



ЭКОНОМЕТРИКА

Лекция 4

Необходимые сведения из
теории вероятности

Количественные характеристики случайных переменных

Математическое ожидание (среднее значение)

Дисперсия и среднее квадратическое отклонение

Ковариация и коэффициент корреляции

Математическое ожидание дискретной случайной переменной

Определение. Математическим ожиданием дискретной случайной переменной называется величина:

$$M(x) = \sum_{i=1}^n P_i x_i \quad (4.1)$$

где: $M(x)$ – математическое ожидание СДП x ,

P_i - вероятность появления в опытах значения x_i ,

x_i - значение дискретной случайной переменной,

n - количество допустимых значений дискретной случайной величины

Математическое ожидание – средневзвешенное значение ДСП, где в качестве веса используется значение вероятности

Дисперсия дискретной случайной переменной

Определение. Дисперсией дискретной случайной переменной называется величина:

$$\sigma^2(x) = \sum_{i=1}^n (x_i - M(x))^2 P(x_i) \quad (4.2)$$

где: $\sigma^2(x)$ – дисперсия случайной переменной x

Дисперсия случайной величины выступает в качестве характеристики разброса возможных ее значений

Положительный корень из дисперсии называют средним квадратическим отклонением или стандартным отклонением, или стандартной ошибкой

Примеры расчета количественных характеристик ДСП

Пример 1. Пусть X_i – результат бросания кубика.

$$A_x = \{1, 2, 3, 4, 5, 6\} \quad P_i = \{1/6, 1/6, 1/6, 1/6, 1/6, 1/6\}$$

Тогда:

$$M(x) = 1/6(1+2+3+4+5+6) = 3.5$$

$$\sigma^2(x) = 1/6[(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2] = 2.92$$

$$\sigma(x) = 1.71$$

Примеры расчета количественных характеристик ДСП

Пример 2. Индикатор случайного события

$$I = \begin{cases} 1 & \text{если событие произошло} \\ 0 & \text{если событие не произошло} \end{cases} \quad P_i(t) = \begin{cases} p & \text{если } t = 1 \\ (1-p) & \text{если } t = 0 \end{cases}$$

Математическое ожидание

$$M(I) = \sum_{j=1}^n p_j I_j = p \cdot 1 + (1-p) \cdot 0 = p$$

Дисперсия

$$\sigma^2(I) = \sum (I_j - M(I))^2 p_j = (1-p)^2 p + (0-p)^2 (1-p) = p(1-p)$$

6

Математическое ожидание непрерывной случайной переменной

Определение. Математическим ожиданием непрерывной случайной величины X с законом распределения $p_x(t)$ называется величина:

$$M(x) = \int_{-\infty}^{\infty} t p_x(t) dt \quad (4.3)$$

Выражение (4.3) называется первым начальным моментом функции $p_x(t)$

Через результаты наблюдений математическое ожидание вычисляется как:

$$M(x) = \frac{1}{n} \sum_{i=1}^n x_i$$

Дисперсия непрерывной случайной переменной

Определение. Дисперсией непрерывной случайной переменной X с функцией плотности вероятности $p_x(t)$ называется выражение:

$$\sigma^2(x) = \int_{-\infty}^{\infty} (t - M(x))^2 p_x(t) dt \quad (4.4)$$

Выражение (4.4) называют вторым центральным моментом функции $p_x(t)$

В общем случае дисперсия случайной переменной определяется как:

$$\sigma^2(x) = M(x - M(x))^2 \quad (4.5)$$

Дисперсия непрерывной случайной переменной

Часто применяется другая формула для вычисления дисперсии

$$\sigma^2(x) = M(x - M(x))^2 = M(x^2 - 2xM(x) + M^2(x)) = M(x^2) - M^2(x)$$

Экспериментальное значение дисперсии может быть вычислено по формуле

$$\sigma^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - M(x))^2$$

Примеры вычисления

Пример 1. Пусть X НСП с равномерным законом распределения.

$$M(x) = \int_a^b t \frac{1}{(b-a)} dt = \frac{1}{(b-a)} \frac{t^2}{2} = \frac{(b+a)}{2}$$

$$\sigma^2(x) = \int_a^b \left(t - \frac{(a+b)}{2} \right)^2 \frac{1}{(b-a)} dt = \frac{1}{3(b-a)} \left(t - \frac{(a+b)}{2} \right)^3 = \frac{(b-a)^2}{12}$$

Самостоятельно вычислить математическое ожидание и дисперсию НСП с нормальным законом распределения

Понятие ковариации двух случайных переменных

По определению ковариацией двух случайных переменных X и Y есть:

$$\text{COV}(x, y) = M((x - M(x))(y - M(y))) \quad (4.6)$$

Значение ковариации отражает наличие связи между двумя случайными переменными

Если $\text{COV}(x, y) > 0$, связь между X и Y положительная

Если $\text{COV}(x, y) < 0$, связь между X и Y отрицательная

Если $\text{COV}(x, y) = 0$, X и Y независимые переменные

Область возможных значений ковариации – вся числовая ось

Понятие коэффициента корреляции двух случайных переменных

Недостатки ковариации в том, что ее значения зависят от масштаба измерения переменных и наличия размерности

Недостатки устраняется путем деления значения ковариации на значения стандартных отклонений переменных:

$$\rho(x, y) = \frac{\text{COV}(x, y)}{\sigma(x)\sigma(y)} \quad (4.7)$$

Выражение (4.7) называют коэффициентом корреляции двух случайных переменных

Коэффициент корреляции изменяется в пределах $[-1; 1]$ и является безразмерной величиной

Основные свойства количественных характеристик

Свойства математического ожидания

$$M(c) = c$$

$$M(c_1x + c_2y) = c_1M(x) + c_2M(y)$$

Пример

$$y = f(x) + u$$

$$M(u|x) = 0$$

$$\sigma^2(u|x) = \sigma_u^2$$

$$M(y) = M(f(x)) + M(u) = f(x)$$

Основные свойства количественных характеристик

2. Свойства дисперсий

$$\sigma^2(c) = 0$$

$$\sigma^2(cX) = c^2 \sigma^2(X)$$

$$\sigma^2(c_1X + c_2Y) = c_1^2 \sigma^2(X) + c_2^2 \sigma^2(Y) + 2c_1c_2 \text{COV}(X, Y)$$

В общем случае

$$\sigma^2\left(\sum_{i=1}^n c_i X_i\right) = C^T \text{COV}(X^T X) C$$

где

$$C = \{c_1, c_2, \dots, c_n\}^T \quad X = \{X_1, X_2, \dots, X_n\}^T$$

Основные свойства количественных характеристик

Свойства ковариаций

$$\text{Cov}(x, y) = \text{Cov}(y, x)$$

$$\text{Cov}(c_1 x_1 + c_2 x_2) = c_1 c_2 \text{Cov}(x_1, x_2)$$

$$\text{Cov}(cx) = 0$$

$$\text{Cov}(x+c, y) = \text{Cov}(x, y)$$

$$\text{Cov}(x+y, z) = \text{Cov}(x, z) + \text{Cov}(y, z)$$

$$\text{Cov}(x, x) = \sigma^2(x)$$

Доказательства этих свойств проведите самостоятельно!

Случайный вектор и его характеристики

Пусть опыт – инвестирование средств на некоторый период времени в рискованные активы $A = \{a_1, a_2, \dots, a_n\}$

Рисковый характер актива означает, что значения доходности на них являются случайными величинами $r(a_1), r(a_2), \dots, r(a_n)$

Определение. Вектор, компонентами которого являются случайные величины, называется случайным вектором

Пример 1. Вектор доходностей по рискованным активам

$$\mathbf{R} = (r_1, r_2, \dots, r_n)^T \quad (4.8)$$

Пример 2. Опыт – бросание игральной кости.

Пусть X – количество очков на верхней грани кости, а Y – количество очков на его нижней грани

Тогда вектор $Z = \{X, Y\}^T$ – пример случайного вектора

Случайный вектор и его характеристики

Пусть $m_i = M(r(a_i))$ – ожидаемое значение доходности актива a_i ,

$\sigma_i^2 = M(r(a_i) - m_i)^2$ – дисперсия доходности актива a_i ,

$\sigma_{ij} = \text{Cov}(r(a_i), r(a_j))$ – ковариация между активами a_i, a_j .

Тогда вектор

$$\overset{\vee}{M} = (m_1, m_2, \dots, m_n) = M(\overset{\vee}{R}) \quad (4.9)$$

является первой основной характеристикой случайного вектора (4.8)

Замечание. Вектор M является константой

Ковариационная матрица

$$\sigma_{rr} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix}$$

является второй основной характеристикой случайного вектора R

Случайный вектор и его характеристики

По предложению Марковца компоненты вектора R рассматривается как характеристики привлекательности каждого рискованного актива, а диагональные элементы ковариационной матрицы – как характеристики риска инвестирования в эти активы

Параметрической моделью Марковца называется следующая тройка:

$$\{A, M, \sigma_{rr}\} \quad (4.10)$$

Для формирования индивидуального пакета акций из списка A ничего больше не требуется

Эта модель является инструментом брокерской деятельности

Основные понятия математической статистики

Задачи математической статистики

1. Оценивание (приближенное определение) параметров законов распределения и самих законов
2. Проверка различных гипотез относительно законов распределения или значений их параметров

Далее будем рассматривать случайные величины с законом распределения $R(t, a_1, a_2, \dots, a_n)$, где $A = \{a_1, a_2, \dots, a_n\}^T$ вектор столбец параметров распределения

Выборка и ее свойства

Определение. Выборка – это случайный вектор, составленный из результатов наблюдений, каждое из которых суть независимая случайная величина

Пусть y_1, y_2, \dots, y_n результаты наблюдения за поведением случайной величины Y с законом распределения $P_y(t, A)$ Тогда выборка есть вектор, собранный из результатов наблюдений $Y = (y_1, y_2, \dots, y_n)^T$

Каждый элемент выборки есть случайная величина и, следовательно, имеет свой закон распределения

$$\begin{aligned} &P_y(y_1, a_1, a_2, \dots, a_k) \\ &P_y(y_2, a_1, a_2, \dots, a_k) \\ &\dots\dots\dots \\ &P_y(y_n, a_1, a_2, \dots, a_k); \end{aligned}$$

Выборка и ее свойства

Свойства случайной выборки

1. Каждый элемент выборки есть случайная величина с тем же законом распределения, что и случайная величина Y
2. Все значения, входящие в выборку независимые величины

Тогда для них справедлива теорема умножения вероятностей:

$$P_y(y_1, y_2, \dots, y_n | A) = P_y(t_1, A) P_y(t_2, A) \dots P_y(t_n, A)$$

Это выражение – закон распределения выборки

Задача заключается в том, чтобы найти процедуры, с помощью которых можно найти значения параметров распределения.

$$A = F(y_1, y_2, \dots, y_n)$$

Свойства оценок параметров распределения

Оценка представляет собой частный случай случайной величины

Например. Рассмотрим оценку математического ожидания в виде среднего значения:

$$\tilde{\bar{X}} = \frac{1}{n} \sum_{i=1}^n x_i \quad (4.11)$$

Замечание

Любую случайную величину можно представить в виде:

$$X_i = \mu + U_i$$

где: U_i – случайная величина

μ – константа равная математическому ожиданию X_i

$$\tilde{\bar{X}} = \frac{1}{n} \sum_{i=1}^n (\mu + u_i) = \mu + \bar{u}$$

Свойства оценок параметров распределения

1. Несмещенность оценки

$$M(\tilde{a}) = M(a) \quad (4.12)$$

Процедуры, которые дают такие оценки будем называть несмещенными

Замечание. Несмещенных процедур может быть много

Пример. Рассмотрим процедуру оценки математического ожидания

$$\tilde{X} = \frac{1}{n} \sum_{i=1}^n x_i$$

Эта процедура несмещенная т.к

$$M(\tilde{X}) = M(\mu + u_i) = \mu + M(u) = \mu$$

Свойства оценок параметров распределения

Вопрос. Можно ли найти иную несмещенную процедуру?

Пусть имеем выборку наблюдений за случайной величиной X с законом распределения $P_x(t)$ из двух значений x_1 и x_2 , следовательно для нее справедливо:

$$M(x_1) = M(x_2) = \mu$$

$$\sigma^2(x_1) = \sigma^2(x_2) = s^2$$

Пусть такой процедурой будет: $Z = \lambda_1 x_1 + \lambda_2 x_2$

Тогда

$$M(\tilde{Z}) = M(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 M(x_1) + \lambda_2 M(x_2) = (\lambda_1 + \lambda_2)\mu$$

Вывод. Все процедуры, для которых $\lambda_1 + \lambda_2 = 1$ дают несмещенные оценки среднего значения.

Свойства оценок параметров распределения

2. Эффективность оценки

Определение. Оценка называется эффективной среди всех оценок параметра, если она имеет минимальную дисперсию среди всех возможных оценок: $\sigma^2(\tilde{a}) = \min$

Задача. При каких значениях λ_1 и λ_2 оценка среднего значения будет эффективной?

Найдем при каких значениях λ_1 и λ_2 достигается минимум дисперсии оценки Z

$$\sigma^2(\tilde{Z}) = \sigma^2(\lambda_1 x_1 + \lambda_2 x_2) = (\lambda_1^2 + \lambda_2^2) s^2$$

Учитывая, что $(\lambda_1 + \lambda_2) = 1$ или $\lambda_2 = (1 - \lambda_1)$, получим:

$$\sigma^2(\tilde{Z}) = (\lambda_1^2 + (1 - \lambda_1)^2) s^2 \quad (4.13)$$

Свойства оценок параметров распределения

Тогда для нахождения минимума выражения (4.13) составляем уравнение

$$\frac{\partial \sigma^2(\tilde{Z})}{\partial \lambda_1} = (2\lambda_1 - 2(1 - \lambda_1))s = 0$$

Откуда следует, что $\lambda_1 = 1/2$

Вторая производная положительна, следовательно, это минимум

Вывод. Оценка (4.11) является несмещенной и эффективной

Аналогичным образом можно показать, что известная оценка дисперсии также не смещена и эффективна

Свойства оценок параметров распределения

Определение. Оценка, достигающая выполнения условий несмещенности и эффективности вне зависимости от объема выборки называется несмещенной и эффективной

Определение. Оценка, достигающая выполнения условий несмещенности и эффективности при неограниченном увеличении объема выборки называется **ассимптотически несмещенной и эффективной**

Определение. Оценка, достигающая выполнения условий несмещенности при неограниченном увеличении объема выборки называется **состоятельной**

Основные понятия математической статистики

Выводы:

1. Методами математической статистики удастся получить оценки параметров законов распределения (моделей)
2. Наилучшими считаются оценки, обладающие свойствами несмещенности и эффективности
3. На практике принимаются оценки, удовлетворяющие свойству состоятельности