Revue Belge de Statistique, d'Informatique et de Recherche Opérationnelle, Vol.22, n°4.
Belgisch Tijdschrift voor Statistiek, Informatica en Operationeel Onderzoek, Vol.22, Nr 4.

EFFICIENT ESTIMATION OF QUANTILES FOR WEIBULL DISTRIBUTION. (*)

by J. TIAGO DE OLIVEIRA

Faculty of Sciences of Lisbon

Center of Statistics and Applications (I.N.I.C.)

Academy of Sciences of Lisbon

ABSTRACT.

It is well known that statistical decision for Weibull distribution when the shape parameter $\alpha \leq 2$ has various dificulties. In this paper, an asymptotically normal estimator of any quantile for probabilities different from $\underline{0}$ and $\underline{1}$ is given and shown that has asymptotic efficiency $\underline{1}$ when $\alpha < 2$; the same estimator for $\alpha = 2$ is shown not to be asymptotically normal and with asymptotic efficiency lower than 1.

KEYWORDS.

WEIBULL DISTRIBUTION, ESTIMATION OF QUANTILES, SHAPE PARAMETER, ASYMPTOTIC BEHAVIOUR, EFFICIENCY.

(*) Paper presented at the VII Jornadas Matemáticas Hispano-Lusitanas, 1980.

I. Introduction.

The estimation of location and dispersion parameters (λ and $\delta > 0$) in Weibull distribution $W_\alpha((x - \lambda)/\delta)$, where $W_\alpha(z) = 0$ if $z < 0$ and $W_\alpha(z) = 1 - \exp(-z^\alpha)$ if $z \geqslant 0$, $\alpha > 0$, when the shape parameter (α) is known, can be split according to $\alpha \leqslant 2$ or $\alpha > 2$. If $\alpha > 2$ the maximum likelihood (ML) method works as usual because regularity conditions (Cramér (1946)) are valid, the ML estimators ($\hat{\lambda}$, $\hat{\delta}$) of (λ, δ) are asymptotically binormal and the estimator $\hat{\theta} = \hat{\lambda} + \chi\hat{\delta}$ of any quantile $\theta = \lambda + \chi\delta$ is asymptotically normal. As regularity conditions fail for $\alpha < 2$ the status is, then, different, with some paradoxical results: if $\alpha \leqslant 1$, the sample minimum is an "hyper-efficient" estimator of λ, in Dubey (1966) expression, as any positive statistic taken for estimator of δ maximizes the likelihood (being infinite for $\hat{\lambda}$ = minimum); see also Mann, Schaper and Singpurwalla (1974); lighter difficulties appear when $1 < \alpha < 2$. Analagous problems appear in the forecasting for a future sample: see Tiago de Oliveira and Littauer (1976) for some details and, also, the consideration of the regularity region ($\alpha > 2$). See Tiago de Oliveira (1980) for a method of decision on the values of $\alpha(\leqslant 2)$.

As for $\alpha > 2$ we have efficient ML estimation we will study, only, the estimation of any quantile $\theta = \lambda + \chi\delta$ with $0 < W_\alpha(\chi) < 1$, i.e., $0 < \chi < +\infty$ if $\alpha \leqslant 2$. For the estimator proposed we will obtain the expression of its asymptotic(variance) efficiency, and show that for $\alpha < 2$ this efficiency is 1.

II. The estimator of the quantile when $\alpha<2$:

Let $(x_1,\ldots,x_n)$ be a sample of independent observations with the distribution $W_\alpha((x - \lambda)/\delta)$ with $\alpha\,(0< \alpha <2)$ known.

If $\lambda = \lambda_0$ is known, the ML estimator ($\hat{\delta}_0$) of δ is given by the equation $\hat{\delta}_0^\alpha\,(x) = \sum_1^n (x_i - \lambda_0)^\alpha/n$. As the moments of the reduced Weibull random variables $z_i = (x_i - \lambda)/\delta$ are given by $\mu'_\rho\,(\alpha) = M(z^\rho) = \int_0^{+\infty} z^\rho \, d\,W_\alpha(z) = \Gamma(1 + \rho/\alpha)$ it is immediate that the mean value of $\hat{\delta}_0^\alpha$ is given by $M(\hat{\delta}_0^\alpha) = \delta^\rho\,\Gamma(2) = \delta^\alpha$

and the variance of δ_0^{α} is given by

$$V(\delta_0^{\alpha}) = \delta^{2\alpha}(\Gamma(3) - \Gamma^2(2))/n = \delta^{2\alpha}/_n.$$

Thus by the central limit theorem we see that $\sqrt{n}\ (\delta_0^{\alpha} - \delta^{\alpha})/\delta^{\alpha}$ is asymptotically standard normal and, by the δ-method, also $\sqrt{n}\ \alpha(\delta_0 - \delta)/\delta$ is asymptotically standard normal.

Consider now the general case (with λ and $\delta > 0$ unknown) with $0<\alpha< 2$, where the ML method results can not be applied.

As $l_n(x) = \min(x_1,\ldots,x_n)$ $(>\lambda)$ has the distribution function $W_{\alpha}(n^{1/\alpha}(l - \lambda)/\delta)$ we see that $M(l_n(x)) = M(\lambda + \delta l_n(z)) = \lambda + \delta\mu_1'(\alpha)/n^{1/\alpha}$ and the variance is

$$V(l_n(x)) = \delta^2\ V(l_n(z)) = (\mu'^2(\alpha) - \mu'^2(\alpha))\ \delta^2/n^{2/\alpha}.$$

Thus $l_n(x) \overset{m.q.}{\rightarrow} \lambda$ and, consequently, $l_n(x) \overset{P}{\rightarrow} \lambda$. The last result can be obtained directly from the distribution function of $l_n(x)$. Note that the variance $V(l_n(x))$ converges to zero more quickly than $1/n$ if $\alpha<2$ which explains the results obtained below about the estimator proposed.

Owing to the quickest convergence of $l_n(x)$ to λ, the expression of δ_0^{α} suggests the consideration of the statistics

$$\delta^{*\alpha}(x) = \sum_1^n (x_i - l_n(x))^{\alpha}/n = \delta^{\alpha}\ \delta^{*\alpha}(z).$$

We have evidently $0 \leqslant \delta^*(z)$. It is sufficient , to prove the asymptotic normality of $\delta^*(z)$, to show that

$$\sqrt{n}\ (\delta_0^{\alpha}(z) - \delta^{*\alpha}(z)) \overset{P}{\rightarrow} 0.$$

Once proved this result we see that $\sqrt{n}\ (\delta^{*\alpha}(x) - \delta^{\alpha})/\delta^{\alpha}$ and, thus, $\sqrt{n}\ \alpha(\delta^*(x) - \delta)/\delta$ are asymptotically standard normal.

For $0<\alpha\leqslant 1$, as $a^{\alpha} + b^{\alpha} \geqslant (a+b)^{\alpha}$ $(a, b > 0)$, we have

$$\sqrt{n}\ (\delta_0^{\alpha}(z) - \delta^{*\alpha}(z)) = \frac{1}{\sqrt{n}} \sum_1^n \{z_i^{\alpha} - (z_i - l_n(z))^{\alpha}\} \leqslant$$
$$\leqslant \sqrt{n}\ l_n^{\alpha}(z) \rightarrow 0,$$

as follows from the distribution $W_{\alpha}(n^{1/\alpha}\ l(z))$ of $l_n(z)$; for $1\leqslant\alpha< 2$, as

$$\sum_1^n \{ z_i^{\alpha} - (z_i - l_n(z))^{\alpha}\} < \alpha\, l_n(z) \sum_1^n (z_i - \bar{l})^{\alpha-1} < \alpha\, l_n(z) \sum_1^n z_i^{\alpha-1}$$

with $0 \leq \bar{l} \leq l_n(z)$ we see that

$$\sqrt{n}\,(\hat{\delta}_0^{\alpha}(z) - \delta^{\alpha}) \leq \sqrt{n}\,\alpha\, l_n(z)\, \frac{1}{n} \sum_1^n z_i^{\alpha-1} \xrightarrow{P} 0$$

because $\frac{1}{n} \sum_1^n z_i^{\alpha-1} \xrightarrow{P} \mu'_{\alpha-1}(\alpha) = \Gamma(2 - 1/\alpha)$, which is finite when $\alpha > 1/2$, and $\sqrt{n}\, l_n(z) \xrightarrow{P} 0$ if $\alpha < 2$.

Thus we see that the statistic $t_n(x) = l_n(x) + \chi_n\, \delta^*(x)$ is an estimator of $\theta = \lambda + \chi\delta$ if $\chi_n \to \chi$.

Then $\sqrt{n}(t_n(x) - \lambda - \chi\delta) - \sqrt{n}\,\chi(\delta^*(x) - \delta) =$

$$= \sqrt{n}\,(l_n(x) - \lambda) + \sqrt{n}\,(\chi_n - \chi)\,\delta^*(x) \xrightarrow{P} 0$$

if $\sqrt{n}(\chi_n - \chi) \to 0$, as $\sqrt{n}(l_n(x) - \lambda) \xrightarrow{P} 0$ as shown before. Thus, the asymptotic distribution of

$\sqrt{n}\,\alpha(t_n(x) - \theta)/\delta\chi$ is the same of the $\sqrt{n}\,\alpha(\delta^*(x) - \delta)/\delta$, i.e., standard normal. The asymptotic variance of $t_n(x)$ is $\chi^2\delta^2/\alpha^2\ n$.

III. <u>A lower bound of the variance</u>: Let, now, $g_n(x)$ be a quasi-linear estimator of $\theta = \lambda + \chi\delta$, i.e., such that

$$g_n(\alpha + \beta x_i) = \alpha + \beta g_n(x_i) \; (\forall\alpha, \forall\beta > 0) \text{ and let}$$
$$M\,(g_n(x)) = \lambda + \delta M(g_n(z)) = \lambda + \xi_n\,\delta, \text{ with } \xi_n \to \chi;$$

$t_n(x)$ is, evidently, a quasi-linear estimator.

From the quasi-linearity of (any) $g_n(x)$ we have

$V\,(g_n(x)) = \delta^2\, V(g_n(z))$ and, consequently, we will obtain a lower bound for $V\,(g_n(z))$.

Denoting by $L_n(x_i | \alpha, \beta)$ the likelihood of a sample $(x_1, \ldots, x_n)$ with parameters of location α and dispersion $\beta (>0)$ we have the relation

$$\int_{-\infty}^{+\infty} \ldots \int (g_n(x_i) - \xi_n)\,(L_n(x_i | \alpha, \beta) - L_n(x_i | 0, 1)) \prod dx_i =$$
$$= \alpha + \xi_n(\beta - 1).$$

Putting in the integral relation $\alpha = 0$ we get

$$\int_{-\infty}^{+\infty}\ldots\int (g_n(x_i) - \xi_n)\,(L_n(x_i|0,\ \beta) - L_n(x_i|0,\ 1))\,\Pi\ dx_i = \xi_n(\beta-1)$$

and by the Schwarz inequality we get

$$\xi_n^{\ 2}(\beta - 1)^2 \leqslant \int_{-\infty}^{+\infty}\ldots\int \ (g_n(x_i) - \xi_n)^2\ L_n(x_i|\ 0,\ 1)\,\Pi\ dx_i \times$$

$$\int_{-\infty}^{+\infty}\ldots\int \frac{(L_n(x_i|0,\ \beta) - L_n(x_i|0,\ 1))^2}{L_n(x_i|0,\ 1)}\,\Pi\ dx_i$$

$$= V\ (g_n(z))\ \{\beta^{\alpha}(2 - \beta^{\alpha}))^{-n} - 1\}.$$

Thus we get $V(g_n(z)) \geqslant \xi_n^2 \lim\limits_{\beta\to 1} \dfrac{(\beta - 1)^2}{(\beta^{\alpha}\ (2 - \beta^{\alpha}))^{-n} - 1} = \dfrac{\xi_n^{\ 2}}{\alpha^2\ n}$

and thus for any estimator $t_n(x)$, with mean value $\lambda + \chi_n\delta$, and asymptotic variance $\chi^2\delta^2/\alpha^2\ \ n$ the asymptotic efficiency is $\lim\limits_{n\to\infty}\ (\xi_n^{\ 2}\delta^2/\alpha^2_{\ n})/(\chi^2\delta^2\ /\delta^2_{\ n}) = 1$.

IV. The case for Rayleigh distribution (α=2): Let us consider now the behaviour of the statistic $t_n(x) = 1_n(x) + \chi_n\delta^*(x)$. It was seen that $1_n(x) \overset{P}{\to} \lambda$ and

$$\delta^*(x) = \delta\ \delta^*(z) = \sqrt{\frac{1}{n}\ \sum_1^n z_i^2 - 2\ 1_n(z)\ \frac{1}{n}\ \sum_1^n z_i + 1_n^2(z)} \overset{P}{\to} \delta$$

because

$$\frac{1}{n}\ \sum_1^n z_i \overset{P}{\to} \mu_2'\ (2) = \Gamma(2) = 1,\ \frac{1}{n}\ \sum_1^n z_i \overset{P}{\to} \mu'_1\ (2) = \Gamma(3/2)$$

and $1_n(z) \overset{P}{\to} 0$ so that $t_n(x) \overset{P}{\to} \lambda + \chi\delta = \theta$.

As $\sqrt{n}(1_n(x) - \lambda)/\delta$ has the standard Rayleigh distribution (α=2) we can not expect tha asymptotic behaviour of $t_n(x)$ to be normal.

Let us obtain the asymptotic distribution of $\sqrt{n}(t_n(x) - \lambda - \chi\delta)/\delta =$ $= \sqrt{n}\ (t_n(z) - \chi)$, passing once more to standard Rayleigh random variables $(z_i = (x_i - \lambda)/\delta)$.

It is easy to show that

$$\sqrt{n}\ (\delta^*(z) - 1 - \frac{1}{2}\ (\frac{1}{n}\ \sum_1^n z_i^2 - 1 - \sqrt{\pi}1_n(z)) \overset{P}{\to} 0$$

as $\mu'_1\ (2) = \Gamma(3/2) = \sqrt{\pi}/2$ and $\mu'_2\ (2) = \Gamma(2) = 1$.

Consequently $\sqrt{n}(t_n(z) - \chi) - \sqrt{n}\ \{(1- \frac{\sqrt{\pi}}{2}\ )\ 1_n(z) + \frac{\chi}{2}\ (\frac{1}{n}\ \sum_1^n z_i^2 - 1)\} \overset{P}{\to} 0$

and the asymptotic distribution of

$\sqrt{n}\ (t_n(x) - \theta)/\delta = \sqrt{n}\ (t_n(z) - \chi)$ is the same as the one of

$$\sqrt{n}\ ((1 - \sqrt{\pi}/2)\ l_n(z) + \frac{\chi}{2}\ (\frac{1}{n}\sum_1^n z_i - 1)).$$

We will show, below, that $\sqrt{n}\ l_n(z)$ and $\frac{1}{n}\sum_1^n z_i - 1$ are asymptotically independent.

Then, as $\sqrt{n}\ l_n(z)$ is a standard Rayleigh random variable and

$\sqrt{n}\ (\frac{1}{n}\sum_1^n z_i - 1)$ is asymptotically standard normal.

we get

$$V(t_n(z)) \sim \{(1 - \sqrt{\pi}/2\ \chi)^2\ (1 - \frac{\pi}{4}) + \frac{\chi^2}{4}\}/n$$

so that the asymptotic (variance) efficiency is

$$\text{eff} \sim \frac{\chi^2/4}{\chi^2/4 + (1 - \sqrt{\pi}_2\chi)^2(1 - \pi/4)} = \frac{\chi^2}{\chi^2 + (4 - \pi)\ (1 - \sqrt{\pi}/2.\chi)^2}$$

whose asymptotic value (as $\chi\to\infty$) is $\frac{4}{4+4\pi - \pi^2} = \frac{4}{8-(\pi-2)^2} = 0.60.$

Recall that, the value of the variance efficiency is misleading because the asymptotic distribution of $\underline{t}$ is not normal, as shown below.

A new estimator must be searched in the Rayleigh case.

It was proved in Tiago de Oliveira (1962) and Rosengard (1966) that the minimum and the average are asymptotically independent. Let us prove it, anew, in a simple and direct way.

Let us consider the moment generating function of $\sqrt{n}\ l_n(z)$ and

$$\sqrt{n}\ (\frac{1}{n}\sum_1^n z_i^2 - 1)$$

$$\phi_n(\theta, \psi) = M\ (e^{\theta\sqrt{n}\ l_n(z) + \psi\sqrt{n}\ (\frac{1}{n}\sum_1^n z_i^2 - 1)}) =$$

$$\int_0^{+\infty}\cdots\int \Pi\, dz_i\ e^{\theta\sqrt{n}\ \min\ (z_i) + \psi\sqrt{n}\ (\frac{1}{n}\sum_1^n z_i^2 - 1)}\ \pi(2\ z_i\ e^{-z_i^2}) =$$

$$= n \int_0^{+\infty}\cdots\int dz_1 \prod_2^n dz_i\ e^{\theta\sqrt{n}\ z_1 + \psi\sqrt{n}\ (\frac{z_1^2}{n} + \frac{1}{n}\sum_2^n z_i^2 - 1)}\times$$

$$2\ z_i\ e^{-z_1^2}\ 2^{n-1} \prod_2^n (z_i .\ e^{-\sum_2^n z_i^2}) \text{ in the domain } 0<z_1<z_i | i>1 =$$

$$= n \int_0^{+\infty} dz_1\ 2z_1\ e^{-z_1^2}\ e^{\theta\sqrt{n}\ z_1 + \frac{\psi}{\sqrt{n}}\ z_1^2 - \psi\sqrt{n}} (\int_{z_1}^{+\infty} 2\ e^{-(1 - \psi/\sqrt{n})z^2}\, dz)^{n-1} =$$

$$= \frac{n\ e^{-\psi\sqrt{n}}}{(1 - \psi/\sqrt{n})^{n-1}} \int_0^{\infty} dz_1\ 2dz_1\ e^{\theta\sqrt{n}\ z_1 - n(1 - \psi/\sqrt{n})z_1^2} \to e^{\psi^2/2} \times \int_0^{+\infty} e^{\theta t - t^2} \times 2t\,dt$$

which shows the asymptotic independence of the two summands.

Thus the distribution of $\sqrt{n}\,(t_n(x) - \theta)/\delta$ is, asymptotically, the one of $(1 - \sqrt{\pi}/2.\ \chi)\ R + \chi/2\ N$ where R and N are independent, R is a standard Rayleigh random variable and N a standard normal random variable.

The asymptotic characteristic function of $(1 - \sqrt{\pi}/2.\chi)\ R + .\chi\ _2\ N$ is, thus,

$$\bar{\phi}(s) = e^{-s^2/2} \times \int_0^{+\infty} e^{ist - t^2} \times 2\ tdt$$

and the corresponding density, obtained by inversion, is

$$f(x) = \frac{1}{2\pi} \int_{\infty}^{+\infty} e^{-i*s}\ \bar{\phi}(s)\ ds = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} t\ e^{-t^2 - (t - \chi)^2/2} dt.$$

REFERENCES.

CRAMÉR, H., Mathematical Methods of Statistics , Princeton University Press, 1946.

DUBEY, S., "Hyper efficient estimator on the location parameter of the Weibull laws", Nav. Res. Log. Quat., vol. 13 (1966).

MANN, N.R., SCHEFER, R.E. and SINGPURWALLA, N.D., Methods for Statistical Analysis of Probability and Life Data , Wiley & Sons, 1974.

ROSENGARD, A., Contributions à l'étude des liaisons limites entre différents statistiques d'un échantillon, Thése, Institut de Statistique de l'Université de Paris, 1966.

TIAGO DE OLIVEIRA, J., "The asymptotic independence of the mean and extreme values",Univ. Lisboa, Rev. Fac. Ciênc. 2(A), 8, (1962).

TIAGO DE OLIVEIRA, J. and LITTAUER, S.B., "Mean square invariant forecasters for the Weibull distribution", Nav. Res. Log. Quat., vol. 23, nº 3, Sept., (1976).

TIAGO DE OLIVEIRA, J., "Initiation of statistical decision for Weibull distribution", presented at the VIII Reunion Nacional de Estadistica, Investigation Operativa y Informatica, Jaca (Spain), (1980).