Computer Modelling and New Technologies, 2008, Vol.12, No.3, 7–15
Transport and Telecommunication Institute, Lomonosova 1, LV-1019, Riga, Latvia

BAYES-FIDUCIAL APPROACH FOR MATHEMATICAL STATISTICS PROBLEM SOLUTION

Yu. Paramonov

Aviation Institute, Riga Technical University
Lomonosova Str. 1, Riga LV-1019, Latvia
Phone: +371 67255394. Fax: +371 67089990. E-mail: rauprm@junik.lv

Review of using of Bayes-Fiducial (BF) approach for solution of some important statistical problem is given. BF decision is always a function of sufficient statistics (even in case when sufficient statistics coincides with the sample itself (for example, for Weibull's distribution)). By contrast with maximum likelihood method BF decision is based on the use of specific loss function. By contrast with Bayes decision it does not need a priori distribution of unknown parameters.

For the distributions with location and/or scale parameter the solutions of the following problems are considered: point estimates of location and scale parameters, p-bound (prediction limit) calculation and specified life nomination with specific loss function. Numerical examples are given.

Keywords: *parameter estimation, testing statistical hypotheses, prediction limit*

1. Introduction

In the name of Bayes-Fiducial (BF) approach the word "Bayes" means that the unknown distribution parameter is considered as random variable (just as in Bayes approach); the word Fiducial means that instead of a priori or a posterior distribution the fiducial distribution is used. Let us remind the main ideas of BF approach, which initially (some version of this approach was initially called quasi-Bayesian) has been offered in 1973 (see [1]). Then it was developed in [2, 3]. In [4] application of this approach to the problem of unbiased estimation when the sufficient statistics coincides with the sample itself was considered.

Let us remind the main idea of this approach for solution of the problem of looking for optimal statistical decision when some loss function is defined. Let's assume that the unknown parameter, θ, of some cumulative distribution function (cdf) of a random vector (rv) $X = (X_1,..., X_n)$, $F(x,\theta)$, is itself a random variable (rv), $\tilde{\theta}$, the probability density function (pdf) of which is fiducial pdf $f_{\tilde{\theta}}(s;x)$. And suppose also that the consequence of taking decision d when the cdf of rv X is $F(x,\theta)$ is a loss, which can be expressed as a nonnegative real number $L(\theta,d)$. Let a decision $d = \delta(x)$, that is, a function whose domain is the set of values of X and whose range is the set of possible decisions. An optimum procedure that minimizes so called Bayes-Fiducial risk of δ

$$R(\theta,\delta) = \int L(s,\delta(x)) f_{\tilde{\theta}}(s;x)ds \tag{1}$$

is called a BF solution of the given decision problem.

Let us remind also a definition of fiducial distribution. It is easy to see that sometimes cdf of random variable *X*, $F_X(x,\theta)$, with one dimensional parameter θ is such that at the fixed *x* the function

$$F_{\tilde{\theta}}(\theta;x) = 1 - F_X(x,\theta)$$

is a cdf of r.v. $\tilde{\theta}$. This distribution is called the fiducial distribution by Fisher and it has been used for construction of interval for unknown parameter similar to confidential interval. Easily we can use this definition of fiducial distribution and for the case when we have some sample of size n>1 if there is one dimensional sufficient statistics. For example, if r.v. Y_i has normal distribution $N(\theta,1)$, $i = 1,...,n$,

$X = \sum_{i=1}^{n} Y_i / n$, then

$$F_X(x,\theta) = \Phi((x-\theta)/\sqrt{1/n})$$

and

$F_{\tilde{\theta}}(s,x) = 1 - \Phi((x-s)/\sqrt{1/n})$

is some cdf of r.v. $\tilde{\theta}$.

In more complex case when sufficient statistics coincides with the sample itself (for example if cdf is the smallest extreme value distribution) but unknown parameters are location and scale parameters are considered in [2, 3].

Let

$$F_{X_i}(x,\theta) = F_{\overset{o}{X}}\left(\frac{x-\theta_0}{\theta_1}\right), \mathrm{i} = 1,\dots,\mathrm{n}, \tag{2}$$

where $F_{\overset{0}{X}}(\cdot)$, are known c.d.f. of $\overset{0}{X}$, θ_0, θ_1 – are unknown location and scale parameters. If both parameters θ_0 and θ_1 are unknown then fiducial pdf of ($\tilde{\theta}_0$, $\tilde{\theta}_1$) is defined by equation

$$f_{\tilde{\theta}_0,\tilde{\theta}_1|x}(s_0,s_1) = h\frac{1}{s_1^{n+1}}\prod\nolimits_{i=1}^{n} f\left(\frac{x_i - s_0}{s_1}\right). \tag{3}$$

If θ_1 is known then without loss of generality we can set $\theta_1 = 1$ then fiducial pdf of $\tilde{\theta}_0$ is defined by equation

$$f_{\hat{\theta}_0|x}(s) = h_0 \prod\nolimits_{i=1}^{n} f\left(x_i - s\right), \tag{4}$$

if θ_0 is known then without loss of generality we can set $\theta_0 = 0$ then fiducial pdf of $\tilde{\theta}_1$ is defined by equation

$$f_{\tilde{\theta}_1|x}(s) = h_1 \frac{1}{s^{n+1}}\prod\nolimits_{i=1}^{n} f\left(x_i / s\right), \tag{5}$$

where h, h_0, h_1 are just normalization factors, $x = (x_1,\dots,x_n)$.

In this paper we consider application of BF method to parameter estimation, p-bound calculation and specified life nomination.

2. Parameter Estimation

Usually for parameter estimation the loss function is defined by equation

$$L(\theta,\hat{\theta}) = (\theta - \hat{\theta})^2. \tag{6}$$

In this case BF risk is equal to

$$R(\theta,\hat{\theta}) = \int (s-\hat{\theta})^2, f_{\tilde{\theta}}(s;x)ds.$$

For the case when there is only one unknown parameter θ_0 this equation takes the following form

$$R(\theta_0,\hat{\theta}_0) = \int (s-\hat{\theta}_0)^2, f_{\tilde{\theta}_0}(s;x)ds, \tag{7}$$

where $f_{\tilde{\theta}_0}(s;x)$ is defined by equation (4).

It is easy to see that minimum of BF risk in this case is reached at

$$\hat{\theta}_0 = \int_{-\infty}^{\infty} s f_{\hat{\theta}_0}(s;x)ds. \tag{8}$$

Using equation (4) we get finally

$$\hat{\theta}_0 = \int_{-\infty}^{\infty} s\prod\nolimits_{i=1}^{n} f\left(x_i - s\right)ds \Big/ \int_{-\infty}^{\infty}\prod\nolimits_{i=1}^{n} f(x_i - s)ds. \tag{9}$$

This estimate coincides with the Pitmen's estimate of location parameter, which provides the minimum of risk in the class of corrected estimate (see Section 3).

Pitmen's estimate of the scale parameter we get if we use the following definition of loss function

$$L(\theta,\hat{\theta}) = (\theta - \hat{\theta})^2 / \theta^2 . \qquad (10)$$

Using the BF approach we get

$$\hat{\theta}_1 = \int_0^{\infty} s^{-(n+2)} (\prod_{i-1}^{n} f(x_i / s))ds / \int_0^{\infty} s^{-(n+3)} (\prod_{i=1}^{n} f(x_i / s))ds .$$

3. P-bound for Random Variables

3.1. Definition of P-Bound for RV

To make possible the common approach for solution of the both problem SL nomination and IP development we need to remind the *p*-set function definition [4]. It is a special statistical decision of P-bound for random variable is a specific case of a *p*-set function, definition of which was introduced by the author [4]. In this paper we limit our self by only application of BF approach for calculation of p-bound.

P-bound is defined in the following way.

Definition. Let Z is a random variable and X is random vector of n dimension and we suppose that it is the known class { P_θ , $\theta \in \Omega$ } to which the probability distribution of the random vector $W = (Z, X)$ is assumed to belong. Of the parameter θ, which labels the distribution, it is assumed known only that it lies in a certain set Ω , the parameter space.

a) Function $\tau(x)$ is called a p-bound for r.v. Z if

$$\sup_{\theta \in \Theta} P_\theta\{Z < \tau(X)\} = p . \qquad (11)$$

b) Function $\tau(x)$ is called a parameter-free (p.f.) p-bound for r.v. Z if

$$P_\theta\{Z < \tau(X)\} = p \text{ for all parameters } \theta \in \Omega . \qquad (12)$$

c) P-bound for r.v. Z is called a right-hand binary (r.h.b. p-bound), if for each possible observation x of r.v. X, function $\tau(x)$ assigns only one of two decisions:

$$\tau(x) = -\infty \text{ if } x \in S; \ \tau(x) = \tau^*, \text{ if } x \in S^*, \qquad (13)$$

where τ^* is some number, S^* and S are two complementary regions of the sample space .

So we see that the definition of p-bound can be considered as some generalization of the definition of prediction limit. But it is some statistical decision function, which covers both prediction limit and, in some way, testing statistical hypotheses.

We can say also that p.f. p-bound $\tau(x)$ is a p-quantile estimate of cdf $F_Z(x)$ and, as function of p, it is an estimate of inverse cumulative distribution function $F_Z^{-1}(p)$, but very specific estimate: expectation value $E(F_Z(\tau(X))) = p$.

It is easy to get $\tau(x)$ for distribution with location and scale parameters. As the main application of the problem under question we'll consider a problem of SL nomination for some fatigue-prone airframe structure. We suppose to have observations of fatigue lives of some identical units of this structure as a result of full-scale fatigue tests. Usually for fatigue life data processing both a lognormal and Weibull's distributions are used. If we use logarithm scale (if we use $X = \ln(T)$ instead of T) then both these distributions become distributions with location and scale parameters. So we can say, that r.v. X has the following structure: $X = \theta_0 + \theta_1 \overset{0}{X}$, where θ_0 , θ_1 are unknown parameters, r.v. $\overset{0}{X}$ has either standard normal c.d.f. $F_{\overset{0}{X}}(x) = \Phi(x)$ or standardized the smallest extreme value (sev) c.d.f. $F_{\overset{0}{X}}(x) = 1 - \exp(-\exp(x))$ for lognormal or Weibull's distributions of T correspondingly. For this case for the specified life nomination problem the following theorem can be used (we give it without proof).

Theorem 1. Let

$$F_{X_i}(x,\theta) = F_{\overset{o}{X}}\left(\frac{x-\theta_0}{\theta_1}\right), \mathrm{i}=1,\ldots,\mathrm{n}, \qquad F_Z(x,\theta) = F_{\overset{o}{Z}}\left(\frac{x-\theta_0}{\theta_1}\right), \tag{14}$$

where $F_{\overset{0}{X}}(\cdot)$, $F_{\overset{0}{Z}}(\cdot)$ are known c.d.f. of $\overset{0}{X}$, $\overset{0}{Z}$, θ_0, θ_1 – are unknown location and scale parameters. And let the random variables, estimations of θ_0, θ_1, as function of $X = (X_1, X_2, \ldots, X_n)$ be described by the similar structural formulas:

$$\hat{\theta}_0 = \theta_0 + \theta_1 \overset{o}{\theta}_0, \quad \hat{\theta}_1 = \theta_1 \overset{o}{\theta}_1, \tag{15}$$

where $\overset{o}{\theta}_0, \overset{o}{\theta}_1$ – are random variables, corresponding to the estimates of θ_0, θ_1 using a sample of the same size n but when $\theta_0 = 0$, $\theta_1 = 1$. We refer to this type of estimates as "correct" estimates.

Then p.f. and r.h.b. p-bounds are described accordingly by formulae

$$\tau_1(x) = \hat{\tau}_1, \quad \tau_2(x) = \begin{cases} -\infty, & \hat{\tau}_2 \le \tau^*, \\ \tau^*, & \hat{\tau}_2 > \tau^*, \end{cases} \tag{16}$$

where $\hat{\tau}_i = \hat{\theta}_0 + t_i \hat{\theta}_1$, $i = 1,2$,

t_1 is p-quantile of r.v. $V_Z = (\overset{0}{Z} - \overset{0}{\theta}_0)/\overset{0}{\theta}_1$, t_2 is the root of equation: $\xi(t) = p$,

$$\xi(\mathrm{t}) = \sup_{\mathrm{c}} F_{\overset{o}{Z}}(c)(1 - F_{\overset{o}{\tau}(t)}(c)) = \sup_{\mathrm{c}} F_{\overset{o}{Z}}(c) F_{V_C}(t),$$

$$\overset{o}{\tau}(t) = \overset{o}{\theta}_0 + \overset{o}{\theta}_1 t, \quad \mathrm{V_C} = (c - \overset{o}{\theta}_0)/\overset{o}{\theta}_1.$$

3.2. Optimality Criterion for P.F. P-bound Used for Aircraft Specified Life Nomination

Now we turn to a discussion of some preference orderings of decision procedures: choice of function $\tau(x)$. In framework of theorem 1 it is really the choice of estimates $\hat{\theta}_0$, $\hat{\theta}_1$ and loss function. Let $X = (X_1, X_2, \ldots, X_n)$, where X_i, $i = 1,\ldots,n$, are fatigue lives of aircraft in (full-scale) laboratory test, $Z = \min(Y_1, Y_2, \ldots, Y_m)$, where Y_j, $j = 1,\ldots,m$, are fatigue lives of aircraft in operation, $F_{X_i}(t) = F_{Y_j}(t)$, $i = 1,\ldots,n$, $j = 1,\ldots,m$; p- allowed probability of failure in operation of at least one aircraft.

In application to the problem of required SL confirmation, when τ^* is required SL, we are interested in increasing of probability that $\tau(x) = \tau^*$. It is something similar to increasing of power of some test in testing some statistical hypothesis.

In application to the problem of some SL nomination we should get the maximum of expectation value of $\tau(X)$ provided that reliability requirements are met, it is if $\tau(X)$ is a p-bound for Z. To study the optimality of $\tau(x)$ we can use the Jensen's inequality. This inequality says that the function of complete sufficient statistics, which is unbiased estimation of its own mathematical expectation, provides the minimal risk if the correspondent loss-function is convex. Consider the simplest case, when θ_1 is a known parameter. Let $\theta_t = \theta_0 + t\theta_1$ is some quantile. Random variable $\hat{\theta}_t = \tau(x) = \hat{\theta}_0 + t\theta_1$ is unbiased estimate of its own expectation (which in general *-case is not equal to θ_t). In problem under question the function $F_Z(\tau)$ can be considered as the loss-function. Then the expectation $E_X\{F_Z(\hat{\theta}_t)\} = P(Z < \tau(X))$ is the risk function. For normal and the smallest extreme value (sev) distributions of Y_j $j = 1,\ldots,m$, $F_Z(\tau)$ is convex (and increasing one) if its value is small enough and we have minimum of $E_X\{F_Z(\hat{\theta}_t)\} = P(Z < \tau(X)) = p$ at the fixed expectation value of $\hat{\theta}_t = \tau(X)$, if $\tau(x)$ is a function of sufficient statistic. And, on the contrary, if $\tau(x)$ is a function of sufficient statistic and $P(Z < \tau(X)) = p$ then we have maximum of expectation value of $\tau(X)$ if p is small enough and probability $P(\tau(X) < c)$ is high enough for such c, that $F_Z(z)$ is convex if $z < c$. For example, for normal distribution $\Phi(z)$ is convex if z < 0. The generalization of the Jensen's inequality for the case of multivariate sufficient statistic can be found in [7].

For the case when sufficient statistics coincides with the sample itself (for example, Weibull's or the smallest extreme value (sev) distribution usually for prediction interval the Monte Carlo (MC) method is used [10]. Here we show that for the problem of p.f. p-bound, $\tau(x)$, calculation analytic solution can be found using Bayes-fiducial (BF) approach.

3.3. Bayes-Fiducial Approach for P.F. P-bound Calculation

Let the problem is to estimate p-quantile $\tau_p(\theta)$ for cdf $F_Z((x-\theta_0)/\theta_1)$ and loss function $L(\tau_\theta(\theta), \tau_X(x)) = (F_Z((\tau_p - \theta_0)/\theta_1) - F_Z((\tau_X(x) - \theta_0)/\theta_1))^2 = (p - F_Z((\tau_X(x) - \theta_0)/\theta_1))^2$

when we have sample $x = (x_1, x_2, ..., x_n)$ from cdf $F_X((x-\theta_0)/\theta_1)$.

Let us denote by $\tau_X(x, p)$ the solution of BF equation, corresponding to the considered loss function

$$E_{\tilde{\theta}}\{F_Z((\tau_X(x, p) - \tilde{\theta}_0)/\tilde{\theta}_1\} = p\,, \tag{17}$$

where $\tilde{\theta} = (\tilde{\theta}_0, \tilde{\theta}_1)$, r.v. $\tilde{\theta}_0$, $\tilde{\theta}_1$ have fiducial distribution. Here $E_X(f(X))$ is expected value of $f(X)$ in accordance with cdf of *X*.

We can simplify solution of Eq. 8. Instead of vector $x = (x_1, ..., x_n)$ without loss of information we can consider vector $\varpi = (\hat{\theta}_0, \hat{\theta}_1, w_1, ..., w_{n-2})$, where $\hat{\theta}_0, \hat{\theta}_1$ are correct parameter estimates (see (15)), $w_i = (x_i - \hat{\theta}_0)/\hat{\theta}_1$, $i = 1, ..., n-2$. Then conditional fiducial distribution (at the fixed invariant $(w_1, ..., w_{n-2})$) of random variables $\tilde{\theta}_0, \tilde{\theta}_1$ can be defined in the following form [3]

$$f_{\tilde{\theta}_0,\tilde{\theta}_1|w_1,...,w_n}(s_0, s_1) = h\frac{\hat{\theta}_1^{\,n-1}}{s_1^{n+1}} \prod_{i=1}^{n} f\left(\frac{\hat{\theta}_0 + \hat{\theta}_1 w_i - s_0}{s_1}\right) ds_0 ds_1\,,$$

where *h* is just normalization factor. (Note: w_{n-1}, w_n, $w_i = (x_i - \hat{\theta}_0)/\hat{\theta}_1$, are functions of vector ϖ).

If in (17) we use new notations:

$$U_0 = (\hat{\theta}_0 - s_0)/\hat{\theta}_1,\quad U_1 = \hat{\theta}_1/s_1\,,\ \overset{0}{\tau}(x, p) = (\tau(x, p) - \hat{\theta}_0)/\hat{\theta}_1\,,$$

then instead of (17) we get equation

$$E_{W_1,...,W_n} E_{U_0 U_1|w_1,...,w_n}\left(F((\overset{0}{\tau}(x, p) - U_0)/U_1)\right) = p\,, \tag{18}$$

where random variables U_0, U_1 has conditional pdf

$$f_{U_0,U_1|w_1,...,w_n}(u_0, u_1) = h_w u_0^{n-2} \prod_{i=1}^{n} f(u_0 + w_i u_1)\,, \tag{19}$$

where h_w is just normalization factor which depends only on invariant vector $w = (w_1, ..., w_{n-2})$.

If $\overset{0}{\tau}(x, p)$ is solution of the equation

$$E_{U_0 U_1|w_1,...,w_n}\left(F((\overset{0}{\tau}(x, p) - U_0)/U_1)\right) = p\,, \tag{20}$$

then

$$\tau_X(x, p) = \hat{\theta}_0 + \overset{0}{\tau}(x, p)\hat{\theta}_1 \tag{21}$$

is solution of Eq. (20) and Eq.(17) because equation (20) takes place for every vector $w = (w_1, ..., w_{n-2})$, cdf of which does not depend on $\theta = (\theta_0, \theta_1)$. So if (20) is true, consequently (17) is true also.

It is very important that $\tau(\overset{0}{x}, p)$ in (20) does not depend on the value of $\theta = (\theta_0, \theta_1)$ and for solution of this equation we can set $\theta_0 = 0$, $\theta_1 = 1$. If $\hat{\theta}_0, \hat{\theta}_1$ have the structures defined by (15), then probability $P(Z < \tau(X, p))$ does not depend on $\theta = (\theta_0, \theta_1)$ and we can find p_1 for which

$$P(Z < \tau(X, p_1)) = p.$$

So $\tau_X(x, p_1)$ is p-bound for random variable Z.

As it is easy to see (see [3; p. 84]) the pdf (19) is conditional pdf of $\hat{\theta}_0, \hat{\theta}_1$ at the fixed $w = (w_1, ..., w_{n-2})$ for the case when $\theta_0 = 0$, $\theta_1 = 1$. This means that the values of p_1 and p coincide.

It is very important also that result does not depend on the choice of the type of correct statistics $\hat{\theta}_0, \hat{\theta}_1$ (see eq. 23.a and 23.b), because vector $x = (x_1, ..., x_n)$ and vector $\varpi = (\hat{\theta}_0, \hat{\theta}_1, w_1, ..., w_{n-2})$ have one-one mapping at any choice of correct statistics.

3.3.1. __Example 1__. P-bound for Lognormal Distribution

Let r.v. T have a lognormal distribution and $t = (t_1, t_2, t_3) = (45\,952,\ 54\,143,\ 65\,440)$ is the sample from the same distribution. Then r.v. $X = \log(T)$ has a normal distribution $N(\theta_0, \theta_1^2)$ and $x = (x_1, x_2, x_3) = (10.735\ 10.899\ \ 11.089)$ is the sample from this distribution. The problem is to calculate the p.f. p-bound for independent r.v. $Z = \min(Y_1, ..., Y_m)$, where r.v. Y_i, $i = 1, ..., m$, has the normal distribution $N(\theta_0, \theta_1^2)$ also. We consider here just the case, when m = 1, because for this case there is a general analytical solution (see, for example, [3; p. 172])

$$\tau(x) = \hat{\theta}_0 + \hat{\theta}_1 t_{n-1,p} (1 + 1/n)^{1/2}, \qquad (22)$$

where

$$\hat{\theta}_0 = \bar{x}\ ,\ \hat{\theta}_1 = \left(\sum (x_i - \bar{x})^2 / (n-1)\right)^{1/2}$$

are estimates of expected value and standard deviation, $t_{k,q}$ is q-quantile from Student's distribution with k degree of freedom. So we can make comparison of this solution with the solution which we get by using new approach.

For considered data, using equation (22) for $p = 0.01$ we calculate t_{St} = exp($\tau(x)$) = 13 162, which is the value of p-bound for r.v. T on the base of observations (t_1, t_2, t_3) .

Now let us consider the new approach. For normal distribution the conditional pdf has the following form

$$f_{U_0, U_1 | w_1, ..., w_n}(u_0, u_1) = h_w u_0^{n-2} \prod_{i=1}^{n} \varphi(u_0 + w_i u_1),$$

where $\varphi(x) = \exp(-x^2/2)/(2\pi)^{1/2}$. After transformation the equation (11) has the following form

$$1 - a(\overset{0}{\tau}, \bar{z}, D_z) / \Gamma((n-1)/2) = p,$$

where

$$a(\overset{0}{\tau}, \bar{z}, D_z) = \int_0^\infty u^{(n-3)/2} \exp(-u) \Phi\left((2u / D_z (n+1))^{1/2} (\bar{z} - \overset{0}{\tau}) \right) du,$$

$$\bar{z} = \sum_1^n z_i / n,\ D_z = \sum_{i=1}^n (z_i - \bar{z})^2 / n,$$

$\Gamma(\cdot)$ is gamma function, $\Phi(\cdot)$ is cdf of standard normal distribution.

Consider two types of statistics $\hat{\theta}_0, \hat{\theta}_1$, which for the considered data have the following values:

a) $\hat{\theta}_0 = \bar{x} = 10.908$, $\hat{\theta}_1 = \left(\sum (x_i - \bar{x})^2 / (n-1)\right)^{1/2} = 0.177$, (23.a)

b) $\hat{\theta}_0 = x_{1,n}$ = 10.735, $\hat{\theta}_1 = x_{n,n} - x_{1,n}$ = 0.354, (23.b)

where $x_{i,n}$ is i-th order statistics of vector $x = (x_1,...,x_n)$.

In case a) we have $\overset{0}{\tau}$ = – 7.889, in case b) we have $\overset{0}{\tau}$ = – 3.560.
Corresponding values of p-bound for r.v. T on the base of observations (t_1, t_2, t_3) are as follows:

$t_a =$ exp($\tau(x)$) = 13 523, t_b = exp($\tau(x)$) = 13 050.

It seems that the difference between t_a , t_b and t_{St} = 13 162 is produced only by the problem to get the required calculation accuracy.

3.3.2. <u>Example 2.</u> P-bound for SEV and WEIBULL'S distribution

Let we have the same sample $t =$ (t_1, t_2, t_3) = (45 952, 54 143, 65 440) or $x =$ (x_1, x_2, x_3) = (10.735 10.899 11.089) but r.v. T has a Weibull's distribution and, correspondingly $X = \log(T)$ has distribution of the smallest extreme value (sev) with cdf $F_X(x) = 1 - \exp(-\exp((x - \theta_0)/\theta_1)$. In this case the equation (20) has the following form

$$1 - a(\overset{0}{\tau}, \bar{z}, D_z) / b(\bar{z}, D_z) = p,$$

where

$$a(\overset{0}{\tau}, \bar{z}, D_z) = \int_0^\infty u^{(n-2)} \Big(\exp(-u\sum_{i=1}^n z_i) / (\sum_{i=1}^n \exp(uz_i) + m\exp(u\overset{0}{\tau}))^n \Big) du,$$

$$b(\bar{z}, D_z) = \int_0^\infty u^{(n-2)} \Big(\exp(-u\sum_{i=1}^n z_i) / (\sum_{i=1}^n \exp(uz_i))^n \Big) du,$$

$$\bar{z} = \sum_1^n z_i / n, \quad D_z = \sum_{i=1}^n (z_i - \bar{z})^2 / n.$$

For m = 1, p = 0.01, using statistics (23.a) we get $\overset{0}{\tau}$ = – 11.929, using statistics (23.b) we get $\overset{0}{\tau}$ = – 5.424.
Corresponding values of p-bound for r.v. T on the base of observations (t_1, t_2, t_3) are as follows:

$t_a =$ exp($\tau(x)$) = 6 616, t_b = exp($\tau(x)$) = 6 752.

For m = 500, p = 0.2 using statistics (23.a) we get $\overset{0}{\tau}$ = – 12.889, using statistics (23.b) we have $\overset{0}{\tau}$ = – 5.970.
Corresponding values of p-bound for r.v. T on the base of observations (t_1, t_2, t_3) are:

$t_a =$ exp($\tau(x)$) = 5 584, t_b = exp($\tau(x)$) = 5 568.

Again, it seems that the difference between t_a and t_b is produced only by the problem to get required calculation accuracy.

Considered data really ware considered in several papers and for m = 500, p = 0.2 Lowless (1973) obtained prediction limit of 5623, Mee and Kushary (1994) – 5225. The Mann and Saunders (1969) result was only 766. For these calculations the Monte Carlo method was used [8].

4. Using Bayes-Fiducial Method for SL Nomination with Optimality Criterion of Economics

Let the income of aircraft successful service during time t is equal to t but in case of failure the loss is equal to some negative value – b, where b is some large positive value. Then the income of one aircraft service, r.v. U, is defined by formula

$$U = \begin{cases} t_{SL}, & if\ T > t_{SL}, \\ T - b, & if\ T \le t_{SL} \end{cases},$$

where T is random fatigue life, t_{SL} is some SL.

Expectation value of U

$$u(t_{SL},\theta, b) = \int_0^{t_{SL}} (t - b)\, d F_T(t, \theta) + t_{SL} (1 - F_T (t,\theta)),$$

where $F_T(t,\theta)$ is c.d.f. of T.

In general case maximum of $u(t_{SL}, \theta, b)$ is reached at t^*_{SL}, which is the root of the equation:

$$bf_T(t)/(1 - F_T(t,\theta)) = 1 .$$

For normal distribution of $X = \ln T$ it can be written in the following way

$$\theta_0 = t^*_{SL} - \theta_1 \lambda^{-1} (t^*_{SL}\, \theta_1 /b),$$

where $\lambda(z) = \varphi(z)/(1-\Phi(z))$ is failure rate function for standard normal distribution, $\lambda^{-1}(.)$ is inverse function. This equation allows very easy to get θ_0 as function of t^*_{SL} at the fixed θ_1 and then to find the inverse function:

$$t^*_{SL} = S^*(\theta_0, \theta_1, b) .$$

For b = 346 000, θ_1 = 0.346 and θ_0 = 9.948 we have: t^*_{SL} = 7936 (flights). It is of interest to note that this value corresponds to the failure probability equal to 0.0026. This can be interpreted in the following way. The failure of 2.6 aircraft (in flight) from 1000 aircraft can be considered as equivalent to the loss of 346000 hours of service time or loss of 346000/7936 = 43.6 aircraft (on the ground) of this types (the value t^*_{SL} = 7936 can be considered as the price of one aircraft of this type). Or in other words, failure of one aircraft (in flight) is equivalent to loss of 43.6/2.6 (approximately 16) aircraft of the same type (on the ground).

But we do not know the parameters of c.d.f. of T and should estimate them using fatigue test data. Usually maximum likelihood estimate is considered as most appropriate. We show here that for considered problem the offered by a Bayes-fiducial approach is much more appropriate.

In accordance with Bayes's approach the parameter θ_0 is r.v. For the case of airframe it can be interpreted in the following way. Design stress analysis of an airframe should meet some standard requirements (FAR, ...). These requirements in fact define only some mean value of θ_0 but of course, in every case there are some "occasional mistakes" and we have some specific (random) value of θ_0 for every aircraft type. And then there is a scatter of r.v. X (specific random fatigue life of some specific aircraft) at this random θ_0. The parameter θ_1 is function of technology level, and if one is not changed, then the parameter θ_1 is not changed as well. So we suppose that θ_1 is known constant but θ_0 is random variable, $\tilde{\theta}_0$. Let $\pi(\theta_0)$ is a priori distribution density for $\tilde{\theta}_0$. Then c.d.f. of r.v. X will be

$$\tilde{F}_X(x) = \int_{-\infty}^{\infty} F_X((x - \theta_0)/\theta_1)\, \pi(\theta_0)\, d\theta_0 .$$

It is well known, that if θ_1 is constant but r.v. $\tilde{\theta}_0$ has normal distribution with known both mean τ_0 and standard deviation τ_1, then distribution of X will be again normal with mean value τ_0 and standard deviation $((\tau_1)^2 + (\theta_1)^2)^{1/2}$. In this case t_{SL} again will be defined by equation (1), but parameter θ_1 should be replaced by $\theta_{1\tau} = ((\tau_1)^2 + (\theta_1)^2)^{1/2}$.

In fact we do not know a priori distribution of $\tilde{\theta}_0$. For this case FB approach is offered. Instead of posterior distribution of $\tilde{\theta}_0$ we offer to use fiducial distribution. In considered case fiducial distribution of $\tilde{\theta}_0$ again is normal with mean $\bar{x}$ and standard deviation $\theta_1/n^{1/2}$. Then for the purpose of calculation t_{SL} we again can use the same equation (1), but θ_0, θ_1 should be replaced by $\hat{\theta}_0 = \bar{x}$ and $\theta_1 (1 + 1/n)^{1/2}$ correspondingly. So using sample $x = (x_1, ..., x_n)$, result of full-scale fatigue test, in case of ML approach the nominated SL is equal to $S^*(\bar{x}, \theta_1, b)$, but for BF approach $t_{SL}(x) = S^*(\bar{x}, \theta_1(1+1/n)^{1/2}, b)$. By the use of Monte Carlo method for θ_0 = 9.948, θ_1 = 0.346, b = 346,000 we have got that the expectation value of r.v. U_X is equal to 2310, 4122, 5571, 6904 for BF approach but it is equal to -8624, 809, 4422, 6935 for ML approach for the same sample sizes n = 1, 2, 4, 100. We see that for small n the expectation value of r.v. U_X is much more for BF than for ML approach.

Conclusions

BF approach has the following advantages:

1. As in a case of using a maximum likelihood (ML) estimates BF solution is always a function of sufficient statistics, but in contrast to ML the BF solution takes into account the specific loss function.

2. By contrast with Bayes's decision it does not need a priori distribution of unknown parameters.

It is given approximate analytical solution of the problem to get the maximum of expected value of SL is given of economics optimality criterion, it is shown also that for the considered type of loss function the BF approach is more preferable than the direct use of ML estimates. Numerical examples are given.

For the distributions with location or scale parameters it is shown that the BF point estimates of these parameters are equal to the corresponding Pitmen's estimates. BF approach can be used also for solution of the problem to get unbiased estimate as function of sufficient statistics when the sufficient statistics coincides with the sample itself.

References

1. Paramonov, Yu. M. *Planning and Processing of High Reliability Article Time Test Results: Thesis for a Doctor's degree*. Riga, 1973. 317 p. (In Russian)
2. Paramonov, Yu. M. *Application of Mathematical Statistics Methods for Estimation and Insurance of Aircraft Reliability*. Riga, 1976. 212 p. (In Russian)
3. Paramonov, Yu. M. *Application of Mathematical Statistics to the Problem of Aircraft Fatigue Life Estimation and Insurance*. Riga: Riga Aviation University, 1992. 248 p. (In Russian)
4. Paramonov, Yu. M. Unbiased Quasi-Bayesian Estimation*, Probability Theory and its Applications,* No 2, 1977, pp. 372–380. (In Russian)
5. Paramonov, Yu. M. P-set Function Application for Specified Life Nomination and Inspection Program Development. In: *Probabilistic Analysis of Rare Events: Theory and Problems of Safety. Insurance and Ruin. International Conference, Jurmala, Latvia, June 28 – July 3, 1999*. Riga: Riga Aviation University, 1999, pp. 202–208.
6. Paramonov, Yu. M., Paramonova, A. Yu. Inspection Program Development by the Use of Approval Test Results*, International Journal of Reliability, Quality and Safety Engineering. World Scientific Publication Company*, Vol. 7, No 4, 2000, pp. 301–308.
7. Klimov, G. P. *Probability Theory and Mathematical Statistics.* Moscow: The University of Moscow, 1983. 328 p. (In Russian)
8. Mee, R. W. Prediction Limits for the Weibull's Distribution Utilizing Simulation, *Computational Statistics&Data Analysis*, Vol. 17, 1994, pp. 327–336.

Received on the 1st of November, 2008