. From (7), it is possible to express the hazard rate of discrete Weibull variable $V$ in an alternate form

$$r_k(q, \alpha) = p_k / \sum_{j=k}^{\infty} p_j \propto \alpha \kappa^{\alpha-1} q^{\kappa^{\alpha} - k^{\alpha}}. \tag{9}$$

This form is similar to hazard rate of continuous Weibull distribution, apart from an additional term. However, this extra term $q^{\kappa^{\alpha} - k^{\alpha}} \to 1$, as $\kappa \to k$ in a discrete Weibull distribution, where consecutive equispaced variate values are close to each other rather than the constant spacing 1 in $\mathbf{N_0}$; in such a situation hazard rate (9) above coincides with that for the Weibull pdf $f(x)$, with $k$ replaced by $x$.
Discrete version of modified Weibull distribution may arise when one considers integer part of the random variables with distribution given by (3). Discrete MWD is appropriate for data

modelling when the parent variables have bathtub hazard rate and the observations are taken on discrete points.
Both continuous and discrete versions of Weibull distribution are used in practice. The continuous version has been used to analyse failure in electronic components, distribution of job characteristics like ovality, eccentricity etc., see e.g., Dasgupta, Ghosh, and RangaRao (1981). Failures of some devices may sometimes depend on the total number of cycles. In such cases, discrete Weibull distribution may provide a good approximation. Two types of discrete Weibull distributions are mainly studied; these are based on either hazard rate or, logarithm of survival function following power law, see e.g., Ali Khan *et al.* (1989), Englehardt and Li (2011), Nakagawa and Osaki (1975), Stein and Dattero (1984).

## 4. Characterization of discrete Weibull distribution

As already mentioned, there are mainly two types of discrete Weibull distributions based on scale invariance of hazard rate / logarithm of survival function following power law over the set of integers $k \in \mathbf{N_0}$. From Remark 2, we see that the logarithm of survival function for discrete Weibull variable $V$ follows power law under appropriate scaling, thus exhibiting invariance under scale change over the set of integers $k \to jk; j, k \in \mathbf{N_0}$.
An arithmetical function $h$, not identically zero, is said to be multiplicative if $h(jk) = h(j)h(k)$ whenever $(j, k) = 1$, and $h$ is completely multiplicative if $h(jk) = h(j)h(k)$ for all $j$ and $k$. The following theorem is for increasing multiplicative functions.

*Theorem (Erdös (1946)).* If $h$ is increasing and multiplicative, then there is a constant $a$ such that $h(k) = k^a$ for all $k \geq 1$.

Extension of the above to decreasing $h$ is possible when $h(3) \neq 0$, see e.g., Howe (1986). Every increasing multiplicative function is completely multiplicative is shown therein.
Let us recall the proof of Theorem 2. Assume that the random variable $W$ has support $\mathbf{N_0}$, and logarithm of the survival function $h(k) = \log P(W \geq k)$ satisfies $h(jk) = h(j)h(k); j, k \in \mathbf{N_0}$, after an appropriate scaling $h(1) = 1$. $h$ is monotonically decreasing. $h(3) \neq 0$, as support of the discrete random variable is $\mathbf{N_0}$. Then from the above theorem, $h$ follows power law and therefore $W$ is a discrete Weibull variable with distribution as mentioned in Remark 2. Hence the following characterization for discrete random variable $W$ holds.

**Theorem 3.** Suppose for a discrete random variable $W$ with support $\mathbf{N_0}$, logarithm of the survival function $h(k) = \log P(W \geq k)$, remains invariant under a change in scale $k \to jk; j, k \in \mathbf{N_0}$. Then $h$ satisfies $h(jk)h(1) = h(j)h(k); j, k \in \mathbf{N_0}$, and consequently $W$ is a discrete Weibull variable with survival function of the form

$$P(W \geq k) = \exp(-\lambda k^\alpha) = q^{k^\alpha}, q = \exp(-\lambda); k \in \mathbf{N_0}. \tag{10}$$

A similar result is possible for characterizing another type of discrete Weibull variable $Z$ with pmf $p_k$ when *hazard rate* $r_k = p_k / \sum_{j=k}^{\infty} p_j, k \in \mathbf{N_0}$, follows power law apart from a multiplicative factor; see (3)-(6) of Stein and Dattero (1984) for the special case with exponent

$\alpha = \beta - 1 \leq 0$. Then the support of the discrete Weibull variable $Z$ is $\mathbf{N_0}$, with hazard rate

$$r_k(\alpha) = ck^{\alpha}, \alpha \leq 0, c \in (0, 1]; k \in \mathbf{N_0}. \tag{11}$$

from Erdös (1946) theorem, with modification made for $c$. Hazard rate of the above form may arise when chance of failing a system is high at an early state with decreasing hazard rate over time. As for example, some plant saplings are likely to die at a tender stage, but it may survive long afterwards crossing that state.
Proceeding in a similar fashion as in Theorem 3, one may obtain the following.

**Theorem 4.** Let the discrete random variable $W$ with pmf $p_k = P(W = k)$ and support $\mathbf{N_0}$, has monotone hazard rate $r_k = p_k / \sum_{j=k}^{\infty} p_j$ that remains invariant under a change in scale $k \to jk; j, k \in \mathbf{N_0}$. Then hazard rate $r$ satisfies $r_{jk} r_1 = r_j r_k; j, k \in \mathbf{N_0}$, and consequently $W$ is a discrete Weibull variable with hazard rate of the form (11), i.e.,

$$P(W \geq k) = \exp(-\lambda k^{\alpha}) = q^{k^{\alpha}}, q = \exp(-\lambda); k \in \mathbf{N_0}$$

The survival function in terms of hazard rate may be written as

$$P(W \geq i) = \Pi_{j=1}^{i-1}(1 - r_j), i \geq 1. \tag{12}$$

Unlike the continuous case, it is not possible to have discrete Weibull distribution with both logarithm of survival function and hazard rate to be invariant under scale change, as the two conditions lead to separate distributions; see (10) for Type 1 discrete Weibull distribution and (11)-(12) for Type 2 distribution. The equation (12) above suggests that for the variable $W$ to have support $\mathbf{N_0}$, hazard rate should be nonincreasing in equation (11), leading to the condition $\alpha \leq 0$.

## 5. Some applications

*Life distribution of Elephant-foot-yam plant.* Yam is a tuber crop that has good market value, especially when supplied early in market before season. It is thus of interest to study the life distribution of yam plant to foresee market supply over time. The hazard rate or, instantaneous failure rate for plant life on maturity, may be interpreted as the rate at which crop is harvested from field to be delivered in market. From probability plots of yam plant lifetime it is seen that to a first approximation Weibull is a reasonable model, under which scale change may proportionately change the hazard rate of plant lifetime measured in terms of days / hours when yams are harvested. Hazard rate or rate of harvest is related to rate of crop arrival in markets. The latter has an impact on price level of the commodity based on supply and demand in market. Sometimes farmers keep unharvested yam underground for some more months to avoid low market price.
The presented data (in days) relates to plant life in five different experiments in different type of plots conducted in Indian Statistical Institute, Giridih farm-land by the riverside Ushri, during the year 2011. In Experiments 1-4 average seed weight used for plantation is 500 g, for Experiment 5 this is 300 g. To a first approximation, Weibull distribution

$F(x) = 1 - \exp(-(x/\sigma)^{\alpha}), \sigma > 0, \alpha > 0, x > 0$; for plant lifetime is seen to be appropriate.

Data on Experiments 1-5 on Plant lifetime (in day) are given in the appendix.

Experiment 1 is conducted in a plot that is moderately fertile after several years of cultivation. The Weibull plot shown in Figure 1 indicates a reasonably good fit. Parameters estimated by the method of maximum likelihood are as follows, $\sigma = 168.0, \alpha = 15.54$.

Agricultural plot of Experiment 2 is fertile. Parameters estimated are $\sigma = 167.0, \alpha = 14.89$, see Figure 2.

The plot of Experiment 3 is situated partly in a shaded region under tall trees causing scarcity of sunlight. Dry leaf accumulated on the ground over time and rain water droplets from trees above fell for a prolonged period. Stagnated water in a part of experimental plot during monsoon, which is quite intense in Giridih, damaged some yam plants. The experiment is partly disturbed, which is reflected in Weibull plot. About 20% of the plants died prematurely and that resulted in poor yield in the affected part of the experimental plot. Data on this experiment may be considered to be an outlier compared to other experiments reported here. Weibull fit is not satisfactory, estimated parameters are $\sigma = 129.6, \alpha = 3.699$; see Figure 3.

Experiment 4 is conducted in an unfertile piece of land, barren and used for the first time in Yam cultivation. Estimated parameters are $\sigma = 154.8, \alpha = 11.09$, see Figure 4. These seem different from the parameters of earlier reported regular experiments. In this harsh experimental environment for crops Weibull model still seems appropriate.

In Experiment 5 although the land is fertile, planted seed corm have weight on an average 60% of those planted in other four experiments. This resulted in undernourished yam plants as in Experiment 4. Estimated parameters are $\sigma = 149.1, \alpha = 8.381$, see Figure 5.
By combining data of Experiment 1 and 2, i.e., those conducted in favorable environment, the estimated parameters are as follows, $\sigma = 167.5, \alpha = 15.20$, see Figure 6. These parameters do not vary much in the combined data compared to individual data sets. In Weibull graph of combined data, only a few points lie outside the 95% confidence interval (CI) band, this is in agreement with 197 data points in combined experiment.
Similarly for combined data of Experiment 4 and 5, those conducted in relatively harsh environment, estimated parameters for Weibull model are $\sigma = 152.3, \alpha = 9.639$, see Figure 7.
Ignoring the data from Experiment 3, which was disturbed; one may check for Weibull fit in combined data from remaining experiments. For combined data of Experiment 1, 2, 4 and 5 with $n = 387$, estimated parameters for Weibull model are $\sigma = 160.8, \alpha = 10.77$, see Figure 8.
Histogram of plant life in all the experiments 1-5 are seen to be negatively skew. Weibull distribution for large value of shape parameter is negatively skew. For combined data of Experiment 1, 2, 4 and 5, the histogram is negatively skew, as shown in Figure 9. About 5% of observations near the start of Figure 8 are seen as separated from main portion and may be interpreted as early arrivals in the market fetching good price, remaining 95% yam that arrive afterwards in market may follow plant lifetime model as Weibull, having proportionate

hazard rate for yam harvest when lifetime is over.
Simulated growth curve of $n = 387$ yam-plant lifetime under Weibull model with parameters $\sigma = 160.8, \alpha = 10.77$, is shown in Figure 10, with assigned seed value 123 in SPLUS program. The pattern of the growth curve is seen to remain stable under different assigned seed values. Simulated curve of Figure 10 shows similarity with actual growth curve of Figure 11, obtained from yam data, assuming that the yam lifetimes are realised from lowest to highest sequentially over time.
Hazard rate of lifetime $h(t) = (\frac{\alpha}{\sigma})(\frac{t}{\sigma})^{\alpha-1}$, estimated from data may be used in forecasting market supply of the crop over time. One needs to scale this with number of plantations in a season and average weight per yam to have the supply value.
Like Weibull distribution, parameters in yam lifetime modelling have standard interpretation; $\sigma$ refers to the spread of variable and $\alpha$ refers to the shape. In the present context of yam lifetime, a large value of $\sigma$ means that different plants survived for widely different lifetimes, thus the subsequent market supply of yam is widely spread over time. A large value of shape parameter for Weibull indicates negatively skew distribution. In the present context $\alpha = 10.77$, this means yam lifetime distribution is clustered towards higher values at the right, indicating congestion in market towards the end of yam yield season.
Modelling hazard rate of yam lifetime and connecting this to specific Weibull distribution helps us to understand the time required for crop maturity and its rate, which has direct implication in marketing the crop. On the other hand, conventional time series model / regression are like general prescription; missing the point that Weibull is a good fit for lifetime. Yam lifetime modelling takes into account the crop cultivation period and crop harvesting time (as time to failure). The specific approach seems satisfactory as seen from the closeness of simulated and observed curve for yam lifetime, suggesting that the model is accurate in this specific case of yam cultivation. The hazard rate or, instantaneous failure rate for plant life on maturity, is interpreted as the rate at which crop is harvested from field to be delivered in market.
This is a power function of time $t$, and for $\alpha > 1$ is increasing in $t$, estimated value of $\alpha$ is 10.77 in the present context.
Analysis presented in Figures 1, 2, 4, 6 show very high (coverage) accuracy in 95% Weibull confidence band. Figures 5, 7, 8 show the same, beyond a threshold value of yam lifetime (beyond a period of early harvest that fetch a good price). Figure 3 refers to a disturbed yam experiment and was not taken into account for estimation of parameters. Once the distributional accuracy is established, subsequent analysis based on that is apparent.

One of the commonly used nonparametric regressions for time series forecasting is Spline technique, see e.g., Caudle and Frey (2012), Huang and Shen (2004). Consider 387 lifetime observations of yam lifetime reported from Experiments 1, 2, 4 and 5 conducted in the year 2011 at Giridih. These observations are to appear in an increasing order of time, as seen in a growth curve, the largest lifetime of yam plant observed being 193 days; see Figure 11; where on the $x$ axis the number of observations $n$ is scaled. Now consider the problem of forecasting the upper part of the curve based on lower 195 observations (about 50% of total observations), from the beginning; with additional information given that the largest lifetime observed is 193 days. Spline regression made in Figure 12 shows a dampened growth. A poor performance of spline in the predicted region is seen especially near the peak of 193, as compared to the simulated curve of Figure 10 having a sharp upturn under Weibull model;

and Figure 10 is closely mimicking the observed lifetime graph shown in Figure 11.

## 6. Discrete Weibull: Yam lifetime and market supply distribution

We modelled lifetime of 387 plants by Weibull variable $U$. Harvesting the crop are done on different days / weeks in a production season. At night harvesting ceases, this is usually the case. In such a situation continuous Weibull variable $U$ of harvest time is discretized to $V = [U]$, a discrete integer valued Weibull variable. Histogram of plant lifetime grouped into class intervals shows negatively skew Weibull distribution, indicating a large value for shape parameter. Discrete Weibull model of plant lifetime may be adopted when a large number of yam plantations is undertaken in a farm. From estimation and testing viewpoint, if a condition like each cell frequency over different days for yam harvest should be greater than 10 is imposed for all possible harvest days, then more than 2000 plantations are required to fit the model, as Elephant foot yam plant may survive more than 200 days in a production season. The estimated pmf of discrete Weibull would reflect region specific characteristics of yam harvest and subsequent market supply pattern on daily / weekly basis.
Arrival of produce to market from farm introduces another time lag component depending on mode of transport, distance from farm etc. Towards end of the production season, crops pile up and time required $t = t(U)$ to reach the huge amount of crop to ultimate consumers may require a proportionate time $aU, (a > 0)$ of $U$. This may be the case when the production is growing very fast but the infrastructure does not react with the same speed, supply is then trapped in congestion. Total time is then an Weibull variable $U(1+a)$ to a first approximation, $T \approx U(1 + a)$. Realized arrival time $T^*$ in market to the buyers measured in day (say), is a discrete version of the above, i.e $T^* = [T] \approx [U(1 + a)]$, which is a discrete Weibull variable of type 1, whose logarithm of survival function is scale invariant on set of positive integers.
The additional parameter $a > 0$ may be estimated by tracking some randomly selected yams $i \in I$ on transit while these arrive at ultimate destinations and then using an efficient ratio estimate,

$$\hat{a} = \frac{\sum_{i\in I} T_i^*}{\sum_{i\in I} V_i} - 1 \tag{13}$$

where $V_i = [U_i]$ and $T_i^* = [T_i]$ are observable discretised harvest time and total time respectively of $i$-th yam recorded in transit. Small values of $a$ indicate efficient marketing system and less wastage on the way to consumer.

*Acknowledgement*: Thanks are due to referee for constructive suggestions.

## 7. Appendix

In this section we provide the yam experimental data from ISI Giridih farm along with Figures associated with data analysis.

**Table 1. Data on Experiment 1: Plant lifetime (in day)**

| |
|---|
| 146,170,165,159,164,163,170,164,159,190,164,164,158,148,174,181,165,139,151, 151,150,150,165,155,139,161,164,177,176,151,165,177,149,149,148,180,174,164, 165,162,179,165,178,164,164,179,161,175,179,145,165,164,159,163,179,112,148, 161,181,181,178,166,155,161,162,172,166,161,161,163,165,176,165,161,147,162, 163,161,170,146,179,179,180,163,148,170,164,163,164,166,170,153,162,141,162, 140,149,149,148,177 |

**Table 2. Data on Experiment 2: Plant lifetime (in day)**

| |
|---|
| 111,166,137,171,164,193,163,163,162,176,161,176,176,179,152,165,164,164,149, 149,170,171,182,159,133,179,149,155,163,178,181,177,164,163,167,166,148,144, 173,163,162,182,164,155,143,166,162,173,174,162,151,139,170,167,162,155,159, 155,163,164,164,161,179,144,152,156,171,161,143,162,151,155,167,133,162,162, 166,158,139,167,176,161,152,152,177,174,151,171,161,174,159,159,158,166,146, 169,154 |

**Table 3. Data on Experiment 3: Plant lifetime (in day)**

| |
|---|
| 98,78,85,123,123,99,94,123,153,161,32,127,86,138,137,124,123,153,150,146,97, 77,124,79,126,126,138,153,137,135,40,43,37,88,152,127,153,120,153,133,33,97, 112,57,136,137,143,153,150,151,75,78,89,130,120,151,152,152,152,151,38,43,88, 77,136,137,150,148,148,151,43,97,77,156,133,146,152,142,151,134,52,32,94,146, 107,144,146,168,154,141,49,40,47,120,164,151,150,107,157,144 |

**Table 4. Data on Experiment 4: Plant lifetime (in day)**

| |
|---|
| 161,153,147,152,153,154,137,138,147,170,152,153,141,153,161,153,149,154,153, 155,153,154,153,165,78,143,155,139,168,153,145,166,168,168,168,153,153,113, 154,153,135,132,148,147,153,150,154,136,147,145,129,138,74,141,142,153,135, 151,170,154,127,166,133,153,162,153,150,114,158,140,151,192,147,151,150,128, 152,153,173,143,161,128,142,135,145,112,144,144,152,149,145,173,168,153,140, 159,164,153,151 |

**Table 5. Data on Experiment 5: Plant lifetime (in day)**

| |
|---|
| 129,151,163,139,162,97,87,136,114,162,155,147,158,155,140,156,67,117,154,109, 153,147,141,143,167,140,156,155,154,78,139,163,160,161,139,155,144,125,149, 143,152,136,152,147,118,111,122,137,151,139,138,153,119,172,173,157,155,156, 160,134,172,144,116,73,150,160,151,109,139,140,137,178,176,132,151,127,154, 132,135,162,125,110,145,131,101,135,153,150,154,139,110 |

## References

Ali Khan M, Khalique A, Abouammoh A (1989). "On Estimating Parameters in a Discrete Weibull Distribution." *IEEE Transactions on reliability*, **33**, 348–350.

Caudle K, Frey M (2012). "Continuous Updates of Penalized Spline Regression for Flow Field Forecasting." *Proceeding of the 32nd Annual International Symposium on Forecasting, Boston, MA 24-27 June*, pp. 1–6.

Cox D (1972). "Regression model and life tables." *JRSS B*, pp. 187–220.

Dasgupta R (2013). "Characterization theorems based on conditional quantiles with applications." *Journal of Environmental Statistics*, **4**. Issue 6.

Dasgupta R, Ghosh JK, RangaRao NTV (1981). "A cutting model and distribution of ovality and related topics." *Proceeding of the ISI golden jubilee conference*, pp. 182–204.

Englehardt J, Li R (2011). "The Discrete Weibull Distribution: An Alternative for Correlated Counts with Confirmation for Microbial Counts in Water." *Risk Analysis*, **31**, 270–381.

Enkhmunkh N, Kim GW, Hwang KJ, Hyun SH (2007). "A Parameter Estimation of Weibull Distribution for Reliability Assessment with Limited Failure Data." *Strategic Technology, IFOST*, pp. 39–42.

Erdös P (1946). "On the distribution function of additive functions." *Ann. of Math.*, **47**, 1–20.

Howe E (1986). "A New Proof of Erdös's Theorem on Monotone Multiplicative Functions." *The American Mathematical Monthly*, **93**, 593–595.

Huang J, Shen H (2004). "Functional Coefficient Regression Models for Non-linear Time Series: A Polynomial Spline Approach." *Scandinavian Journal of Statistics*, **31**, 515–534.

Joh H, Kim J, Malaiya Y (2008). "Vulnerability Discovery Modeling using Weibull Distribution." *19th International Symposium on Software Reliability Engineering.* Http://www.cs.colostate.edu/ malaiya/pub/weibull08.pdf.

Johnson N, Kotz S, Balakrishnan N (1994). *Continuous univariate distributions*, volume 1. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics (2nd ed.), New York.

Lai C, Xie M, Murthy D (2003). "A Modified Weibull Distribution." *IEEE Transactions on reliability*, **52**, 33–37.

Murthy D, Xie M, Jiang R (2004). *Weibull Models.* John Wiley. New York.

Nakagawa T, Osaki S (1975). "The Discrete Weibull Distribution." *IEEE Transactions on reliability*, **R-24, 5**, 300–301.

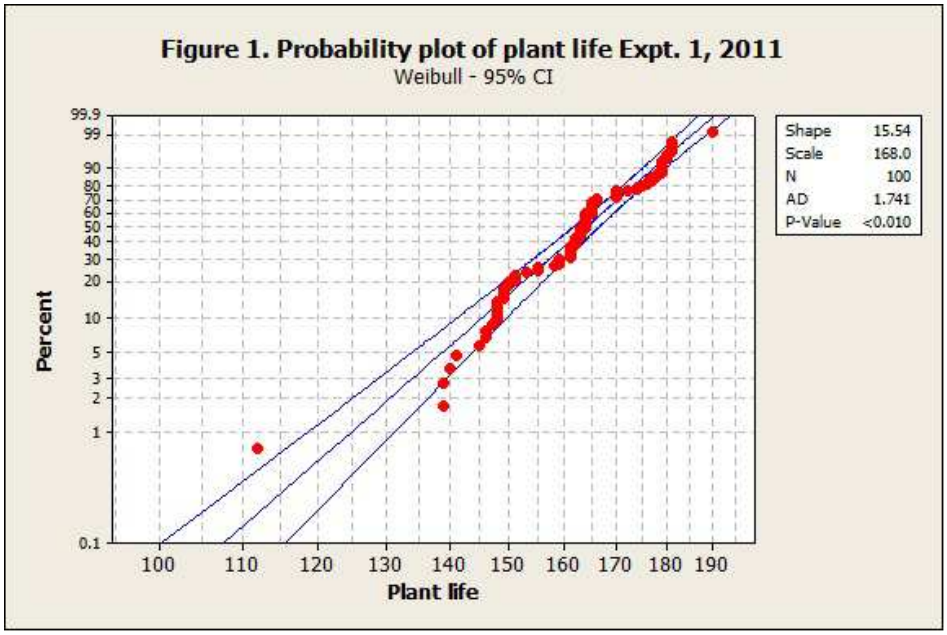
Qin X, she Zhang J, dong Yan X (2012). "Two Improved Mixture Weibull Models for the Analysis of Wind Speed Data." *J. Appl. Meteor. Climatol.*, **51**, 1321–1332.

Stein W, Dattero R (1984). "A new discrete Weibull distribution." *IEEE Transactions on reliability*, **R-33**, 196–197.

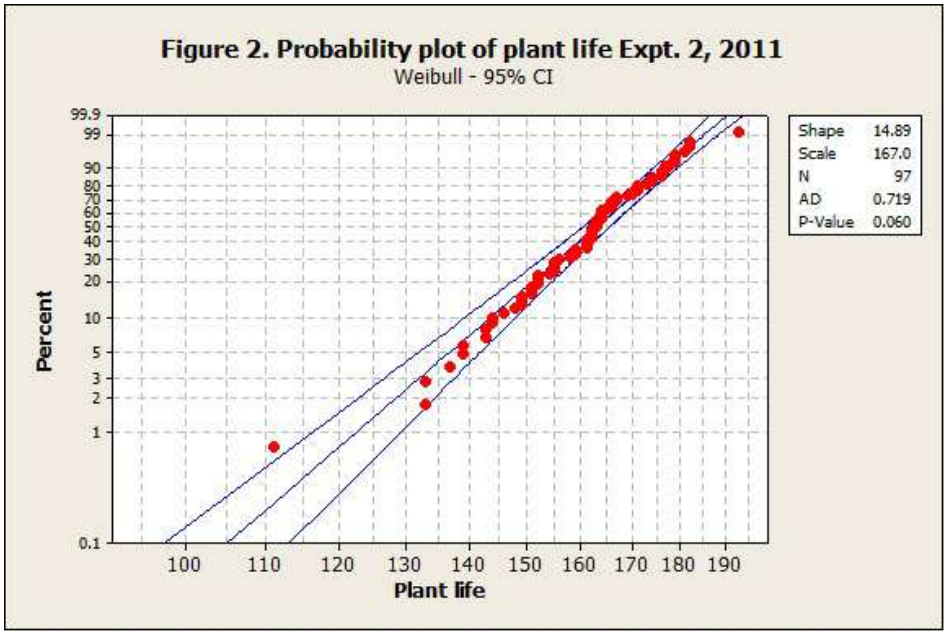
Wang Lj, Hong-we W, Xi-chao S (2012). "Inventory control model for fresh agricultural products on Weibull distribution under inflation and delay in payment." *Kybernetes*, **41**, 1277–1288.

Weibull W (1951). "A statistical distribution function of wide applicability." *J. Appl. Mech.-Trans. ASME*, **18**, 293–297.

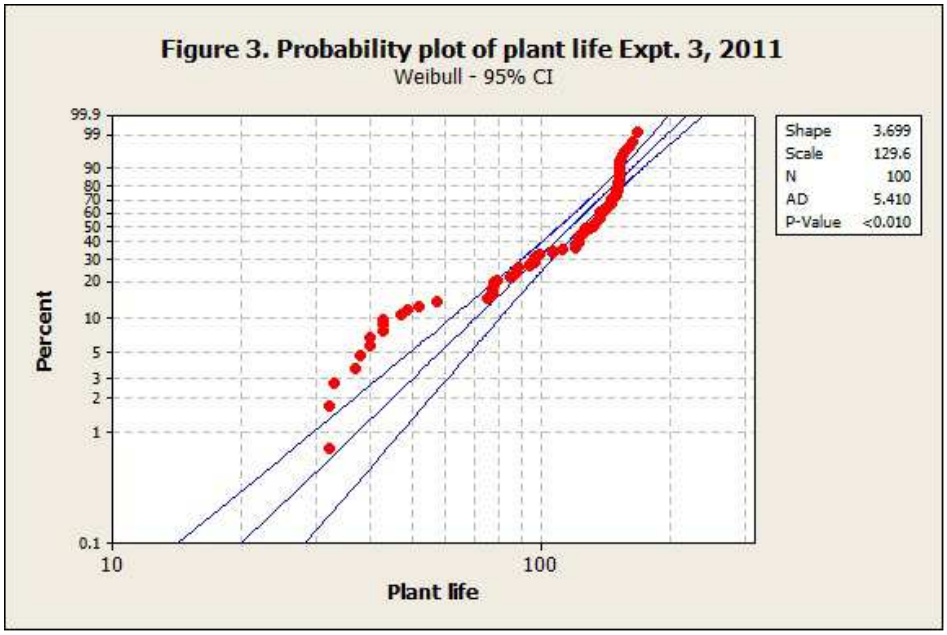
Zhu HP, Xia X, Yu CH, Adnan A, Liu SF, Du YK (2011). "Application of Weibull model for survival of patients with gastric cancer." *BMC Gastroenterology*, **11(1)**. DOI: 10.1186/1471-230X-11-1.



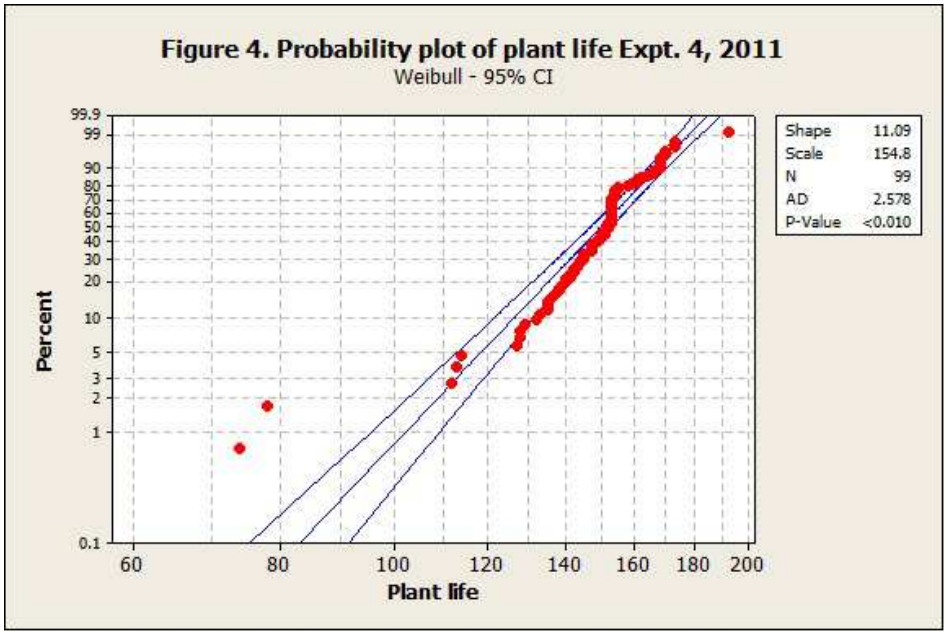

Experiment 1 is conducted in a plot that is moderately fertile after several years of cultivation. The Weibull plot shown in Figure 1 indicates a reasonably good fit. Parameters estimated by the method of maximum likelihood are as follows, $\sigma = 168.0, \alpha = 15.54$.



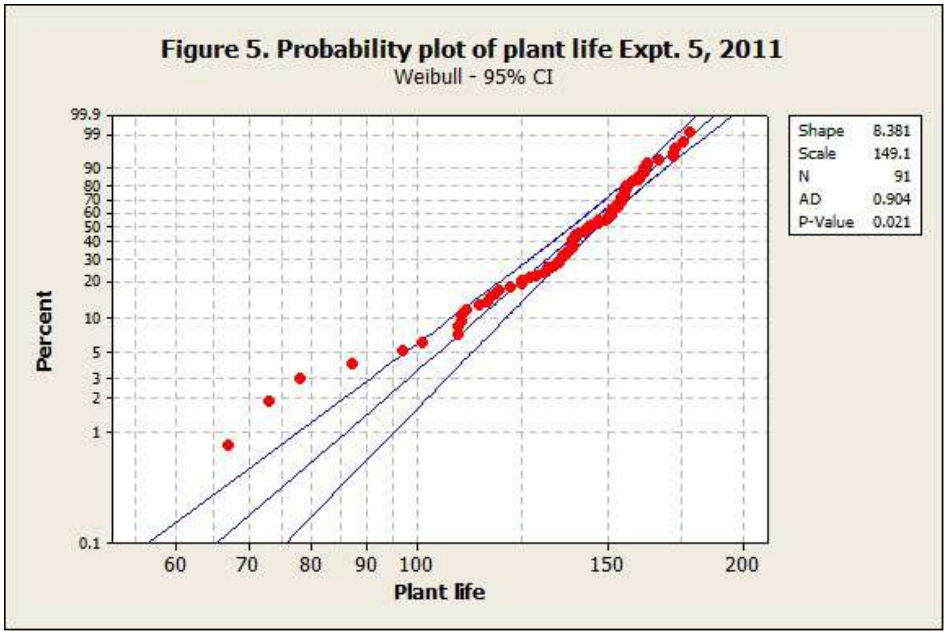

Agricultural plot of Experiment 2 is fertile. Parameters estimated are $\sigma = 167.0, \alpha = 14.89$.



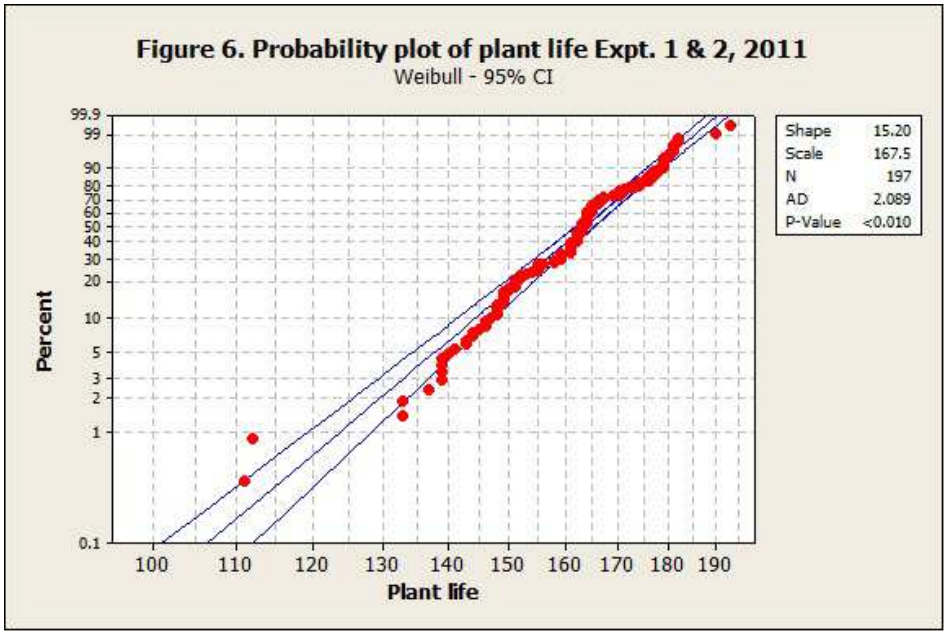

The plot of Experiment 3 is situated partly in a shaded region under tall trees causing scarcity of sunlight. Dry leaf accumulated on the ground over time and rain water droplets from trees above fell for a prolonged period. Stagnated water in a part of experimental plot during monsoon, which is quite intense in Giridih, damaged some yam plants. The experiment is partly disturbed, which is reflected in Weibull plot. About 20% of the plants died prematurely and that resulted in poor yield in the affected part of the experimental plot. Data on this experiment may be considered to be an outlier compared to other experiments reported here. Weibull fit is not satisfactory, estimated parameters are $\sigma = 129.6, \alpha = 3.699$.



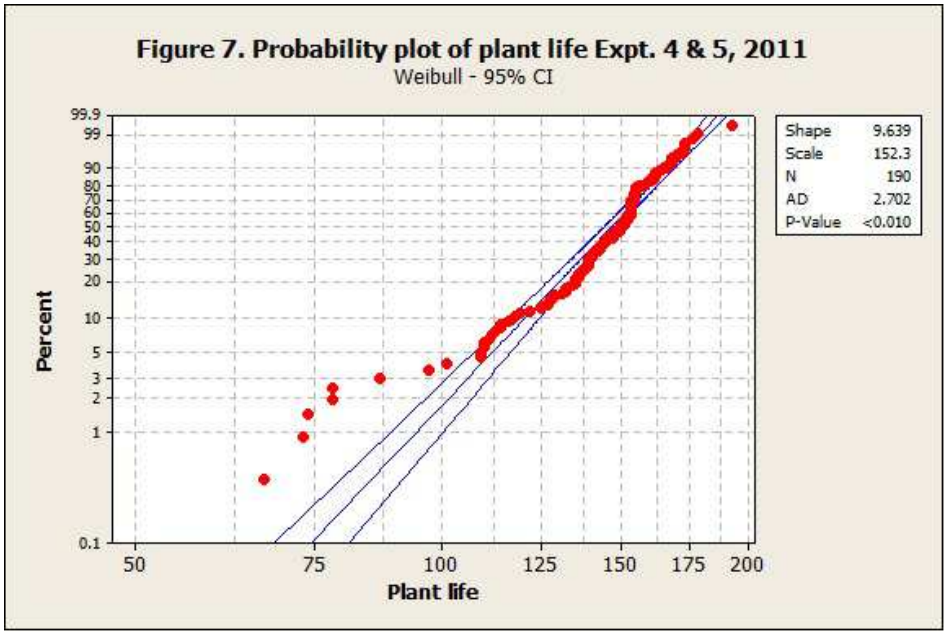

Experiment 4 is conducted in an unfertile piece of land, barren and used for the first time in Yam cultivation. Estimated parameters are $\sigma = 154.8, \alpha = 11.09$. These seem different from the parameters of earlier reported regular experiments. In this harsh experimental environment for crops Weibull model still seems appropriate.



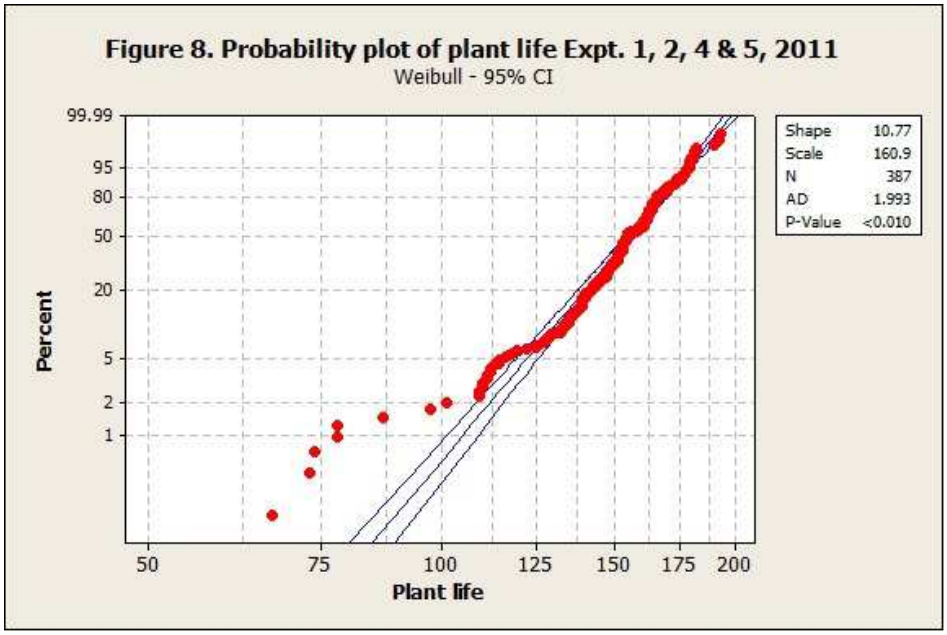

In Experiment 5 although the land is fertile, planted seed corm have weight on an average 60% of those planted in other four experiments. This resulted in undernourished yam plants as in Experiment 4. Estimated parameters are $\sigma = 149.1, \alpha = 8.381$.




By combining data of Experiment 1 and 2, i.e., those conducted in favorable environment, the estimated parameters are as follows, $\sigma = 167.5, \alpha = 15.20$, see Figure 6. These parameters do not vary much in the combined data compared to individual data sets. In Weibull graph of combined data, only a few points lie outside the 95% confidence interval (CI) band, this is in agreement with 197 data points in combined experiment.

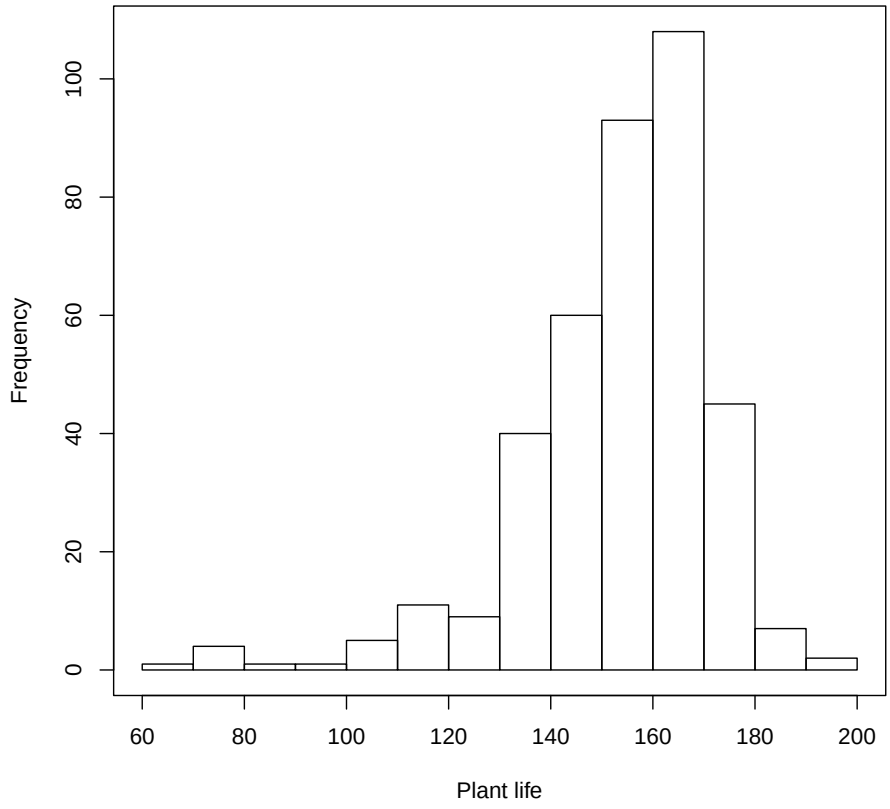



For combined data of Experiment 4 and 5, those conducted in relatively harsh environment, estimated parameters for Weibull model are $\sigma = 152.3, \alpha = 9.639$.



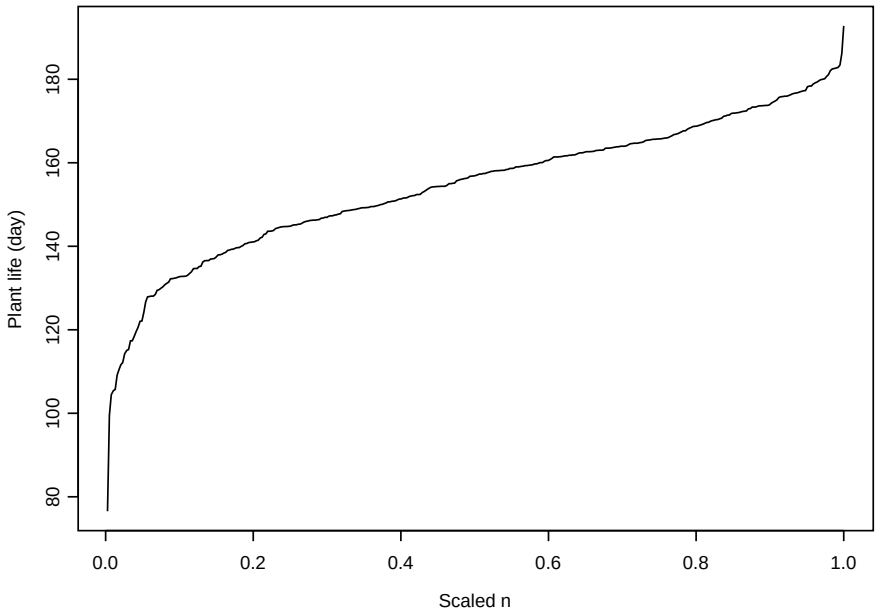

Ignoring the data from Experiment 3, which was disturbed; one may check for Weibull fit in combined data from remaining experiments. For combined data of Experiment 1, 2, 4 and 5 with $n = 387$, estimated parameters for Weibull model are $\sigma = 160.8, \alpha = 10.77$. About 5% of observations near the start of Figure 8 are seen as separated from main portion and may be interpreted as early arrivals in the market fetching good price, remaining 95% yam that arrive afterwards in market may follow plant lifetime model as Weibull, having proportionate hazard rate for yam harvest when lifetime is over.

**Figure 9. Histogram of yam life Expt. 1, 2, 4 & 5, yr. 2011**



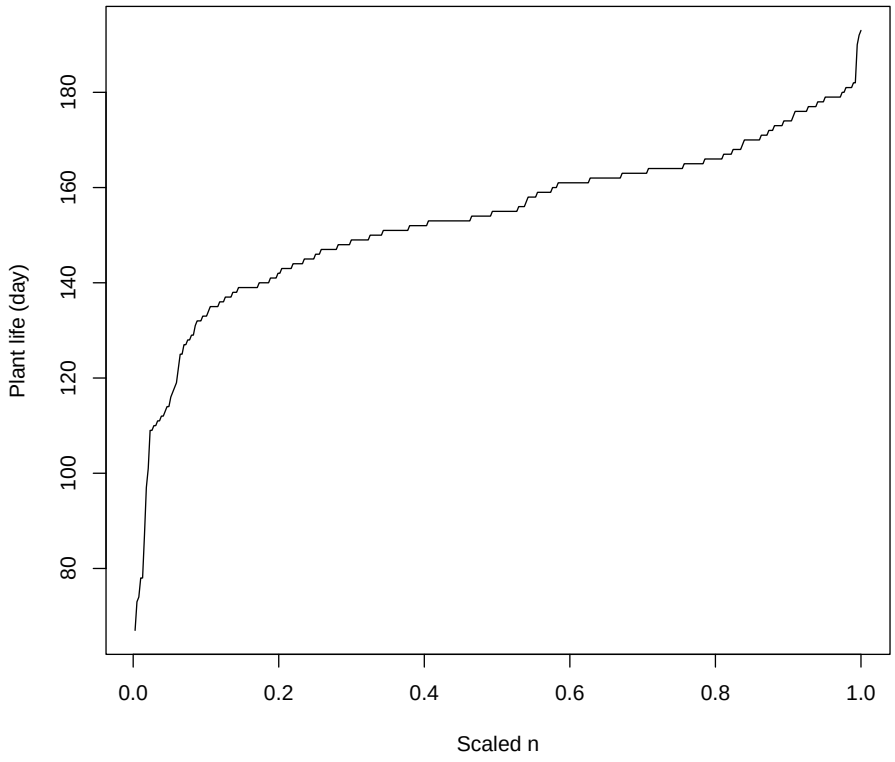

Histogram of plant life in all the experiments 1-5 are seen to be negatively skew. Weibull distribution for large value of shape parameter is negatively skew. For combined data of Experiment 1, 2, 4 and 5, the histogram is negatively skew, as shown in Figure 9.

Figure 10. Simulated growth curve of yam plant lifetime



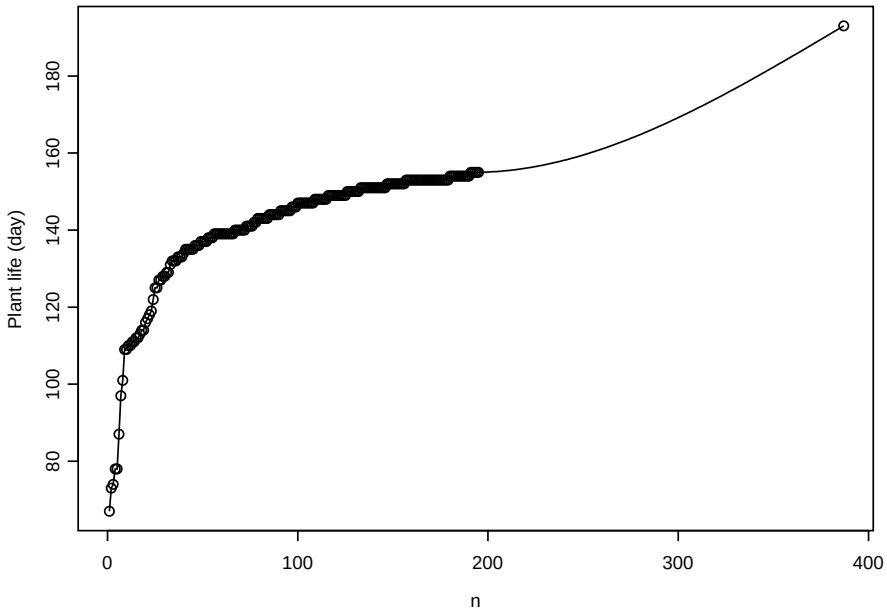

Simulated growth curve of $n = 387$ yam-plant lifetime under Weibull model with parameters $\sigma = 160.8, \alpha = 10.77$, is shown in Figure 10, with assigned seed value 123 in SPLUS program. The pattern of the growth curve is seen to remain stable under different assigned seed values.

**Figure 11. Growth curve of yam plant life time**




Simulated curve of Figure 10 shows similarity with actual growth curve of Figure 11, obtained from yam data, assuming that the yam lifetimes are realised from lowest to highest sequentially over time.

Figure 12. Spline regression of growth curve: yam plant lifetime




Consider the problem of forecasting the upper part of the curve based on lower 195 observations (about 50% of total observations), from the beginning; with additional information given that the largest lifetime observed is 193 days. Spline regression made in Figure 12 shows a dampened growth. A poor performance of spline in the predicted region is seen especially near the peak of 193.

**Affiliation:**

Ratan Dasgupta
Indian Statistical Institute
Theoretical Statistics and Mathematics Unit
Calcutta 700 108, INDIA
E-mail: ratandasgupta@gmail.com
rdgupta@isical.ac.in