**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

Нижегородский государственный университет им. Н.И. Лобачевского

ЗАДАЧИ ОЦЕНИВАНИЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

Учебно-методическое пособие

Рекомендовано методической комиссией факультета ВМК
для студентов ННГУ, обучающихся по направлению подготовки
09.03.03 «Прикладная информатика»

Нижний Новгород
2015

УДК 519.21
ББК В171(Я73-4)

ЗАДАЧИ ОЦЕНИВАНИЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ. Составитель Сморкалова В.М.: Учебно-методическое пособие. – Нижний Новгород: Нижегородский госуниверситет, 2015. - 51 с.

Рецензент: к.т.н., доцент **П.Д. Басалин**

Учебно-методическое пособие предназначено для студентов третьего курса факультета вычислительной математики и кибернетики, обучающихся по направлению подготовки «Прикладная информатика». Оно включает четыре практических занятия по разделу «Математическая статистика» общего курса «Теория вероятностей и математическая статистика». Темы занятий: «Некоторые свойства точечных оценок», «Задачи точечного оценивания. Метод моментов», «Метод максимального правдоподобия», «Задачи интервального оценивания». По каждой теме занятий в пособии дается необходимая теоретическая информация и приводится подробное решение типовых задач.

УДК 519.21
ББК В171(Я73-4)

Занятие 1. Некоторые свойства точечных оценок

Цель занятия: освоить способы проверки несмещенности, эффективности (с использованием неравенства Рао-Крамера), состоятельности точечных оценок.

Постановка задачи точечного оценивания неизвестных параметров распределения. Пусть $x_1, x_2, \ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, функция распределения которой - $F(x;\vec{\theta})$ ($x \in R^1$) - известна с точностью до параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, $\vec{\theta} \in \Theta \subset R^k$. Требуется по выборке найти точечную оценку параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$.

При независимых повторных наблюдениях элементы выборки $x_1, x_2, \ldots, x_n$ трактуют как значения n независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, каждая из которых имеет ту же функцию распределения, что и ξ, т.е. $F(x;\vec{\theta})$.

Пусть $k=1$, тогда $\vec{\theta} = \theta$ - скалярная величина.

Определение 1.1. *Точечной оценкой неизвестного параметра* θ, $(\theta \in \Theta \subset R^1)$ *в распределении* $F(x;\theta)$ ($x \in R^1$) *называют измеримую функцию от* $\xi_1, \xi_2, \ldots, \xi_n$ *(статистику), значения которой, при* $\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n$, *используют вместо неизвестного параметра* θ *в качестве его приближения.*

Обычно оценку параметра θ обозначают $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$. Таким образом, если $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$- оценка θ, то $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$ - измеримая функция и $\theta \approx \hat{\theta}_n(x_1, \ldots, x_n)$. Из определения оценки следует, что она может быть выбрана неоднозначно. Рассмотрим некоторые свойства точечных оценок.

Несмещенные оценки. ***Определение 1.2***. *Оценка* $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$ *неизвестного параметра* θ *в распределении* $F(x;\theta)$ *называется несмещенной, если ее математическое ожидание* $\mathrm{M}(\hat{\theta}_n(\xi_1, \ldots, \xi_n)) = \theta$ *для* $\forall \theta \in \Theta$ *и для* $\forall$ n.

Определение 1.3. *Смещением оценки* $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$ *называется величина* $b(\theta) = \mathrm{M}(\hat{\theta}_n(\xi_1, \ldots, \xi_n)) - \theta$.

Пусть $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$ - несмещенная оценка неизвестного параметра θ в распределении $F(x;\theta)$. Произведем N независимых повторных выборок объема n из распределения $F(x;\theta)$ и по каждой из них найдем значение оценки. В результате получим N значений оценки

$\widehat{\theta}_n\left(x_1^{(1)},\ldots,x_n^{(1)}\right)$, $\widehat{\theta}_n\left(x_1^{(2)},\ldots,x_n^{(2)}\right)$, …, $\widehat{\theta}_n\left(x_1^{(N)},\ldots,x_n^{(N)}\right)$, где $x_1^{(i)},\ldots,x_n^{(i)}$ - выборка с номером *i, i=1,2,…,N.*

Так как выборки повторные, то элементы выборки с номером i (*i=1,2,…,N)* можно рассматривать как значения независимых случайных величин $\xi_1^{(i)},\ldots,\xi_n^{(i)}$, каждая из которых распределена так же, как и ξ. Следовательно, $\widehat{\theta}_n\left(x_1^{(i)},\ldots,x_n^{(i)}\right)$, *i=1,2,…,N* будут представлять собой значения независимых одинаково распределенных случайных величин $\widehat{\theta}_n\left(\xi_1^{(i)},\ldots,\xi_n^{(i)}\right)$.

Если математическое ожидание и дисперсия оценки конечны, то есть $\mathrm{M}\left(\widehat{\theta}_n\left(\xi_1^{(i)},\ldots,\xi_n^{(i)}\right)\right)=\theta<\infty$, $D\left(\widehat{\theta}_n\left(\xi_1^{(i)},\ldots,\xi_n^{(i)}\right)\right)<\infty$, то, по закону больших чисел, $\frac{1}{N}\sum\limits_{i=1}^{N}\widehat{\theta}_n\left(\xi_1^{(i)},\ldots,\xi_n^{(i)}\right)\xrightarrow[N\to\infty]{}\theta$ по вероятности. Последний результат означает, что *несмещенность является желательным свойством точечных оценок.*

Рассмотрим несколько задач, в которых требуется определить, является ли та или иная оценка неизвестного параметра несмещенной.

Задача 1.1

Пусть $x_1,x_2,\ldots,x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, распределенной по закону Пуассона с неизвестным параметром λ $(\lambda>0)$. Для оценки λ выбрали следующие функции от результатов наблюдений:

$$a)\ \widehat{\lambda}_n(x_1,\ldots,x_n)=2x_1-x_2; \qquad b)\ \widehat{\lambda}_n(x_1,\ldots,x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i;$$

$$c)\ \widehat{\lambda}_n(x_1,\ldots,x_n)=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{j=1}^{n}x_j\right)^2$$

Требуется определить, являются ли соответствующие оценки несмещенными.

Решение. Найдем для каждой из оценок $\mathrm{M}\left(\widehat{\lambda}_n(\xi_1,\ldots,\xi_n)\right)$.

a) $\mathrm{M}(2\xi_1-\xi_2)=2\mathrm{M}\xi_1-\mathrm{M}\xi_2=2\lambda-\lambda=\lambda$. Оценка несмещенная.

b) $\mathrm{M}\left(\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i\right)=\frac{1}{n}\sum\limits_{i=1}^{n}\mathrm{M}(\xi_i)=\lambda$. Оценка несмещенная.

$$c)\ \mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\frac{1}{n}\sum_{j=1}^{n}\xi_j\right)^2\right)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{M}\left((\xi_i-\lambda)-\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-\lambda\right)\right)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{M}(\xi_i-\lambda)^2+\mathrm{M}\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-\lambda\right)^2-2\mathrm{M}\left((\xi_i-\lambda)\left(\frac{1}{n}\sum_{j=1}^{n}(\xi_j-\lambda)\right)\right)\right)=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(D(\xi_i)+D\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j\right)-\frac{2}{n}D(\xi_i)\right)=\frac{1}{n}\sum_{i=1}^{n}\left(\lambda+\frac{1}{n^2}\sum_{j=1}^{n}D(\xi_j)-\frac{2}{n}\lambda\right)=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(\lambda-\frac{\lambda}{n}\right)=\lambda\left(\frac{n-1}{n}\right)\neq\lambda.$$ Оценка смещенная.

Задача 1.2

Пусть $x_1, x_2, \dots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, функция распределения которой - $F(x;\theta)$ ($x\in R^1$) - известна с точностью до параметра θ, $\theta\in\Theta\subset R^1$. И пусть $\hat{\theta}_n(\xi_1,\dots,\xi_n)$ - несмещенная оценка θ, причем $D(\hat{\theta}_n(\xi_1,\dots,\xi_n))<\infty$. Определить, является ли $\left(\hat{\theta}_n(\xi_1,\dots,\xi_n)\right)^2$ несмещенной оценкой θ^2.

Решение. $\mathrm{M}\left(\hat{\theta}_n(\xi_1,\dots,\xi_n)\right)^2=D(\hat{\theta}_n(\xi_1,\dots,\xi_n))+\left(\mathrm{M}\left(\hat{\theta}_n(\xi_1,\dots,\xi_n)\right)\right)^2=$

$=D(\hat{\theta}_n(\xi_1,\dots,\xi_n))+\theta^2\neq\theta^2$.

Следовательно, $\left(\hat{\theta}_n(\xi_1,\dots,\xi_n)\right)^2$ является смещенной оценкой θ^2.

Задача 1.3

Пусть $x_1, x_2, \dots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, функция распределения которой - $F(x;a,\sigma^2)$ ($x\in R^1$) - известна с точностью до параметров a,σ^2, где $a=\mathrm{M}\xi$, $\sigma^2=D\xi$. Для оценки параметра σ выбрали функцию $\hat{\sigma}_n(x_1,\dots,x_n)=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$, где

$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$. Определить, является ли выбранная оценка параметра σ несмещенной.

Решение. Предположим, что $\widehat{\sigma}_n(\xi_1,\ldots,\xi_n)$ является несмещенной оценкой σ. Тогда, как было показано в предыдущей задаче, квадрат этой оценки, то есть $\left(\widehat{\sigma}_n(\xi_1,\ldots,\xi_n)\right)^2$, будет смещенной оценкой σ^2. Найдем

$$\mathrm{M}\left(\widehat{\sigma}_n(\xi_1,\ldots,\xi_n)\right)^2=\mathrm{M}\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i-\frac{1}{n}\sum_{j=1}^{n}\xi_j\right)^2\right)=$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\mathrm{M}\left((\xi_i-a)-\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-a\right)\right)^2=$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\left(D(\xi_i)+D\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j\right)-\frac{2}{n}D(\xi_i)\right)==\frac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2+\frac{1}{n^2}\sum_{j=1}^{n}D(\xi_j)-\frac{2}{n}\sigma^2\right)=$$

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\sigma^2-\frac{\sigma^2}{n}\right)=\sigma^2.$$

Мы показали, что $\left(\widehat{\sigma}_n(\xi_1,\ldots,\xi_n)\right)^2$ является несмещенной оценкой σ^2. Следовательно, предположение о том, что $\widehat{\sigma}_n(\xi_1,\ldots,\xi_n)$ - несмещенная оценка σ, неверно, то есть $\widehat{\sigma}_n(\xi_1,\ldots,\xi_n)$ является смещенной оценкой σ.

Задача 1.4

Пусть x_1, x_2 - результаты повторных независимых наблюдений над случайной величиной ξ, имеющей нормальное распределение с параметрами $a=\mathrm{M}\xi=0,\ \ \sigma^2=D\xi<\infty$. В качестве оценки параметра σ выбрали статистику $\widehat{\sigma}_n(\xi_1,\xi_2)=\frac{\sqrt{\pi}}{2}|\xi_1+\xi_2|$. Определить, является ли данная оценка несмещенной.

Решение. Найдем $\mathrm{M}\left(\widehat{\sigma}_n(\xi_1,\xi_2)\right)=\mathrm{M}\left(\frac{\sqrt{\pi}}{2}|\xi_1+\xi_2|\right)$. Так как ξ_1 и ξ_2 независимые случайные величины, распределенные так же, как и ξ, то, по свойствам нормального распределения, случайная величина $\eta=\xi_1+\xi_2$ будет иметь нормальное распределение с параметрами $\mathrm{M}\eta=0,\ \ D\eta=2\sigma^2$.

Используя определение математического ожидания функции от случайной величины, получим $M\left(\frac{\sqrt{\pi}}{2}\left|\xi_1+\xi_2\right|\right)=M\left(\frac{\sqrt{\pi}}{2}|\eta|\right)=\frac{\sqrt{\pi}}{2}\frac{2}{2\sigma\sqrt{\pi}}\int\limits_0^{+\infty} x e^{-\frac{x^2}{4\sigma^2}}dx=\sigma$.

Следовательно, $\hat{\sigma}_n(\xi_1,\xi_2)=\frac{\sqrt{\pi}}{2}\left|\xi_1+\xi_2\right|$ является несмещенной оценкой σ.

Задача 1.5

Пусть $x_1, x_2, \ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ с плотностью распределения $f_\xi(x;\theta)=\begin{cases}\frac{1}{b-\theta}, & x\in(\theta,b)\\ 0, & x\notin(\theta,b)\end{cases}$ и функцией распределения $F_\xi(y)=\begin{cases}0, & y\le\theta\\ \frac{y-\theta}{b-\theta}, & \theta<y\le b\\ 1, & y>b\end{cases}$, где b – известно, а θ - неизвестный параметр, $\theta<b$.

Для оценки неизвестного параметра θ выбрали функцию $\hat{\theta}_n(x_1,\ldots,x_n)=\min\{x_1,x_2,\ldots,x_n\}$. Определить, является ли выбранная оценка несмещенной.

Решение. Найдем функцию распределения случайной величины $\hat{\theta}_n(\xi_1,\ldots,\xi_n)=\min\{\xi_1,\xi_2,\ldots,\xi_n\}$. $F(y)=P\{\hat{\theta}_n(\xi_1,\ldots,\xi_n)<y\}=$

$$=1-P\{\hat{\theta}_n(\xi_1,\ldots,\xi_n)\ge y\}=1-P\{\xi_1\ge y,\xi_2\ge y,\ldots,\xi_n\ge y\}=1-\prod_{i=1}^{n}P\{\xi_i\ge y\}=$$

$=1-\prod\limits_{i=1}^{n}(1-P\{\xi_i<y\})=1-(1-F_\xi(y))^n$. Найдем плотность распределения $f(y)$ случайной величины $\hat{\theta}_n(\xi_1,\ldots,\xi_n)$. $f(y)=\frac{dF(y)}{dy}=n(1-F_\xi(y))^{n-1}f_\xi(y)$.

Следовательно, $f(y)=n\left(1-\frac{y-\theta}{b-\theta}\right)^{n-1}\frac{1}{b-\theta}=\frac{n}{(b-\theta)^n}(b-y)^{n-1},\quad y\in(\theta,b)$ и $f(y)=0,\quad y\notin(\theta,b)$.

Найдем математическое ожидание $\hat{\theta}_n(\xi_1,\ldots,\xi_n)$.

$$\mathrm{M}\left(\widehat{\theta}_n\left(\xi_1,\ldots,\xi_n\right)\right)=\int\limits_{\theta}^{b} y\frac{n}{(b-\theta)^n}(b-y)^{n-1}dy=\frac{n}{(b-\theta)^n}\left(-y\frac{(b-y)^n}{n}\right)\Bigg|_{\theta}^{b}+$$

$$+\frac{n}{(b-\theta)^n}\int\limits_{\theta}^{b}\frac{(b-y)^n}{n}dy=\frac{n}{(b-\theta)^n}\left(\frac{\theta(b-\theta)^n}{n}-\left(\frac{(b-y)^{n+1}}{n(n+1)}\right)\Bigg|_{\theta}^{b}\right)=$$

$$=\frac{n}{n+1}\theta+\frac{b}{n+1}.$$

Таким образом, $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)=\min\{\xi_1,\xi_2,\ldots,\xi_n\}$ является смещенной оценкой параметра θ.

Эффективные оценки. Предположим, что есть две несмещенные оценки $\widehat{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$ и $\breve{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$ параметра θ $\left(\theta\in\Theta\subset R^1\right)$ в распределении $F(x;\theta)$. Если дисперсии оценок конечны и удовлетворяют условию $D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\le D\left(\breve{\theta}_n(\xi_1,\ldots,\xi_n)\right)$ для любого фиксированного θ из области допустимых значений, то следует предпочесть оценку $\widehat{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$, поскольку ее разброс относительно параметра θ меньше, чем разброс оценки $\breve{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$.

Определение 1.4. *Несмещенную оценку* $\widehat{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$ *неизвестного параметра* θ, $\left(\theta\in\Theta\subset R^1\right)$ *в распределении* $F(x;\theta)$ $(x\in R^1)$ *называют эффективной (или несмещенной оценкой с минимальной дисперсией), если для всякой другой несмещенной оценки* $\breve{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$ $D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\le D\left(\breve{\theta}_n(\xi_1,\ldots,\xi_n)\right)$ *для любого* $\theta\in\Theta$ *(предполагается, что дисперсии оценок конечны).*

Очевидно, что *эффективная оценка является наиболее предпочтительной среди всех несмещенных оценок параметра* θ, поскольку разброс ее значений относительно θ минимален.

Для проверки эффективности оценок в некоторых случаях может быть использовано неравенство Рао-Крамера.

Рассмотрим случай, когда $x_1,x_2,\ldots,x_n$ - результаты n повторных независимых наблюдений над непрерывной случайной величиной ξ, плотность распределения которой - $f(x;\theta)$ $(x\in R^1)$ - известна с точностью до параметра θ $\left(\theta\in\Theta\subset R^1\right)$.

Теорема 1.1. *Пусть при любом фиксированном* θ *из области допустимых значений выполнены условия [4]*

$$\mathrm{M}\left|\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right|<\infty\ ;\quad \mathrm{M}\left|\frac{\partial^2}{\partial\theta^2}\ln f(\xi;\theta)\right|<\infty\ ;\quad \mathrm{M}\left|\frac{\partial}{\partial\theta}\ln f(\xi;\theta)\right|^2<\infty,$$

и $g(\theta)$ *- дифференцируемая функция на* Θ*, для которой существует несмещенная оценка* $\hat{\theta}_n\left(\xi_1,\ldots,\xi_n\right)$ *с конечной дисперсией, удовлетворяющая условию*

$$\int\limits_{-\infty}^{+\infty}\cdots\int\limits_{-\infty}^{+\infty}\left|\hat{\theta}_n(x_1,\ldots,x_n)\frac{\partial}{\partial\theta}\breve{f}(x_1,\ldots,x_n;\theta)\right|dx_1\cdots dx_n<\infty,\quad \forall\theta\in\Theta\ ,$$

где $$\breve{f}(x_1,\ldots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta).$$

Тогда, для любого $\theta\in\Theta$, $D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)=\mathrm{M}\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)-g(\theta)\right)^2\geq\dfrac{(g'(\theta))^2}{nI(\theta)}$,

где $I(\theta)=\mathrm{M}\left|\dfrac{\partial}{\partial\theta}\ln f(\xi;\theta)\right|^2>0$ *- количество информации Фишера о параметре* θ*, содержащееся в единичном наблюдении.*

Неравенство

$$D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\geq\frac{(g'(\theta))^2}{nI(\theta)} \qquad (1.1)$$

называется неравенством Рао-Крамера.

Аналогичное неравенство имеет место и для случая, когда наблюдаемая случайная величина ξ является дискретной. Количество информации Фишера в этом случае определяется выражением $I(\theta)=\mathrm{M}\left|\dfrac{\partial}{\partial\theta}\ln P(\xi;\theta)\right|^2$, где $P(x;\theta)=P\{\xi=x\}$. В частном случае, когда $g(\theta)=\theta$, неравенство (1.1) принимает вид

$$D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\geq\frac{1}{nI(\theta)}. \qquad (1.2)$$

Если для закона распределения наблюдаемой случайной величины выполнены условия приведенной выше теоремы, то та несмещенная оценка, для которой в неравенстве Рао-Крамера достигается равенство, является эффективной.

Рассмотрим примеры решения задач, в которых требуется определить, является ли та или иная оценка эффективной.

Задача 1.6

Пусть $x_1, x_2, \dots, x_n$ - результаты n повторных независимых наблюдений над непрерывной случайной величиной ξ, плотность распределения которой - $f(x;\theta)$ $(x \in R^1)$ - известна с точностью до параметра θ $(\theta > 0)$ и имеет вид $f(x;\theta) = \begin{cases} \theta e^{-\theta x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$.

В качестве оценки функции $g(\theta) = \frac{1}{\theta}$ выбрана статистика $\hat{\theta}_n(\xi_1, \dots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n} \xi_i$, математическое ожидание которой $\mathrm{M}(\hat{\theta}_n(\xi_1, \dots, \xi_n)) = \frac{1}{\theta}$.

Определить, является ли данная статистика эффективной оценкой $g(\theta) = \frac{1}{\theta}$.

Решение. Можно показать, что условия Теоремы 1.1. в данном случае выполняются, поэтому для решения поставленной задачи воспользуемся неравенством Рао-Крамера. Найдем дисперсию оценки: $D(\hat{\theta}_n(\xi_1, \dots, \xi_n)) = \mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n} \xi_i - \frac{1}{\theta}\right)^2 = \frac{1}{n\,\theta^2}$. В целях вычисления количества информации Фишера, найдем $\ln f(x;\theta) = \ln\theta - \theta x$ и $\frac{\partial}{\partial\theta}\ln f(x;\theta) = \frac{1}{\theta} - x$. Тогда $I(\theta) = \mathrm{M}\left(\xi - \frac{1}{\theta}\right)^2 = \frac{1}{\theta^2}$ и $\frac{(g'(\theta))^2}{nI(\theta)} = \frac{\theta^2}{\theta^4 n} = \frac{1}{n\theta^2}$. Таким образом, мы показали, что $D(\hat{\theta}_n(\xi_1, \dots, \xi_n)) = \frac{(g'(\theta))^2}{nI(\theta)}$. Следовательно, $\hat{\theta}_n(\xi_1, \dots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n} \xi_i$ - эффективная оценка $g(\theta) = \frac{1}{\theta}$.

Задача 1.7

Пусть $x_1, x_2, \dots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, имеющей нормальное распределение с параметрами a, θ, где $a = \mathrm{M}\xi$ - известно, а $\theta = D\xi$ - неизвестный параметр. В качестве оценки параметра θ выбрана статистика $\hat{\theta}_n(\xi_1, \dots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n} (\xi_i - a)^2$, математическое ожидание которой $\mathrm{M}(\hat{\theta}_n(\xi_1, \dots, \xi_n)) = \theta$. Определить, является ли данная оценка эффективной.

Решение. Плотность распределения вероятностей наблюдаемой случайной величины ξ имеет вид $f(x;\theta)=\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-a)^2}{2\theta}}, \quad -\infty<x<+\infty$. Можно показать, что условия Теоремы 1.1. в данном случае выполняются, поэтому для решения поставленной задачи воспользуемся неравенством Рао-Крамера.

Вычислим дисперсию оценки:

$$D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right) = D\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_i-a)^2\right) = \frac{1}{n^2}\sum_{i=1}^{n}D(\xi_i-a)^2 = \frac{D(\xi-a)^2}{n}$$

$D(\xi-a)^2 = D\left(\frac{(\xi-a)}{\sqrt{\theta}}\sqrt{\theta}\right)^2 = D\left(\theta\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2\right) = \theta^2 D\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2$. Так как $\xi\in N(a,\theta)$, то $\frac{\xi-a}{\sqrt{\theta}}\in N(0,1)$. Следовательно, случайная величина $\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2$ имеет Хи-квадрат распределение с одной степенью свободы и $D\left(\frac{\xi-a}{\sqrt{\theta}}\right)^2=2$. Таким образом, $D(\xi-a)^2=2\theta^2$, а $D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right) = \frac{2\theta^2}{n}$.

Найдем количество информации Фишера.

$$\ln f(x;\theta)=\ln\left(\frac{1}{\sqrt{2\pi\theta}}e^{-\frac{(x-a)^2}{2\theta}}\right)=\ln\left(\frac{1}{\sqrt{2\pi}}\right)-\frac{1}{2}\ln\theta-\frac{(x-a)^2}{2\theta},$$

$\frac{\partial}{\partial\theta}\ln f(x;\theta)=-\frac{1}{2\theta}+\frac{(x-a)^2}{2\theta^2}$. Тогда $I(\theta)=\mathrm{M}\left(\frac{(\xi-a)^2}{2\theta^2}-\frac{1}{2\theta}\right)^2=$
$=\frac{1}{4\theta^4}D(\xi-a)^2=\frac{1}{4\theta^4}2\theta^2=\frac{1}{2\theta^2}$, а $\frac{1}{nI(\theta)}=\frac{2\theta^2}{n}$. Так как $D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right) = \frac{1}{nI(\theta)}$, то $\hat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-a)^2$ - эффективная оценка θ.

Задача 1.8

Пусть $x_1, x_2, \ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, имеющей биномиальное распределение с параметрами m, θ, где m – число испытаний по схеме Бернулли, а θ - вероятность “успеха” в одном испытании – неизвестный параметр, $0<\theta<1$. В качестве оценки функции $g(\theta)=m\theta$ выбрана статистика $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$, математическое ожидание которой $\mathrm{M}\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)=m\theta$. Определить, является ли данная оценка эффективной.

Решение. Для наблюдаемой случайной величины ξ $P(k;\theta)=P\{\xi=k\}=C_m^k\theta^k(1-\theta)^{m-k}, \quad k=0,\ldots,m.$ Можно показать, что условия Теоремы 1.1. в данном случае выполняются, поэтому для решения поставленной задачи воспользуемся неравенством Рао-Крамера.

Найдем дисперсию оценки.

$$D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right) = D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}D(\xi_i)=\frac{m\theta(1-\theta)}{n}.$$

Найдем количество информации Фишера.

$$\ln P(k;\theta)=\ln C_m^k+k\ln\theta+(m-k)\ln(1-\theta),$$

$$\frac{\partial}{\partial\theta}\ln P(k;\theta)=\frac{k}{\theta}-\frac{m-k}{1-\theta}=\frac{k-m\theta}{\theta(1-\theta)}. \qquad I(\theta)=\mathrm{M}\left(\frac{\partial}{\partial\theta}\ln P(\xi;\theta)\right)^2=$$

$$=\frac{1}{\theta^2(1-\theta)^2}\mathrm{M}(\xi-m\theta)^2=\frac{m\theta(1-\theta)}{\theta^2(1-\theta)^2}=\frac{m}{\theta(1-\theta)}.$$

Тогда $\dfrac{(g'(\theta))^2}{nI(\theta)}=\dfrac{m^2\theta(1-\theta)}{nm}=\dfrac{m\theta(1-\theta)}{n}$.

Следовательно, $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ - эффективная оценка $g(\theta)=m\theta$.

Задача 1.9

Пусть $x_1, x_2, \ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, имеющей распределение Пуассона с неизвестным параметром $\theta = \mathrm{M}\xi,\ \theta > 0$. В качестве оценки параметра θ выбрана статистика $\widehat{\theta}_n(\xi_1, \ldots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, $\mathrm{M}(\widehat{\theta}_n(\xi_1, \ldots, \xi_n)) = \theta$. Определить, является ли данная оценка эффективной.

Решение. Для наблюдаемой случайной величины ξ $P(k;\theta) = P\{\xi = k\} = \frac{\theta^k}{k!}e^{-\theta}, \quad k = 0,1,\ldots$. Можно показать, что условия Теоремы 1.1. в данном случае выполняются, поэтому для решения поставленной задачи воспользуемся неравенством Рао-Крамера. Найдем дисперсию оценки. $D(\widehat{\theta}_n(\xi_1, \ldots, \xi_n)) = D\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}D(\xi_i) = \frac{\theta}{n}$. Найдем количество информации Фишера. $\ln P(k;\theta) = -\ln(k!) + k\ln\theta - \theta$,

$$\frac{\partial}{\partial\theta}\ln P(k;\theta) = \frac{k}{\theta} - 1 = \frac{k-\theta}{\theta}. \qquad I(\theta) = \mathrm{M}\left(\frac{\partial}{\partial\theta}\ln P(\xi;\theta)\right)^2 = \quad = \frac{1}{\theta^2}\mathrm{M}(\xi - \theta)^2 = \frac{1}{\theta}.$$

Тогда $\frac{1}{nI(\theta)} = \frac{\theta}{n}$. Следовательно, $\widehat{\theta}_n(\xi_1, \ldots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n}\xi_i$ - эффективная оценка θ.

Следует отметить, что не для всякого закона распределения может быть использовано неравенство Рао-Крамера. Так, например, с помощью данного неравенства нельзя проверить является ли эффективной та или иная оценка параметра θ, если наблюдаемая случайная величина распределена равномерно на интервале $(0,\theta)$.

Состоятельные оценки. Свойства несмещенности и эффективности оценок рассматриваются при фиксированном объеме выборки n. Пусть теперь $n \to \infty$.

Определение 1.5. *Оценка* $\widehat{\theta}_n(\xi_1, \ldots, \xi_n)$ *называется состоятельной оценкой параметра* θ, *если* $\widehat{\theta}_n(\xi_1, \ldots, \xi_n) \underset{n\to\infty}{\overset{p}{\to}} \theta$. Иными словами, для состоятельной оценки $\widehat{\theta}_n(\xi_1, \ldots, \xi_n)$ отклонение ее от θ на величину ε $(\varepsilon > 0)$ и более становится маловероятным при большом объеме выборки. Это свойство оценки является очень важным, так как несостоятельные оценки на практике не используются.

При определении состоятельности оценок могут оказаться полезными следующие теоремы.

Теорема 1.2. *Пусть* $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ *- состоятельная оценка параметра* θ *в распределении* $F(x,\theta)$, *а* $g(\theta)$*- непрерывная функция, тогда* $g(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))$*- состоятельная оценка* $g(\theta)$.

Теорема 1.3. *Пусть* $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ *- несмещенная оценка параметра* θ *в распределении* $F(x,\theta)$, *то есть* $\mathrm{M}(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))=\theta$ *для всякого* θ *из области допустимых значений и для всякого* n, *и пусть* $D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)) \xrightarrow[n\to\infty]{} 0$, *тогда* $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ *- состоятельная оценка* θ.

Обобщением теоремы 1.3. является следующая теорема.

Теорема 1.4. *Пусть* $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ *- оценка параметра* θ *в распределении* $F(x,\theta)$ *такая, что* $\mathrm{M}(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)) \xrightarrow[n\to\infty]{} \theta$ и $D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)) \xrightarrow[n\to\infty]{} 0$, *тогда* $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ *- состоятельная оценка* θ.

Рассмотрим несколько задач, в каждой из которых требуется определить, является ли та или иная оценка состоятельной.

Задача 1.10

Пусть $x_1, x_2,\ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, математическое ожидание и дисперсия которой неизвестны. В качестве оценки дисперсии выбрана статистика $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\widehat{a}_n)^2$, где $\widehat{a}_n=\frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$.

В предположении, что случайная величина ξ имеет моменты до четвертого порядка включительно, покажем, что данная оценка является состоятельной.

Решение. Введем обозначения $a=\mathrm{M}\xi$, $\theta=D\xi$ и найдем

$$\mathrm{M}(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))=\mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\widehat{a}_n)^2\right)=\mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}((\xi_i-a)-(\widehat{a}_n-a))^2\right)=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{M}(\xi_i-a)^2+\mathrm{M}(\widehat{a}_n-a)^2-2\mathrm{M}\left((\xi_i-a)\left(\frac{1}{n}\sum_{j=1}^{n}\xi_j-a\right)\right)\right)=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(\theta+\frac{\theta}{n}-2\frac{\theta}{n}\right)=\frac{\theta(n-1)}{n}=\theta-\frac{\theta}{n}\xrightarrow[n\to\infty]{}\theta .$$

Найдем дисперсию $D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)$, но предварительно покажем, что $\hat{\theta}_n\left(\breve{\xi}_1,\ldots,\breve{\xi}_n\right)=\hat{\theta}_n(\xi_1,\ldots,\xi_n)$, где $\breve{\xi}_i=\xi_i-a,\quad i=1,\ldots,n$.

$$\hat{\theta}_n\left(\breve{\xi}_1,\ldots,\breve{\xi}_n\right)=\frac{1}{n}\sum_{i=1}^{n}\left((\xi_i-a)-\frac{1}{n}\sum_{j=1}^{n}\left(\xi_j-a\right)\right)^2=\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-a-\frac{1}{n}\sum_{j=1}^{n}\xi_j+a\right)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\hat{a}_n\right)^2=\hat{\theta}_n(\xi_1,\ldots,\xi_n).$$

Следовательно, $D\left(\hat{\theta}_n\left(\breve{\xi}_1,\ldots,\breve{\xi}_n\right)\right)=D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)$. Так как $\mathrm{M}\left(\breve{\xi}_i\right)=0$, то без ограничения общности при вычислении $D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)$ можно положить $\mathrm{M}(\xi_i)=a=0$.

Представим $\hat{\theta}_n(\xi_1,\ldots,\xi_n)$ в виде

$$\hat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\hat{a}_n\right)^2=\frac{1}{n}\sum_{i=1}^{n}\xi_i^2-2\hat{a}_n\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)+\left(\hat{a}_n\right)^2=$$

$$=\frac{1}{n}\sum_{i=1}^{n}\xi_i^2-\left(\hat{a}_n\right)^2. \tag{1.3}$$

По свойству дисперсии случайной величины $D\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)=\mathrm{M}\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)^2-\left(\mathrm{M}\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\right)^2$. В свою очередь, второй момент оценки с учетом (1.3) примет вид $\mathrm{M}\left(\hat{\theta}_n(\xi_1,\ldots,\xi_n)\right)^2=\mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^2\right)^2+\mathrm{M}\left(\hat{a}_n\right)^4-2\mathrm{M}\left(\left(a_n{}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^2\right)\right)$.

Найдем выражения для трех слагаемых в правой части последнего равенства, используя свойства математического ожидания и независимость случайных величин $\xi_1,\ldots,\xi_n$, каждая из которых имеет математическое ожидание $a=0$ и дисперсию θ.

$$\mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^2\right)^2=\frac{1}{n^2}\mathrm{M}\left(\left(\sum_{i=1}^{n}\xi_i^4\right)+\left(\sum_{\substack{i,j=1\\ i\neq j}}^{n}\xi_i^2\xi_j^2\right)\right)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{M}\left(\xi_i^4\right)+$$

$$+\frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n} \mathrm{M}\xi_i^2 \times \mathrm{M}\xi_j^2 = \frac{1}{n^2}\sum_{i=1}^{n} m_4 + \frac{1}{n^2}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\theta^2 = \frac{1}{n}m_4 + \frac{n-1}{n}\theta^2,$$

где $m_4 = \mathrm{M}\xi^4$.

Чтобы вычислить третье слагаемое, преобразуем его:

$$\mathrm{M}\left(\hat{a}_n^2\sum_{i=1}^{n}\xi_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\left(\sum_{j=1}^{n}\xi_j\right)^2\sum_{i=1}^{n}\xi_i^2\right) =$$

$$= \frac{1}{n^2}\mathrm{M}\left(\left(\sum_{j=1}^{n}\xi_j^2 + \sum_{\substack{j,k=1\\ j\neq k}}^{n}\xi_j\xi_k\right)\sum_{i=1}^{n}\xi_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\sum_{j=1}^{n}\xi_j^2\sum_{i=1}^{n}\xi_i^2\right) +$$

$$+\frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n}\xi_i^2\sum_{\substack{j,k=1\\ j\neq k}}^{n}\xi_j\xi_k\right).$$

Так как $\xi_1,\ldots,\xi_n$ - независимые случайные величины, то

$$\mathrm{M}\left(\sum_{\substack{j=1\\ j\neq k}}^{n}\xi_j\xi_k\right) = \sum_{\substack{j=1\\ j\neq k}}^{n}\mathrm{M}\xi_j\mathrm{M}\xi_k = 0, \quad k=\overline{1,n} \quad \text{и} \quad \mathrm{M}\left(\sum_{i=1}^{n}\xi_i^2\sum_{\substack{j,k=1\\ j\neq k}}^{n}\xi_j\xi_k\right) =$$

$$= \sum_{\substack{i,j,k=1\\ j\neq k, i\neq k, i\neq j}}^{n}\mathrm{M}\left(\xi_i^2\xi_j\xi_k\right) + 2\sum_{\substack{i,j=1\\ i\neq j}}^{n}\mathrm{M}\left(\xi_i^3\xi_j\right) = 0.$$

Следовательно,

$$\mathrm{M}\left(\hat{a}_n^2\sum_{i=1}^{n}\xi_i^2\right) = \frac{1}{n^2}\mathrm{M}\left(\sum_{i=1}^{n}\xi_i^4 + \sum_{\substack{i,j=1\\ i\neq j}}^{n}\xi_i^2\xi_j^2\right) = \frac{1}{n^2}\left(n\,m_4 + n(n-1)\theta^2\right).$$

Аналогично можно показать, что $\mathrm{M}\hat{a}_n^4 = \frac{1}{n^4}\mathrm{M}\left(\sum_{i=1}^{n}\xi_i\right)^4 = \frac{m_4 + 3(n-1)\theta^2}{n^3}$.

В итоге получаем:

$$\mathrm{M}\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)^2=\mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^2\right)^2+\mathrm{M}\left(\widehat{a}_n\right)^4-2\mathrm{M}\left(\left(a_n{}^2\right)\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i^2\right)\right)=$$

$$=\frac{1}{n}m_4+\frac{n-1}{n}\theta^2+\frac{m_4+3(n-1)\theta^2}{n^3}-$$

$$-\frac{2}{n}\left(\frac{1}{n^2}\left(n\,m_4+n(n-1)\theta^2\right)\right)=\theta^2+\frac{m_4-3\theta^2}{n}-\frac{2m_4-5\theta^2}{n^2}+\frac{m_4-3\theta^2}{n^3}.$$

Тогда, $D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)=\mathrm{M}\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)^2-\left(\mathrm{M}\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\right)^2=$

$$=\theta^2+\frac{m_4-3\theta^2}{n}-\frac{2m_4-5\theta^2}{n^2}+\frac{m_4-3\theta^2}{n^3}-\frac{\theta^2(n-1)^2}{n^2}=$$

$$=\frac{m_4-\theta^2}{n}-\frac{2m_4-4\theta^2}{n^2}+\frac{m_4-3\theta^2}{n^3}.$$

Так как $D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)\xrightarrow[n\to\infty]{}0$, то, в соответствии с теоремой 1.4., $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ является состоятельной оценкой дисперсии случайной величины ξ.

Задача 1.11

Пусть $x_1,x_2,\ldots,x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, математическое ожидание и дисперсия которой неизвестны. В качестве оценки дисперсии выбрана статистика $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n-1}\sum_{i=1}^{n}(\xi_i-\widehat{a}_n)^2$, где $\widehat{a}_n=\frac{1}{n}\sum_{i=1}^{n}\xi_i$.

В предположении, что случайная величина ξ имеет моменты до четвертого порядка включительно, покажем, что данная оценка является состоятельной.

Решение. Представим $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ в виде $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)=\frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\widehat{a}_n)^2\right)$. Тогда, с учетом результатов, полученных при решении предыдущей задачи, имеем:

$$\mathrm{M}\left(\widehat{\theta}_n\left(\xi_1,\dots,\xi_n\right)\right)=\frac{n}{n-1}\mathrm{M}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\widehat{a}_n\right)^2\right)=\frac{n}{n-1}\left(\frac{n-1}{n}\theta\right)=\theta,$$

$$D\left(\widehat{\theta}_n\left(\xi_1,\dots,\xi_n\right)\right)=\frac{n^2}{(n-1)^2}D\left(\frac{1}{n}\sum_{i=1}^{n}\left(\xi_i-\widehat{a}_n\right)^2\right)\xrightarrow[n\to\infty]{}0.$$

По теореме 1.3., $\widehat{\theta}_n\left(\xi_1,\dots,\xi_n\right)$ - состоятельная оценка дисперсии случайной величины ξ.

Задача 1.12

Пусть $x_1, x_2,\dots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ с плотностью

распределения $f_{\xi}(x;\theta)=\begin{cases}\dfrac{1}{b-\theta}, & x\in(\theta,b)\\ 0, & x\notin(\theta,b)\end{cases}$, где b – известно, а θ - неизвестный параметр, $\theta<b$. Для оценки неизвестного параметра θ выбрана функция $\widehat{\theta}_n\left(x_1,\dots,x_n\right)=\min\left\{x_1,x_2,\dots,x_n\right\}$. Определить, является ли выбранная оценка состоятельной.

Решение. По обобщенному неравенству Чебышева, для всякого $\varepsilon>0$

$$\mathrm{P}\left\{\left|\widehat{\theta}_n-\theta\right|>\varepsilon\right\}\le\frac{\mathrm{M}\left(\widehat{\theta}_n-\theta\right)^2}{\varepsilon^2}. \tag{1.4}$$

Представим $\mathrm{M}\left(\widehat{\theta}_n-\theta\right)^2$ в виде $\mathrm{M}\left(\widehat{\theta}_n-\theta\right)^2=\mathrm{M}\left(\widehat{\theta}_n\right)^2-2\theta\,\mathrm{M}\left(\widehat{\theta}_n\right)+\theta^2$ и, с учетом результатов решения задачи 1.5, вычислим $\mathrm{M}\left(\widehat{\theta}_n\right)^2$.

$$\mathrm{M}\left(\widehat{\theta}_n\right)^2=\int_{\theta}^{b}y^2\frac{n}{\left(b-\theta\right)^n}\left(b-y\right)^{n-1}dy=\left(-y^2\frac{n}{\left(b-\theta\right)^n}\frac{\left(b-y\right)^n}{n}\right)\Bigg|_{\theta}^{b}+$$

$$+\frac{n}{\left(b-\theta\right)^n}\int_{\theta}^{b}2y\frac{\left(b-y\right)^n}{n}dy=\theta^2+\frac{2}{\left(b-\theta\right)^n}\left(\left(-y\frac{\left(b-y\right)^{n+1}}{n+1}\right)\Bigg|_{\theta}^{b}+\int_{\theta}^{b}\frac{\left(b-y\right)^{n+1}}{n+1}dy\right)=$$

$$=\theta^2+\frac{2\theta(b-\theta)}{n+1}-\left(\frac{\left(b-y\right)^{n+2}}{\left(n+1\right)\left(n+2\right)}\frac{2}{\left(b-\theta\right)^n}\right)\Bigg|_{\theta}^{b}=\theta^2+\frac{2\theta(b-\theta)}{n+1}+\frac{2\left(b-\theta\right)^2}{(n+1)(n+2)}.$$

В задаче 1.5 было найдено $\mathrm{M}\left(\widehat{\theta}_n\right) = \frac{n}{n+1}\theta + \frac{b}{n+1}$. Следовательно, $\mathrm{M}\left(\widehat{\theta}_n - \theta\right)^2 = \mathrm{M}\left(\widehat{\theta}_n\right)^2 - 2\theta\mathrm{M}\left(\widehat{\theta}_n\right) + \theta^2 = \theta^2 + \frac{2\theta(b-\theta)}{n+1} + \frac{2(b-\theta)^2}{(n+1)(n+2)} -$

$- \frac{2n}{n+1}\theta^2 - \frac{2b\theta}{n+1} + \theta^2 = \frac{2(b-\theta)^2}{(n+1)(n+2)}$. Так как $\mathrm{M}\left(\widehat{\theta}_n - \theta\right)^2 = = \frac{2(b-\theta)^2}{(n+1)(n+2)} \xrightarrow[n\to\infty]{} 0$, то, в соответствии с неравенством (1.4), $\widehat{\theta}_n(\xi_1,\ldots,\xi_n) = \min\{\xi_1,\xi_2,\ldots,\xi_n\}$ - состоятельная оценка параметра θ.

Занятие 2. Задачи точечного оценивания. Метод моментов

Цель занятия: научиться решать задачи точечного оценивания неизвестных параметров распределений с использованием метода моментов.

Постановка задачи точечного оценивания неизвестных параметров распределения. Пусть $x_1, x_2, \ldots, x_n$ - результаты n независимых повторных наблюдений над случайной величиной ξ, функция распределения которой - $F(x;\vec{\theta})$ ($x \in R^1$) - известна с точностью до неизвестного параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, $\vec{\theta} \in \Theta \subset R^k$. Требуется по выборке найти точечную оценку параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, то есть указать вектор $\widehat{\theta}_n = (\widehat{\theta}_{n,1}, \widehat{\theta}_{n,2}, \ldots, \widehat{\theta}_{n,k})$ такой, что $\widehat{\theta}_{n,i} = \widehat{\theta}_{n,i}(\xi_1, \xi_2, \ldots, \xi_n)$ - измеримая функция от независимых одинаково распределенных случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ ($\xi_i \sim \xi$), и $\theta_i \approx \widehat{\theta}_{n,i}(x_1, \ldots, x_n)$ для всякого $i=1,2,\ldots,k$.

В математической статистике разработано большое число методов оценивания неизвестных параметров по повторной выборке, из которых на практике наиболее часто используются метод моментов и метод максимального правдоподобия.

Метод моментов. Метод моментов был предложен английским статистиком К. Пирсоном и является одним из первых регулярных методов оценивания. Он состоит в следующем.

Будем предполагать, что у случайной величины ξ существуют первые k моментов $m_i = M(\xi^i)$, $i = \overline{1,k}$. Так как закон распределения ξ известен с точностью до параметра $\vec{\theta}$, величины m_i будут функциями параметра $\vec{\theta}$, то есть

$$m_i = f_i(\vec{\theta}), \quad i = \overline{1,k}. \tag{2.1}$$

Предположим, что система уравнений (2.1) разрешима относительно $\theta_1, \theta_2, \ldots, \theta_k$, то есть существуют такие функции $\Psi_i(m_1, m_2, \ldots, m_k)$, $i = \overline{1,k}$, что

$$\theta_i = \Psi_i(m_1, \ldots, m_k), \quad i = \overline{1,k}. \tag{2.2}$$

Вычислим $m_i^* = \frac{1}{n}\sum_{j=1}^{n} x_j^i$, $i = \overline{1,k}$. Тогда вектор $\hat{\theta}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2}, \ldots, \hat{\theta}_{n,k})$, где $\hat{\theta}_{n,i} = \Psi_i(m_1^*, m_2^*, \ldots, m_k^*)$, $i = \overline{1,k}$, будет представлять собой значение оценки по методу моментов неизвестного параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ в распределении $F(x;\vec{\theta})$.

Рассмотрим примеры решения задач точечного оценивания с использованием метода моментов.

Задача 2.1

Случайная величина ξ - число семян сорняков в пробе зерна – распределена по закону Пуассона с неизвестным параметром λ $(\lambda > 0)$, то есть $\xi \in \{0,1,2,\ldots\}$, $P\{\xi = k\} = \frac{\lambda^k}{k!}e^{-\lambda}$, $k = 0,1,2,\ldots$. Было взято n=1000 независимых проб зерна. Распределение семян сорняков в этих пробах представлено в виде следующего статистического ряда:

x_i:	0	1	2	3	4	5	6
n_i:	405	366	175	40	8	4	2

Найти значение оценки неизвестного параметра распределения Пуассона с использованием метода моментов.

Решение. Вычислим первый момент случайной величины ξ: $m_1 = M\xi = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda$. Следовательно, $\lambda = m_1$. Вычислим первый выборочный момент: $m_1^* = \frac{1}{1000}(366 + 175\times 2 + 40\times 3 + 8\times 4 + 4\times 5 + 2\times 6)$=0,9. Тогда $\hat{\lambda}_n = 0{,}9$ - значение оценки по методу моментов неизвестного параметра λ в распределении Пуассона, то есть $\lambda \approx 0{,}9$.

Задача 2.2

Случайная величина ξ - число «успехов» в m=5 испытаниях по схеме Бернулли – имеет биномиальное распределение с неизвестным параметром p, где p – вероятность «успеха» в одном испытании. То есть, $\xi \in \{0,1,2,3,4,5\}$, $P\{\xi = k\} = C_5^k p^k (1-P)^{5-k}$, $k = \overline{0,5}$. Эмпирическое распределение числа «успехов» в n=10 независимых опытах представлено в виде следующего статистического ряда:

x_i:	0	1	2	3	4
n_i:	5	2	1	1	1

Найти значение оценки неизвестного параметра p методом моментов.

Решение. Найдем первый момент случайной величины ξ**:** $m_1 = M\xi = \sum_{k=0}^{5} k C_5^k p^k (1-p)^{5-k} = 5p$. Следовательно, $p = \frac{m_1}{5}$. Вычислим выборочный первый момент: $m_1^* = \frac{1}{10}(2+2+3+4) = 1{,}1$. Тогда $\widehat{p}_n = \frac{1{,}1}{5} = 0{,}22$ - значение оценки по методу моментов неизвестного параметра p, т.е. $p \approx 0{,}22$.

Задача 2.3

Случайная величина ξ - время работы элемента имеет показательное распределение с неизвестным параметром λ $(\lambda > 0)$, т.е. плотность распределения вероятностей данной случайной величины имеет вид

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0 \quad , & x < 0 \end{cases}.$$

Эмпирическое распределение времени работы n=200 независимых элементов представлено в виде следующего статистического ряда

x_i:	2,5	7,5	12,5	17,5	22,5	27,5
n_i:	133	45	15	4	2	1

Найти значение оценки неизвестного параметра λ методом моментов.

Решение. Найдем первый момент ξ: $m_1 = M\xi = \lambda \int_0^\infty x e^{-\lambda x} dx = \frac{1}{\lambda}$.

Следовательно, $\lambda = \frac{1}{m_1}$. Вычислим первый выборочный момент:

$m_1^* = \frac{1}{200}(5\times 45 + 10\times 15 + 15\times 4 + 20\times 2 + 25) + 2{,}5 = 5$. Тогда, $\widehat{\lambda}_n = \frac{1}{5} = 0{,}2$ - значение оценки по методу моментов неизвестного параметра λ, то есть $\lambda \approx 0{,}2$.

При вычислении первого выборочного момента (выборочного среднего) мы воспользовались свойством, состоящим в том, что если от всех элементов выборки отнять одну и ту же константу, то выборочное среднее уменьшится на эту же константу.

Задача 2.4

Случайная величина ξ - ошибка измерения дальности радиодальномером – имеет равномерное распределение на (a,b), где a,b - неизвестные параметры. Эмпирическое распределение ошибки n=200 независимых измерений дальности имеет вид:

x_i:	3	5	7	9	11	13	15	17	19	21
n_i:	21	16	15	26	22	14	21	22	18	25

Найти методом моментов значения точечных оценок неизвестных параметров a,b.

Решение. Плотность распределения вероятностей случайной величины, равномерно распределенной на интервале (a,b), имеет вид

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a,b) \\ 0, & x \notin (a,b) \end{cases}.$$

Найдем первый и второй моменты случайной величины ξ. $m_1 = M\xi = \frac{1}{b-a}\int_a^b x\,dx = \frac{a+b}{2}$; $m_2 = M\,\xi^2 = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{a^2+ab+b^2}{3}$. Решим систему уравнений

$$\begin{cases} m_1 = \frac{a+b}{2} \\ m_2 = \frac{a^2+ab+b^2}{3} \end{cases}$$

относительно a,b. В результате получим $a = m_1 - \sqrt{3}\times\sqrt{m_2 - m_1^2}$; $b = m_1 + \sqrt{3}\times\sqrt{m_2 - m_1^2}$. Вычислим первый и второй выборочные моменты: $m_1^* = 12{,}31$; $m_2^* = 185{,}32$. Тогда $\widehat{a}_n = m_1^* - \sqrt{3}\times\sqrt{m_2^* - (m_1^*)^2} = 2{,}24$ и $\widehat{b}_n = m_1^* + \sqrt{3}\times\sqrt{m_2^* - (m_1^*)^2} = 22{,}38$ являются значениями оценок по методу моментов неизвестных параметров a, b, т.е. $a \approx 2{,}24; \quad b \approx 22{,}38$.

Задача 2.5

Пусть $x_1, x_2, \ldots, x_n$- результаты n независимых повторных наблюдений над случайной величиной ξ, плотность распределения которой имеет вид

$$f(x;\theta) = p\, f_1(x;\theta) + (1-p)\, f_2(x;\theta), \text{ где } f_1(x;\theta) = \begin{cases} \dfrac{1}{\theta}, & 0 < x \le \theta \\ 0, & x - другое \end{cases},$$

$$f_2(x;\theta) = \begin{cases} \dfrac{1}{(1-\theta)}, & \theta < x < 1 \\ 0, & x - другое \end{cases},$$ p, θ - неизвестные параметры, $0 \le p \le 1$.

Найти оценки неизвестных параметров p, θ методом моментов.

Решение. Вычислим первый и второй моменты случайной величины ξ**:**

$$m_1 = M\xi = \int_0^1 x f(x;\theta)\, dx = \int_0^\theta x\, f_1(x;\theta)\, dx + \int_\vartheta^1 x\, f_2(x;\theta)\, dx = \frac{p}{\theta}\int_0^\theta x\, dx + \frac{1-p}{1-\theta}\int_\theta^1 x\, dx =$$
$$= \frac{1+\theta-p}{2}$$

$$m_2 = M\,\xi^2 = \frac{p}{\theta}\int_0^\theta x^2\, dx + \frac{1-p}{1-\theta}\int_\vartheta^1 x^2\, dx = p \times (\frac{\theta^2}{3}) + (1-p)(\frac{\theta^2+\theta+1}{3}). \qquad (2.3)$$

При вычислении первого и второго моментов ξ, можно воспользоваться тем обстоятельством, что эти моменты равны взвешенной сумме (с весами p и 1-p) соответственно первых и вторых моментов двух случайных величин, одна из которых равномерно распределена в интервале $(0,\theta)$, а другая в интервале $(\theta,1)$. Преобразовав правую часть выражения (2.3), получим следующую систему уравнений:

$$\begin{cases} m_1 = \dfrac{1+\theta-p}{2}, \\ m_2 = \dfrac{1+\theta-p+\theta^2-p\theta}{3}. \end{cases}$$

Решая ее, находим $\theta = \dfrac{3m_2 - 2m_1}{2m_1 - 1}; \quad p = \dfrac{3m_2 - 1}{2m_1 - 1} - 2m_1$. Следовательно, величины оценок по методу моментов будут иметь вид $\widehat{\theta}_n = \dfrac{3m_2^* - 2m_1^*}{2m_1^* - 1}, \quad \widehat{p}_n = \dfrac{3m_2^* - 1}{2m_1^* - 1} - 2m_1^*$

$\left(m_1^* = \dfrac{1}{n}\sum_{i=1}^n x_i, \quad m_2^* = \dfrac{1}{n}\sum_{i=1}^n x_i^2 \right)$ при условии, что $0 \le \widehat{p}_n \le 1$.

Задача 2.6

Пусть $x_1, x_2, \ldots, x_n$ - результаты n независимых повторных наблюдений над случайной величиной ξ, плотность распределения которой имеет вид $f(x;a,b) = b\, f_1(x) + (1-b)\, f_1(x-a)$, где $f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, a, b - неизвестные параметры, $0 \le b \le 1$.

Найти по методу моментов оценки параметров a, b.

Решение. Вычислим первый момент случайной величины ξ.

$$m_1 = M\xi = \int_{-\infty}^{+\infty} x\, f(x;a,b)dx = b\int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + (1-b)\int_{-\infty}^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}} dx.$$

Первый интеграл в правой части полученного выражения есть математическое ожидание стандартной нормальной случайной величины, а оно равно нулю. Второй интеграл в правой части представляет собой математическое ожидание нормальной случайной величины, дисперсия которой равна 1, а математическое ожидание равно a. Следовательно, $m_1 = (1-b)a$.

Вычислим второй момент случайной величины ξ.

$$m_2 = M\,\xi^2 = \int_{-\infty}^{+\infty} x^2\, f(x;a,b)dx = b\int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + (1-b)\int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}} dx$$

В правой части полученного выражения мы видим сумму вторых моментов (с весами *b* и *1-b*) двух случайных величин, одна из которых имеет стандартное нормальное распределение, а другая - нормальное распределение с математическим ожиданием, равным a и дисперсией, равной 1. С учетом того, что для произвольной случайной величины η $M\eta^2 = D\eta + (M\eta)^2$, получим $\int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1;$ $\int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2}} dx = 1 + a^2$. Следовательно, $m_2 = b + (1-b)(1+a)^2$.

В результате мы получили следующую систему уравнений

$$\begin{cases} m_1 = (1-b)a, \\ m_2 = 1 + a^2(1-b). \end{cases}$$

Решая ее, находим a, b: $a = \frac{m_2 - 1}{m_1}$, $b = 1 - \frac{m_1^2}{m_2 - 1}$. Следовательно, величины оценок по методу моментов будут иметь вид: $\hat{a}_n = \frac{m_2^* - 1}{m_1^*}$, $\hat{b}_n = 1 - \frac{(m_1^*)^2}{m_2^* - 1}$ $\left(m_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad m_2^* = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \right)$ при условии, что $0 \le \hat{b}_n \le 1$.

Задача 2.7

Пусть $x_1, x_2, \ldots, x_n$ - результаты n независимых повторных наблюдений над дискретной случайной величиной ξ, принимающей значения из множества $\{0,1,2,\ldots\}$ с вероятностями $\mathrm{P}\{\xi = x_i\} = \frac{1}{2}\frac{\lambda_1^{x_i} e^{-\lambda_1}}{x_i!} + \frac{1}{2}\frac{\lambda_2^{x_i} e^{-\lambda_2}}{x_i!}$, $i=1,2,\ldots$, где $\lambda_1 > 0,\ \lambda_2 > 0, \lambda_2 > \lambda_1$.

Найти методом моментов значения точечных оценок параметров λ_1, λ_2, если n=327, а статистический ряд выборки имеет вид

z_i:	0	1	2	3	4	5	6	7	8	9	10
n_i:	28	47	81	67	53	24	13	8	3	2	1

Решение. Вычислим первый и второй моменты случайной величины ξ.

$$m_1 = \mathrm{M}\xi = \frac{1}{2}\sum_{k=1}^{\infty} k\frac{\lambda_1^k e^{-\lambda_1}}{k!} + \frac{1}{2}\sum_{k=1}^{\infty} k\frac{\lambda_2^k e^{-\lambda_2}}{k!} = \frac{1}{2}\lambda_1 + \frac{1}{2}\lambda_2.$$

$$m_2 = \mathrm{M}\xi^2 = \frac{1}{2}\sum_{k=1}^{\infty} k^2\frac{\lambda_1^k e^{-\lambda_1}}{k!} + \frac{1}{2}\sum_{k=1}^{\infty} k^2\frac{\lambda_2^k e^{-\lambda_2}}{k!} = \frac{1}{2}(\lambda_1 + \lambda_1^2) + \frac{1}{2}(\lambda_2 + \lambda_2^2).$$

В результате получили систему уравнений

$$m_1 = \frac{\lambda_1 + \lambda_2}{2}, \quad m_2 = \frac{\lambda_1 + \lambda_2}{2} + \frac{\lambda_1^2 + \lambda_2^2}{2},$$

или $\lambda_1 + \lambda_2 = 2m_1$, $\lambda_1^2 + \lambda_2^2 = 2m_2 - 2m_1$. Решая данную систему, найдем выражения для λ_1 и λ_2. С учетом того, что $0 < \lambda_1 < \lambda_2$, $\lambda_1 = m_1 - \sqrt{m_2 - m_1 - m_1^2}$, $\lambda_2 = m_1 + \sqrt{m_2 - m_1 - m_1^2}$.

Вычислив первый и второй выборочные моменты $\left(m_k^* = \frac{1}{327}\sum_{i=1}^{11} z_i^k n_i;\quad m_1^* = 2{,}838;\quad m_2^* = 11{,}425\right)$ и подставив их вместо соответствующих моментов случайной величины ξ в выражения для λ_1 и λ_2, получим значения оценок: $\widehat{\lambda}_1 = m_1^* - \sqrt{m_2^* - m_1^* - \left(m_1^*\right)^2} = 2{,}11$; $\widehat{\lambda}_2 = m_1^* + \sqrt{m_2^* - m_1^* - \left(m_1^*\right)^2} = 3{,}57$.

Занятие 3. Метод максимального правдоподобия

Цель занятия: научиться решать задачи точечного оценивания неизвестных параметров распределений с использованием метода максимального правдоподобия.

Одним из наиболее универсальных методов оценивания является метод максимального правдоподобия, предложенный Р.Фишером. Рассмотрим формальное описание данного метода.

Случай 1. Пусть $x_1, x_2, \ldots, x_n$ - n повторных независимых наблюдений над непрерывной случайной величиной ξ, плотность распределения которой $f(x;\vec{\theta})$ известна с точностью до параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$, $\vec{\theta} \in \Theta \subset R^k$.

При независимых повторных наблюдениях элементы выборки $x_1, x_2, \ldots, x_n$ трактуют как значения независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, распределенных так же, как и ξ. Найдем совместную плотность распределения случайных величин $\xi_1, \xi_2, \ldots, \xi_n$. При каждом фиксированном $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ из области допустимых значений совместная плотность распределения будет иметь вид: $\breve{f}(y_1, y_2, \ldots, y_n; \vec{\theta}) = \prod_{i=1}^{n} f(y_i; \vec{\theta})$. Функцию $L(\vec{\theta}) = \breve{f}(x_1, \cdots, x_n; \vec{\theta})$, где $x_1, x_2, \ldots, x_n$ - элементы выборки, а $\vec{\theta} \in \Theta \subset R^k$, называют ***функцией правдоподобия.*** Таким образом, функция $\prod_{i=1}^{n} f(y_i; \vec{\theta})$ при фиксированном значении $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ является совместной плотностью распределения случайных величин $\xi_1, \xi_2, \ldots, \xi_n$, а при фиксированных значениях $y_i = x_i$, $i = \overline{1, n}$ представляет собой функцию правдоподобия. Функцию $\ln L(\vec{\theta}) = \sum_{i=1}^{n} \ln f(x_i; \vec{\theta})$ называют ***логарифмической функцией правдоподобия.***

По определению ***оценкой максимального правдоподобия*** неизвестного параметра $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ в распределении $f(x;\vec{\theta})$ называют такой вектор $\widehat{\theta}_n = (\widehat{\theta}_{n,1}, \widehat{\theta}_{n,2}, \ldots, \widehat{\theta}_{n,k})$, который для любой повторной выборки $x_1, x_2, \ldots, x_n$ (из распределения $f(x;\vec{\theta})$) удовлетворяет условию $L(\widehat{\theta}_n) = \max_{\vec{\theta} \in \Theta} L(\vec{\theta})$.

Поиск оценки максимального правдоподобия упрощается, если искать точку максимума не функции правдоподобия, а логарифмической функции правдоподобия, так как максимумы этих двух функций достигаются при одном и том же $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$.

Если функция $\ln L(\vec{\theta}) = \sum_{i=1}^{n} \ln f(x_i; \vec{\theta})$ дифференцируема по θ_i, $i = \overline{1,k}$, то необходимым условием существования экстремума является равенство нулю ее частных производных по θ_i, $i = \overline{1,k}$. В этом случае, для нахождения оценки максимального правдоподобия, необходимо решить систему уравнений

$$\frac{\partial \ln L(\vec{\theta})}{\partial \theta_i} = 0, \quad i = \overline{1,k}, \tag{3.1}$$

а затем отобрать то ее решение, на котором достигается максимум функции $\ln L(\vec{\theta})$.

Уравнения (3.1) называют ***уравнениями правдоподобия.***

Рассмотрим примеры нахождения оценок неизвестных параметров непрерывных распределений с помощью метода максимального правдоподобия.

Задача 3.1

Пусть x_1, x_2, x_3, x_4 - результаты независимых повторных наблюдений над случайной величиной ξ, плотность распределения которой

$$f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{(x-4)}{\beta}}, & x > 4 \\ 0, & x \le 4 \end{cases}$$

известна с точностью до параметра β $(\beta > 0)$.

Найти значение оценки максимального правдоподобия параметра β, если $x_1 = 8{,}2;$ $x_2 = 9{,}1;$ $x_3 = 10{,}6;$ $x_4 = 4{,}9$.

Решение. Построим функцию правдоподобия

$L(\beta) = \breve{f}(x_1, x_2, x_3, x_4; \beta) = \prod_{i=1}^{4} \frac{1}{\beta} e^{-\frac{(x_i-4)}{\beta}} = \frac{1}{\beta^4} \prod_{i=1}^{4} e^{-\frac{(x_i-4)}{\beta}}$, $x_i > 4,$ $i = \overline{1,4}$. Тогда $\ln L(\beta) = -4\ln\beta - \sum_{i=1}^{4} \frac{(x_i - 4)}{\beta}$. Построим уравнение правдоподобия: $\frac{d \ln L(\beta)}{d\beta} = -\frac{4}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{4} (x_i - 4) = 0$. Решая его, получим: $\widehat{\beta}_n = \frac{1}{4} \sum_{i=1}^{4} (x_i - 4)$.

Покажем, что $\beta = \widehat{\beta}_n$ является точкой максимума функции $\ln L(\beta) = -4\ln\beta - \sum_{i=1}^{4} \frac{(x_i - 4)}{\beta}$. Для этого найдем вторую производную от этой функции по β и вычислим ее в точке $\beta = \widehat{\beta}_n$.

$$\frac{d^2\ln L(\beta)}{d\beta^2}=\left(\frac{4}{\beta^2}-\frac{2}{\beta^3}\sum_{i=1}^{4}(x_i-4)\right)\Bigg|_{\beta=\widehat{\beta}_n}=-\frac{4}{\widehat{\beta}_n^{\,2}}<0.$$ Из того, что значение второй производной в точке $\beta=\widehat{\beta}_n$ меньше нуля, следует, что данная точка является точкой максимума функции $\ln L(\beta)$. Таким образом, $\widehat{\beta}_n=\frac{1}{4}\sum_{i=1}^{4}(x_i-4)=4{,}2$ - значение оценки максимального правдоподобия неизвестного параметра β, т.е. $\beta\approx 4{,}2$.

Задача 3.2

Пусть $x_1, x_2, \ldots, x_n$ - результаты n независимых повторных наблюдений над случайной величиной ξ, плотность распределения которой имеет вид

$$f(x;\theta)=\begin{cases}\sqrt{\theta}\,x^{\sqrt{\theta}-1}, & 0\le x\le 1\\ 0, & x-\textit{другое}\end{cases},$$

где θ $(\theta>0)$ - неизвестный параметр. Найти оценку максимального правдоподобия параметра θ.

Решение. Построим функцию правдоподобия

$L(\theta)=\breve{f}(x_1,x_2,\ldots,x_n;\theta)=\prod_{i=1}^{n}(\sqrt{\theta}\,x_i^{\sqrt{\theta}-1})$, $0\le x_i\le 1$, $i=\overline{1,n}$. Тогда

$\ln L(\theta)=\frac{n}{2}\ln\theta+\sqrt{\theta}\sum_{i=1}^{n}\ln x_i-\sum_{i=1}^{n}\ln x_i$. Уравнение правдоподобия будет иметь вид

$\frac{d\ln L(\theta)}{d\theta}=\frac{n}{2\theta}+\frac{1}{2\sqrt{\theta}}\sum_{i=1}^{n}\ln x_i=0\Rightarrow\quad n+\sqrt{\theta}\sum_{i=1}^{n}\ln x_i=0$. Решая его, получим

$\widehat{\theta}_n=\frac{n^2}{\left(\sum_{i=1}^{n}\ln x_i\right)^2}$. Покажем, что решение уравнения действительно является точкой максимума функции правдоподобия. Для этого вычислим значение второй производной логарифмической функции правдоподобия в точке $\theta=\widehat{\theta}_n$.

$$\frac{d^2L(\theta)}{d\theta^2}=\left(-\frac{n}{2\theta^2}-\frac{1}{4\theta\sqrt{\theta}}\sum_{i=1}^{n}\ln x_i\right)\Bigg|_{\theta=\widehat{\theta}_n}=-\frac{n}{2\widehat{\theta}_n^2}+\frac{n}{4\widehat{\theta}_n^2}<0.$$

Таким образом, $\widehat{\theta}_n = \dfrac{n^2}{\left(\sum\limits_{i=1}^{n} \ln \xi_i\right)^2}$ - оценка максимального правдоподобия неизвестного параметра θ.

Задача 3.3

Пусть $x_1, x_2, \ldots, x_n$ - результаты независимых повторных наблюдений над случайной величиной $\xi \in N(a,1)$, a - неизвестный параметр. Требуется найти оценку максимального правдоподобия параметра a.

Решение. Построим функцию правдоподобия $L(a) = \breve{f}(x_1, x_2, \ldots, x_n; a) = \prod\limits_{i=1}^{n} \dfrac{1}{\sqrt{2\pi}} e^{-\frac{(x_i - a)^2}{2}} = (2\pi)^{-\frac{n}{2}} \prod\limits_{i=1}^{n} e^{-\frac{(x_i - a)^2}{2}}$. Тогда $\ln L(a) = -\dfrac{n}{2} \ln 2\pi - \sum\limits_{i=1}^{n} \dfrac{(x_i - a)^2}{2}$. Построим уравнение правдоподобия $\dfrac{d \ln L(a)}{da} = \sum\limits_{i=1}^{n} (x_i - a) = \sum\limits_{i=1}^{n} x_i - na = 0$. Решая его, получим $\widehat{a}_n = \dfrac{1}{n} \sum\limits_{i=1}^{n} \xi_i$ - оценка максимального правдоподобия ($\dfrac{d^2 \ln L(a)}{da^2} = -n < 0$).

Задача 3.4

Пусть $x_1, x_2, \ldots, x_n$ - результаты независимых повторных наблюдений над случайной величиной $\xi \in N(a, \theta)$, $\theta = D\xi$, a, θ - неизвестные параметры. Требуется найти оценки максимального правдоподобия параметров a, θ.

Решение. Построим функцию правдоподобия $L(a, \theta) = \breve{f}(x_1, x_2, \ldots, x_n; a, \theta) = \prod\limits_{i=1}^{n} \dfrac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x_i - a)^2}{2\theta}} = (2\pi)^{-\frac{n}{2}} \theta^{-\frac{n}{2}} \prod\limits_{i=1}^{n} e^{-\frac{(x_i - a)^2}{2\theta}}$. Тогда $\ln L(a, \theta) = -\dfrac{n}{2} \ln 2\pi - \dfrac{n}{2} \ln \theta - \sum\limits_{i=1}^{n} \dfrac{(x_i - a)^2}{2\theta}$. Составим уравнения правдоподобия:

$$\frac{\partial \ln L(a, \theta)}{\partial a} = \frac{1}{\theta} \sum_{i=1}^{n} (x_i - a) = 0 \quad \Rightarrow \sum_{i=1}^{n} x_i - na = 0;$$

$$\frac{\partial \ln L(a, \theta)}{\partial \theta} = -\frac{n}{2\theta} + \frac{1}{2\theta^2} \sum_{i=1}^{n} (x_i - a)^2 = 0 \Rightarrow \quad -n\theta + \sum_{i=1}^{n} (x_i - a)^2 = 0.$$

В результате получили следующую систему уравнений правдоподобия

$$\begin{cases} \sum\limits_{i=1}^{n} x_i - na = 0, \\ n\theta - \sum\limits_{i=1}^{n} (x_i - a)^2 = 0. \end{cases}$$

Ее решение имеет вид $\hat{a}_n = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$; $\hat{\theta}_n = \frac{1}{n}\sum\limits_{i=1}^{n}(x_i - \hat{a}_n)^2$. Покажем, что $(\hat{a}_n, \hat{\theta}_n)$ действительно является точкой максимума логарифмической функции правдоподобия. Для этого построим матрицу вторых производных данной функции и покажем, что при $a = \hat{a}_n$, $\theta = \hat{\theta}_n$ она является отрицательно определенной.

$$\begin{pmatrix} \frac{\partial^2 \ln L(a,\theta)}{\partial a^2} & \frac{\partial^2 \ln L(a,\theta)}{\partial a\,\partial\theta} \\ \frac{\partial^2 \ln L(a,\theta)}{\partial a\,\partial\theta} & \frac{\partial^2 \ln L(a,\theta)}{\partial\theta^2} \end{pmatrix} = \left.\begin{pmatrix} -\frac{n}{\theta} & -\frac{\sum\limits_{i=1}^{n}(x_i - a)}{\theta^2} \\ -\frac{\sum\limits_{i=1}^{n}(x_i - a)}{\theta^2} & \frac{n}{2\theta^2} - \frac{1}{\theta^3}\sum\limits_{i=1}^{n}(x_i - a)^2 \end{pmatrix}\right|_{\substack{a=\hat{a}_n \\ \theta=\hat{\theta}_n}} =$$

$$= \begin{pmatrix} -\frac{n}{\hat{\theta}_n} & 0 \\ 0 & -\frac{n}{2\hat{\theta}_n^2} \end{pmatrix} \Rightarrow$$ матрица отрицательно определенная, т.к. ее собственные значения являются отрицательными величинами. Следовательно, $\hat{a}_n = \frac{1}{n}\sum\limits_{i=1}^{n}\xi_i$ и $\hat{\theta}_n = \frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i - \hat{a}_n)^2$ - оценки максимального правдоподобия неизвестных параметров a, θ.

Случай 2. Пусть $x_1, x_2, \ldots, x_n$- результаты n независимых повторных наблюдений над дискретной случайной величиной ξ, принимающей значения из множества $Y = \{y_1, y_2, \ldots\}$ с вероятностями $P\{\xi = y_i\} = P(y_i; \vec{\theta})$ $(i = 1,2,\ldots)$, где $\vec{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ - неизвестный параметр, $\vec{\theta} \in \Theta \subset R^k$. Формальное описание метода максимального правдоподобия применительно к данному случаю отличается от приведенного ранее (см. *Случай 1*) только тем, что функция правдоподобия $L(\vec{\theta})$ и логарифмическая функция правдоподобия $\ln L(\vec{\theta})$ в данном случае будут иметь вид:

$$L(\vec{\theta}) = P\{\xi_1 = x_1, \xi_2 = x_2, \dots, \xi_n = x_n\} = \prod_{i=1}^{n} P\{\xi = x_i\} = \prod_{i=1}^{n} P(x_i, \vec{\theta});$$

$$\ln L(\vec{\theta}) = \sum_{i=1}^{n} \ln P(x_i; \vec{\theta}).$$

Рассмотрим примеры нахождения оценок неизвестных параметров дискретных распределений с помощью метода максимального правдоподобия.

Задача 3.5

Пусть $x_1, x_2, \dots, x_n$ результаты n независимых повторных наблюдений над дискретной случайной величиной ξ, принимающей значения из множества $Y=\{0,1\}$ с вероятностями: $P\{\xi = 1\} = p, \quad P\{\xi = 0\} = 1 - p, \quad 0 < p < 1$, p – неизвестный параметр. Найти величину оценки максимального правдоподобия неизвестного параметра p, если n=100, а $\sum_{i=1}^{100} x_i = 65$.

Решение. Так как элементы полученной выборки равны либо нулю, либо единице, то $P\{\xi = x_i\} = p^{x_i}(1-p)^{(1-x_i)}, \quad i = \overline{1,n}$. Поэтому функция правдоподобия будет иметь вид $L(p) = \prod_{i=1}^{n} (p^{x_i}(1-p)^{(1-x_i)})$. Следовательно, $\ln L(p) = (\sum_{i=1}^{n} x_i) \ln p + (\sum_{i=1}^{n} (1 - x_i)) \ln(1-p)$. Обозначим $m = \sum_{i=1}^{n} x_i$, тогда $\ln L(p) = m \ln p - (n-m)\ln(1-p)$. Построим уравнение правдоподобия $\frac{dL(p)}{dp} = \frac{m}{p} - \frac{n-m}{(1-p)} = 0.$ Решая его, получим $\widehat{p}_n = \frac{m}{n} = 0{,}65$ - значение оценки максимального правдоподобия $\left(\frac{d^2 L(p)}{dp^2} = -\frac{m}{p^2} - \frac{n-m}{(1-p)^2} < 0 \right)$, т.е. $p \approx 0{,}65$.

Задача 3.6

Пусть $x_1, x_2, \dots, x_n$ результаты n независимых повторных наблюдений над дискретной случайной величиной ξ, принимающей значения из множества $Y=\{0,1\}$ с вероятностями: $P\{\xi = 1\} = \frac{1+\theta}{2}, \quad P\{\xi = 0\} = \frac{1-\theta}{2}, \quad -1 < \theta < 1, \quad \theta$ –

неизвестный параметр. Найти оценку максимального правдоподобия неизвестного параметра θ.

Решение. По аналогии с предыдущей задачей $P\{\xi = x_i\} = \left(\frac{1+\theta}{2}\right)^{x_i}\left(\frac{1-\theta}{2}\right)^{(1-x_i)}$, $i = \overline{1,n}$. Функция правдоподобия будет иметь вид $L(\theta) = \prod_{i=1}^{n}\left(\frac{1+\theta}{2}\right)^{x_i}\left(\frac{1-\theta}{2}\right)^{(1-x_i)}$. Следовательно,

$$\ln L(\theta) = (\sum_{i=1}^{n} x_i)\ln(1+\theta) + (\sum_{i=1}^{n}(1-x_i))\ln(1-\theta) + c, \quad c = const.$$

Построим уравнение правдоподобия: $\frac{dL(\theta)}{d\theta} = \frac{\sum_{i=1}^{n} x_i}{1+\theta} - \frac{\sum_{i=1}^{n}(1-x_i)}{(1-\theta)} = 0.$

Решая его, получим $\hat{\theta}_n = -1 + \frac{2}{n}\sum_{i=1}^{n}\xi_i$ - оценка максимального правдоподобия параметра θ $\left(\frac{d^2L(\theta)}{d\theta^2} = -\frac{\sum_{i=1}^{n} x_i}{(1+\theta)^2} - \frac{\sum_{i=1}^{n}(1-x_i)}{(1-\theta)^2} < 0\right)$.

Задача 3.7

Пусть $x_1, x_2, \ldots, x_n$ результаты n независимых повторных наблюдений над дискретной случайной величиной ξ, ряд распределения которой имеет вид

x_i:	1	2	3
$P\{\xi = x_i\}$:	$1-\theta$	$\theta(1-\theta)$	θ^2

и θ - неизвестный параметр.

Пусть m_i - число элементов выборки, равных *i*, *i=1,2,3*. Найти оценку максимального правдоподобия параметра θ.

Решение. Найдем область допустимых значений параметра θ. θ должно удовлетворять следующим неравенствам: $0 < (1-\theta) < 1;\quad 0 < \theta(1-\theta) < 1;\quad \theta^2 < 1 \quad\Rightarrow$ $0 < \theta < 1$.

Построим функцию правдоподобия $L(\theta)=\prod_{i=1}^{m_1}(1-\theta)\prod_{i=1}^{m_2}\theta(1-\theta)\prod_{i=1}^{m_3}\theta^2=(1-\theta)^{m_1+m_2}\theta^{m_2+2m_3}$. Тогда $\ln L(\theta)=(m_1+m_2)\ln(1-\theta)+(m_2+2m_3)\ln\theta$. Построим уравнение правдоподобия: $\frac{dL(\theta)}{d\theta}=-\frac{m_1+m_2}{1-\theta}+\frac{m_2+2m_3}{\theta}=0$. Решая его, получим $\widehat{\theta}_n=\frac{m_2+2m_3}{m_1+2m_2+2m_3}$ - величина оценки максимального правдоподобия параметра θ $\left(\frac{d^2L(\theta)}{d\theta^2}=-\frac{m_1+m_2}{(1-\theta)^2}-\frac{m_2+2m_3}{\theta^2}<0\right)$.

Задача 3.8

Пусть ξ - случайная величина с распределением Паскаля. То есть, $\xi\in\{a,a+1,a+2,\ldots\}$, $a\geq 2$, $P\{\xi=x\}=C_{x-1}^{a-1}\theta^a(1-\theta)^{x-a}$, $0<\theta<1$, θ - неизвестный параметр. По одному наблюдению над случайной величиной ξ найти оценку максимального правдоподобия параметра θ.

Решение. Пусть y – результат наблюдения над случайной величиной ξ, тогда функция правдоподобия $L(\theta)$ и логарифмическая функция правдоподобия $\ln L(\theta)$ будут иметь вид $L(\theta)=C_{y-1}^{a-1}\theta^a(1-\theta)^{y-a}$, $\ln L(\theta)=\ln C_{y-1}^{a-1}+a\ln\theta+(y-a)\ln(1-\theta)$.

Составим уравнение правдоподобия: $\frac{dL(\theta)}{d\theta}=-\frac{y-a}{1-\theta}+\frac{a}{\theta}=0$. Решая его, получим $\widehat{\theta}_n=\frac{a}{\xi}$ - оценка максимального правдоподобия параметра θ $\left(\frac{d^2L(\theta)}{d\theta^2}=-\frac{(y-a)}{(1-\theta)^2}-\frac{a}{\theta^2}<0\right)$.

Задача 3.9

Пусть $x_1,x_2,\ldots,x_n$ результаты n независимых повторных наблюдений над дискретной случайной величиной ξ, ряд распределения которой имеет вид

z_i:	z_1	z_2	z_3	z_4	z_5
$P\{\xi=z_i\}$:	$\alpha\frac{1-\theta_1}{1-\alpha}$	$\left(\frac{1}{2}-\alpha\right)\frac{1-\theta_1}{1-\alpha}$	$\left(\frac{1}{2}-\alpha\right)\frac{1-\theta_1}{1-\alpha}$	$\theta_1\theta_2$	$\theta_1(1-\theta_2)$

α $\left(0<\alpha<\frac{1}{2}\right)$ - известно, а θ_1, θ_2 - неизвестные параметры, $\theta_1, \theta_2 \in (0,1)$. Пусть n_i - число элементов выборки, равных z_i, $i=\overline{1,5}$. Найти оценки максимального правдоподобия параметров θ_1, θ_2.

Решение. Построим функцию правдоподобия

$$L(\theta_1,\theta_2)=\prod_{i=1}^{n_1}\left(\alpha\frac{1-\theta_1}{1-\alpha}\right)\prod_{i=1}^{n_2+n_3}\left(\frac{1}{2}-\alpha\right)\left(\frac{1-\theta_1}{1-\alpha}\right)\prod_{i=1}^{n_4}\theta_1\theta_2\prod_{i=1}^{n_5}\theta_1(1-\theta_2)=$$

$$=\frac{\alpha^{n_1}}{(1-\alpha)^{n_1+n_2+n_3}}(1-\theta_1)^{n_1+n_2+n_3}\left(\frac{1}{2}-\alpha\right)^{n_2+n_3}\theta_1^{(n_4+n_5)}(\theta_2)^{n_4}(1-\theta_1)^{n_5}.$$

$$\ln L(\theta_1,\theta_2)=(n_1+n_2+n_3)\ln(1-\theta_1)+(n_4+n_5)\ln\theta_1+n_4\ln\theta_2+n_5\ln(1-\theta_2)+c,$$

где $c=const$. Составим систему уравнений правдоподобия $\frac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_1}=-\frac{n_1+n_2+n_3}{1-\theta_1}+\frac{n_4+n_5}{\theta_1}=0$; $\frac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_2}=-\frac{n_5}{1-\theta_2}+\frac{n_4}{\theta_2}=0$. Решая ее, получим $\hat{\theta}_{n,1}=\frac{n_4+n_5}{n}$; $\hat{\theta}_{n,2}=\frac{n_4}{n_4+n_5}$.

Покажем, что найденное решение действительно является точкой максимума функции правдоподобия. Для этого построим матрицу вторых производных логарифмической функции правдоподобия и покажем, что при значениях $\theta_1=\hat{\theta}_{n,1}$, $\theta_2=\hat{\theta}_{n,2}$ она является отрицательно определенной.

$$\begin{pmatrix} -\frac{n_1+n_2+n_3}{(1-\theta_1)^2}-\frac{n_4+n_5}{\theta_1^2} & 0 \\ 0 & -\frac{n_4}{\theta_2^2}-\frac{n_5}{(1-\theta_2)^2}\end{pmatrix}$$

- отрицательно определенная матрица при любых значениях θ_1, θ_2. Следовательно, $\hat{\theta}_{n,1}=\frac{n_4+n_5}{n}$, $\hat{\theta}_{n,2}=\frac{n_4}{n_4+n_5}$ - величины оценок максимального правдоподобия параметров θ_1, θ_2.

Занятие 4. Задачи интервального оценивания

Цель занятия: освоить методы интервального оценивания неизвестных параметров распределений.

Постановка задачи интервального оценивания неизвестного параметра распределения. Пусть $x_1, x_2, \ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, функция распределения которой - $F(x;\theta)$ $(x \in R^1)$ - известна с точностью до параметра θ, $\theta \in \Theta \subset R^1$. Требуется по выборке указать такой интервал $(\underline{\theta}(\xi_1,\xi_2,\ldots,\xi_n)$, $\overline{\theta}(\xi_1,\xi_2,\ldots,\xi_n))$ ($\xi_1,\ldots,\xi_n$ - независимые случайные величины, распределенные так же, как и ξ), который с вероятностью, близкой к 1, накрывает истинное значение неизвестного параметра.

Один из способов построения такого интервала, основанный на использовании обобщенного неравенства Чебышева, заключается в следующем.

Пусть $\widehat{\theta}_n(\xi_1,\ldots,\xi_n)$ - несмещенная точечная оценка неизвестного параметра θ, а $D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)$ - дисперсия этой оценки. По обобщенному неравенству Чебышева, для $\forall\,\varepsilon > 0$ $0 \le \mathrm{P}\left(\left|\widehat{\theta}_n(\xi_1,\ldots,\xi_n) - \theta\right| \ge \varepsilon\right) \le \dfrac{D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)}{\varepsilon^2}$,

или $\mathrm{P}\left(\left|\widehat{\theta}_n(\xi_1,\ldots,\xi_n) - \theta\right| < \varepsilon\right) \ge 1 - \dfrac{D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)}{\varepsilon^2}$.

Зададим $\varepsilon = 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))}$, тогда $\mathrm{P}\left(\left|\widehat{\theta}_n(\xi_1,\ldots,\xi_n) - \theta\right| < 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))}\right) \ge 1 - \dfrac{D\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n)\right)}{\varepsilon^2} = \dfrac{8}{9}$. Таким образом, получили $\mathrm{P}\left(\left|\widehat{\theta}_n(\xi_1,\ldots,\xi_n) - \theta\right| < 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))}\right) \ge \dfrac{8}{9} \approx 0{,}89$.

Раскроем неравенство в круглых скобках: $-3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))} < \widehat{\theta}_n(\xi_1,\ldots,\xi_n) - \theta < 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))} \Rightarrow$

$\widehat{\theta}_n(\xi_1,\ldots,\xi_n) - 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))} < \theta < \widehat{\theta}_n(\xi_1,\ldots,\xi_n) + 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))}$.

Следовательно, интервал вида

$$\left(\widehat{\theta}_n(\xi_1,\ldots,\xi_n) - 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))};\ \widehat{\theta}_n(\xi_1,\ldots,\xi_n) + 3\sqrt{D(\widehat{\theta}_n(\xi_1,\ldots,\xi_n))}\right) \qquad (4.1)$$

с вероятностью, большей $\frac{8}{9} \approx 0{,}89$, накрывает истинное значение неизвестного параметра θ.

Основной недостаток данного метода состоит в том, что нельзя точно указать вероятность, с которой интервал (4.1) накрывает значение θ.

Воспользуемся рассмотренным методом для решения следующей задачи.

Задача 4.1

Пусть $x_1, x_2, \ldots, x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, имеющей нормальное распределение с параметрами a, σ^2, где σ^2 - известно, а a - неизвестный параметр. Указать интервал, который с вероятностью, близкой к 1 накрывает истинное значение неизвестного параметра.

Решение. В качестве несмещенной точечной оценки неизвестного параметра $a = \mathrm{M}\xi$ выберем статистику $\hat{a}_n(\xi_1, \ldots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n}\xi_i$, где $\xi_1, \ldots, \xi_n$ - независимые случайные величины $\xi_i \in \mathrm{N}(a, \sigma^2)$, $i = \overline{1,n}$. По свойствам нормального распределения $\hat{a}_n(\xi_1, \ldots, \xi_n) \in \mathrm{N}\left(a, \frac{\sigma^2}{n}\right)$. Тогда, в соответствии с выражением (4.1), $\mathrm{P}\left(\hat{a}_n(\xi_1, \ldots, \xi_n) - 3\frac{\sigma}{\sqrt{n}} < a < \hat{a}_n(\xi_1, \ldots, \xi_n) + 3\frac{\sigma}{\sqrt{n}}\right) \geq \frac{8}{9}$.

В данном случае можно найти точное значение вероятности, с которой интервал $\left(\hat{a}_n(\xi_1, ..., \xi_n) - 3\frac{\sigma}{\sqrt{n}}; \hat{a}_n(\xi_1, ..., \xi_n) + 3\frac{\sigma}{\sqrt{n}}\right)$ накрывает истинное значение параметра a.

Действительно, так как $\hat{a}_n(\xi_1, \ldots, \xi_n) \in \mathrm{N}\left(a, \frac{\sigma^2}{n}\right)$, то $\left(\hat{a}_n(\xi_1, \ldots, \xi_n) - a\right) \in \mathrm{N}\left(0, \frac{\sigma^2}{n}\right)$ и $\mathrm{P}\left(-\varepsilon < \hat{a}_n(\xi_1, \ldots, \xi_n) - a < \varepsilon\right) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\int_{-\varepsilon}^{\varepsilon} e^{-\frac{t^2}{2\sigma^2}n}\,dt =$
$= \Phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\sqrt{n}\right)$. Следовательно,

$$\mathrm{P}\left(-\varepsilon < \hat{a}_n(\xi_1, \ldots, \xi_n) - a < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right) - 1, \qquad (4.2)$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$. Для $\varepsilon = 3\frac{\sigma}{\sqrt{n}}$,

$$\mathrm{P}\left(-3\frac{\sigma}{\sqrt{n}} < \hat{a}_n(\xi_1, \ldots, \xi_n) - a < 3\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(3) - 1 \approx 0{,}9974.$$

Из рассмотренного примера видно, что интервал, построенный на основе неравенства Чебышева очень широк, так как вероятность того, что он накрывает истинное значение неизвестного параметра, практически равна 1.

При решении задачи 4.1 можно воспользоваться и другим методом. Зададим вероятность α $(0 < \alpha < 1)$ и будем искать такое ε, что

$\mathrm{P}\left(-\varepsilon<\widehat{a}_n(\xi_1,\ldots,\xi_n)-a<\varepsilon\right)=2\Phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right)-1=1-\alpha$. Тогда $\Phi\left(\frac{\varepsilon}{\sigma}\sqrt{n}\right)=1-\frac{\alpha}{2}$. Решением этого уравнения является $u_{1-\frac{\alpha}{2}}$ - *квантиль стандартного нормального распределения уровня* $1-\frac{\alpha}{2}$. Следовательно, $\frac{\varepsilon}{\sigma}\sqrt{n}=u_{1-\frac{\alpha}{2}}$ и $\varepsilon=u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. В результате получим интервал $\left(\widehat{a}_n(\xi_1,\ldots,\xi_n)-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}};\ \widehat{a}_n(\xi_1,\ldots,\xi_n)+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$, для которого $\mathrm{P}\left(\widehat{a}_n(\xi_1,\ldots,\xi_n)-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<a<\widehat{a}_n(\xi_1,\ldots,\xi_n)+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$. Таким образом, задав α $(0<\alpha<1)$, мы построили интервал, который с заданной вероятностью $1-\alpha$ накрывает истинное значение неизвестного параметра a.

Определение 4.1. *Интервал* $(\underline{\theta}(x_1,x_2,\ldots,x_n),\overline{\theta}(x_1,x_2,\ldots,x_n))$, *который с заданной вероятностью* $1-\alpha$ *накрывает истинное значение неизвестного параметра* θ $(\theta\in\Theta\subseteq R^1)$ *в распределении* $F(x;\theta)$, *называется доверительным интервалом, а вероятность* $1-\alpha$ - *доверительной вероятностью или надежностью доверительного интервала.*

Центральный метод построения доверительных интервалов. Пусть $\varphi(x_1,\ldots,x_n;\theta)$ - измеримая функция.

Определение 4.2. $\varphi(\xi_1,\ldots,\xi_n;\theta)$ *называется центральной статистикой, если: а) закон распределения* $\varphi(\xi_1,\ldots,\xi_n;\theta)$ *известен и не зависит ни от* θ, *ни от других неизвестных параметров; б)* $\varphi(\xi_1,\ldots,\xi_n;\theta)$ *непрерывна и строго монотонна по* θ.

Обозначим функцию распределения центральной статистики $\varphi(\xi_1,\ldots,\xi_n;\theta)$ через $F_\varphi(x)$ и рассмотрим случай, когда $\varphi(\xi_1,\ldots,\xi_n;\theta)$ - непрерывная случайная величина (при всяком фиксированном θ из области допустимых значений), тогда $\mathrm{P}\left\{u_{p_1}\le\varphi(\xi_1,\ldots,\xi_n;\theta)<u_{p_2}\right\}=F_\varphi(u_{p_2})-F_\varphi(u_{p_1})=p_2-p_1$, где u_{p_1},u_{p_2} - квантили распределения $F_\varphi(x)$, уровней p_1, p_2 $(0<p_1<p_2<1)$ соответственно. Так как статистика $\varphi(\xi_1,\ldots,\xi_n;\theta)$ непрерывна и строго монотонна по θ, то неравенство в фигурных скобках может быть разрешено относительно θ. То есть $\mathrm{P}\left\{u_{p_1}\le\varphi(\xi_1,\ldots,\xi_n;\theta)<u_{p_2}\right\}=\mathrm{P}\left\{\underline{\theta}(\xi_1,\ldots,\xi_n)\le\theta<\overline{\theta}(\xi_1,\ldots,\xi_n)\right\}=p_2-p_1$. Зададим

p_1, p_2 такими, чтобы $p_2 - p_1 = 1-\alpha$, где $1-\alpha$ - заданная вероятность. Тогда интервал $(\underline{\theta}(x_1, x_2, \ldots, x_n), \overline{\theta}(x_1, x_2, \ldots, x_n))$ будет представлять собой доверительный интервал надежности $1-\alpha$ для неизвестного параметра θ. Так как p_1, p_2, удовлетворяющие равенству $p_2 - p_1 = 1-\alpha$, можно выбрать неоднозначно, то и доверительный интервал для заданной доверительной вероятности $1-\alpha$ можно построить не единственным образом. Поскольку длина интервала характеризует точность, с которой локализовано значение параметра θ, то хотелось бы из всех возможных интервалов надежности $1-\alpha$ выбрать тот, длина которого минимальна. В математической статистике доказывается, что, если $\varphi(\xi_1, \ldots, \xi_n; \theta)$ (при фиксированном θ) является непрерывной случайной величиной с плотностью распределения симметричной относительно оси ординат, то при $p_1 = \frac{\alpha}{2}$, $p_2 = 1 - \frac{\alpha}{2}$ длина доверительного интервала будет минимальна.

Решим несколько задач, в которых требуется построить доверительный интервал надежности $1-\alpha$ для неизвестного параметра распределения.

Задача 4.2

Пусть $x_1, x_2, \ldots, x_n$ - результаты $n=25$ повторных независимых наблюдений над случайной величиной ξ, имеющей нормальное распределение с параметрами a, σ^2, где $\sigma^2=25$, а a - неизвестный параметр. Указать интервал, который с вероятностью 0,95 накрывает истинное значение неизвестного параметра, если $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 14$.

Решение. Пусть $\hat{a}_n(\xi_1, \ldots, \xi_n) = \frac{1}{n}\sum_{i=1}^{n} \xi_i$, где $\xi_1, \ldots, \xi_n$ - независимые случайные величины $\xi_i \in \mathrm{N}(a, \sigma^2)$, $i = \overline{1,n}$, n=25, $\sigma^2 = 25$. Тогда $\hat{a}_n(\xi_1, \ldots, \xi_n) \in \mathrm{N}\left(a, \frac{\sigma^2}{n}\right)$, а $\varphi(\xi_1, \ldots, \xi_n; a) = \frac{\hat{a}_n(\xi_1, \ldots, \xi_n) - a}{\sigma}\sqrt{n} \in N(0;1)$, то есть является центральной статистикой.

В соответствии с центральным методом построения доверительных интервалов

$$\mathrm{P}\{u_{p_1} \le \varphi(\xi_1, \ldots, \xi_n; a) < u_{p_2}\} = \mathrm{P}\left\{u_{p_1} \le \frac{\hat{a}_n(\xi_1, \ldots, \xi_n) - a}{\sigma}\sqrt{n} < u_{p_2}\right\} =$$

$=\mathrm{P}\left(\hat{a}_n(\xi_1,...,\xi_n)-u_{p_2}\frac{\sigma}{\sqrt{n}}\le a<\hat{a}_n(\xi_1,...,\xi_n)-u_{p_1}\frac{\sigma}{\sqrt{n}}\right)=p_2-p_1$, где u_{p_1}, u_{p_2} - квантили стандартного нормального распределения уровней p_1, p_2 $(0<p_1<p_2<1)$ соответственно. Положив $p_1=\frac{\alpha}{2}$, $p_2=1-\frac{\alpha}{2}$, получим $\mathrm{P}\left(\hat{a}_n(\xi_1,...,\xi_n)-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le a<\hat{a}_n(\xi_1,...,\xi_n)-u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$. (4.3)

Для стандартного нормального распределения $u_{\frac{\alpha}{2}}=-u_{1-\frac{\alpha}{2}}$, поэтому выражение (4.3) примет вид $\mathrm{P}\left(\hat{a}_n(\xi_1,...,\xi_n)-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le a<\hat{a}_n(\xi_1,...,\xi_n)+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$.

По условию задачи $1-\alpha=0{,}95$, следовательно $1-\frac{\alpha}{2}=0{,}975$.

По таблице квантилей стандартного нормального распределения найдем квантиль уровня 0,975: $u_{0{,}975}=1{,}96$. Тогда, с учетом всех известных параметров, искомый интервал будет иметь вид $(12{,}04;\ 15{,}96)$.

Задача 4.3

Для определения вертикального угла ориентира используют среднее арифметическое нескольких замеров угла при помощи секстанта. Для углов, измеряемых секстантом, среднеквадратическое отклонение принимается равным 1,5 минуты. Найти минимальное количество замеров, которое надо произвести, чтобы погрешность результата с вероятностью 0,99 не превосходила 1 минуту. Предполагается, что измерения имеют нормальное распределение.

Решение. Пусть $\hat{a}_n(\xi_1,\dots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i$, где $\xi_1,\dots,\xi_n$ - независимые случайные величины $\xi_i\in\mathrm{N}(a,\sigma^2)$, $i=\overline{1,n}$, $\sigma=1{,}5'$. Тогда $\hat{a}_n(\xi_1,\dots,\xi_n)\in\mathrm{N}\left(a,\frac{\sigma^2}{n}\right)$. По условию задачи, нужно найти такое *n*, чтобы $\mathrm{P}\{|\hat{a}_n(\xi_1,\dots,\xi_n)-a|\le 1\}=0{,}99$. С учетом (4.3), *n* может быть найдено из условия $u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}=1$. Так как $1-\alpha=0{,}99$, то $1-\frac{\alpha}{2}=0{,}995$. По таблице квантилей стандартного нормального распределения найдем квантиль уровня 0,995: $u_{0,995}=2{,}576$. Тогда $2{,}576\times\frac{1{,}5}{\sqrt{n}}=1$, откуда $n\geq(3{,}864)^2\Rightarrow n=15$.

Задача 4.4

Содержание углевода в единице продукта подчиняется нормальному распределению с неизвестным математическим ожиданием и неизвестной дисперсией. По результатам 25 независимых измерений $x_1,\ldots,x_{25}$ были получены несмещенные оценки математического ожидания $\bar{x}=\frac{1}{25}\sum_{i=1}^{25}x_i=18\ (г)$ и дисперсии $S^2=\frac{1}{24}\sum_{i=1}^{25}\left(x_i-\bar{x}\right)^2=16\ (г^2)$.

Построить доверительный интервал надежности 0,99 для неизвестного математического ожидания.

Решение. Пусть $\hat{a}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i$, $\hat{D}_n\left(\xi_1,\ldots,\xi_n\right)=\frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i-\hat{a}_n\left(\xi_1,\ldots,\xi_n\right)\right)^2$, где $\xi_1,\ldots,\xi_n$ - независимые случайные величины $\xi_i\in \mathrm{N}\left(a,\sigma^2\right)$, $i=\overline{1,n}$, n=25.

В математической статистике доказывается, что случайная величина $\varphi\left(\xi_1,\ldots,\xi_n;a\right)=\frac{\hat{a}_n\left(\xi_1,\ldots,\xi_n\right)-a}{\sqrt{\hat{D}_n\left(\xi_1,\ldots,\xi_n\right)}}\sqrt{n}$ имеет распределение Стьюдента с *n-1* степенью свободы. Кроме того, функция $\varphi(\xi_1,\ldots,\xi_n;a)$ непрерывна и строго монотонна по параметру a, следовательно, является центральной статистикой.

В соответствии с центральным методом построения доверительных интервалов

$$\mathrm{P}\left\{t_{n-1,p_1}\leq\varphi\left(\xi_1,\ldots,\xi_n;a\right)<t_{n-1,p_2}\right\}=$$

$$=\mathrm{P}\left\{t_{n-1,p_1}\leq\frac{\hat{a}_n\left(\xi_1,\ldots,\xi_n\right)-a}{\sqrt{\hat{D}_n\left(\xi_1,\ldots,\xi_n\right)}}\sqrt{n}<t_{n-1,p_2}\right\}=$$

$$=\mathrm{P}\left(\hat{a}_n(\xi_1,\ldots,\xi_n)-t_{n-1,p_2}\frac{\sqrt{\hat{D}_n\left(\xi_1,\ldots,\xi_n\right)}}{\sqrt{n}}\leq a<\hat{a}_n(\xi_1,\ldots,\xi_n)-t_{n-1,p_1}\frac{\sqrt{\hat{D}_n\left(\xi_1,\ldots,\xi_n\right)}}{\sqrt{n}}\right)=$$

$$=1-\alpha \tag{4.4}$$

В выражении (4.4) $t_{n-1,p_1},\ t_{n-1,p_2}$ - квантили распределения Стьюдента с *n-1* степенью свободы уровней $p_1,\ p_2$ $(0<p_1<p_2<1)$ соответственно.

Положив $p_1=\dfrac{\alpha}{2},\ p_2=1-\dfrac{\alpha}{2},$ с учетом того, что для распределения Стьюдента $t_{n-1,\frac{\alpha}{2}}=-t_{n-1,1-\frac{\alpha}{2}},$ получим

$$\mathrm{P}\left(\hat{a}_n(\xi_1,\ldots,\xi_n)-t_{n-1,1-\frac{\alpha}{2}}\frac{\sqrt{\hat{D}_n(\xi_1,\ldots,\xi_n)}}{\sqrt{n}}\le a<\hat{a}_n(\xi_1,\ldots,\xi_n)-t_{n-1,1-\frac{\alpha}{2}}\frac{\sqrt{\hat{D}_n(\xi_1,\ldots,\xi_n)}}{\sqrt{n}}\right)=$$

$$=1-\alpha \qquad (4.5)$$

По условию задачи $1-\alpha=0{,}99$, следовательно $1-\dfrac{\alpha}{2}=0{,}995$.

По таблице квантилей распределения Стьюдента найдем квантиль уровня 0,995: $t_{24;\,0{,}995}=2{,}797$. Тогда, с учетом всех известных параметров, доверительный интервал надежности 0,995 будет иметь вид $\left(18-2{,}797\times\dfrac{4}{5};\ 18+2{,}797\times\dfrac{4}{5}\right)$, или $(15{,}7624;\ 20{,}2376)$.

Задача 4.5

Результаты измерений диаметра вала имеют нормальное распределение с неизвестным математическим ожиданием и неизвестной дисперсией. Было произведено 10 независимых измерений диаметра вала $x_1,\ldots,x_{10}$ и найдены несмещенные оценки математического ожидания $\bar{x}=\dfrac{1}{10}\sum\limits_{i=1}^{10}x_i=30\ (\text{мм})$ и дисперсии $S^2=\dfrac{1}{9}\sum\limits_{i=1}^{10}(x_i-\bar{x})^2=9\ (\text{мм}^2)$.

Построить доверительный интервал надежности 0,9 для неизвестной дисперсии.

Решение. Пусть $\hat{a}_n(\xi_1,\ldots,\xi_n)=\dfrac{1}{n}\sum\limits_{i=1}^{n}\xi_i,$

$\hat{D}_n(\xi_1,\ldots,\xi_n)=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}\left(\xi_i-\hat{a}_n(\xi_1,\ldots,\xi_n)\right)^2,$ где $\xi_1,\ldots,\xi_n$ - независимые случайные величины $\xi_i\in\mathrm{N}(a,\sigma^2),\ i=\overline{1,n},\ n=10$. В математической статистике доказывается, что случайная величина $\varphi(\xi_1,\ldots,\xi_n;\sigma^2)=\dfrac{(n-1)\times\hat{D}_n(\xi_1,\ldots,\xi_n)}{\sigma^2}$ имеет Хи-квадрат распределение с *n-1* степенью свободы. Кроме того, функция

$\varphi(\xi_1,\dots,\xi_n;\sigma^2)$ непрерывна и строго монотонна по параметру σ^2, следовательно, является центральной статистикой.

В соответствии с центральным методом построения доверительных интервалов

$$\mathrm{P}\left\{ \chi^2_{n-1,\, p_1} \le \varphi\left(\xi_1,\dots,\xi_n;\sigma^2\right) < \chi^2_{n-1,\, p_2} \right\} =$$

$$= \mathrm{P}\left\{ \chi^2_{n-1,\, p_1} \le \frac{(n-1)\times \widehat{D}_n\left(\xi_1,\dots,\xi_n\right)}{\sigma^2} < \chi^2_{n-1,\, p_2} \right\} =$$

$$= \mathrm{P}\left(\frac{(n-1)\times\sqrt{\widehat{D}_n\left(\xi_1,\dots,\xi_n\right)}}{\chi^2_{n-1,\,p_2}} < \sigma^2 \le \frac{(n-1)\times\sqrt{\widehat{D}_n\left(\xi_1,\dots,\xi_n\right)}}{\chi^2_{n-1,\,p_1}} \right) = 1-\alpha,$$ где

$\chi^2_{n-1,\,p_1}$, $\chi^2_{n-1,\,p_2}$ - квантили Хи-квадрат распределения с *n-1* степенью свободы уровней p_1, p_2 $\left(0 < p_1 < p_2 < 1\right)$ соответственно.

Положив $p_1 = \frac{\alpha}{2}$, $p_2 = 1 - \frac{\alpha}{2}$, получим

$$\mathrm{P}\left(\frac{(n-1)\times\sqrt{\widehat{D}_n\left(\xi_1,\dots,\xi_n\right)}}{\chi^2_{n-1,1-\frac{\alpha}{2}}} < \sigma^2 \le \frac{(n-1)\times\sqrt{\widehat{D}_n\left(\xi_1,\dots,\xi_n\right)}}{\chi^2_{n-1,\frac{\alpha}{2}}} \right) = 1-\alpha. \qquad (4.6)$$

По условию задачи $1-\alpha = 0{,}9$, следовательно $\frac{\alpha}{2} = 0{,}05$, $1-\frac{\alpha}{2} = 0{,}95$. Из таблицы квантилей Хи-квадрат распределения найдем квантили соответствующих уровней: $\chi^2_{9;\,0{,}05} = 3{,}33$; $\chi^2_{9;\,0{,}95} = 16{,}9$. С учетом всех известных параметров, из (4.6) получим доверительный интервал надежности 0,9 для неизвестной дисперсии: $\left(\frac{9\times 3}{16{,}9}; \frac{9\times 3}{3{,}33}\right)$, или (1,6; 8,1).

Задача 4.6

Пусть 51, 62, 53, 52, 63 – результаты пяти повторных независимых наблюдений над случайной величиной $\xi \in N\left(a,\sigma_1^2\right)$, а $x_1,\dots,x_7$ - результаты семи повторных независимых наблюдений над случайной величиной $\eta \in N\left(b,\sigma_2^2\right)$, причем ξ и η- независимые случайные величины. По выборке

$x_1, \dots, x_7$ получены следующие несмещенные оценки математического ожидания и дисперсии η: $\bar{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 60{,}3$; $S^2 = \frac{1}{6}\sum_{i=1}^{7}(x_i - \bar{x})^2 = 36{,}06$.

Построить доверительный интервал надежности 0,9 для отношения дисперсий $\frac{\sigma_2^2}{\sigma_1^2}$ случайных величин ξ и η.

Решение. Пусть $\widehat{a}\,(\xi_1, \dots, \xi_5) = \frac{1}{5}\sum_{i=1}^{5}\xi_i$, $\widehat{b}\,(\eta_1, \dots, \eta_7) = \frac{1}{7}\sum_{i=1}^{7}\eta_i$, $\widehat{D}_1(\xi_1, \dots, \xi_5) = \frac{1}{4}\sum_{i=1}^{5}\left(\xi_i - \widehat{a}\,(\xi_1, \dots, \xi_5)\right)^2$, $\widehat{D}_2(\eta_1, \dots, \eta_7) = \frac{1}{6}\sum_{i=1}^{7}\left(\eta_i - \widehat{b}\,(\eta_1, \dots, \eta_7)\right)^2$, где $\xi_1, \dots, \xi_5, \eta_1, \dots, \eta_7$ - независимые случайные величины, $\xi_i \in \mathrm{N}\left(a, \sigma_1^2\right)$, $i = \overline{1,5}$, $\eta_i \in \mathrm{N}\left(b, \sigma_2^2\right)$, $i = \overline{1,7}$. В математической статистике доказывается, что случайная величина $\varphi\left(\xi_1, \dots, \xi_5, \eta_1, \dots, \eta_7; \sigma_1^2, \sigma_2^2\right) = \frac{\widehat{D}_1\left(\xi_1, \dots, \xi_5\right) \times \sigma_2^2}{\widehat{D}_2\left(\eta_1, \dots, \eta_7\right) \times \sigma_1^2}$ имеет распределение Фишера с 4 и 6 степенями свободы. Кроме того, она непрерывна и строго монотонна по переменной $w = \frac{\sigma_2^2}{\sigma_1^2}$, следовательно, является центральной статикой.

В соответствии с центральным методом построения доверительных интервалов

$$\mathrm{P}\left(\left\{ X_{4,6,p_1} \le \varphi\left(\xi_1, \dots, \xi_5, \eta_1, \dots, \eta_7; \sigma_1^2, \sigma_2^2\right) < X_{4,6,p_2} \right\}\right) =$$

$$= \mathrm{P}\left(X_{4,6,p_1} \le \frac{\widehat{D}_1\left(\xi_1, \dots, \xi_5\right) \times \sigma_2^2}{\widehat{D}_2\left(\eta_1, \dots, \eta_7\right) \times \sigma_1^2} < X_{4,6,p_2} \right) =$$

$$= \mathrm{P}\left(X_{4,6,p_1} \times \frac{\widehat{D}_2\left(\eta_1, \dots, \eta_7\right)}{\widehat{D}_1\left(\xi_1, \dots, \xi_5\right)} \le \frac{\sigma_2^2}{\sigma_1^2} < X_{4,6,p_2} \times \frac{\widehat{D}_2\left(\eta_1, \dots, \eta_7\right)}{\widehat{D}_1\left(\xi_1, \dots, \xi_5\right)} \right) = 1 - \alpha,$$

где $X_{4,6,p_1}$, $X_{4,6,p_2}$ - квантили распределения Фишера с 4 и 6 степенями свободы уровней p_1, p_2 $\left(0 < p_1 < p_2 < 1,\ p_2 - p_1 = 1 - \alpha\right)$ соответственно.

Положив $p_1 = \frac{\alpha}{2}$, $p_2 = 1 - \frac{\alpha}{2}$, получим

$$\mathrm{P}\left(X_{4,6;\frac{\alpha}{2}} \times \frac{\widehat{D}_2\left(\eta_1, \dots, \eta_7\right)}{\widehat{D}_1\left(\xi_1, \dots, \xi_5\right)} \le \frac{\sigma_2^2}{\sigma_1^2} < X_{4,6;1-\frac{\alpha}{2}} \times \frac{\widehat{D}_2\left(\eta_1, \dots, \eta_7\right)}{\widehat{D}_1\left(\xi_1, \dots, \xi_5\right)} \right) = 1 - \alpha. \qquad (4.7)$$

Найдем несмещенные точечные оценки математического ожидания и дисперсии случайной величины ξ. $\widehat{a} = \frac{1}{5}(51+62+53+52+63)=56{,}2$; $\widehat{D}_1 = \frac{1}{4}\left((51-56{,}2)^2 + (62-56{,}2)^2 + (53-56{,}2)^2 + (52-56{,}2)^2 + (63-56{,}2)^2\right) = 33{,}7.$

По условию задачи $1-\alpha = 0{,}9$, следовательно $\frac{\alpha}{2} = 0{,}05$, $1-\frac{\alpha}{2} = 0{,}95$. Из таблицы квантилей распределения Фишера найдем квантили соответствующих уровней: $X_{4,6;0,05} = \frac{1}{X_{6,4;0,95}} = \frac{1}{6{,}16}$; $X_{4,6;0,95} = 4{,}53$. С учетом всех известных параметров, из (4.7) получим доверительный интервал надежности 0,9 для отношения дисперсий $\frac{\sigma_2^2}{\sigma_1^2}$: $\left(\frac{1}{6{,}16}\times\frac{36{,}06}{33{,}7};\ 4{,}53\times\frac{36{,}06}{33{,}7}\right)$, или $(0{,}1737;\ 4{,}8472)$.

Задача 4.7

Амплитуда колебаний определялась двумя лаборантами. Первый лаборант по 10 независимым наблюдениям получил выборочное среднее амплитуды, равное 81 мм, а второй по 15 независимым наблюдениям получил выборочное среднее, равное 84 мм. В предположении, что результаты измерений первого и второго лаборантов подчиняются нормальным распределениям с дисперсиями 64 $(мм)^2$ и 81 $(мм)^2$ соответственно, построить доверительный интервал надежности 0,99 для разности математических ожиданий случайных величин с соответствующими распределениями.

Решение. По условиям задачи мы имеем две независимые повторные выборки $x_1,\dots,x_m$ ($m=10$) и $y_1,\dots,y_n$ ($n=15$) из нормальных распределений с параметрами $a_1, \sigma_1=8$ и $a_2, \sigma_2=9$ соответственно. По выборкам требуется построить доверительный интервала надежности $1-\alpha=0{,}99$ для параметра $\breve{a} = a_1 - a_2$. Рассмотрим статистики $\widehat{a}_1(\xi_1,\dots,\xi_m) = \frac{1}{m}\sum_{i=1}^{m}\xi_i$ и $\widehat{a}_2(\eta_1,\dots,\eta_n) = \frac{1}{n}\sum_{i=1}^{n}\eta_i$, где $\xi_1,\dots,\xi_m,\eta_1,\dots,\eta_n$ - независимые случайные величины; ξ_i имеют нормальное распределение с параметрами $a_1, \sigma_1=8$, $i=1,\dots,m$; η_j имеют нормальное распределение с параметрами $a_2, \sigma_2=9$, $j=1,\dots,n$. По свойствам нормального распределения

$\widehat{a}_1(\xi_1,\ldots,\xi_m)\in \mathrm{N}\left(a_1,\frac{\sigma_1^2}{m}\right)$, $\widehat{a}_2(\eta_1,\ldots,\eta_n)\in \mathrm{N}\left(a_2,\frac{\sigma_2^2}{n}\right)$, а разность $\widehat{a}_1(\xi_1,\ldots,\xi_m)-\widehat{a}_2(\eta_1,\ldots,\eta_n)$ имеет нормальное распределение с параметрами a_1-a_2 и $\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}$. Следовательно, статистика $Y(\xi_1,\ldots,\xi_m,\eta_1,\ldots,\eta_n;a_1,a_2)=\frac{(\widehat{a}_1(\xi_1,\ldots,\xi_m)-\widehat{a}_2(\eta_1,\ldots,\eta_n))-(a_1-a_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}$ будет иметь стандартное нормальное распределение. Кроме того, эта статистика непрерывна и строго монотонна по параметру $\breve{a}=a_1-a_2$, то есть является центральной. В соответствии с центральным методом, $\mathrm{P}\left\{-u_{1-\frac{\alpha}{2}}<Y(\xi_1,\ldots,\xi_m,\eta_1,\ldots,\eta_n;a_1,a_2)<u_{1-\frac{\alpha}{2}}\right\}=1-\alpha$, где $u_{1-\frac{\alpha}{2}}$ - квантиль стандартного нормального распределения уровня $1-\frac{\alpha}{2}$. Или

$$\mathrm{P}\left\{-u_{1-\frac{\alpha}{2}}<\frac{(\widehat{a}_1(\xi_1,\ldots,\xi_m)-\widehat{a}_2(\eta_1,\ldots,\eta_n))-(a_1-a_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}<u_{1-\frac{\alpha}{2}}\right\}=1-\alpha.$$

С учетом всех известных параметров получим: $81-84-u_{0,995}\sqrt{\frac{64}{10}+\frac{81}{15}}<a_1-a_2<81-84+u_{0,995}\sqrt{\frac{64}{10}+\frac{81}{15}}$. Так как $u_{0,995}=2,576$, то $-11,848<a_1-a_2<5,848$.

Приближенные интервальные оценки. Рассмотрим следующие частные задачи.

Задача 4.8

Пусть $x_1,x_2,\ldots,x_n$ - результаты n повторных независимых наблюдений над случайной величиной ξ, дисперсия которой σ^2 - конечна и известна, а $\mathrm{M}\xi=a$ - неизвестный параметр. Требуется построить доверительный интервал надежности $1-\alpha$ для неизвестного параметра a, если объем выборки *n>>1*.

Решение. Так как выборка повторная, то ее элементы можно трактовать как значения независимых случайных величин $\xi_1,\ldots,\xi_n$, каждая из которых имеет тот же закон распределения что и ξ.

Пусть $\widehat{a}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i$. Поскольку, по условиям задачи, объем выборки велик, то, в соответствии с центральной предельной теоремой, закон распределения вероятностей статистики $\varphi\left(\xi_1,\ldots,\xi_n;a\right)=\frac{\widehat{a}_n\left(\xi_1,\ldots,\xi_n\right)-a}{\sigma}\sqrt{n}$ будет близок к нормальному распределению с параметрами 0 и 1. Следовательно, для достаточно больших n $\mathrm{P}\left\{u_{p_1}\le\frac{\widehat{a}_n\left(\xi_1,\ldots,\xi_n\right)-a}{\sigma}\sqrt{n}<u_{p_2}\right\}\approx p_2-p_1$, где u_{p_1}, u_{p_2} - квантили стандартного нормального распределения уровней p_1 и p_2 соответственно. Если задать $p_1=\frac{\alpha}{2}$, а $p_2=1-\frac{\alpha}{2}$, то

$$\mathrm{P}\left(\widehat{a}_n(\xi_1,\ldots,\xi_n)-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le a<\widehat{a}_n(\xi_1,\ldots,\xi_n)+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)\approx 1-\alpha .$$

Задача 4.9

Из большой партии электроламп было отобрано случайным образом 400 штук для определения средней продолжительности горения. Выборочное среднее продолжительности горения оказалось равным 1220 часам. Построить доверительный интервал надежности 0,99 для средней продолжительности горения электролампы во всей партии, если среднеквадратическое отклонение продолжительности горения известно и равно 35 часам.

Решение. Так как объем выборки достаточно велик (n=400), то для решения данной задачи можно воспользоваться результатами предыдущей. Тогда $\mathrm{P}\left(\widehat{a}_n(\xi_1,\ldots,\xi_n)-u_{0,995}\frac{35}{20}\le a<\widehat{a}_n(\xi_1,\ldots,\xi_n)+u_{0,995}\frac{35}{20}\right)\approx 0,99$, где $\widehat{a}_n(\xi_1,\ldots,\xi_n)=\frac{1}{n}\sum_{i=1}^{n}\xi_i$, а $u_{0,995}$=2,576 - квантиль стандартного нормального распределения уровня 0,995. С учетом того, что $\widehat{a}_n(x_1,\ldots,x_n)=1220$, искомый доверительный интервал будет иметь вид (1215,49; 1224,51).

Задача 4.10

Пусть проводится n последовательных независимых испытаний, каждое из которых имеет два возможных исхода A и $\overline{A}$. Вероятность исхода A от испытания к испытанию не меняется и равна p $(0<p<1)$, где p – неизвестный параметр. При каждом испытании фиксируется, какой из исходов имел место, и формируется выборка $x_1, x_2, \dots, x_n$, в которой $x_i = 1$, если в результате i - го испытания произошло *A*, и $x_i = 0$, в противном случае.

По результатам наблюдений, требуется построить доверительный интервал надежности $1-\alpha$ для неизвестного параметра *p*, если число испытаний *n>>1*.

Решение. Из условий задачи следует, что элементы выборки $x_1, x_2, \dots, x_n$ можно трактовать, как значения независимых случайных величин $\xi_1, \dots, \xi_n$, где $\xi_i \in \{0,1\}$, $P\{\xi_i = 1\} = p$, $P\{\xi_i = 0\} = 1-p$, *i=1,...,n*. Рассмотрим статистику $K(\xi_1, \dots, \xi_n) = \sum\limits_{i=1}^{n} \xi_i$ - число "успехов" в *n* испытаниях по схеме Бернулли. Математическое ожидание $K(\xi_1, \dots, \xi_n)$ равно *np*, а дисперсия – *np(1-p)*. В соответствии с предельной теоремой Муавра-Лапласа данная статистика, при *n>>1,* имеет закон распределения, близкий к нормальному распределению с параметрами *np* и *np(1-p)*. Следовательно, для всяких p_1, p_2, таких, что $0 < p_1 < p_2 < 1$, при n>>1, можно записать $\mathrm{P}\left\{ u_{p_1} \le \frac{K(\xi_1, \dots, \xi_n) - np}{\sqrt{np(1-p)}} < u_{p_2} \right\} \approx p_2 - p_1$, где u_{p_1}, u_{p_2} - квантили стандартного нормального распределения уровней p_1 и p_2 соответственно. Если задать $p_1 = \frac{\alpha}{2}$, $p_2 = 1 - \frac{\alpha}{2}$, то

$$\mathrm{P}\left\{ -u_{1-\frac{\alpha}{2}} \le \frac{K(\xi_1, \dots, \xi_n) - np}{\sqrt{np(1-p)}} < u_{1-\frac{\alpha}{2}} \right\} \approx 1-\alpha. \qquad (4.8)$$

Разрешая неравенство в фигурных скобках относительно параметра *p*, получаем

$$\mathrm{P}\left\{ \frac{a(\xi_1, \dots, \xi_n) - b(\xi_1, \dots, \xi_n)}{2\left(n^2 + n\left(u_{1-\frac{\alpha}{2}}\right)^2\right)} < p < \frac{a(\xi_1, \dots, \xi_n) + b(\xi_1, \dots, \xi_n)}{2\left(n^2 + n\left(u_{1-\frac{\alpha}{2}}\right)^2\right)} \right\} \approx 1-\alpha, \qquad (4.9)$$

где $a(\xi_1, \dots, \xi_n) = 2nK(\xi_1, \dots, \xi_n) + n\left(u_{1-\frac{\alpha}{2}}\right)^2$,

$$b(\xi_1, \dots, \xi_n) = \sqrt{\left(a(\xi_1, \dots, \xi_n)\right)^2 - 4\left(K(\xi_1, \dots, \xi_n)\right)^2\left(n^2 + n\left(u_{1-\frac{\alpha}{2}}\right)^2\right)}.$$

Пусть *n*=50, $K(x_1, \dots, x_n) = \sum\limits_{i=1}^{50} x_i = 25$, $1-\alpha = 0{,}9$. Тогда доверительный интервал надежности 0,9 для неизвестной вероятности будет иметь вид (0,3867; 0,6133).

Следует отметить, что выражение (4.8) часто представляют в виде

$$\mathrm{P}\left\{\frac{1}{n}K(\xi_1,\dots,\xi_n)-\frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{p(1-p)}\le p<\frac{1}{n}K(\xi_1,\dots,\xi_n)+\frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{p(1-p)}\right\}\approx 1-\alpha$$

и затем в качестве доверительного интервала берут $(\underline{p}(\xi_1,\dots,\xi_n);\ \overline{p}(\xi_1,\dots,\xi_n))$,

где $$\underline{p}(\xi_1,\dots,\xi_n)=\left(\frac{K(\xi_1,\dots,\xi_n)}{n}-\frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\frac{K(\xi_1,\dots,\xi_n)}{n}\times\left(1-\frac{(K(\xi_1,\dots,\xi_n))}{n}\right)}\right),$$

$$\overline{p}(\xi_1,\dots,\xi_n)=\left(\frac{K(\xi_1,\dots,\xi_n)}{n}+\frac{1}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\sqrt{\frac{K(\xi_1,\dots,\xi_n)}{n}\times\left(1-\frac{(K(\xi_1,\dots,\xi_n))}{n}\right)}\right).$$

Данный прием связан с двукратным использованием приближения, а именно: закон распределения статистики $K(\xi_1,\dots,\xi_n)$ заменяется нормальным и, кроме того, в выражения для границ доверительного интервала вместо неизвестной вероятности подставляется ее несмещенная точечная оценка. При малых и средних объемах выборки применение такого приема может приводить к значительным ошибкам, поэтому использовать его следует с осторожностью.

Найдем $(\underline{p}(x_1,\dots,x_n);\ \overline{p}(x_1,\dots,x_n))$ для n=50, $K(x_1,\dots,x_n)=\sum_{i=1}^{50}x_i=25$, $1-\alpha=0{,}9$.

$$\underline{p}=\left(\frac{25}{50}-\frac{1{,}645}{\sqrt{50}}\sqrt{\frac{25}{50}\times\left(1-\frac{25}{50}\right)}\right)\approx\frac{1}{2}-\frac{1{,}645}{5\times1{,}414}\times\frac{1}{2}\approx 0{,}5-0{,}1162=0{,}3838;$$

$$\overline{p}=\left(\frac{25}{50}+\frac{1{,}645}{\sqrt{50}}\sqrt{\frac{25}{50}\times\left(1-\frac{25}{50}\right)}\right)\approx\frac{1}{2}+\frac{1{,}645}{5\times1{,}414}\times\frac{1}{2}\approx 0{,}5+0{,}1162=0{,}6162.$$

Таким образом, доверительный интервал для неизвестной вероятности, надежность которого приблизительно равна 0,9, будет иметь вид (0,3838;0,6162). Сравнивая его с доверительным интервалом, найденным с помощью выражения (4.9), видим, что в данном случае замена точного значения вероятности на его несмещенную точечную оценку не привела к значительным ошибкам, а вычисление границ интервала значительно упростилось.

Задача 4.11

В результате 10000 сеансов игры с автоматом выигрыш появился 4000 раз. Найти доверительный интервал надежности 0,95 для вероятности выигрыша в одном сеансе.

Решение. Воспользуемся результатами предыдущей задачи. Тогда

$$\underline{p}=\frac{4000}{10000}-\frac{1}{100}u_{0{,}975}\sqrt{\frac{4000}{10000}\times\frac{6000}{10000}}=0{,}4-0{,}01\times1{,}96\sqrt{0{,}24}=$$

$=0{,}4-0{,}0196\times0{,}49=0{,}4-0{,}0096\approx0{,}39$. $\overline{p}=0{,}4+0{,}0096\approx0{,}41$. Таким образом, доверительный интервал надежности 0,95 для неизвестной вероятности выигрыша имеет вид (0,39; 0,41).

Литература

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Учебник для вузов. – М.: ЮНИТИ, 1988. – 1022 с.
2. Боровков А.А. Математическая статистика. – М.: Наука, 1984. – 472 с.
3. Браунли К.А. Статистическая теория и методология в науке и технике. – М.: Наука, 1977. – 407 с.
4. Гихман И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика. – Киев: Вища школа, 1979. – 408 с.
5. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие для студентов втузов. – М.: Высшая школа, 1979. – 400 с.
6. Ивченко Г.И., Медведев Ю.И. Математическая статистика: Учеб. пособие для втузов. – М.: Высшая школа, 1984. – 248 с.
7. Климов Г.П. Теория вероятностей и математическая статистика. – М.: Изд-во Моск. ун-та, 1983. – 328 с.
8. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648с.
9. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2000. – 543 с.
10. Математическая статистика: Учеб. для вузов/ В.Б. Горяинов, И.В. Павлов, Г.М. Цветкова и др. Под ред. В.С. Зарубина, А.П. Крищенко. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2001. – 424 с.
11. Сборник задач по математике для втузов. Ч.3. Теория вероятностей и математическая статистика: Учеб. пособие для втузов/ Под ред. А.В. Ефимова – М.: Наука, 1990. – 428 с.
12. Тихов М.С., Агеев В.В., Зорин В.А. Сборник задач по прикладной теории вероятностей и математической статистике: Уч. пос. – Горький: ГГУ, 1986. – 80 с.
13. Федоткин М.А. Основы прикладной теории вероятностей и статистики: Учебник/ М.А. Федоткин. – М.: Высш. шк., 2006. – 368 с.

Содержание

ЗАДАЧИ ОЦЕНИВАНИЯ НЕИЗВЕСТНЫХ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЙ

Составитель
Валентина Михайловна **Сморкалова**

Учебно-методическое пособие

Федеральное государственное автономное образовательное учреждение высшего образования
«Нижегородский государственный университет им. Н.И. Лобачевского»
603950, Нижний Новгород, пр. Гагарина, 23