ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

2011 Управление, вычислительная техника и информатика № 2(15)

УДК 519.25

Ж.Н. Зенкова

УЧЕТ ИНФОРМАЦИИ ОБ S^{α}-РАВНОПЛЕЧНОЙ СИММЕТРИИ ПРИ ОБРАБОТКЕ ЦЕНЗУРИРОВАННЫХ ДАННЫХ

Вводится понятие S^{α}-равноплечной симметрии функции распределения, показано, что данным свойством обладает произвольная непрерывная функция распределения, в том числе экспоненциальная и логнормальная. Получены оценки функции распределения, модифицированные с учетом априорной информации об S^{α}-равноплечной симметрии, как для полной, так и для цензурированной справа выборки, изучены их свойства. Методом подстановки получены оценки математического ожидания.

Ключевые слова: *прогрессивно цензурированная выборка*, *функция распределения*, *математическое ожидание*, *симметризация*, *априорная информация.*

При обработке статистических данных нередко возникает необходимость построения оценки неизвестной функции распределения исследуемой случайной величины и/или ее числовых характеристик. При этом на практике имеют место ситуации, когда существует дополнительная информация об исследуемой случайной величине, например о непрерывности, симметричности, значениях моментов или квантилей распределений и пр. Источником этой информации могут служить условия эксперимента, результаты предыдущих исследований, теоретические выводы, физический смысл анализируемой случайной величины и т.д. Следовательно, возникают вопросы учета имеющейся дополнительной информации при разработке статистических процедур, а также исследовании свойств получаемых при этом статистик [1, 2].

Рассматриваемая проблема становится еще более важной в случае, если выборка является неполной, например цензурированной. Данные такого рода встречаются в практической работе довольно часто, особенно в теории надежности, при проведении медицинских, биологических, демографических, экономических исследований и пр. [3 – 6]. Цензурирование приводит к существенным потерям информации, поэтому привлечение дополнительных сведений о распределении – важная и перспективная задача.

Кроме того, проведение многих экспериментов затратно или требует много времени для получения результатов, поэтому возникает задача привлечения априорных сведений о распределении для сокращения количества испытаний и продолжительности опытов.

В работе рассматривается задача привлечения априорной информации об S^{α}-равноплечной симметрии случайной величины типа времени жизни при статистической обработке прогрессивно I типа однократно цензурированных справа выборок. Получены модифицированные оценки функции распределения неизвестной случайной величины, проанализированы ее свойства, показано, что оценка является несмещенной, при этом ее использование повышает точность оценивания неизвестной функции распределения в смысле минимизации дисперсии. Полученные результаты позволяют утверждать, что привлечение дополнительной информации дает возможность получать оценки, которые по точности превосходят эмпирическую функцию распределения, построенную по полной выборке без цензурирования, однако уступают по точности симметризованной эмпирической функции распределения, при этом точность оценивания падает с ростом доли цензурированных наблюдений.

Также в работе методом подстановки найдены модифицированные оценки математического ожидания и дисперсии, изучены их свойства. Полученные формулы обобщают классические результаты.

1. Оценка функции распределения по цензурированной выборке

В данной работе рассматривались прогрессивно цензурированные выборки (ц.в.), цензурирование I типа.

Определение. Под *прогрессивно цензурированной выборкой* (п.ц.в. [3 – 6]) понимается такая ц.в., характер цензуры которой неслучаен.

Определение. Будем говорить, что имеет место цензурирование I типа, если i-й опыт (цензурирован или нет) производится независимо от других опытов.

Пусть τ – случайная величина (с.в.) типа времени жизни, $\tau \in [0, T]$, с функцией распределения (ф.р.) $F(t), (X, I) = \{(X_1, I_1), (X_2, I_2), ..., (X_N, I_N)\}$ – п.ц.в. объема N, где для $i = \overline{1, N}$

$$I_i = \begin{cases} 0, & X_i - \text{полная наработка;} \\ 1, & X_i - \text{неполная наработка до цензурирования;} \end{cases}$$

при этом момент цензурирования T_1 неслучаен.

Рассмотрим п.ц.в., построенные по следующей схеме цензурирования: количество неполных наработок в интервале $(T_1, T]$ – с.в., численно равная доле g, $0 < g < 1$, от числа исправных изделий в конце интервала $[0, T_1]$. Тогда оценка функции распределения определяется формулой [3,5]

$$F_N^{ц}(t) = \begin{cases} 0, & t < 0, \\ \dfrac{1}{N} \sum\limits_{i=1}^{N} \mathrm{I}_{[0,t)}(X_i)\bar{I}_i, & 0 \le t \le T_1, \\ \left.\begin{matrix} \dfrac{r}{N} + \dfrac{1}{(1-g)N} \sum\limits_{i=1}^{N} \mathrm{I}_{[T_1,t)}(X_i)\bar{I}_i, & N_1 > 0 \\ \dfrac{r}{N}, & N_1 = 0 \end{matrix}\right\} & T_1 < t \le T, \\ 1, & t > T, \end{cases} \tag{1}$$

где для $i = \overline{1, N}$ $\bar{I}_i = 1 - I_i$, r – число полных наработок в интервале $[0, T_1]$, $I_A(x) = \{0 : x \notin A, 1 : x \in A\}$ – индикаторная функция, $N_1 = (N - r)(1 - g)$.

Оценка (1) асимптотически несмещенная, при этом

$$\sigma_{ц}^2(t) = \begin{cases} 0, & t \notin [0, T], \\ F(t)(1 - F(t)), & t \in [0, T_1], \\ F(t)(1 - F(t)) + \dfrac{g(F(t) - p)(1 - F(t))}{(1 - p)(1 - g)}, & t \in (T_1, T], \end{cases}$$

где $p = F(T_1)$, $p \in (0,1)$,

$$\sigma_{\text{ц}}^2(t) = \lim_{N\to\infty} NDF_N^{\text{ц}}(t). \quad (2)$$

Здесь $D\xi$ – дисперсия с.в. ξ.

2. S^α-равноплечная симметрия функции распределения

Определение. Будем говорить, что с.в. $\tau \in R$ является S^α*-равноплечно симметричной* относительно центра симметрии α, если ф.р. $F(t)$ удовлетворяет условию

$$F(t) = 1 - F(S(t)+0), \quad t \in R, \quad (3)$$

где функция $S(t)$ является непрерывной, монотонно убывающей и удовлетворяет требованиям

$$(S)^{-1}(t) = S(t), \quad S(\alpha) = \alpha.$$

Здесь $(S)^{-1}(t)$ – обратная к $S(t)$ функция, $F(\alpha) = 0{,}5$. Заметим, что в случае $S(t) = 2\alpha - t$ получим обычную симметрию ф.р. относительно медианы

$$F(t) = 1 - F(2\alpha - t + 0).$$

Теорема 1. *Если ф.р. $F(t)$ является непрерывно возрастающей, то $F(t)$ обладает свойством* (3), *при этом*

$$S(t) = F^{-1}(1 - F(t)), \quad (4)$$

$$\alpha = F^{-1}(0.5),$$

где $F^{-1}(t)$ – обратная к $F(t)$ функция.

Доказательство. Так как ф.р. $F(t)$ непрерывно возрастает, то обратная функция $F^{-1}(t)$ также является непрерывной и возрастающей, в то время как с ростом t значение выражения $1–F(t)$ убывает. Из этого следует, что функция $S(t)$, определяемая по формуле (4), является убывающей.

Нетрудно убедиться, что $(S)^{-1}(t) = S(t)$, при этом для $\alpha = F^{-1}(0{,}5)$

$$S(\alpha) = F^{-1}(1 - F(\alpha)) = F^{-1}(1 - 0{,}5) = F^{-1}(0{,}5) = \alpha.$$

Выполнение условия (3) таким образом очевидно. Теорема доказана.

Следствие 1. *Экспоненциальная ф.р.* $F(t;\lambda) = 1 - e^{-\lambda t}$, $t > 0$, с *параметром* $\lambda > 0$ *обладает свойством* (3), *при этом*

$$S(t) = -\frac{\alpha \ln\left(1 - e^{-\lambda}\right)}{\ln 2}, \ \alpha = \frac{\ln 2}{\lambda}.$$

Следствие 2. *Равномерная ф.р.* $F(x;a,b) = \dfrac{x-a}{b-a}$, $x \in [a,b]$, *с параметрами* $-\infty < a < b < +\infty$ *обладает свойством* (3), *где* $\alpha = \dfrac{a+b}{2}$, $S(x) = 2\alpha - x = a + b - x$.

Рассмотрим с.в. ξ с логнормальной ф.р. $F_{\xi}\left(x;\theta,\sigma^2\right)$. С.в. $\eta=\ln\xi$ является нормальной с.в. с параметрами θ и σ^2, следовательно, η обладает свойством симметрии относительно центра θ

$$F_{\eta}(x)=1-F_{\eta}(2\theta-x),$$

где $F_{\eta}(x)=P(\eta<x)$ – ф.р. с.в. η. Тогда

$$F_{\xi}(x)=P(\xi<x)=P(\ln\xi<\ln x)=F_{\eta}(\ln x)=1-F_{\eta}(2\theta-\ln x)=$$
$$=1-P(\eta<2\theta-\ln x)=1-F_{\xi}\left(e^{2\theta-\ln x}\right)=1-F_{\xi}\left(\frac{e^{2\theta}}{x}\right).$$

Таким образом, для $x>0$ $F_{\xi}(x)=1-F_{\xi}\left(\frac{e^{2\theta}}{x}\right)$.

Рассмотрим $S(x)=\frac{e^{2\theta}}{x}$, $x>0$. Данная функция является непрерывной и монотонно убывающей при $x>0$, при этом очевидно, что $(S)^{-1}(t)=S(t)$. Найдем α, удовлетворяющее условию $S(\alpha)=\frac{e^{2\theta}}{\alpha}=\alpha$. Отсюда получим $\alpha=e^{\theta}$, при этом

$$F_{\xi}(\alpha)=F_{\eta}(\ln\alpha)=F_{\eta}\left(\ln e^{\theta}\right)=F_{\eta}(\theta)=0,5.$$

Таким образом, ф.р. логнормальной с.в. с параметрами θ и σ^2 обладает свойством S^{α}-равноплечной симметрии относительно центра $\alpha=e^{\theta}$ с функцией $S(x)=\frac{\alpha^2}{x}$, т.е. для всех $x>0$ выполняется равенство

$$F_{\xi}(x)=1-F_{\xi}\left(\frac{\alpha^2}{x}\right)=1-F_{\xi}\left(\frac{e^{2\theta}}{x}\right).$$

3. Учет информации об S^{α}-равноплечной симметрии при оценивании функции распределения

Пусть известно, что с.в. $\tau\in[0,T]$ является S^{α}-равноплечно симметричной относительно известного центра $\alpha=\frac{T}{2}$, т.е. ф.р. $F(t)$ удовлетворяет условию (3) для $t\in[0,T]$.

Построим оценку ф.р. $F(t)$ с использованием информации о свойстве (3) на основе S^{α}-равноплечно симметризованной выборки

$$\left(X^S,I^S\right)=\left\{\left(X_1^S,I_1^S\right),\left(X_2^S,I_2^S\right),\ldots,\left(X_N^S,I_N^S\right)\right\},$$

где для $i=\overline{1,N}$

$$X_i^S=\min\{X_i,S(X_i)\},\quad I_i^S=I_i,$$
$$X_{N+i}^S=\max\{X_i,S(X_i)\},\quad I_{N+i}^S=I_i,$$

по формуле

$$F_N^{цS}(t)=\frac{1}{2\left(r+(N-r)(1-g)\right)}\sum_{i=1}^{2N} I_{[0,t)}\left(X_i^S\right)\overline{I_i^S} \ . \tag{5}$$

Здесь $\overline{I_i^S}=1-I_i^S$.

Найдем математическое ожидание (м.о.) оценки (5). Воспользуемся формулой

$$M\xi=M\{M\{\xi\,|\,\eta\}\}, \tag{6}$$

где ξ,η – произвольные с.в., $M\{M\{\xi\,|\,\eta\}\}$ – условное м.о.

Теорема 2. *Пусть с.в.* ξ *имеет ф.р.* $F(x)$, *обладающую свойством* S^{α}*-равноплечной симметрии* (3), *тогда с.в.* $\eta=\min\{\xi,S(\xi)\}$ *имеет ф.р.*

$$F_{\eta}(x)=\min\{1,2F(x)\}. \tag{7}$$

Доказательство. Рассмотрим

$$P(\eta<x)=P\left(\min\{\xi,S(\xi)\}<x\right)=1-P\left(\min\{\xi,S(\xi)\}\ge x\right)=$$

$$=1-P\left(\{\xi\ge x\}\cap\{S(\xi)\ge x\}\right)=1-P\left(\{\xi\ge x\}\cap\left\{\xi\le S^{-1}(x)\right\}\right).$$

Так как $S(x)=S^{-1}(x)$ монотонно убывает, то

$$P(\eta<x)=1-P\left(\{\xi\ge x\}\cap\{\xi\le S(x)\}\right)=$$

$$=\min\{1,1-P(x\le\xi\le S(x))\}=\min\{1,1-F(S(x)+0)+F(x)\}.$$

Воспользуемся свойством (3), получим

$$P(\eta<x)=\min\{1,F(x)+F(x)\}=\min\{1,2F(x)\}.$$

Теорема доказана.

Следствие 2.1. *Пусть с.в.* ξ *имеет ф.р.* $F(x)$, *обладающую свойством* S^{α}*-равноплечной симметрии* (3), *и* $\eta=\min\{\xi,S(\xi)\}$, *тогда*

$$M\eta=M\xi-\int_{\alpha}^{+\infty}(x-S(x))\,dF(x). \tag{8}$$

Доказательство. Рассмотрим

$$M\eta=2\int_{-\infty}^{\alpha}xdF(x)=\int_{-\infty}^{\alpha}xdF(x)+\int_{-\infty}^{\alpha}xdF(x).$$

Воспользуемся свойством (3), будем иметь

$$M\eta=\int_{-\infty}^{\alpha}xdF(x)+\int_{-\infty}^{\alpha}xd\left(1-F(S(x))\right)=\int_{-\infty}^{\alpha}xdF(x)-\int_{-\infty}^{\alpha}xdF(S(x)).$$

Произведя замену $y=S(x)$ с учетом свойств $S(x)$, будем иметь

$$M\eta=\int_{-\infty}^{\alpha}xdF(x)+\int_{\alpha}^{+\infty}S(x)dF(x)=\int_{-\infty}^{+\infty}xdF(x)-\int_{\alpha}^{+\infty}(x-S(x))\,dF(x)=$$

$$=M\xi-\int_{\alpha}^{+\infty}(x-S(x))\,dF(x).$$

Следствие доказано.

Следствие 2.2. *Пусть с.в.* ξ *имеет ф.р.* $F(x)$, *обладающую свойством* S^{α}*-равноплечной симметрии* (3), *и* $\eta = \min\{\xi, S(\xi)\}$, *тогда*

$$D\eta = D\xi + \int_{\alpha}^{+\infty} \left((S(x) - M\xi)^2 - (x - M\xi)^2 \right) dF(x). \tag{9}$$

Доказательство. Воспользуемся формулами (7) и (8), получим

$$D\eta = \int_{-\infty}^{+\infty} (x - M\eta)^2 \, dF_{\eta}(x) = 2 \int_{-\infty}^{\alpha} \left(x - M\xi + \int_{\alpha}^{+\infty} (x - S(x)) dF(x) \right)^2 dF(x) =$$

$$= 2 \int_{-\infty}^{\alpha} (x - M\xi)^2 \, dF(x) + 4 \int_{-\infty}^{\alpha} (x - M\xi) dF(x) \int_{\alpha}^{+\infty} (x - S(x)) dF(x) +$$

$$+2 \left(\int_{\alpha}^{+\infty} (x - S(x)) dF(x) \right)^2 = D\xi + \int_{-\infty}^{\alpha} (x - M\xi)^2 \, dF(x) - \int_{\alpha}^{+\infty} (x - M\xi)^2 \, dF(x) + S,$$

где
$$S = 4 \int_{-\infty}^{\alpha} (x - M\xi) dF(x) \int_{\alpha}^{+\infty} (x - S(x)) dF(x) + 2 \left(\int_{\alpha}^{+\infty} (x - S(x)) dF(x) \right)^2 =$$

$$= 2 \int_{\alpha}^{+\infty} (x - S(x)) dF(x) \left\{ 2 \int_{-\infty}^{\alpha} (x - M\xi) dF(x) + \int_{\alpha}^{+\infty} (x - S(x)) dF(x) \right\}.$$

Раскрыв скобки, а также произведя во втором интеграле замену $y = S(x)$, получим

$$S = 2 \int_{\alpha}^{+\infty} (x - S(x)) dF(x) \left\{ 2 \int_{-\infty}^{\alpha} x dF(x) - 2F(\alpha) M\xi + \int_{\alpha}^{+\infty} x dF(x) + \int_{-\infty}^{\alpha} y dF(S(y)) \right\}.$$

Следствие доказано.

Теорема 3. *Пусть с.в.* ξ *имеет ф.р.* $F(x)$, *обладающую свойством* S^{α}*-равноплечной симметрии* (3), *тогда с.в.* $\eta = \max\{\xi, S(\xi)\}$ *имеет ф.р.*

$$F_{\eta}(x) = \max 0\{1, 2F(x) - 1\}. \tag{10}$$

Доказательство. Рассмотрим

$$P(\eta < x) = P(\max\{\xi, S(\xi)\} < x) = P(\{\xi < x\} \cap \{S(\xi) < x\}).$$

Так как $S(x) = S^{-1}(x)$ – монотонно убывает, то

$$P(\eta < x) = P\left(\{\xi < x\} \cap \left\{ \xi > S^{-1}(x) \right\} \right) = \max\{0, P(S(x) < \xi < x)\} =$$
$$= \max\{0, F(x) - F(S(x) + 0)\}.$$

Воспользуемся свойством (3), получим

$$P(\eta < x) = \max\{0, F(x) - 1 + F(x)\} = \max\{0, 2F(x) - 1\}.$$

Теорема доказана.

Следствие 3.1. *Пусть с.в.* ξ *имеет ф.р.* $F(x)$, *обладающую свойством* S^{α}*-равноплечной симметрии* (3), *тогда для с.в.* $\eta = \max\{\xi, S(\xi)\}$

$$M\eta = M\xi + \int\limits_{\alpha}^{+\infty} (x - S(x))\,dF(x). \qquad (11)$$

Доказательство аналогично доказательству следствия 2.1.

Следствие 3.2. *Пусть с.в.* ξ *имеет ф.р.* $F(x)$, *обладающую свойством* S^{α}*-равноплечной симметрии* (3), *тогда для с.в.* $\eta = \max\{\xi, S(\xi)\}$

$$D\eta = D\xi - \int\limits_{\alpha}^{+\infty} \left\{(S(x) - M\xi)^2 - (x - M\xi)^2\right\} dF(x). \qquad (12)$$

Доказательство аналогично доказательству следствия 2.2.

Так как с.в. r распределена биномиально с параметрами $p = 0,5$ и N, то $Mr = pN$, поэтому для $t \in [0, T_1]$ с учетом I типа цензурирования и теоремы 2

$$M\left(F_N^{\text{ц}S}(t)\middle| r\right) = \frac{1}{2(r + (N-r)(1-g))}\sum_{i=1}^{N} 2F(t)\overline{I_i^S} = \frac{F(t)(r + (N-r)(1-g))}{(r + (N-r)(1-g))} = F(t).$$

Для $t \in [T_1, T]$

$$M\left(F_N^{\text{ц}S}(t)\middle| r\right) = \frac{1}{2} + \frac{1}{2(r + (N-r)(1-g))}\sum_{i=1}^{N}(2F(t) - 1)\overline{I_i^S} =$$
$$= \frac{1}{2} + \frac{(2F(t) - 1)(r + (N-r)(1-g))}{(r + (N-r)(1-g))} = F(t).$$

Таким образом, для $t \in [0, T]$

$$M\left(F_N^{\text{ц}S}(t)\right) = M\left(M\left(F_N^{\text{ц}S}(t)\right)\middle| r\right) = F(t).$$

Следовательно, оценка (5) обладает свойством несмещенности.

Найдем дисперсию оценки (5). Воспользуемся формулой

$$D\xi = M\{D\{\xi \mid \eta\}\} + D\{M\{\xi \mid \eta\}\}. \qquad (13)$$

Выше было получено, что для $t \in [0, T]$

$$M\left(F_N^{\text{ц}S}(t)\middle| r\right) = F(t),$$

следовательно,

$$D\left\{M\left(F_N^{\text{ц}S}(t)\middle| r\right)\right\} = D\{F(t)\} = 0.$$

Для $t \in [0, T_1]$ рассмотрим

$$D\left(F_N^{\text{ц}S}(t)\middle| r\right) = D\left\{\frac{1}{2(r + (N-r)(1-g))}\sum_{i=1}^{N} I_{[0,t)}\left(X_i^S\right)\overline{I_i^S}\middle| r\right\}.$$

С учетом I типа цензурирования и теоремы 2

$$D\left(F_N^{\text{ц}S}(t)\middle| r\right) = \frac{r + (N-r)(1-g)}{4(r + (N-r)(1-g))^2} D\left\{I_{[0,t)}\left(X_i^S\right)\right\} = \frac{F(t)(1 - 2F(t))}{2(r + (N-r)(1-g))}.$$

Таким образом,

$$D\left(F_N^{\text{ц}S}(t)\right) = M\left\{\frac{F(t)\left(0,5-F(t)\right)}{r+(N-r)(1-g)}\right\} = F(t)\left(0,5-F(t)\right)\sum_{k=0}^{N}\frac{C_N^k p^k (1-p)^{N-k}}{k+(N-k)(1-g)} =$$

$$= \frac{F(t)\left(0,5-F(t)\right)}{pN+(N-Np)(1-g)} = \frac{F(t)\left(0,5-F(t)\right)}{N(1-0,5g)}.$$

Для случая $t \in \left[T_1, T\right]$ аналогично можно получить, что

$$D\left(F_N^{\text{ц}S}(t)\right) = \frac{(1-F(t))\left(0,5-F(t)\right)}{N(1-0,5g)}.$$

В итоге для $t \in \left[0, T\right]$ дисперсию оценки можно представить в следующем виде:

$$\sigma_{\text{ц}S}^2(t) = ND\left(F_N^{\text{ц}S}(t)\right) = \frac{F\left(t_{\min}\right)\left(0,5-F\left(t_{\min}\right)\right)}{N(1-0,5g)}, \tag{14}$$

где

$$t_{\min} = \min\left\{t, S(t)\right\}. \tag{15}$$

Сравним дисперсии оценок (1) и (5). Для $t \in \left[0, T_1\right]$

$$\frac{\sigma_{\text{ц}}^2(t)}{\sigma_{\text{ц}S}^2(t)} = \frac{(1-0,5g)F(t)(1-F(t))}{F(t)\left(0,5-F(t)\right)} = \frac{(1-0,5g)(1-F(t))}{0,5-F(t)} \geq 1,$$

если

$$0,5-F(t) \leq (1-F(t))(1-0,5g) = 1-F(t)-0,5g+0,5gF(t)$$

или, что то же самое,

$$0,5-0,5g+0,5gF(t) \geq 0,$$

$$1-g+F(t) \geq 0,$$

что истинно, так как $g \in (0,1)$ и $F(t) \in [0,1]$.

Для $t \in \left[T_1, T\right]$ с учетом того, что $p = 0,5$,

$$\frac{\sigma_{\text{ц}}^2(t)}{\sigma_{\text{ц}S}^2(t)} = \frac{(1-0,5g)F(t)(1-F(t))}{(1-F(t))(F(t)-0,5)} + \frac{g(1-0,5g)(F(t)-0,5)(1-F(t))}{0,5(1-g)(1-F(t))(F(t)-0,5)} =$$

$$= \frac{F(t)(2-g)}{2F(t)-1} + \frac{g(2-g)}{(1-g} = \frac{(2-g)\left(g+(1-g)F(t)\right)}{(1-g)(2F(t)-1)} \geq 1,$$

если

$$(2-g)\left(g+(1-g)F(t)\right) \geq (1-g)(2F(t)-1),$$

т.е.

$$g(1-g)-gF(t)+g^2F(t)+1 \geq 0.$$

С учетом того, что здесь $F(t) \in \left(0,5,\ 1\right]$ и $g \in (0,1)$, получим

$$g(1-g)-gF(t)+g^2F(t)+1 \geq g(1-g)-1+g^2F(t)+1 = g(1-g)+g^2F(t) \geq 0.$$

Таким образом, для $t \in \left[0, T\right]$ $\sigma_{\text{ц}S}^2(t) \leq \sigma_{\text{ц}}^2(t).$

В [5] была предложена оценка неизвестной ф.р. с.в. типа времени жизни с учетом свойства (3), которая определяется формулой

$$F_N^{\text{ц}S*}(t) = \frac{F_N^{\text{ц}}(t) + 1 - F_N^{\text{ц}}(S(t)+0)}{2}, \tag{16}$$

при этом было показано, что $MF_N^{\text{ц}S*}(t) = F(t)$,

$$\sigma^2_{\text{ц}S*}(t) = \lim_{N\to\infty} NDF_N^{\text{ц}S*}(t) = \frac{(2-g)F(t_{\min})(0,5 - F(t_{\min}))}{2(1-g)}, \tag{17}$$

где $t_{\min}$ определяется формулой (15).

Сравним нормированные дисперсии оценок (5) и (16). Нетрудно получить, что для любого $t \in [0,T]$

$$\frac{\sigma^2_{\text{ц}S}(t)}{\sigma^2_{\text{ц}S*}(t)} = \frac{4(1-g)}{(2-g)^2} \le 1,$$

если $$4(1-g) = 4 - 4g \le (2-g)^2 = 4 - 4g + g^2,$$

т.е. $g^2 \ge 0$, что истинно, так как $g \in (0,1)$.

При отсутствии цензурирования в качестве оценки неизвестной ф.р. используется эмпирическая ф.р. (э.ф.р.), которая определяется следующим образом:

$$F_N(t) = \frac{1}{N}\sum_{i=1}^{N} I_{[0,t)}(X_i). \tag{18}$$

Известно, что э.ф.р. – несмещенная оценка ф.р., при этом [6]

$$\sigma^2(t) = NDF_N(t) = F(t)(1-F(t)).$$

Заметим, что при $g = 0$, т.е. при отсутствии цензурирования, э.ф.р. (18) и оценка (1) совпадают, при этом значение $\sigma^2(t)$ есть частный случай (2).

Нетрудно убедиться, что для любого $t \in [0,T]$

$$\frac{\sigma^2_{\text{ц}S}(t)}{\sigma^2(t)} = \frac{F(t_{\min})(0,5 - F(t_{\min}))}{(1-0,5g)F(t)(1-F(t))} \le 1.$$

В [1] рассматривалась оценка ф.р. $F(t)$, построенная на основе полной выборки с учетом свойства (3),

$$F_N^S(t) = \frac{F_N(t) + 1 - F_N(S(t)+0)}{2}, \tag{19}$$

которая является несмещенной, при этом

$$\sigma_N^S(t) = NDF_N^S(t) = F(t_{\min})(0,5 - F(t_{\min})). \tag{20}$$

Заметим, что формула (20) совпадает с (14) при $g = 0$, т.е. при отсутствии цензурирования. Очевидно, что

$$\frac{\sigma^2_{\text{ц}S}(t)}{\sigma^2_S(t)} = \frac{1}{1-0,5g} \ge 1,$$

при этом рассматриваемое соотношение возрастает с ростом g.

На рис. 1 – 3 приведены графики нормированных дисперсий оценок (1), (5), (16), (18) и (19) для случая $F(t) = R_{[0,1]}(t)$ – равномерного в [0,1] распределения, $g = 0,3$ (рис.1), $g = 0,6$ (рис. 2), $g = 0,9$ (рис. 3).

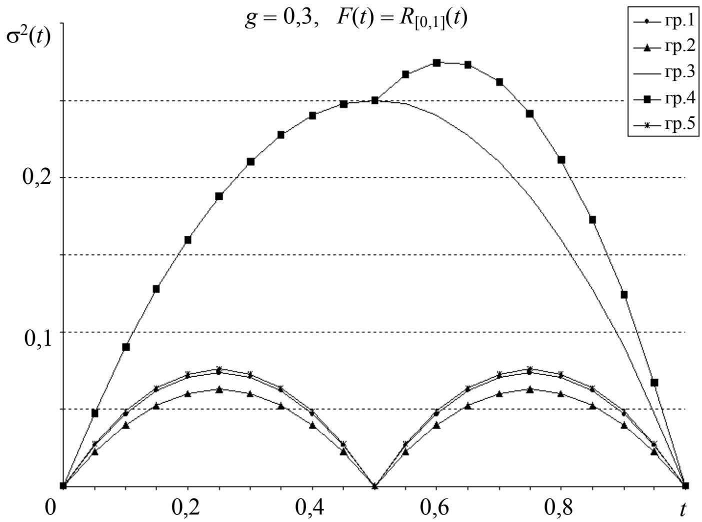

Рис. 1. Графики $\sigma^2_{цS}(t)$ (гр. 1), $\sigma^2_{S}(t)$ (гр. 2), $\sigma^2(t)$ (гр. 3), $\sigma^2_{ц}(t)$ (гр. 4), $\sigma^2_{цS*}(t)$ (гр. 5) для $F(t) = R_{[0,1]}(t)$, $g = 0,3$

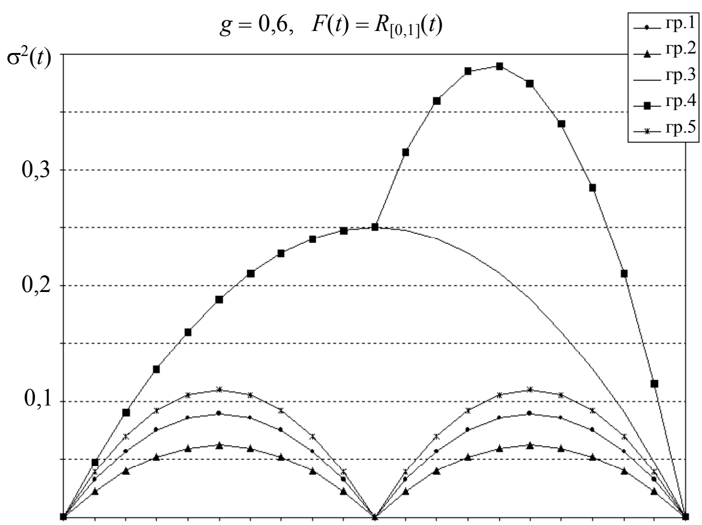

Рис. 2. Графики $\sigma^2_{цS}(t)$ (гр. 1), $\sigma^2_{S}(t)$ (гр. 2), $\sigma^2(t)$ (гр. 3), $\sigma^2_{ц}(t)$ (гр. 4), $\sigma^2_{цS*}(t)$ (гр. 5) для $F(t) = R_{[0,1]}(t)$, $g = 0,6$

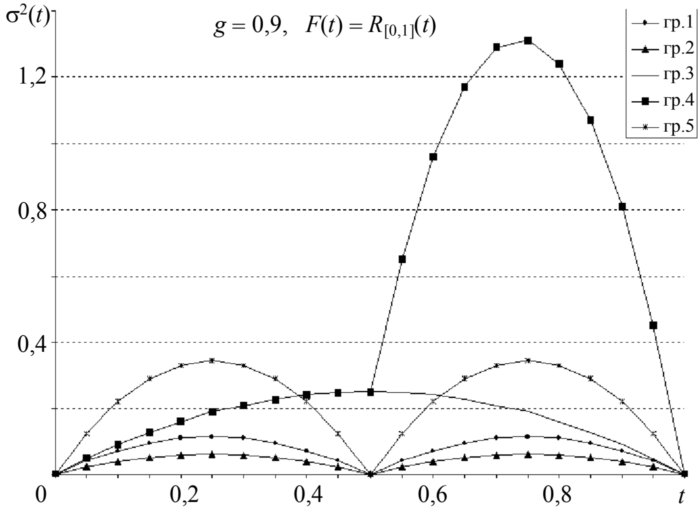

Рис. 3. Графики $\sigma^2_{\text{ц}S}(t)$ (гр. 1), $\sigma^2_S(t)$ (гр. 2), $\sigma^2(t)$ (гр. 3), $\sigma^2_{\text{ц}}(t)$ (гр. 4), $\sigma^2_{\text{ц}S*}(t)$ (гр. 5) для $F(t) = R_{[0,1]}(t)$, $g = 0{,}9$

Таким образом, предложенная оценка (5) является несмещенной и эффективной оценкой ф.р., построенной на основе однократно цензурированной выборки с использованием информации об S^α-равноплечной симметрии относительно известного центра $\alpha = \dfrac{T}{2} = T_1$. При этом чем меньше доля цензурированных наблюдений в выборке, тем ближе ее дисперсия к дисперсии оценки (19), построенной с использованием свойства (3) по полной выборке.

4. Оценка математического ожидания по симметризованной прогрессивно цензурированной выборке

Традиционно в качестве оценки математического ожидания исследуемой с.в. τ, $M\tau = \theta$, используется выборочное среднее [7 – 9]

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i, \tag{21}$$

где X_i, $i = \overline{1, N}$, – выборочные значения. Если выборка является неполной, то $\overline{X}$ применять некорректно.

Пусть имеется п.ц.в. объема N, построенная по уже упомянутой выше схеме. Тогда оценку м.о. можно найти с помощью метода подстановки [7, 8]:

$$\theta_N^{\text{ц}} = \int_0^T t dF_N^{\text{ц}}(t) = \frac{1}{N}\sum_{i=1}^{N} X_{(i)} + \frac{1}{(1-g)N}\sum_{i=r+1}^{r+(N-r)(1-g)} X_{(i)}, \tag{22}$$

где $F_N^{\text{ц}}(t)$ – оценка (1), $X_{(i)}$, $i = \overline{1, r + (1-g)(N-r)}$, – члены вариационного ряда, построенного по полным наработкам п.ц.в. (X, I).

Заметим, что при $g=0$, т.е. при отсутствии цензурирования, классическая оценка (21) совпадает с (22).

Покажем, что оценка (22) является несмещенной. Используя формулу (6), получим

$$M\theta_N^{ц} = M\left\{M\left\{\theta_N^{ц}\middle|r\right\}\right\} = M\left\{\frac{r}{N}\theta + \frac{(N-r)(1-g)}{(1-g)N}\theta\right\} = M\theta = \theta.$$

Найдем дисперсию оценки (22) с помощью формулы (13). Получим

$$D\left\{M\left\{\theta_N^{ц}\middle|r\right\}\right\} = D\theta = 0.$$

Если $D\tau = S^2 < +\infty$, то

$$D\left\{\theta_N^{ц}\middle|r\right\} = \frac{r}{N^2}S^2 + \frac{(N-r)(1-g)}{(1-g)^2N^2}S^2 = \frac{S^2}{N^2}\left(r + \frac{N-r}{1-g}\right).$$

Зная, что $Mr = pN = 0.5N$, получим

$$D\theta_N^{ц} = M\left\{D\left\{\theta_N^{ц}\middle|r\right\}\right\} = \frac{1-0,5g}{N(1-g)}S^2.$$

Так как $S^2 < +\infty$, то $\lim\limits_{N\to+\infty} D\theta_N^{ц} = 0.$

Пусть известно, что ф.р. $F(t)$ обладает свойством (3). Найдем оценку м.о. с.в. r методом подстановки:

$$\theta_N^{цS} = \int_0^T t dF_N^{цS}(t),$$

где $F_N^{цS}(t)$ определяется формулой (5).Тогда

$$\theta_N^{цS} = \frac{1}{2K}\sum_{i=1}^{K}\left(X_{(i)}^S + X_{(K+i)}^S\right), \tag{23}$$

где $K = r + (N-r)(1-g)$, $X_{(i)}^S$, $i=\overline{1,2K}$, – члены вариационного ряда, построенного по полным наработкам симметризованной выборки $\left(X^S, I^S\right)$.

Используя следствия из теорем 2 и 3, нетрудно показать, что оценка (23) является несмещенной, при этом

$$D\theta_N^{цS} = M\left\{D\left\{\theta_N^{цS}\middle|r\right\}\right\} + D\theta = M\left\{\frac{K}{4K^2}\left(DX_{(i)}^S + 2\operatorname{cov}\left(X_{(i)}^S, X_{(K+i)}^S\right) + DX_{(K+i)}^S\right)\right\} =$$

$$= M\left\{\frac{1}{K}\left(DX_{(i)}^S + 2\operatorname{cov}\left(X_{(i)}^S, X_{(K+i)}^S\right) + DX_{(K+i)}^S\right)\right\} =$$

$$= M\left\{\frac{1}{2K}\right\}\left(S^2 + \int_0^T (x-\theta)(S(x)-\theta)dF(x)\right) = \frac{S^2+Z}{N(1-0,5g)},$$

где $S^2 = D\tau$,

$$Z = \int_0^T (x-\theta)(S(x)-\theta)dF(x).$$

Заметим, что $Z<0$, так как функция $S(x)$ убывает по $x\in[0,T]$.

Сравним дисперсии оценок м.о. (22) и (23), получим

$$\frac{D\theta_N^{цS}}{D\theta_N^S} = \frac{1-g}{(1-0,5g)}\left(1+\frac{Z}{S^2}\right) < 1,$$

так как $1-g \le 1-g+0,25g^2$ для $g\in(0,1)$ и $Z<0$. Здесь $S^2 = D\tau$.

Таким образом, учет информации об S^α-равноплечной симметрии позволил получить модифицированную оценку м.о., построенную по п.ц.в., при этом новая оценка является несмещенной и эффективной.

5. Оценка дисперсии по симметризованной прогрессивно цензурированной выборке

С помощью метода подстановки нетрудно также получить оценку дисперси

$$S_N^{ц2} = \frac{1}{N}\sum_{i=1}^{r}\left(X_{(i)}-\theta_N^ц\right)^2 + \frac{1}{(1-g)N}\sum_{i=r+1}^{r+(N-r)(1-g)}\left(X_{(i)}-\theta_N^ц\right)^2, \tag{24}$$

где $\theta_N^ц$ – оценка математического ожидания (22), $X_{(i)}$, $i=\overline{1, r+(N-r)(1-g)}$, – члены вариационного ряда, построенного по полным наработкам п.ц.в. (X, I).

Заметим, что при $g=0$, т.е. при отсутствии цензурирования, оценка (24) совпадает с классической выборочной дисперсией

$$S_N^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_{(i)}-\overline{X}\right)^2, \tag{25}$$

Найдем математическое ожидание (24), используя формулу (6), получим

$$M\left\{S_N^{ц2}\middle|r\right\} = M\left\{\frac{1}{N}\sum_{i=1}^{r}\left(\left(X_{(i)}-\theta\right)+\left(\theta-\theta_N^ц\right)\right)^2 + \frac{1}{(1-g)N}\sum_{i=r+1}^{r+(N-r)(1-g)}\left(\left(X_{(i)}-\theta\right)+\left(\theta-\theta_N^ц\right)\right)^2\middle|r\right\} =$$

$$= \left(\frac{r}{N}+\frac{(N-r)(1-g)}{(1-g)N}\right)\left(M\left\{\left(X_{(i)}-\theta\right)^2\middle|r\right\} - M\left\{\left(\theta_N^ц-\theta\right)^2\middle|r\right\}\right),$$

из чего следует, что

$$MS_N^{ц2} = M\left\{M\left\{S_N^{ц2}\middle|r\right\}\right\} = M\left\{M\left\{\left(X_{(i)}-\theta\right)^2\middle|r\right\} - M\left\{\left(\theta_N^ц-\theta\right)^2\middle|r\right\}\right\} =$$

$$= S^2 - D\theta_N^ц = S^2 - \frac{1-0,5g}{N(1-g)}S^2 = S^2\left(1-\frac{1-0,5g}{N(1-g)}\right).$$

Таким образом,

$$MS_N^{ц2} = S^2\left(1-\frac{1-0,5g}{N(1-g)}\right), \tag{26}$$

где $S^2 = D\tau$.

Заметим, что оценка дисперсии смещенная, при этом в случае отсутствия цензурирования, т.е. при $g=0$ формула (26) принимает значение

$$MS_N^{ц2} = S^2\left(1-\frac{1}{N}\right) = MS_N^2,$$

т.е. совпадает с м.о. классической оценки (25) [2], при этом с ростом объема наблюдений N смещение оценки (24) сходится к нулю.

Заключение

В данной работе рассмотрен вопрос привлечения априорной информации об S^{α}-равноплечной симметрии при обработке прогрессивно однократно I типа справа цензурированных данных, в частности, построена оценка неизвестной функции распределения по цензурированной выборке, методом подстановки получены оценки математического ожидания и дисперсии, изучено качество этих оценок, проведено сравнение с классическими результатами. Знание симметрии позволило значительно улучшить качество оценивания и восполнить потерянную при цензурировании информацию.

ЛИТЕРАТУРА

1. *Дмитриев Ю.Г.*, *Зенкова Ж.Н.* Об оценивании симметричного распределения по цензурированной выборке // Труды X юбилейного симпозиума по непараметрическим и робастным статистическим методам в кибернетике. Томск: Изд-во НТЛ, 2004. С. 22–30.
2. *Дмитриев Ю.Г*, *Устинов Ю.К.* Статистическое оценивание распределений вероятностей с использованием дополнительной информации. Томск: Изд-во ТГУ, 1988. 194 с.
3. *Klein John P.*, *Moeschberger Melvin L.* Survival Analysis: Techniques for Censored and Truncated Data. Springer, 2005. P. 547.
4. *Marubini E.*, *Valsecchi M.G.* Analysing Survival Data from Clinical Trials and Observational Studies. Willes, 2004. P. 417.
5. *Анализ* надежности технических систем по цензурированным выборкам / В.М. Скрипник, А.Е. Назин, Ю.Г. Приходько, Ю.Н. Благовещенский. М.: Радио и связь, 1988. 84 с.
6. *Скрипник В.М.*, *Назин А.Е.* Оценка надежности технических систем по цензурированным выборкам / под ред. А.И. Широкова. Минск: Наука и техника, 1981. 144 с.
7. *Тарасенко Ф.П.* Непараметрическая статистика. Томск: Изд-во ТГУ. 1976.
8. *Боровков А.А.* Математическая статистика. Новосибирск: Наука; Изд-во Института математики, 1997. 772 с.
9. *Айвазян С.А.*, *Мхитарян В.С.* Теория вероятностей и прикладная статистика. М.: ЮНИТИ-ДАТА, 2001. 656 с.

Зенкова Жанна Николаевна
Томский государственный университет
E-mail: thankoffjean@mail.ru

Поступила в редакцию 1 ноября 2010 г.