Методическая разработка

для проведения лекции

Занятие 5. Методы статистического анализа и особенности их реализации в системах аналитической обработки данных

Учебные вопросы занятия:
1. Типовые задачи статистического анализа данных и методы их решения
2. Особенности многомерного статистического анализа данных

Заключительная часть

1. Типовые задачи статистического анализа данных и методы их решения

Часто в процессе обработки различной информации возникает задача измерения ее некоторых параметров. При этом, как правило, обрабатываемая информация имеет различную степень достоверности и носит вероятностный характер.

Повышение достоверности информации может быть достигнуто разработкой оптимального в некотором смысле алгоритма анализа и обработки наблюдаемого случайного процесса. Так как наиболее полным описанием различных случайных последовательностей является функция распределения вероятностей ее значений, то задача анализа в общем случае сводится к получению эмпирических вероятностных характеристик по доступным выборочным данным и проверке гипотез об их соответствии некоторым стандартным характеристикам, определяющим различные классы случайных последовательностей и отдельные их свойства.

Обобщенный алгоритм анализа случайной последовательности может включать следующие этапы.

1. Определение эмпирических вероятностных характеристик анализируемой случайной последовательности (математического ожидания, дисперсии, корреляционного момента, вероятностей событий и функции распределения вероятностей). Важно, чтобы качество полученных эмпирических оценок соответствовало выдвигаемым априорно требованиям к допустимому отклонению от истинных значений характеристик (доверительному интервалу и доверительной вероятности), а также определялось требуемым для этого размером выборки. На основе полученных характеристик могут быть установлены свойства симметрии распределения (совпадение значений среднего, моды и медианы,

либо равенство значений вероятностей превышения и не превышения среднего значения) и близости его формы к некоторому стандартному распределению, например, к нормальному.

2. Построение гистограммы вероятностей и восстановление эмпирического распределения случайной последовательности на основе полученных вероятностных характеристик и выдвижение гипотезы о виде распределения случайной последовательности.

3. Проверка верности выдвинутой гипотезы по критериям соответствия (согласия) эмпирических и аналитических вероятностных характеристик, а также определение класса и основных свойств случайной последовательности с оценкой показателей качества полученных оценок и решений.

Проанализируем основные этапы оценивания параметров случайных последовательностей в предположении выполнения условия стационарности выборочных данных.

При этом необходимо учитывать, что оценки бывают точечные и интервальные. *Интервальные* оценки предполагают отыскание области (доверительного интервала), в которую попадает оцениваемый параметр с заданной вероятностью. При *точечных* оценках значение параметра определяется точкой, с последующим нахождением погрешности оценивания.

Согласно положениям математической статистики, оценка является случайной величиной и должна удовлетворять требованиям состоятельности, несмещенности и эффективности.

Оценка называется состоятельной, если она сходится по вероятности к оцениваемому параметру при неограниченном увеличении размера выборки

$$\lim_{n\to\infty} P\left\{\left|\hat{x}(t)-x(t)\right|\geq\gamma\right\}=0, \qquad \forall\gamma>0. \tag{1}$$

где $x(t)$ - истинное значение оцениваемого параметра;

$\hat{x}(t)$ - оценка параметра;

n - объем выборки;

γ - бесконечно малая случайная величина.

Состоятельность оценки означает (в соответствии с законом больших чисел), что выборочное среднее сходится по вероятности к априорному среднему.

Оценка называется несмещенной, если ее среднее значение по совокупности выборок данного размера в точности равно оцениваемому параметру

$\bar{x}(t)=x(t)$, где $\bar{x}(t)=\dfrac{1}{N}\sum x_i(t)$.

Оценка называется эффективной, если она имеет наименьшую среди возможных оценок дисперсию, то есть удовлетворяет неравенству

$$\sigma^2\geq\sigma^2_{эф} , \tag{2}$$

где $\sigma^2_{эф}$ - дисперсия эффективной оценки. Если неравенство (2) выполняется лишь при $n\to\infty$, то такая оценка называется асимптотически эффективной.

Рассмотрим задачу оценивания *вероятности* случайного события. Для оценивания указанной характеристики проводится серия из n испытаний по схеме Бернулли (независимых и однородных) и подсчитывается число m испытаний, в которых анализируемой событие появилось.

Отношение

$$\tilde{p}(x) = \frac{\hat{m}}{n} \tag{3}$$

называется частотой события $x(t)$ в серии испытаний и его статистической вероятностью.

Проанализируем свойства статистической вероятности $\tilde{p}(x)$.

Поскольку число m появления события в независимых и однородных испытаниях подчинено биномиальному закону распределения, то

$$M[\tilde{p}(x)] = \frac{1}{n} M[\hat{m}] = \frac{np}{n} = p \tag{4}$$

и, следовательно, частота $\tilde{p}(x)$ является несмещенной оценкой вероятности p.

Поскольку согласно теореме Бернулли

$$\lim_{n \to \infty} P\{|\tilde{p}(x) - p| < \gamma\} = 1, \tag{5}$$

т.е. частота $\tilde{p}(x)$ сходится по вероятности к вероятности p, то это состоятельная оценка вероятности p.

Дисперсия частоты $\tilde{p}(x)$ равна

$$D[\tilde{p}(x)] = \frac{1}{n^2} D[\hat{m}] = \frac{npq}{n^2} = \frac{pq}{n}, \text{ где } q = 1 - p. \tag{6}$$

Поскольку при $n \to \infty$ дисперсия $D[\tilde{p}(x)] \to 0$, то частота $\tilde{p}(x)$ - асимптотически эффективная оценка вероятности p. Таким образом, частота $\tilde{p}(x)$ события $x(t)$ в серии независимых однородных испытаний есть подходящее значение его вероятности p и, значит, наилучшая его точечная оценка.

Для определения объема выборки, необходимого для оценивания значений частоты $\tilde{p}(x)$, в качестве исходных данных выступают максимально вероятная погрешность статистической оценки закона распределения ε и доверительная вероятность β. Число проводимых испытаний n_p оценивалось по формуле

$$n_p \geq \frac{2p(1-p)}{\varepsilon^2} (\Phi^{-1}(\beta))^2,$$

где $\Phi(\beta)$ - функция Лапласа.

Рассмотрим задачу оценивания *математического ожидания* $M(x)$ эргодического случайного процесса $x(t)$. Стационарный случайный процесс называется эргодическим, если все его статистические характеристики, полученные в результате усреднения по ансамблю реализаций (множество всех возможных реализаций, заданных вместе с их распределением вероятностей), могут быть вычислены путем усреднения по времени данных, выбранных из одной реализации процесса.

Эргодическое свойство стационарного случайного процесса заключается в том, что любая его реализация обладает одними и теми же статистическими свойствами и на достаточно большом интервале T (аргумента t) ведет себя, в среднем, так же как и все другие реализации.

Будем оценивать математическое ожидание $M(x)$ путем усреднения по времени на конечном интервале

$$\tilde{M}(x) = \frac{1}{T}\int\limits_{t=0}^{T} x(t)dt. \tag{7}$$

Необходимо отметить, что хотя $\tilde{M}(x)$ - некоторое число в каком-либо эксперименте, эта величина также является случайной, так как мы получили бы другое число, если бы наблюдалась другая реализация. Таким образом, $\tilde{M}(x)$ не будет тождественно равно истинному математическому ожиданию $M(x)$.

Погрешность приближения оценки $\tilde{M}(x)$ к $M(x)$, равная

$$\Delta = \tilde{M}(x) - M(x), \tag{8}$$

является также случайной величиной.

Аналитическое выражение для оценивания анализируемого параметра желательно выбирать так, чтобы выполнялись следующие условия

1. Математическое ожидание Δ равно нулю:

$$M[\Delta] = 0. \tag{9}$$

2. $\lim\limits_{n\to\infty} P(\Delta \ge \gamma) = 0$, $\forall\gamma > 0$ (10')

3. Дисперсия Δ стремится к нулю с увеличением n

$$\lim_{n\to\infty} D(\Delta) = 0. \tag{10}$$

При выполнении условия (9) оценка является *несмещенной,* условия (10') – *состоятельной и* (10) – *эффективной.*

Вследствие случайного характера погрешности (8) для характеристики точности приближенного равенства $\tilde{M}(x) = M(x)$ необходимо располагать вероятностью β того, что абсолютное значение погрешности не превзойдет некоторого предела

$$P(|\Delta| \le \varepsilon) = \beta. \tag{11}$$

Интервал от $\tilde{M}(x) - \Delta$ до $\tilde{M}(x) + \Delta$, в котором с вероятностью β находится истинное значение $M(x)$, называется *доверительным интервалом,* а его границы - *доверительными границами.*

Если число экспериментальных данных n достаточно велико, то погрешность (8) состоятельной оценки $\tilde{M}(x)$ можно практически считать распределенной нормально с математическим ожиданием (9), дисперсией $D(\Delta) = D(\tilde{M}(x)) = D_\theta$ и средним квадратическим отклонением $\sigma(\Delta) = \sigma(\tilde{M}(x)) = \sigma_\theta = \sqrt{D_\theta}$. При этом выражение (11) имеет вид:

$$\beta = \Phi\left(\frac{\varepsilon}{\sigma_\theta}\right) - \Phi\left(\frac{-\varepsilon}{\sigma_\theta}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma_\theta}\right) = 2\Phi(t)\ , \tag{12}$$

где $\Phi(t)$- функция Лапласа, $t=\dfrac{\varepsilon}{\sigma_\theta}$.

С помощью этой формулы решается задача определения доверительной вероятности β по известным данным $\varepsilon, \sigma_\theta$.

Функция Лапласа $\tau=\Phi(t)$ выражает зависимость τ от t. Обратная $t=\Phi^{-1}(\tau)$ выражает зависимость t от τ. При $t=\dfrac{\varepsilon}{\sigma_\theta}$, $\tau=\dfrac{\beta}{2}$ имеем

$$\frac{\varepsilon}{\sigma_\theta}=\Phi^{-1}(\tau). \tag{13}$$

С помощью формулы (12) и обратной функции Лапласа решается задача определения доверительного интервала ε по известным β и σ_θ и необходимого числа испытаний по известным β и ε.

При решении первой задачи согласно (12) определяется ε. При решении второй задачи согласно (12) определяется σ_θ, а затем *n*.

Анализируя свойства оценки $\tilde{M}[x]$, получим:

$$M\left[\tilde{M}(x)\right]=M\left[\frac{1}{T}\int_0^T x(t)dt\right]=\frac{1}{T}\int_0^T M[x(t)]dt=\frac{1}{T}\left[M[x]\cdot t\Big|_0^T\right]=M[x]. \tag{14}$$

Отсюда следует, что $\tilde{M}(x)$ имеет математическое ожидание, равное математическому ожиданию $x(t)$.

На практике операция интегрирования редко может быть выполнена аналитически, поскольку $x(t)$ не может быть выражено в явном виде. Альтернативой является численное интегрирование выборок значений случайного процесса $x(t)$, наблюдаемых через равноотстоящие промежутки времени. Таким образом, если $x_1=x(\Delta t)$, $x_2=x(2\Delta t)$, ..., $x_n=x(n\Delta t)$, то оценка случайной величины $\hat{x}$ может быть представлена как

$$\tilde{M}[x]=\frac{1}{n}\sum_{i=1}^{n}x_i\,, \qquad [15]$$

$$M\left[\tilde{M}[x]\right]=M\left[\frac{1}{n}\sum_{i=1}^{n}x_i\right]=\frac{1}{n}\sum_{i=1}^{n}M[x_i]=M[x].$$

Следовательно, и в этом случае оценка параметра имеет математическое ожидание, равное его истинному значению.

Математическое ожидание погрешности оценки среднего равно

$$M\left[\Delta\tilde{M}\right]=M\left[\tilde{M}[x]-M[x]\right]=M\left[\frac{1}{n}\sum_{i=1}^{n}x_i\right]-M[x]=\frac{1}{n}nM[x]-M[x]=0. \tag{16}$$

Дисперсия погрешности оценки среднего равна

$$D[\Delta\tilde{M}]=D[\Delta\tilde{M}]=D[\tilde{M}]=D\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2}D\left(\sum_{i=1}^{n}x_i\right)=\frac{1}{n^2}nD[x]=\frac{D[x]}{n}=\frac{\sigma^2[x]}{n}. \tag{17}$$

Среднее квадратическое отклонение оценки математического ожидания (15)

$$\sigma[\tilde{M}] = \sigma[x]/\sqrt{n} \approx \tilde{\sigma}[x]/\sqrt{n}\,. \qquad (18)$$

Как видно из (16,17) оценка (15) – несмещенная, состоятельная и эффективная.

Объем наблюдений, требуемый для оценивания математического ожидания, определяется в соответствии с выражением

$$n_M \geq 2\left(\frac{\tilde{\sigma}[x]\,\Phi^{-1}(\beta)}{\varepsilon}\right)^2,$$

где

$$\tilde{\sigma}[x] = \sqrt{\frac{\sum_{i=1}^{n_M}(x - \tilde{M}\,[x])^2}{n_M - 1}}\,.$$

Учитывая тот факт, что в выражении фигурирует оценка СКО $\tilde{\sigma}[x]$, рассчитываемая на основе результатов наблюдений и зависящая от объема выборки n_M, для оценивания объема наблюдений n_M используется метод последовательных приближений.

Оценивание дисперсии случайной величины и ее среднего квадратического отклонения

Оценка дисперсии $\tilde{D}[x]$ как экспериментальное значение второго центрального момента случайной величины $x(t)$ может быть вычислена по формуле

$$\tilde{D}[x] = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - M[x]\right)^2. \qquad (19)$$

Так как значение $M[x]$ априори неизвестно, то принимают $M[x] \approx \tilde{M}[x]$ и тогда

$$\tilde{D}[x] = \frac{1}{n}\sum_{i}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i}^{n} x_i\right)^2. \qquad (20)$$

Математическое ожидание погрешности оценки равно

$$M\left[\Delta\tilde{D}\right] = M\left[\tilde{D}[x] - D[x]\right] = -\frac{D[x]}{n}, \qquad (21)$$

что означает, что оценка (19) является *смещенной.*

Смещение пропорционально $D[x]$ и обратно пропорционально n. Это означает, что оценка $\tilde{D}[x]$, полученная согласно (19), - *состоятельная.* Смещение устраняется с переходом к $\tilde{D}^*[x] = \dfrac{n}{n-1}\tilde{D}[x]$.

При этом вместо (19) имеем

$$\tilde{D}[x] = \frac{1}{n-1}\sum_{i}^{n} x_i^2 - \frac{n}{n-1}\tilde{M}^2[x]. \qquad (22)$$

При больших значениях n результаты расчета по формулам (19) и (22) практически будут одинаковыми.

Выражение для дисперсии оценки, равной дисперсии погрешности $\Delta\tilde{D}[x] = \tilde{D}[x] - D[x]$, при нормальном виде закона распределения $x(t)$ (для худшего случая) можно получить следующее:

$$D[\tilde{D}_x] = \frac{2}{n-1} D^2[x]. \quad (23)$$

Зависимость среднего квадратического отклонения $\sigma[\tilde{D}_x]$ от его точного значения $\sigma[x]$ определяется выражением

$$\sigma[\tilde{D}_x] = \sqrt{\frac{2}{n-1}}\sigma[x]. \quad (24)$$

Оценивание корреляционного момента и коэффициента корреляции

Экспериментальное значение корреляционного момента R_{xy} как оценка смешанного центрального момента m_{11} системы двух случайных величин равно

$$R_{xy} = \frac{1}{n}\sum_{1}^{n}(x_i - M_x)(y_i - M_y). \quad (25)$$

Так как значения M_x, M_y неизвестны, то принимают $M_x \approx \tilde{M}_x$, $M_y \approx \tilde{M}_y$ и тогда

$$\tilde{R}_{xy} = \frac{1}{n}\sum_{1}^{n}(x_i - \tilde{M}_x)(y_i - \tilde{M}_y) = \frac{1}{n}\sum_{1}^{n}x_i y_i - \left(\frac{1}{n}\sum_{1}^{n}x_i\right)\left(\frac{1}{n}\sum_{1}^{n}y_i\right). \quad (26)$$

Погрешность оценки $\tilde{R}_{xy}$

$$\Delta\tilde{R}_{xy} = \tilde{R}_{xy} - R_{xy}. \quad (27)$$

Математическое ожидание погрешности (27)

$$M\left[\Delta\tilde{R}_{xy}\right] = M\left[\frac{1}{n}\sum_{1}^{n}x_i y_i\right] - M\left[\left(\frac{1}{n}\sum_{1}^{n}x_i\right)\left(\frac{1}{n}\sum_{1}^{n}y_i\right)\right] = -\frac{R_{xy}}{n}$$

Это означает, что оценка (25) - смещена и равна

$$\tilde{R}_{xy} = \frac{n-1}{n}R_{xy}. \quad (28)$$

Можно показать, что она является и состоятельной. Смещение устраняется с переходом от $\tilde{R}_{xy}$ к $\tilde{R}^*_{xy} = \frac{n}{n-1}\tilde{R}_{xy}$. При этом вместо (25) имеем

$$\tilde{R}_{xy} = \frac{1}{n-1}\sum_{1}^{n}x_i y_i - \frac{n}{n-1}\tilde{M}_x\tilde{M}_y. \quad (29)$$

Для дисперсии оценки (25), равной дисперсии $D\left(\Delta\tilde{R}_{xy}\right)$ погрешности (27), можно получить

$$D_{\tilde{R}_{xy}} = \frac{\left(\mu_{2,2} + R_{xy}^2\right)n}{(n-1)^2} - \frac{2\mu_{2,2} - 3R_{xy}^2 - D_x D_y}{(n-1)^2} + \frac{2\mu_{2,2} - 2R_{xy}^2 - D_x D_y}{n(n-1)^2}, \qquad (30)$$

где $\mu_{2,2}$ - четвертый смешанный центральный момент системы ($x(t)$, $y(t)$). При $x(t) = y(t)$ выражения (29) и (30) превращаются в (22), (23). Если система ($x(t)$, $y(t)$) распределена нормально, то $\mu_{2,2} = 2R_{xy}^2 - D_x D_y$ и

$$D_{\tilde{R}_{xy}} = \left(R_{xy}^2 + D_x D_y\right)/(n-1). \qquad (31)$$

Так как значения R_{xy}, D_x, D_y неизвестны, то практически используется приближение

$$D_{\tilde{R}_{xy}} \approx \left(\tilde{R}_{xy}^2 + D_x D_y\right)/(n-1). \qquad (32)$$

Среднее квадратическое значение погрешности (27) равно:

$$\sigma_{\tilde{R}_{xy}} \approx \sqrt{\left(\tilde{R}_{xy}^2 + D_x D_y\right)/(n-1)}. \qquad (33)$$

Оценка коэффициента корреляции определяется согласно

$$\tilde{r}_{xy} = \tilde{R}_{xy} / \left(\tilde{\sigma}_x \tilde{\sigma}_y\right). \qquad (34)$$

Объем наблюдений, требуемый для оценивания дисперсии и коэффициента корреляции случайной величины x, рассчитывается по формуле вида

$$n_D \geq 4\left(\frac{\tilde{D}[x]\,\Phi^{-1}(\beta)}{\varepsilon}\right)^2 + 1.$$

Следует отметить, что в этом выражении фигурирует оценка дисперсии $\tilde{D}[x]$, получаемая на основе результатов наблюдений. Так как она зависит от объема обучающей выборки n, то объем n_D определяется методом последовательных приближений и равен наименьшему из значений, обеспечивающих выполнение неравенства.

2. Особенности многомерного статистического анализа данных

Решение задач многомерного статистического анализа предусматривает формирование описаний анализируемого процесса, представленного случайными наборами признаков, в виде совместного закона распределения этих признаков с соответствующими параметрами. Причем в случае одномерной нормально распределенной случайной величины x плотность вероятности может быть представлена следующим выражением:

$$\omega(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}. \qquad (35)$$

Основными параметрами указанного закона распределения являются среднее значение M и дисперсия D, поэтому часто его представляют в виде функции $N(M, D)$.

Образы, характеризующиеся нормальным распределением, проявляют тенденцию к группировке вокруг среднего значения, а их рассеяние − пропорционально среднеквадратическому отклонению σ. Около 95 % объектов, извлеченных из совокупности с нормальным распределением, попадут в интервал, равный 2σ и имеющий в качестве центра среднее значение *M*.

При этом вероятность попадания реализации случайной величины *x* в интервал $[\alpha,\beta]$ равна:

$$P(\alpha < x \le \beta) = \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(\frac{-\left(x - M_x\right)^2}{2\sigma_x^2} \right) dx . \qquad (36)$$

При замене переменной $t = \frac{x - M_x}{\sqrt{\sigma_x^2}}$ вычисление интеграла сводится к вычислению табулированного интеграла вероятности:

$$P(\alpha < x \le \beta) = [\Phi(B) - \Phi(A)],$$

$$A = \frac{\alpha - M_x}{\sqrt{\sigma_x^2}}; B = \frac{\beta - M_x}{\sqrt{\sigma_x^2}}; \qquad (37)$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \Phi(x) = -\Phi(-x).$$

Рассмотрим нормальный закон распределения двух случайных величин $\left(x_1, x_2\right)$

$$\omega(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r_{12}^2}} \exp\left[-\frac{1}{2(1 - r_{12}^2)} \left(\frac{\left(x_1 - M_1\right)^2}{\sigma_1^2} + \frac{\left(x_2 - M_2\right)^2}{\sigma_2^2} - 2r_{12} \frac{\left(x_1 - M_2\right)}{\sigma_1} \frac{\left(x_2 - M_2\right)}{\sigma_2} \right) \right].$$

Можно показать, что закон определяется следующими пятью параметрами: M_1, M_2 - математическими ожиданиями; σ_1, σ_2 - средне квадратическими отклонениями и r_{12} - коэффициентом корреляции величин x_1 и x_2. В случае $r_{12} = 0$ (некоррелированности компонентов) совместная плотность $\omega(x_1, x_2)$ может быть представлена произведением одномерных нормальных распределений вероятностей значений независимых компонентов x_1 и x_2.

Рассмотрим общий случай, когда одному классу сигналов принадлежит совокупность выборок случайных величин. Тогда реализации сигналов этого класса описываются *l*-мерными случайными величинами x_1, x_2, ..., x_l. Совокупность *l* случайных величин будем обозначать вектором или матрицей-столбцом $x = \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_l \end{bmatrix}$. Будем рассматривать случайные величины, которые имеют *l*-мерную нормальную плотность распределения:

$$\omega(\vec{x}) = \frac{1}{(2\pi)^{l/2} \left|C\right|^{1/2}} e^{-\frac{1}{2}\left(\vec{x} - M[\vec{x}]\right)^T C^{-1} \left(\vec{x} - M[\vec{x}]\right)},$$

где $\left|C\right|$ – определитель ковариационной матрицы C;

$(\vec{x} - M[\vec{x}])^T$ – транспонированная матрица $(\vec{x} - M[\vec{x}])$.

Каждая условная плотность распределения полностью определена вектором средних значений $M[\vec{x}]$ и ковариационной матрицей C. При этом элементы ковариационной матрицы и ковариационная матрица в целом определяются следующим образом:

$$c_{jk} = \operatorname{cov}(x_j x_k) = M[(x_j - M[x_j])(x_k - M[x_k])],$$

где j,k=1, …, l – номер случайной величины;

$$C = M[(\vec{x} - M[\vec{x}])(\vec{x} - M[\vec{x}])^T],$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1l} \\ c_{21} & c_{22} & \dots & c_{2l} \\ \dots & & & \\ c_{l1} & c_{l2} & \dots & c_{ll} \end{bmatrix}.$$

Ковариационная матрица C является симметрической и положительно полуопределенной. Ее диагональный элемент c_{kk} есть дисперсия k-го элемента вектора образов. Элемент c_{jk}, не стоящий на диагонали матрицы, представляет собой ковариацию случайных переменных x_j и x_k (*ковариация* – смешанный момент второго порядка $\operatorname{cov}(x_j, x_k)$). Если переменные x_j и x_k независимы, то элемент $c_{jk} = 0$. Многомерная плотность нормального распределения сводится к произведению одномерных плотностей нормальных распределений, если все недиагональные элементы ковариационной матрицы равны нулю.

Заключительная часть.

В заключение необходимо отметить, что все используемые в настоящее время методы анализа случайных последовательностей не выходят за рамки представленного общего подхода, однако в некоторых случаях позволяют существенно упростить процедуры их классификации, если учитывают специфику анализируемых последовательностей.

Рекомендованная литература:

1. Иоффе А.Я., Петухов Г.Б., Миров Ю.Н. Лекции по математической статистике. – Л.: ВИКА, 1970.

2. Статистические методы обработки результатов наблюдений / под ред. Р.М. Юсупова. – Л.: МО СССР, 1984.