НОВЫЙ МЕТОДОЛОГИЧЕСКИЙ ПОДХОД К ПРОВЕРКЕ СООТВЕТСТВИЯ УРОВНЯ БЕЗОПАСНОСТИ ПОЛЕТОВ ЗАДАННЫМ ТРЕБОВАНИЯМ

В.А. Осташкевич, С.П. Тарасов

Бурное развитие компьютерной техники привело к появлению нового многообещающего направления в математической статистике – компьютерной статистики. Компьютерная статистика позволяет более полно и точно проводить статистические оценивания при ограниченном объеме исходных данных в тех или иных случайных величинах.

Обычно применяемое точечное оценивание параметров БП при ограниченном объеме исходных статистических данных является малоэффективным и в силу того, что истинные значения параметров БП остаются всегда неизвестными. Для определения эффективности точечных оценок необходимо знать их распределения. Казалось бы, что это невозможно без знания истинных значений параметров БП. Однако недавние научные открытия в области компьютерной статистики полностью опровергли это сомнение. С помощью ЭВМ удалось выработать некоторые искусственные распределения, которые имитируют ненаблюдаемые распределения отклонений оцениваемых случайных параметров от их неизвестных истинных значений.

Рассмотрим некоторую общую схему сбора данных, в которую укладывается и сбор данных по БП.

Пусть $[\Omega,\ W(\Omega),\ P_o(.)]$ будет базисным вероятностным пространством, где Ω - множество элементарных событий w, $W(\Omega)$ - измеримое множество всех элементарных событий и Po(.) - истинное вероятностное распределение на $[\Omega,\ W(\Omega)]$. Сбор данных понимается как последовательность экспериментов E_1, … E_2, … Результат i-го эксперимента представляется точкой $x_i \in \aleph$, где $\aleph$ - множество всех значений x_i, ($x_i = x_1$, ... x_i,).

Вероятность того, что $x_i \in B(B \in \aleph)$, задается как $P_{\theta\, 0,i}(B) = P_0(x_i(w) \in B)$, где $P_{\theta\, 0,i}(.)$ принадлежит некоторому семейству распределений, а θ_0 - истинное значение оцениваемого параметра.

Пусть, далее, $\theta_n = S_n(x_{i,} \ldots x_2, \ldots, x_n)$ будет состоятельной точечной оценкой θ_0, т.е. $\theta_n \to \theta_0$ при $n \to \infty$. В этом случае можно рассмотреть закон распределений отклонений $\theta_{n}\text{-}\theta_0$:$L_{\Theta 0n} = L[\theta_{n}\text{-}\theta_0]$.

Если интересующий нас параметр $I_o = I(\theta_0)$, то закон распределений отклонений I-I_0 есть $L_{I(\Theta 0),n} = L[I_n - I_0]$, где $I_n = T_n(x_1, \ldots, x_n)$ - точечная оценка $I_o = I(\Theta_0)$.

Пусть эксперименты $E_{1,}$ …, E_n выполняются в произвольном порядке, допустим в таком: E_{i1}, E_{i2}, … E_{in}. Пусть, далее, первоначальные данные (x_{i1}, …, x_{in}), здесь $\{i_{1,}, \ldots, i_n\}$ - перестановки из {1, …, n}. Таким образом, E_i, …, E_n есть случайная выборка из более общего объема возможных экспериментов. Эти данные можно размножить путем выборки с заменой из E_i, …, E_n, т.е. осуществить генерацию копии статистических данных путем случайной переработки с заменой из первоначальных статистических данных (x_1, … ,x_n) тем самым получим больше знаний о распределении отклонений I_n-I_o.

Укажем последовательность действий, позволяющих получить (это удобно делать с помощью компьютера) некоторое искусственное распределение $L^*_{n.k}$, которое (при больших n) будет асимптотически стремиться к неизвестному распределению $L_{I(\Theta 0),n}$. Итак,

1) Находят состоятельную оценку $I_n = T_n(x_1, ..., x_n) = T_n(x_n)$.

2) Берут К случайных выборок с перевыборкой из первоначальных данных $X_n = (x_1 ... x_n)$ такого же объема n:

$$x^{*1}_n = (x^{*1}_1, x^{*1}_2, \ldots, x^{*1}_n)$$
$$x^{*2}_n = (x^{*2}_1, x^{*2}_2, \ldots, x^{*2}_n)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x^{*k}_n = (x^{*k}_1, x^{*k}_2, \ldots, x^{*k}_n).$$

Здесь $x^{*q}_n (1 \le q \le n)$ - перевыборочная копия первоначальных данных.

3) Если $I_o = I(\Theta_0)$ - интересующий нас параметр и состоятельная оценка $I_n = T_n(X_n) = T_n(x_n)$, то для каждой перевыборной копии находят $I_n^{*1} = T_n(x_n^{*1})$, $I_n^{*2} = T_n(x_n^{*2})$, …, $I_n^{*k} = T_n(x_n^{*k})$.

4) Далее определяют отклонения оцениваемых величин пункта 3) от I_n, т.е.

$$(I_n^{*1} - I_n,\ I_n^{*2} - I_n,\ \ldots,\ I_n^{*k} - I_n).$$

5) Строят закон распределения отклонений от I_n, задаваемых через $x_1, ..., x_n$:

$$L^*_{n,k} = L[(I_n^* - I_n)/x_1, \ldots, x_n]$$

ассимпотически (при $n \to \infty$) приближающийся к $L_{\Theta 0n}$. Это строго доказал Ю.Н.Беляев.

Покажем, как применить описанный выше принципиально новый статистический подход к оценке параметров БП воздушных судов ГА при ограниченном объеме исходных данных (n - порядка нескольких десятков).

Пусть имеем ограниченную реальную статистику по авиационным происшествиям (АП) i -го типа ВС за весь срок его эксплуатации. Обозначим число АП за этот период через $A_{АП}$. Предположим, что $A_{АП}$ имеет порядок 2-4 десятка.

Пусть число ВС данного типа в эксплуатации равно M, наработку каждого ВС (в часах налета или в млн. км) обозначим через T_i ($i = 1, \ldots, M$), а наработку ВС до АП - через T_j ($j = I, ..., n_{АП}$).

Тогда по этим ограниченным данным мы можем определить величину $T^*_{0к}$ наработки до отказа как:

$$T^*_{0к} = \frac{\sum_{i=1}^{M} T_i + \sum_{j=1}^{n_{АП}} T_j}{n_{АП}} \qquad (1)$$

Величина $T^*_{0к}$ в силу ограниченного значения $n_к$ является случайным отклонением от истинного значения T_o наработки на АП (по всей генеральной совокупности), которое всегда остается неизвестным.

На практике неправомерно заданное требование по наработке на АП $T_{озад}$ сравнивали с числом T^*_0 и принимали решение по результатам такого сравнения о соответствии заданным требованиям по безопасности полетов. Ошибочность такого подхода очевидна.

Покажем, как можно по данной ограниченной статистике все же выполнить процедуру проверки на соответствие заданным требованиям.

Пронумеруем все $n_{АП}$ и M значений наработок ВС к рассматриваемому моменту времени (к этому времени уже прошло $n_{АП}$) и будем по ним случайным образом производить перевыборки (выборки с возвращением), каждый раз получая разные выборки одного и того же объема $n_{АП} + M$. Пусть Q - число таких перевыборок (сколь угодно большая величина, т.к. все перевыборки осуществляются с помощью компьютера).

Для каждой перевыборки по формуле (1) получаем Q значений ($T^*_{0к} = I, \ldots, Q$).
Оказалось, что случайная величина, имеющая Q реализаций:

$$T^*_{ок} - T_{o1};\ T^*_{ок} - T_{o1};\ \ldots;\ T^*_{ок} - T_{o1}$$

асимптотически распределена так же, как случайная величина T_0 - T_0к.

Этот удивительный результат в самом общем виде был доказан отечественным математиком Ю.К. Беляевым.

Значит, по реализациям $T^*_{ок} - T^*_{0i}$ ($i = I, \ldots, Q$) можно сколько угодно точно построить распределение случайной величины $T_{ок}$ - T_o^* и по этому распределению построить верхнюю и нижнюю доверительные границы, определить, попала или нет заданная величина наработки на одно АП $T_{озад}$ в область значений между верхней и нижней границами доверительного интервала.

Если появляется дополнительные сведения об АП (увеличилось число $n_{АП}$), то процедуру перевыборки следует повторить и получить более узкий доверительный интервал.