ACTUARIAL RESEARCH CLEARING HOUSE
1985 VOL. 2

Tolerance Intervals in Risk Theory*

Wolf-Rüdiger Heilmann

Institut für Mathematische Stochastik
Universität Hamburg
West Germany

Keywords: Credibility theory, Percentile principle, Tolerance interval, Failure rate.

* Lecture presented to the 19th Actuarial Research Conference, University of California, Berkeley, October 1984

Tolerance Intervals
in Risk Theory

Wolf-Rüdiger Heilmann
University of Hamburg

1. Abstract. Up to now, the main aim of credibility theory has been to provide statistical models which allow for estimating (net) risk premiums appropriately.
In the present note, a simple credibility model based on the percentile principle is introduced. It turns out that there are close connections between the resulting credibility premiums and statistical tolerance limits.

2. The credibility model. We consider the following simple model of credibility theory: Let $\Theta, X_1, X_2, \ldots$ be random variables on some probability space $(\Omega, \mathcal{A}, P)$ such that the "risks" $X_1, X_2, \ldots$ are real-valued, and given $\Theta = \vartheta$, $X = X_1, X_2, \ldots$ are iid with distribution-function F_ϑ, i.e.

$$F_\vartheta(x) = P^{X|\Theta=\vartheta}((-\infty, x]) = P(X \leq x | \Theta = \vartheta), \quad x \in \mathbb{R}.$$

Moreover, we assume that for all ϑ

$$F_\vartheta(0-) = P(X<0 | \Theta=\vartheta) = 0.$$

Now let H denote a real functional on some set of distribution functions containing the F_ϑ (a "premium calculation principle") and let $L: \mathbb{R}^2 \to \mathbb{R}$ be a "loss function". Then for fixed n, we are looking for a function $\pi : \mathbb{R}^n \to \mathbb{R}$ such that

$$\int L(H(F_\vartheta),\pi(x_1,\dots,x_n))P^{(\Theta,X_1,\dots,X_n)}(d(\vartheta,x_1,\dots,x_n))$$
$$= \int L(H(F_\vartheta),\pi(x_1,\dots,x_n))F_\vartheta(dx_1)\dots F_\vartheta(dx_n)P^\Theta(d\vartheta)$$
$$\overset{!}{=} \text{Min}.$$

(All necessary measurability conditions are assumed to hold.)
The minimum is achieved by the function π^* if and only if for $P^{(X_1,\dots,X_n)}$-almost all $(x_1,\dots,x_n)$ π^* minimizes the expression

$$\int L(H(F_\vartheta),\pi(x_1,\dots,x_n))P^{\Theta|(X_1,\dots,X_n)=(x_1,\dots,x_n)}(d\vartheta). \quad (1)$$

If $H(F_\vartheta) = E_\vartheta X = \int x F_\vartheta(dx)$ ("net premium principle"), an appropriate loss function is $L(a,b) = (a-b)^2$, and the usual credibility model is obtained.

3. The percentile principle. For a distribution function F, let F^{-1} denote the pseudo-inverse of F, i.e.

$$F^{-1}(p) = \inf\{t \in \mathbb{R} : F(t) \geq p\}, \quad p \in [0,1].$$

Then

$$F^{-1}(p) \leq x \Leftrightarrow p \leq F(x),$$
$$F(F^{-1}(p)) \geq p, \quad p \in (0,1),$$
$$F^{-1}(F(t)) \leq t, \quad t \in \mathbb{R}.$$

If F is continuous,

$$F(F^{-1}(p)) = p, \quad p \in (0,1),$$

if F is strictly increasing,

$$F^{-1}(F(t)) = t, \quad t \in \mathbb{R}.$$

Similar results can be obtained for other types of monotone functions.
For some fixed $\varepsilon \in (0,1)$ (usually close to 0), we now put

$$H(F) = F^{-1}(1-\varepsilon) \quad \text{(the } (1-\varepsilon)\text{-quantile of F)} \quad (2)$$

to obtain the percentile principle. Despite its formal simplicity and intuitive appeal, this principle is not very customary. The main reason for this seems to be that the percentile premium does not have the form "net premium plus loading". In general, H does not provide for a nonnegative safety loading and is neither additive (not even sub- or super-additive) nor iterative. It is consistent, homogeneous and fulfills the "no ripoff"-condition (cf. GERBER [6]). A nonnegative safety-loading could of course be achieved by simple modifications of (2), e.g.

$$H^*(F) = \begin{cases} EX & \text{if } EX \geq F^{-1}(1-\varepsilon) \\ F^{-1}(1-\varepsilon) & \text{if } EX < F^{-1}(1-\varepsilon) \end{cases} .$$

<u>Remark.</u> Let $\mu = EX = \int_0^\infty x\,F(dx) = \int_0^\infty (1-F(x))\,dx = \int_0^1 F^{-1}(s)\,ds$ be positive and finite, and for $p \in (0,1)$

$$l(p) = \frac{1}{\mu}\int_0^p F^{-1}(s)\,ds \quad (\text{"Lorenz curve"}),$$

$$t(p) = \frac{1}{\mu}\int_0^{F^{-1}(p)} (1-F(s))\,ds = l(p) + \frac{(1-p)F^{-1}(p)}{\mu} \tag{3}$$

("scaled total time on test transform"),

cf. Heilmann [9].

Then $H(F) \geq \mu \Leftrightarrow \dfrac{F^{-1}(1-\varepsilon)}{\mu} \geq 1$

$$\Leftrightarrow t(1-\varepsilon) - l(1-\varepsilon) \geq \varepsilon . \tag{4}$$

As an example, consider the Pareto-type distribution given by the failure rate function

$$r(t) = \frac{\beta}{\alpha+t}, \quad t \geq 0, \ \alpha > 0, \ \beta > 1, \tag{5}$$

which obviously is of type DFR.
Then

$$t(p) = 1-(1-p)^{(\beta-1)/\beta},$$
$$l(p) = \beta(1-(1-p)^{(\beta-1)/\beta}) - p(\beta-1).$$

For $\beta=2$, we obtain

$$t(p) = 1 - \sqrt{1-p},$$
$$l(p) = (1-\sqrt{1-p})^2,$$

and (4) is fulfilled if and only if $\varepsilon \leq \frac{1}{4}$. □

For $\alpha \in (0,1)$, let $d^{\vartheta}(\alpha)$ denote the α-quantile of F_{ϑ}, and $Z(\vartheta) = d^{\vartheta}(1-\varepsilon) = H(F_{\vartheta})$.

Now for real numbers a,b let

$$L(a,b) = \begin{cases} k_o(a-b) & \text{if } a \geq b \\ k_1(b-a) & \text{if } a < b \end{cases} \tag{6}$$

with positive constants k_o, k_1. With $q = \frac{k_o}{k_o+k_1}$, (1) is minimized if $\pi(x_1,\ldots,x_n)$ equals $d(q)$, the q-quantile of $P^{Z\circ\Theta|(X_1,\ldots,X_n)=(x_1,\ldots,x_n)}$ (cf. BERGER [4], FERGUSON [5]). Let $x^{(n)} = (x_1,\ldots,x_n)$, and assume that ϑ is a real parameter such that Z is continuous and strictly decreasing in ϑ. Then, since $Z(\vartheta) \leq d(q) \Leftrightarrow \vartheta \geq Z^{-1}(d(q))$,

$$P^{\Theta|(X_1,\ldots,X_n)=x^{(n)}}(\{\vartheta : Z(\vartheta) \leq d(q)\}) = q,$$

$\Leftrightarrow Z^{-1}(d(q))$ equals $d^{(n)}(1-q)$,

the (1-q)-quantile of $P^{\Theta|(X_1,\ldots,X_n)=x^{(n)}}$

$\Leftrightarrow d(q) = Z(d^{(n)}(1-q))$.

Thus,

$$\pi(x^{(n)}) = d^{d^{(n)}(1-q)}(1-\varepsilon) \tag{7}$$

is a solution to the above minimization problem and the desired credibility premium.

<u>Example.</u> Let

$$P^{X|\Theta=\vartheta} = \Gamma_{\vartheta,\nu} \quad \text{(gamma distribution)}$$

with density function

$$f_{\vartheta,\nu}: x \to \frac{\vartheta^{\nu}}{\Gamma(\nu)} e^{-\vartheta x} x^{\nu-1} 1_{(0,\infty)}(x),$$

hence

$$F_{\vartheta}(x) = \frac{1}{\Gamma(\nu)} \int_0^{\vartheta x} t^{\nu-1} e^{-t}\, dt, \quad x \geq 0. \tag{8}$$

Obvious, $Z(\vartheta) = F_{\vartheta}(1-\varepsilon)$ is continuous and strictly decreasing in ϑ. Let $\Gamma_{\vartheta,\nu}(q)$ (resp. $\chi^2_m(q)$) denote the q-quantile of $\Gamma_{\vartheta,\nu}$ (resp. $\chi^2_m = \Gamma_{\frac{1}{2},\frac{m}{2}}$, where $m \in \mathbb{N}$, and, of course, χ^2_m denotes the chi-squared distribution with m degrees of freedom).

Then

$$Z(\vartheta) = \Gamma_{\vartheta,\nu}(1-\varepsilon) = \frac{1}{2\vartheta}\,\Gamma_{\frac{1}{2},\frac{2\nu}{2}}(1-\varepsilon)$$

$$\left(= \frac{1}{2\vartheta}\,\chi^2_{2\nu}(1-\varepsilon) \text{ in case } 2\nu \in \mathbb{N}\right).$$

Now let

$$P^{\Theta} = \Gamma_{a,b}\,,$$

then

$$P^{\Theta|(X_1,\ldots,X_n)=x^{(n)}} = \Gamma_{a+z,b+n\nu}$$

where $z = x_1 + \ldots + x_n$.

Then

$$\begin{aligned} d^{(n)}(1-q) &= \Gamma_{a+z,b+n\nu}(1-q) \\ &= \frac{1}{2(a+z)}\,\Gamma_{\frac{1}{2},\frac{2(b+n\nu)}{2}}(1-q) \\ &\left(= \frac{1}{2(a+z)}\,\chi^2_{2(b+n\nu)}(1-q) \text{ if } 2(b+n\nu) \in \mathbb{N}\right). \end{aligned}$$

Hence

$$\begin{aligned} \pi(x^{(n)}) &= Z(d^{(n)}(1-q)) \\ &= \frac{1}{2d^{(n)}(1-q)}\,\Gamma_{\frac{1}{2},\frac{2\nu}{2}}(1-\varepsilon) \\ &= (a+z)\,\frac{\Gamma_{\frac{1}{2},\frac{2\nu}{2}}(1-\varepsilon)}{\Gamma_{\frac{1}{2},\frac{2(b+n\nu)}{2}}(1-q)} \\ &\left(= (a+z)\,\frac{\chi^2_{2\nu}(1-\varepsilon)}{\chi^2_{2(b+n\nu)}(1-q)}\,, \text{ if } 2\nu,\ 2(b+n\nu) \in \mathbb{N}\right), \end{aligned} \tag{10}$$

where, as introduced before, ε denotes the tolerated probability of loss (or "ruin"), and $q = \frac{k_0}{k_0+k_1}$ is given by the coefficients of the loss function (6). Notice that $\pi(x^{(n)})$ is increasing in q, hence decreasing in k_1, which is in agreement with (6) - the greater the loss incurred for positive deviations $\pi(x^{(n)})-H(F_\vartheta)$, the smaller the premium $\pi(x^{(n)})$.

In the special case $\nu=1$, i.e. $P^{X|\Theta=\vartheta}$ is an exponential distribution with distribution function

$$F_\vartheta(x) = 1-e^{-\vartheta x}, \quad x \geq 0,$$

$Z(\vartheta)$ is simply the solution of the equation

$$1-e^{-\vartheta x} = 1-\varepsilon ,$$

i.e. $\quad Z(\vartheta) = -\frac{\ln \varepsilon}{\vartheta} ,$

hence $\quad \pi(x^{(n)}) = -\ln \varepsilon \dfrac{2(a+z)}{\Gamma_{\frac{1}{2},\frac{2(b+n)}{2}}(1-q)} .$ □

4. Tolerance intervals. The concept of a confidence interval, i.e. an interval based on a random sample constructed to capture an unknown parameter of the underlying distribution, is very familiar in statistics. In many cases, however, one is mainly interested in some information on future observations from the same population. This leads to the concept of a tolerance (or prediction) interval.

To be specific, let the assumptions on $\Theta, X_1, X_2, \ldots$ be as in the second paragraph. Then a function $u: \mathbb{R}^n \to \mathbb{R}$ is said to be an upper β-content tolerance limit at level γ if

$$P^{(X_1,\ldots,X_n)|\Theta=\vartheta}(\{(x_1,\ldots,x_n) : \\ : P^{X_{n+1}|\Theta=\vartheta}(\{x : x \leq u(x_1,\ldots,x_n)\}) \geq \beta\}) \geq \gamma \quad (11)$$

where the lhs does not depend on ϑ.
The function u is said to be a Bayesian upper β-content tolerance limit at level γ if

$$P^{\Theta|(X_1,\ldots,X_n)=x^{(n)}}(\{\vartheta : P^{X_{n+1}|\Theta=\vartheta}(\{x : x \leq u(x^{(n)})\}) \geq \beta\}) \geq \gamma . \quad (12)$$

Now if the function $Z_\beta : \vartheta \to d^\vartheta(\beta)$ is continuous and strictly decreasing, we may conclude as in the third paragraph that

$$u(x^{(n)}) = Z_\beta(d^{(n)}(1-\gamma)) = d^{d^{(n)}(1-\gamma)}(\beta)$$

is a Bayesian upper β-content tolerance limit at level γ (cf. AITCHISON [1]).
Thus, if k_o, k_1 are chosen such that $\gamma = \frac{k_o}{k_o+k_1}$ and if $1-\varepsilon$ equals β, the credibility premium equals the Bayesian upper β-content tolerance limit at level γ.
In the above example, the corresponding (non-Bayesian) upper β-content tolerance limit at level γ is

$$u(x^{(n)}) = z \frac{\Gamma_{\frac{1}{2},\frac{2\nu}{2}}(\beta)}{\Gamma_{\frac{1}{2},\frac{2n\nu}{2}}(1-\gamma)}$$

$$\left(= -\ln(1-\beta) \frac{2z}{\chi^2_{2n}(1-\gamma)} \quad \text{if } \nu=1\right) \tag{13}$$

(cf. GUENTHER [7]), which corresponds to the case a=b=0 of the Bayesian result.

In risk theory, DFR distributions are particularly important (cf. HEILMANN [8]). Tolerance limits for distributions of this class (and other classes of distributions based on failure rate properties) have been derived by BARLOW/PROSCHAN [3]. E.g., for an ordered sample $0 = x_{(0)} \le x_{(1)} \le \dots \le x_{(n)}$ and $r = 1,2,\dots,n$, let

$$T_{r,n} = \frac{1}{r} \sum_{i=1}^{r} (n-i+1)(x_{(i)} - x_{(i-1)})$$

$$= \frac{1}{r} (x_{(1)} + \dots + x_{(r-1)} + (n-r+1)x_{(r)})$$

(where the term in brackets is the "total time on test up to the r-th failure" connected to the transform (3)).

Then for the class of DFR distributions

$$u(x^{(n)}) = T_{r,n} \cdot \min\left(\frac{r}{n}, -\ln(1-\beta) \frac{2r}{\chi^2_{2r}(1-\gamma)}\right) \tag{14}$$

is an upper β-content tolerance limit at level γ.

If r=n (i.e., there is no "censoring") and the minimum in (14) is achieved by the second term, we obtain

$$u(x^{(n)}) = -\ln(1-\beta) \frac{2z}{\chi^2_{2n}(1-\gamma)}$$

which coincides with (13); the exponential distribution underlying (13) is both IFR and DFR.

A distribution-free upper β-content tolerance limit at level γ is a function $u^* : \mathbb{R}^n \to \mathbb{R}$ with

$$P^{(X_1,\dots,X_n)}(\{(x_1,\dots,x_n) : P^{X_{n+1}}(\{x : x \le u^*(x_1,\dots,x_n)\}) \ge \beta\}) \ge \gamma .$$

Such a limit can be obtained in the following way. Choose

$$u^*(x_1,\dots,x_n) = x_{(j)}$$

where the index j is such that

$$\sum_{k=n-j+1}^{n} \binom{n}{k} (1-\beta)^k \beta^{n-k} \geq \gamma \tag{15}$$

(cf. GUENTHER [7]). The condition (15) is equivalent to

$$\frac{n-j+1}{j} \frac{\beta}{1-\beta} \leq F_{2j;2(n-j+1)}(\gamma) \tag{16}$$

where the rhs of (16) is the γ-quantile of the F-distribution with $2j$ and $2(n-j+1)$ degrees of freedom.

5. The "collective premium". This is defined to be $H(G)$ where G is the distribution function of X, i.e.

$$G(x) = \int F_\vartheta(x) P^\Theta(d\vartheta).$$

We confine ourselves to the above example with (8),(9) where G has the density function

$$\begin{aligned} g(x) &= \int f_{\vartheta,\nu}(x) f_{a,b}(\vartheta) d\vartheta \\ &= \frac{a^b}{B(b,\nu)} \frac{x^{\nu-1}}{(x+a)^{b+\nu}}, \quad x \geq 0, \end{aligned} \tag{17}$$

where $B(b,\nu) = \frac{\Gamma(b)\Gamma(\nu)}{\Gamma(b+\nu)}$.

A distribution with density

$$h(x) = \frac{1}{B(b,\nu)} \frac{x^{\nu-1}}{(1+x)^{b+\nu}}, \quad x \geq 0, \tag{18}$$

is sometimes called (second kind) Beta. If X has density (17), then $\frac{1}{a}X$ has density (18), and if Y has density (18) with $b,\nu \in \mathbb{N}$, then $\frac{b}{\nu}Y$ is distributed according to an F-distribution with 2ν and $2b$ degrees of freedom. Hence for $b,\nu \in \mathbb{N}$

$$H(G) = \frac{a \cdot \nu}{b} F_{2\nu;2b}(1-\varepsilon). \tag{19}$$

Since, symbolically,

$$F_{2\nu;2b} \sim \frac{\chi^2_{2\nu}/2\nu}{\chi^2_{2b}/2b} = \frac{b}{\nu} \frac{\chi^2_{2\nu}}{\chi^2_{2b}}$$

(where $\chi^2_{2\nu}$ and χ^2_{2b} are independent), formula (19) corresponds in a sense to (10) (with $n=0$, hence $z=0$, and $\varepsilon=q$).

Remark. If, instead of g, we introduce g*, where

$$g^*(x) = \int f_{\vartheta,\nu}(x)\, P^{\Theta|(X_1,\ldots,X_n)=x^{(n)}}(d\vartheta), \quad x \geq 0,$$

we obtain

$$g^*(x) = \frac{(a+z)^{b+n\nu}}{B(b+n\nu,\nu)} \frac{x^{\nu-1}}{(x+az)^{b+(n+1)\nu}} \tag{20}$$

with distribution function G^*, and

$$H(G^*) = \frac{(a+z)\nu}{b+n\nu} F_{2\nu,2(b+n\nu)}(1-\varepsilon), \tag{21}$$

cf. (10).

In case $\nu=1$,

$$g^*(x) = \frac{b+n}{a+z} \left(\frac{a+z}{x+a+z}\right)^{b+n+1},$$

$$1 - G^*(x) = \left(\frac{a+z}{x+a+z}\right)^{b+n},$$

$$r^*(x) = \frac{b+n}{x+a+z} \quad \left(= \frac{b}{x+a} \text{ if } n=0\right), \tag{22}$$

i.e. a Pareto-type failure rate (5).

This is an example of the well-known phenomenon that the mixture of IFR distributions is DFR, resolving in a way the "failure rate paradox" mentioned by BARLOW [2]. □

If $b \leq 1$, $EX = \int x\, G(dx)$ is not finite, i.e., in the sense of premium principles based on EX, X is not insurable, whereas H(G) is finite (but goes to infinity if b goes to 0).

6. References.

[1] J. AITCHISON. Bayesian Tolerance Regions. J. Royal Statistical Society, Ser.B, 26, 1964, 161-175.

[2] R.E. BARLOW. A Bayesian Explanation of an Apparent Failure Rate Paradox. Operations Research Center Research Report 83-13. Berkeley, 1983.

[3] R.E. BARLOW / F. PROSCHAN. Tolerance and Confidence Limits for Classes of Distributions Based on Failure Rate. Ann. Math. Statistics 37, 1966, 1593-1601.

[4] J.O. BERGER. Statistical Decision Theory. Springer, 1980.

[5] TH.S. FERGUSON. Mathematical Statistics. Academic Press, 1967.

[6] H.U. GERBER. An Introduction to Mathematical Risk Theory. Irwin, 1979.

[7] W.C. GUENTHER. Tolerance Intervals for Univariate Distributions. Nav. Res. Log. Quarterly 19, 1972, 309-333.

[8] W.-R. HEILMANN. Characterizations of Claim Distributions by Reliability Techniques. Blätter der DGVM 16, 1983, 13-19.

[9] W.-R. HEILMANN. Transformations of Distributions. Lecture given at the Oberwolfach Meeting on Risk Theory, 1984.