Вестник УГАТУ
Уфа : УГАТУ, 2011 Т. 15, № 2 (42). С. 22–28
УПРАВЛЕНИЕ, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И ИНФОРМАТИКА
УДК 62-19:519.2

В. Е. Гвоздев, Г. И. Таназлы, А. Ю. Хасанов, М. А. Абдрафиков

АНАЛИЗ НАДЕЖНОСТИ ТЕХНИЧЕСКИХ СИСТЕМ НА ОСНОВЕ МАТЕМАТИКО-СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Статья посвящена вопросам математико-статистического моделирования в задачах управления надежностью. Основой моделирования в этих задачах является оценивание закона распределения случайных величин, в том числе – порядковых статистик, по выборочным данным. Особое внимание уделено исследованию влияния априорного знания о случайной величине на оценки статистических характеристик в зависимости от свойств выборки. *Надежность*; *точность*; *устойчивость*; *малая выборка*; *порядковая статистика*; *закон распределения*

Математико-статистическое моделирование занимает центральное место среди методов математического обеспечения системы информационной поддержки управления надежностью технических изделий. Основное назначение этих методов – статистическое оценивание критериев и показателей надежности по экспериментальным данным.

Особое место математико-статистических моделей среди других математических моделей обусловлено сложностью исследования процессов, протекающих в технических изделиях, недостаточной изученностью механизмов изменения их состояния.

1. ЗАДАЧИ, СВЯЗАННЫЕ С УПРАВЛЕНИЕМ НАДЕЖНОСТЬЮ

Основу математико-статистического моделирования в задачах управления надежностью составляет оценка закона распределения случайной величины $\hat{F}(x)$, определяемая по выборке $\{x\}$ объема N. Особенностью задач надежности является то, что достаточно часто объем исходных данных не превышает нескольких десятков единиц. В известной литературе вопросам оценивания законов распределения случайных величин, в том числе по выборкам малого объема, уделено достаточно внимания. При этом в общем случае задача сводится к построению оператора A, преобразующего выборочные данные $\{x\}$ в оценку закона распределения $\hat{F}(x)$:

$$A:\{x\}\to \hat{F}(x). \tag{1}$$

По поводу термина «малая выборка» отметим следующее. На наш взгляд, отнесение выборки к классу «малых» либо «больших» зависит от двух обстоятельств: во-первых, от требований к точности результата, во-вторых, от используемого способа обработки данных, то есть от того, насколько математический аппарат настроен на «выжимание» информации из выборочных данных.

Решение задачи (1) создает основу для решения «производных» задач, связанных с управлением надежностью. Примерами таких «производных» задач являются:

- Оценка сопротивляемости технических устройств.

Решение задачи сводится к реализации следующих преобразований:

$$A^{(1)}:\ \{u\}\to \hat{F}(u),$$
$$A^{(2)}:\ \{x\}\to \hat{F}(x),$$
$$A^{(3)}:\ \hat{F}(u),\hat{F}(x)\ \to\ P(x<u).$$

Здесь $\{u\}$ – выборочные данные, характеризующие нагрузку; $\{x\}$ – выборочные данные, характеризующие сопротивление; $\hat{F}(u),\hat{F}(x)$ – законы распределения нагрузки и сопротивления соответственно; $P(x < u)$ – вероятность того, что сопротивляемость оказывается меньше нагрузки.

- Прогнозирование отказов технических изделий

Решение задачи сводится к реализации следующих преобразований

$$A^{(1)}:\ \{u(t)\},\{x(t)\}\ \to\ \left\{\overset{*}{t}\right\},$$
$$A^{(2)}:\ \left\{\overset{*}{t}\right\}\ \to\ P(t<T).$$

Контактная информация: 8-917-44-55-980

Приведенные результаты получены в рамках проведения исследований по гранту РФФИ 10-09-00359-а.

Здесь $\{u(t)\}$, $\{x(t)\}$ – временные ряды, характеризующие изменение нагрузки и сопротивляемости во времени; $P(u < x, \overset{*}{t})$ - вероятность того, что во временном сечении $\overset{*}{t}$ нагрузка окажется меньше сопротивляемости.

- Пересчет характеристик надежности на разные условия испытаний

Решение задачи сводится к реализации следующих преобразований:

$$A^{(1)} \ : \ \left\{x^{(\varepsilon_1)}\right\} \ \to \ F(x, \varepsilon_1)$$
$$\left\{x^{(\varepsilon_2)}\right\} \ \to \ F(x, \varepsilon_2),$$
$$A^{(2)} \ : \ F(x,\varepsilon_1), F(x,\varepsilon_2) \ \to \ x^{\varepsilon_2} = \varphi(x^{(\varepsilon_1)}),$$

здесь $\left\{x^{(\varepsilon_1)}\right\}$ – выборочные значения характеристик надежности, полученные в условиях ε_1;

$\left\{x^{(\varepsilon_2)}\right\}$ – выборочные значения характеристик надежности, полученные в условиях ε_2;

$F(x, \varepsilon_1)$, $F(x, \varepsilon_2)$ – законы распределения характеристик надежности, соответствующие разным условиям;

$x^{\varepsilon_2} = \varphi(x^{(\varepsilon_1)})$ – правила пересчета данных, соответствующих условиям ε_1, в данные, соответствующие условиям ε_2.

- Оценка технологической зрелости процесса производства технических изделий

Основу решения этой задачи составляет следующее утверждение: технологическая зрелость характеризуется разбросом характеристик надежности технических изделий.

Известно, что существует функциональная взаимосвязь между такими показателями разброса, как среднеквадратическое отклонение и размах.

Выборочное значение размаха, особенно при малом числе исходных данных, определяется с низкой точностью. К тому же при определении размаха как расстояния между максимальным и минимальным элементами выборки, теряется информация, содержащаяся в остальных выборочных данных. В связи с этим для оценки размаха по выборке объема N целесообразно использовать следующую схему решения задачи:

$$A^{(1)} \ : \ \{x\} \ \to \ \hat{F}_N(x)$$
$$A^{(2)} \ : \ \hat{F}_N(x) \ \to \ \hat{F}_N(x_{\max}), \hat{F}_N(x_{\min}),$$
$$A^{(3)} \ : \ \hat{F}_N(x_{\max}), \hat{F}_N(x_{\min}) \ \to$$
$$\to \ \hat{M}[x_{\max}], \hat{M}[x_{\min}], \hat{d}_N,$$

здесь $\hat{F}_N(x)$ – оценка закона распределения случайной величины, получаемая по выборке объема N;

$\hat{F}_N(x_{\max}), \hat{F}_N(x_{\min})$ – оценки законов распределения минимального и максимального элементов выборки объема N;

$\hat{M}[x_{\max}], \hat{M}[x_{\min}]$ – оценки математических ожиданий минимального и максимального элементов в выборке объема N;

$\hat{d}_N$ – оценка размаха по выборке объема N.

Перечисленные задачи не покрывают всего множества задач, основанных на статистическом моделировании надежности.

2. СТАТИСТИЧЕСКИЕ МОДЕЛИ

Статистическая модель $\hat{F}_N(x)$, получаемая в результате решения задачи (1), ориентирована на получение «средней» оценки закона распределения случайной величины, так как в ее основе лежит использование всех выборочных данных. В то же время решение ряда задач основано на использовании законов распределения m-х порядковых статистик $\varphi_N(x_m)$ в выборке объема N. Примером является описанная выше задача оценки технологической зрелости процесса производства технических изделий.

К настоящему времени в известной литературе по статистическому моделированию в теории надежности аспект, посвященный использованию аппарата порядковых статистик, не получил должного развития. Причиной этого, на наш взгляд, является то, что в основе построения дифференциальной функции распределения m-й порядковой статистики в выборке объема N лежит известное соотношение:

$$\varphi_N(x_m) = C_N^m [F(x)]^{m-1} [1 - F(x)]^{N-m} f(x) \ . \qquad (2)$$

Из него следует, что получить $\varphi_N(x_m)$ можно лишь зная $F(x)$. Как было отмечено ранее, построение оценки $\hat{F}_N(x)$ по выборке малого объема в случае, когда тип закона распределения априорно неизвестен, представляет достаточно непростую задачу. В случае, когда тип закона распределения $F(x)$ априорно известен, оценка закона распределения $\hat{F}_N(x)$ сводится к задаче оптимального оценивания параметров по выборочным данным. В обоих случаях решению задачи (1) посвящено немалое количество работ.

Как видно из выражения (2), плотность распределения m-й порядковой статистики зависит от закона распределения $F(x)$ и объема выборки N. Однако, как будет показано далее, оценивание закона распределения $\hat{F}_N(x)$ и оценивание функции распределения m-й порядковой статистики $\Phi_N(x_m)$ – две разные задачи со своими особенностями, причем первая из них является подзадачей при решении второй.

Поскольку на практике оценивание показателей надежности по экспериментальным данным в большинстве случаев осуществляется параметрическим методом (при известном типе закона распределения, например наработки на отказ), в рамках проводившихся исследований рассматривался случай, когда закон распределения случайной величины априорно известен. Целью исследования являлся анализ влияния априорного знания закона распределения случайной величины на точность и эффективность оценивания статистических характеристик в зависимости от свойств выборки. Под свойствами выборки понимается ее объем и точность регистрации выборочных данных.

3. АНАЛИЗ ТОЧНОСТИ И ЭФФЕКТИВНОСТИ ОЦЕНИВАНИЯ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ПОРЯДКОВЫХ СТАТИСТИК В ЗАВИСИМОСТИ ОТ ОБЪЕМА И ТОЧНОСТИ РЕГИСТРАЦИИ ИСХОДНЫХ ДАННЫХ

Влияние объемов выборочных данных на точность и эффективность оценивания статистических характеристик

Для исследования точности и эффективности оценок законов распределения порядковых статистик в зависимости от объема исходных данных выполнялся машинный эксперимент. Суть эксперимента состояла в генерации выборок объема N, оценивании по выборкам параметров распределения $\hat{F}(x)$ при априорно известном типе закона распределения $F(x)$, оценивании интегральных функций крайних порядковых статистик $\hat{\Phi}_N(x_{\min})$ и $\hat{\Phi}_N(x_{\max})$, расчете показателей, характеризующих среднюю погрешность оценивания $M[D_N]$ и среднеквадратическое отклонение погрешности оценивания $Sko[D_N]$.

В ходе исследования рассматривались различные типы распределений: экспоненциальный, Гаусса, Вейбулла и Гамма.

Обобщенная схема вычислительного эксперимента представлена на рис. 1.

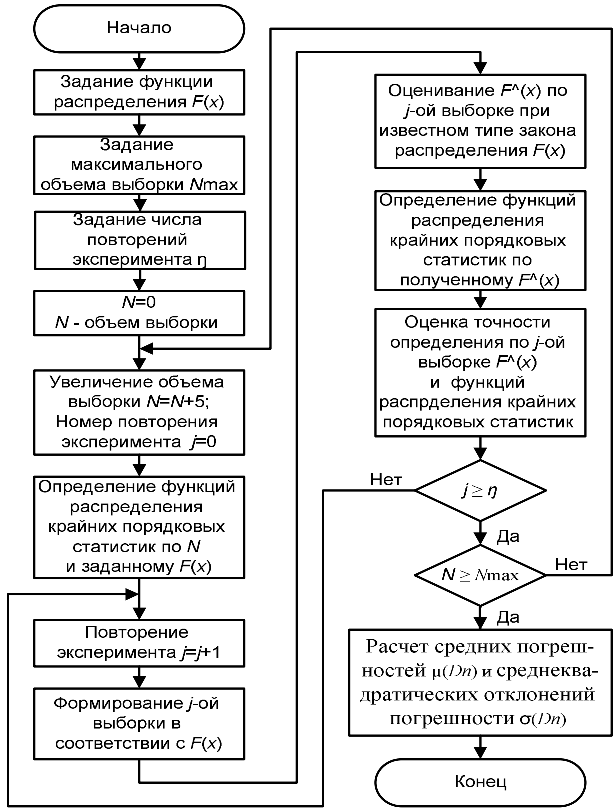

Рис. 1. Обобщенная схема исследования влияния объема исходных данных на точность и эффективность оценивания законов распределения порядковых статистик

Для того чтобы как можно более полно охарактеризовать влияние объема и точности регистрации исходных данных на точность и устойчивость оценок законов распределения порядковых статистик, недостаточно знать значение величин D_N в отдельности для каждого испытания. Только совокупность величин $\{D_N\}$ за несколько циклов испытаний действительно позволяет оценить это влияние. В соответствии с рекомендациями, приведенными в известных литературных источниках, число проведения эксперимента η принималось равным 50.

В качестве показателей точности оценивания статистических характеристик использовалась метрика доминирования:

- точность оценивания

$$F(x)\text{: } D_N^{(\text{Д})}\_F(x) = \max_{x>0}\left|F(x) - \hat{F}(x)\right|;$$

- точность оценивания $\Phi(x_{\min})$:

$$D_N^{(\text{Д})}\_\Phi(x_{\min}) = \max_{x>0}\left|\Phi(x_{\min}) - \hat{\Phi}(x_{\min})\right|;$$

- точность оценивания $\Phi(x_{\max})$:

$$D_N^{(Д)}\_\Phi(x_{\max}) = \max_{x>0}\left|\Phi(x_{\max}) - \hat{\Phi}(x_{\max})\right|.$$

Обработка результатов эксперимента сводится к вычислению показателей точности, характеризуемой величиной математического ожидания совокупности значений

$$M\left[D_N\right] = \frac{1}{\eta}\sum_{j=1}^{\eta} D_N^j ,$$

и устойчивости, характеризуемой величиной среднеквадратического отклонения совокупности значений:

$$Sko\left[D_N\right] = \sqrt{\frac{1}{\eta}\sum_{j=1}^{\eta}\left(M\left[D_N\right] - D_N^j\right)^2} .$$

Оценки точности, получаемые в результате обработки случайных выборок, также являются случайными. В связи с этим, для сглаживания полученных результатов использовалась модель вида $z = e^{A + BN}$, где в качестве z выступает $M[D_N]$; $Sko[D_N]$. Параметры модели A и B определялись в соответствии с рекомендациями [8] методом наименьших квадратов. Для наглядности зависимости представлялись в виде непрерывных кривых.

На рис. 2, 3 представлены результаты эксперимента для экспоненциального закона распределения при $N = \overline{5;100}$.

Практическое совпадение зависимостей точности и устойчивости оценивания законов распределения минимальной порядковой статистики и функции распределения от объема исходных данных объясняется следующим образом.

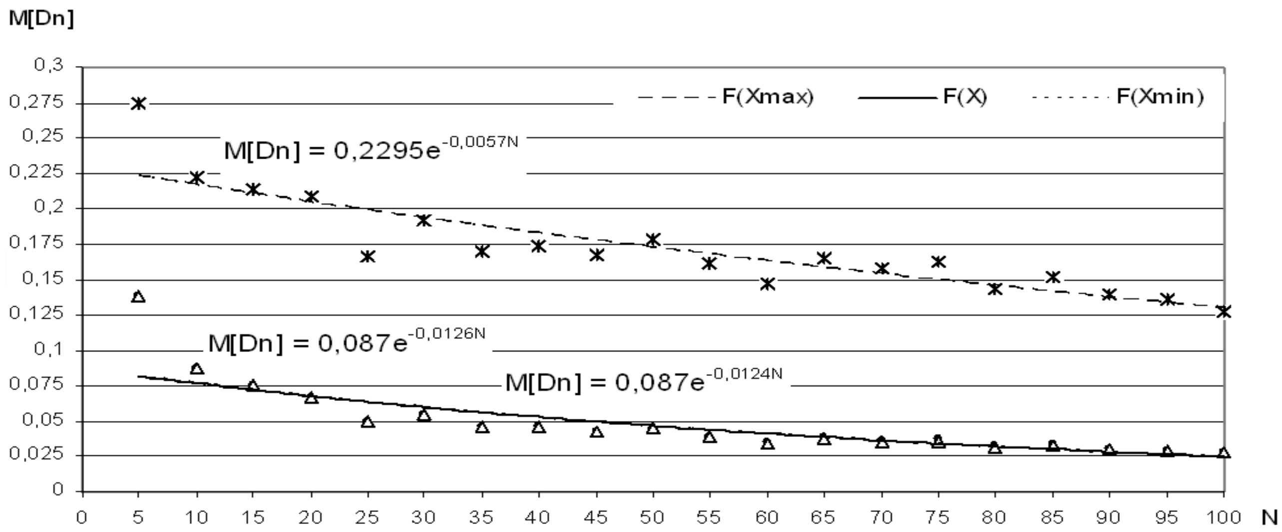

Рис. 2. Точность оценивания $F(X)$, $F(X_{\max})$, $F(X_{\min})$ при экспоненциальном законе распределения

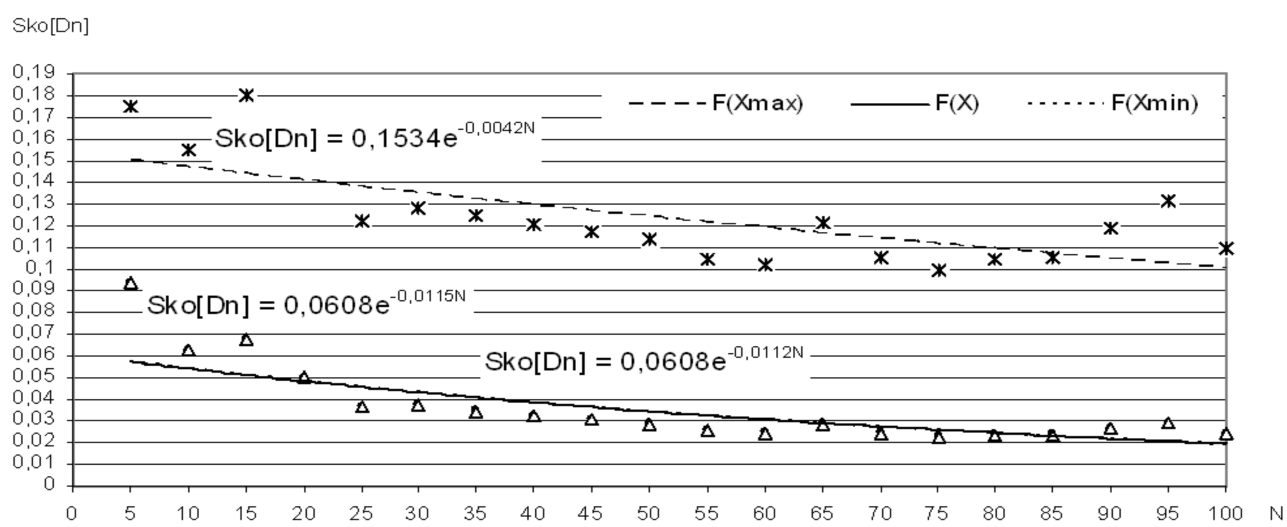

Рис. 3. Эффективность оценивания $F(X)$, $F(X_{\max})$, $F(X_{\min})$ при экспоненциальном законе распределения

Интегральная функция распределения для экспоненциального распределения определяется выражением $F(x) = 1 - e^{-\lambda x}$. Выразим для нее интегральную функцию распределения минимальной (первой) порядковой статистики:

$$\Phi_N(x_{\min}) = \int \varphi_N(x_{\min}) dx = \int \frac{N!}{(N-1)!}\left[1-(1-e^{-\lambda x})\right]^{N-1} \cdot e^{-\lambda x} dx =$$

$$= \int N \cdot e^{-\lambda N x} dx = 1 - e^{-\lambda N x}.$$

Точность и эффективность оценивания $F(x)$ и $\Phi_N(x_{\min})$ зависит от точности и устойчивости оценивания параметра λ по выборочным данным. Поэтому точность и устойчивость оценивания $F(x)$ и $\Phi_N(x_{\min})$ по одним и тем же выборкам совпадают.

Точность и эффективность оценивания как функции распределения, так и крайних порядковых статистик улучшаются с увеличением объема выборок. Чем больше объем выборки, тем точнее можно оценить функцию распределения и крайние порядковые статистики.

В рамках проводившихся исследований было установлено, что для экспоненциального закона распределения точность оценивания максимальной порядковой статистики (средняя погрешность $M[D_N]^{\max}$) ниже точности оценивания минимальной порядковой статистики (средняя погрешность $M[D_N]^{\min}$) в 2÷4,7 раза $\left(\frac{M[D_N]^{\max}}{M[D_N]^{\min}} = 2 \div 4{,}7\right)$; для закона распределения Вейбулла с параметром формы $C = 2$ $\frac{M[D_N]^{\max}}{M[D_N]^{\min}} = 1{,}1 \div 2$; для закона распределения Гамма с параметром формы $C = 2$ $\frac{M[D_N]^{\max}}{M[D_N]^{\min}} = 0{,}8 \div 1{,}9$; для закона распределения Гамма с параметром формы $C = 3$ $\frac{M[D_N]^{\max}}{M[D_N]^{\min}} = 0{,}7 \div 1{,}2$; для закона распределения Гаусса $M[D_N]^{\text{мач}}$ и $M[D_N]^{\min}$ практически совпадают.

Чем более ассиметричен вид закона распределения, тем ниже точность оценивания максимальной порядковой статистики относительно точности оценивания минимальной порядковой статистики. С увеличением симметричности закона распределения, разница между точностью оценивания максимальной порядковой статистики и точностью оценивания минимальной порядковой статистики уменьшается. Так, например, для экспоненциального закона распределения точность оценивания максимальной порядковой статистики ниже точности оценивания минимальной порядковой статистики почти в 2÷4,7 раза, а для закона распределения Гаусса они практически совпадают.

Влияние точности регистрации выборочных данных на точность оценивания статистических характеристик

Для исследования точности оценок законов распределения порядковых статистик в зависимости от точности регистрации исходных данных выполнялся машинный эксперимент, аналогичный вышеописанному (рис. 1). Все отличие заключалось в преобразовании выборочных данных, получаемых с датчиков случайных чисел и соответствующих заданному закону распределения $F(x)$. В ходе эксперимента анализировалось влияние аддитивной помехи, то есть выборочные данные, получаемые с датчиков случайных чисел и соответствующие закону распределения $F(x)$, преобразовались по следующему правилу:

$$x_i^{\text{пом}} = x_i + \xi_i \quad (i = \overline{1;N}),$$

где ξ_i – случайная величина (помеха), соответствующая закону распределения $F_{\text{пом}}(\xi)$. В ходе исследования в качестве $F_{\text{пом}}(\xi)$ рассматривался закон распределения Гаусса.

Параметры закона распределения $F_{\text{пом}}(\xi)$ подбирались таким образом, чтобы соблюдалось условие:

$$\sigma_\xi = \alpha \cdot \sigma_x,$$

где σ_ξ – среднеквадратическое отклонение помехи (математическое ожидание помехи принималось равным нулю);

σ_x – значение среднеквадратического отклонения, соответствующее $F(x)$.

α – коэффициент, характеризующий соотношение масштабов помехи σ_ξ и случайной величины σ_x. В ходе исследований α варьировалось в диапазоне $\alpha \in [0;1]$.

В ходе эксперимента рассматривалось 5 значений коэффициента α, характеризующего соотношение масштабов помехи σ_ξ и случайной величины σ_x: α = 0,00; 0,25; 0,50; 0,75; 1,00.. Объем выборки N менялся в диапазоне $N = \overline{5;100}$.

На рис. 4, 5, 6 представлены результаты эксперимента для экспоненциального закона распределения.

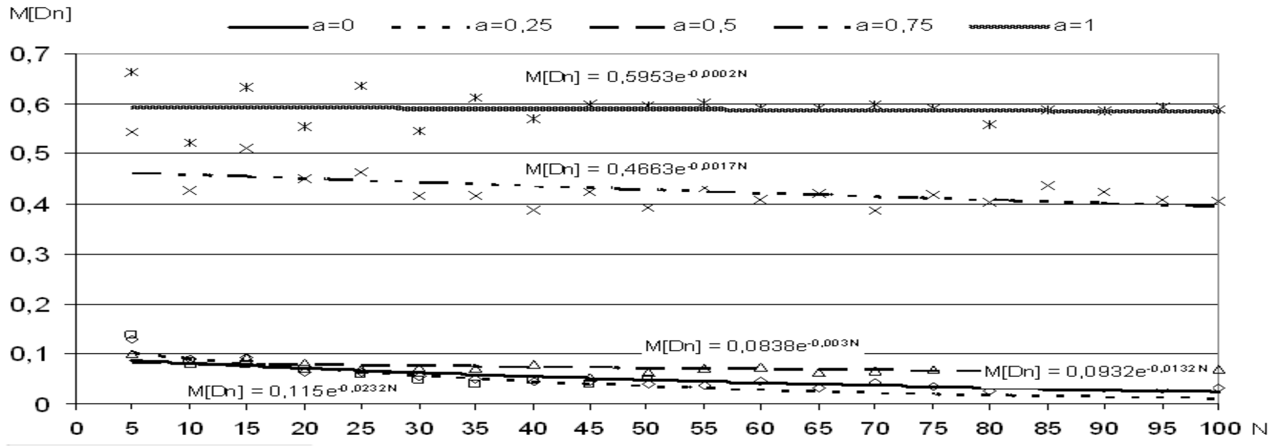

Рис. 4. Оценивание $F(x)$ при экспоненциальном законе распределения и разных коэффициентах помехи a

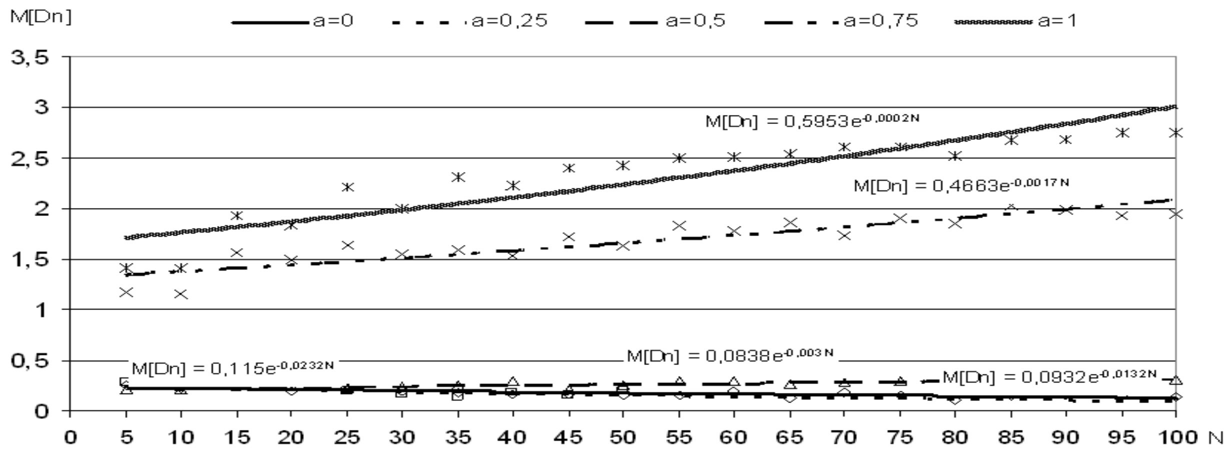

Рис. 5. Оценивание $F(x_{\max})$ при экспоненциальном законе распределения и разных коэффициентах помехи a

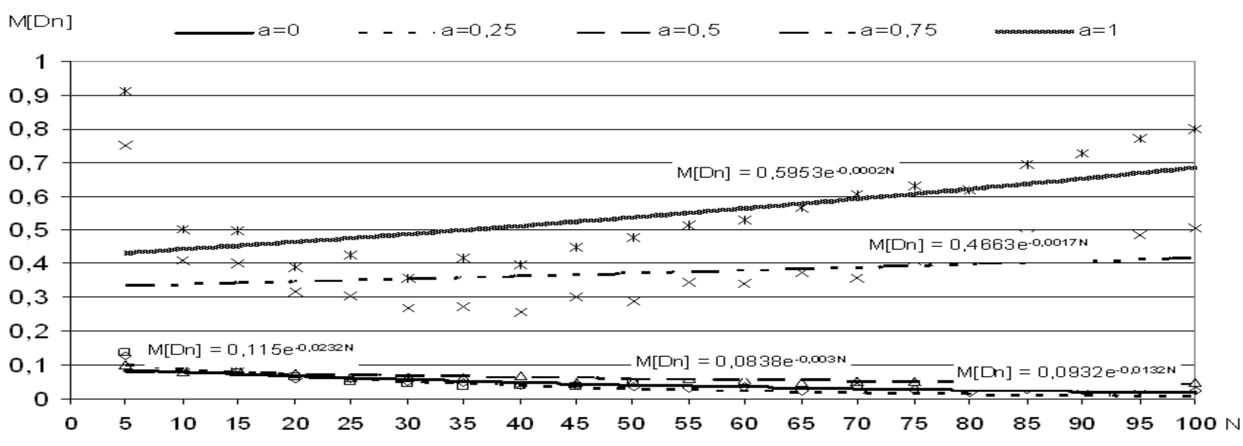

Рис. 6. Оценивание $F(x_{\min})$ при экспоненциальном законе распределения и разных коэффициентах помехи a

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Полученные результаты позволяют заключить, что с увеличением значения коэффициента α, характеризующего соотношение масштабов помехи σ_ξ и случайной величины σ_x, точность оценивания функции распределения $F(x)$ перестает улучшаться с увеличением объема выборки, а при $\alpha = 1$ вообще не зависит от объема выборки. Точность оценивания крайних порядковых статистик при $\alpha \geq 0{,}75$ (для максимальной порядковой статистики $\Phi_N(x_{\max})$ – при $\alpha \geq 0{,}5$) с увеличением объема выборки не только не улучшается, она ухудшается. Это объясняется тем, что при $\alpha \geq 0{,}75$ выборка настолько «засорена» помехой, что увеличение объема выборки N только ухудшает точность оценивания – каждый новый элемент выборки содержит в себе больше помехи, чем полезной информации.

ВЫВОДЫ

При априорно известном законе распределения случайной величины точность оценивания его параметров с увеличением объема выборки повышается независимо от точности регистрации исходных данных. Однако при оценивании крайних порядковых статистик данное утверждение неверно. Точность оценивания крайних порядковых статистик с увеличением объема выборки повышается лишь до определенного уровня точности регистрации исходных данных. При точности регистрации исходных данных ниже данного уровня увеличение объема выборки приводит к снижению точности оценивания крайних порядковых статистик. Поэтому при отсутствии уверенности в достаточной точности регистрации исходных данных даже при априорном знании вида функции распределения случайной величины, целесообразно воспользоваться унифицированной моделью оценивания функции распределения [9].

СПИСОК ЛИТЕРАТУРЫ

1. **Смирнов Н. В., Дунин-Барковский И. В.** Курс теории вероятности и математической статистики. М.: Наука, 1969. 512 с.

2. **Гаскаров Д. В., Шаповалов В. И.** Малая выборка. М.: Статистика, 1978. 248 с.

3. **Дружинин Г. В.** Надежность автоматизированных систем. М.: Энергия, 1977. 536 с.

4. **Барлоу Р., Прошан Ф.** Математическая теория надежности. М.:Советское радио, 1969. 488 с.

5. **Гумбель Э.** Статистика экстремальных значений. М.: Мир, 1965. 450 с.

6. **Дэйвисон М.** Многомерное шкалирование. Методы наглядного представления данных. М.: Финансы и статистика, 1987. 254 с.

7. **Рябинин И. А.** Основы теории и расчета надежности судовых электроэнергетических систем. Л.: Судостроение, 1971. 362 с.

8. **Демиденко Е. З.** Линейная и нелинейная регрессии. М.: Финансы и статистика, 1981. 320 с.

9. Статистическое исследование территориальных систем / М. Б. Гузаиров [и др.]. М.: Машиностроение, 2008. 187 с.

ОБ АВТОРАХ

Гвоздев Владимир Ефимович, зав. каф. автоматизации проектирования инф. систем. Дипл. инженер по электронике (УАИ). Д-р техн. наук по автоматизированным системам управлениям (УГАТУ, 2002). Иссл. в обл. мат. моделирования, прикладной статистики, управления сложными объектами

Таназлы Георгий Иванович, доц. той же каф. Дипл. магистр по информатике и вычисл. технике (УГАТУ, 2002). Канд. техн. наук по матем. моделированию, численным методам и комплексам программ (УГАТУ, 2005). Иссл. в обл. матем. моделирования.

Хасанов Айрат Юлаевич, доц. той же каф. Дипл. математик (МГУ, 1976). Канд. техн. наук по системн. анализу и автоматическому управлению (УАИ, 1986). Иссл. в обл. матем. моделирования, матем. статистики

Абдрафиков Михаил Асхатович, асп. той же каф. Дипл. инженер по системам автоматизированного проектирования (УГАТУ). Иссл. в обл. матем. моделирования.