



**Центральный банк
Республики Узбекистан**

**Краткосрочные прогнозы инфляции:
подробное изложение применения
байесовского авторегрессионного
подхода (BVAR)**

Июль, 2021 год

Содержание

I. Актуальность исследования	2
II. Введение в модель	4
III. Внедрение модели для Узбекистана и прогнозы	8
IV. Вопрос оценки параметров модели	11
V. Применение метода Байеса	12
VI. Определение апостериорного распределения	16
VII. Генератор выборки Гиббса	25
VIII. Априор Миннесоты	29
Приложение 1. Основы матричной алгебры	32
Приложение 2. Основы теории вероятности	40
Приложение 3. Теоремы пределов	55
Приложение 4. Некоторые важные распределения вероятности.	60
Приложение 5. Цепь Маркова	64
Приложение 6. Метод симуляций Монте Карло	67
Список литературы	70

I. Актуальность исследования

Для эффективной реализации режима инфляционного таргетирования, решения по денежно-кредитной политике принимаются на основе среднесрочного прогноза инфляции, который формируется исходя из краткосрочных прогнозов на горизонт 2-3 квартала.

На сегодняшний день используются краткосрочные прогнозы инфляции, выведенные с помощью следующих прогнозных моделей:

ARIMA (*англ.* autoregressive integrated moving average) – интегрированная модель авторегрессии скользящего среднего;

DFM (*англ.* dynamic factor model) – динамическая факторная модель;

FAVAR (*англ.* factor augmented vector autoregressive) – факторно-расширенная векторная авторегрессионная модель.

В ряд этих моделей планируется включить краткосрочные модели **BVAR** и **SVAR**, которые находятся на стадии совершенствования и разработки соответственно.

Каждый из вышеприведенных подходов имеет свои преимущества и недостатки, которые дополняют друг друга.

На практике трудно найти идеальную модель, которая полностью и наиболее точно описывает фактические и будущие тенденции в экономике, поэтому в целях обеспечения высокой точности прогнозов большинство центральных банков используют несколько моделей одновременно.

К примеру, **Центральный банк Норвегии** для краткосрочного прогнозирования инфляции использует различные модели VAR, AR, векторную модель коррекции ошибок (VECM) и факторные модели (Aastveit, Gerdrup и Jore, 2011). При этом, для краткосрочного прогнозирования инфляции основным критерием выбора моделей является степень их соответствия реальным данным, а также точность вневыборочных прогнозов.

Банк России применяет комбинированный метод краткосрочного прогнозирования инфляции¹, который даёт возможность использовать большой объем данных. Так, для каждого субиндекса ИПЦ (индекс потребительских цен) строятся прогнозы с использованием различных моделей (VAR, BVAR², RW, LSTAR, UC) и каждая из моделей дает свой обобщенный прогноз инфляции. Затем для получения финального прогноза инфляции прогнозы, полученные из различных моделей агрегируются с весами, пропорциональными их точности.

Среди прочих моделей краткосрочного прогнозирования инфляции **байесовские векторные авторегрессионные модели (BVAR)** стали широко используемыми инструментами для прогнозирования и анализа инфляции.

Так, в исследовании, проведенном **Центральным банком Турецкой Республики** оценивается эффективность прогнозирования инфляции с помощью BVAR³ согласно альтернативным спецификациям. Результаты показывают, что BVAR даёт точные точечные прогнозы при использовании переменных в форме логарифмической разности. В качестве переменных выбраны цена нефти, импортные цены в долларах, обменный курс, разрыв выпуска, реальные затраты на рабочую силу, инфляционные ожидания и ИПЦ.

Вместе с тем, департаментом исследований **Центрального банка Албании** опубликован рабочий документ⁴, в котором проводится сравнительный анализ BVAR и VAR моделей. Результаты анализа показали, что BVAR позволяет охватывать больше эндогенных переменных и решает проблемы чрезмерной параметризации, присущие стандартной модели VAR.

¹Прогнозирование инфляции методом комбинирования прогнозов в Банке России. Серия докладов об экономических исследованиях. Август 2016. Андрей Андреев.

²Моделирование Российской экономики с помощью BVAR. Гайдаровский форум. 2018. Ломиворотов Родион.

³A Bayesian VAR Approach to Short-Term Inflation Forecasting. Central Bank of the Republic of Turkey 2019.

⁴Forecasting the Albanian short-term inflation through a Bayesian VAR model. Bank of Albania 2019.

Таким образом, расширение портфеля инструментов краткосрочного прогнозирования инфляции, включая модель BVAR даёт возможность всестороннего анализа инфляции и совершенствования модельного инструментария Центрального банка.

II. Введение в модель

BVAR (байесовская векторная авторегрессия) использует метод Байеса для оценки векторной авторегрессии (VAR). Для глубокого понимания сущности BVAR модели необходимо изучение исходной – VAR модели.

Векторная авторегрессия (далее VAR) является статистической моделью, которая используется для определения связи между несколькими переменными величинами посредством их изменения во времени. VAR является обобщением авторегрессионной модели (AR) с одной зависимой переменной поскольку охватывает многомерные временные ряды. Как и в случае с моделью авторегрессии, для каждой переменной будет существовать уравнение, моделирующее ее формирование с течением времени. Это уравнение состоит из предыдущих значений наблюдаемой переменной (лаги⁵), лагов других переменных и ошибок.

Например, для оценки взаимозависимости двух переменных, допустим инфляции (π) и обменного курса (s) с одним лагом, записывается следующая система уравнений:

$$\begin{cases} \pi_t = c_1 + a_{1,1} \cdot \pi_{t-1} + a_{1,2} \cdot s_{t-1} + \varepsilon_{1,t} \\ s_t = c_2 + a_{2,1} \cdot \pi_{t-1} + a_{2,2} \cdot s_{t-1} + \varepsilon_{2,t} \end{cases}$$

где, t – период времени, c_i - константа (intercept), $a_{i,j}$ – параметры модели, $\varepsilon_{i,t}$ - ошибки (error terms).

⁵Лаг (от *англ.* “lag” – “запаздывание”) – это временной показатель, отражающий отставание или опережение во времени одного явления. Например, если инфляцию в текущем году обозначить π_t , то в годовом выражении инфляция предыдущего года обозначается π_{t-1}

Вышеприведенную систему уравнений можно преобразовать в одно уравнение, используя действия с матрицами⁶:

$$\begin{pmatrix} \pi_t \\ s_t \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} \pi_{t-1} \\ s_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}. \quad (1)$$

Введем следующие обозначения:

$$Y_t = \begin{pmatrix} \pi_t \\ s_t \end{pmatrix}, \quad c = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}, \quad \varepsilon_t = \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}.$$

Тогда, уравнение (1) можно будет записать в линейном выражении:

$$Y_t = c + A \cdot Y_{t-1} + \varepsilon_t.$$

Это и есть модель векторной авторегрессии порядка 1 – VAR (1), то есть с одним лагом. Данную модель можно дополнить ещё другими переменными и их лагами.

Например, модель VAR с n количеством зависимых (эндогенных), d количеством независимых (экзогенных) переменных и p -ым количеством лагов можно записать в следующей **обобщенной форме**:

$$Y_t = c + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + D z_t + \varepsilon_t, \quad t = p + 1, \dots, S. \quad (2)$$

где

Y_t – $n \times 1$ размерный вектор значений зависимых переменных в периоде t ,

Y_{t-i} – значения переменной Y_t за предыдущий i -ый период времени; “ i -лаг” ($i = 1, \dots, p$),

z_t – $d \times 1$ размерный вектор значений независимых переменных в периоде t ,

A_i и D матрицы параметров размерами $n \times n$ и $n \times d$ соответственно,

S – количество наблюдений.

⁶Смотрите: Приложение 1 определения 6-8.

Для удобства введем такие обозначения, как $T = S - p$.

Матричный вид этого уравнения (*вид без экзогенных переменных*):

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} + \begin{pmatrix} a_{1,1}^1 & a_{1,2}^1 & \cdots & a_{1,n}^1 \\ a_{2,1}^1 & a_{2,2}^1 & \cdots & a_{2,n}^1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}^1 & a_{n,2}^1 & \cdots & a_{n,n}^1 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{n,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} a_{1,1}^p & a_{1,2}^p & \cdots & a_{1,n}^p \\ a_{2,1}^p & a_{2,2}^p & \cdots & a_{2,n}^p \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1}^p & a_{n,2}^p & \cdots & a_{n,n}^p \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{n,t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix}.$$

При записи в виде отдельных **уравнений регрессии**: (*вид без экзогенных переменных*):

$$\begin{aligned} y_{1,t} &= c_1 + a_{1,1}^1 y_{1,t-1} + a_{1,2}^1 y_{2,t-1} + \cdots + a_{1,n}^1 y_{n,t-1} + \cdots + a_{1,1}^p y_{1,t-p} + a_{1,2}^p y_{2,t-p} + \cdots + a_{1,n}^p y_{n,t-p} + \varepsilon_{1,t} \\ y_{2,t} &= c_2 + a_{2,1}^1 y_{1,t-1} + a_{2,2}^1 y_{2,t-1} + \cdots + a_{2,n}^1 y_{n,t-1} + \cdots + a_{2,1}^p y_{1,t-p} + a_{2,2}^p y_{2,t-p} + \cdots + a_{2,n}^p y_{n,t-p} + \varepsilon_{2,t} \\ &\vdots \\ y_{n,t} &= c_n + a_{n,1}^1 y_{1,t-1} + a_{n,2}^1 y_{2,t-1} + \cdots + a_{n,n}^1 y_{n,t-1} + \cdots + a_{n,1}^p y_{1,t-p} + a_{n,2}^p y_{2,t-p} + \cdots + a_{n,n}^p y_{n,t-p} + \varepsilon_{n,t}. \end{aligned}$$

Уравнение (2) можно записать в более краткой форме:

$$Y_t = X_t \beta + \varepsilon_t \quad (3)$$

где⁷:

$X_t = W_{t-1} \otimes I_n$ и его размер $n \times [n(np + d + 1)]$,

$W_{t-1} = (1 \quad Y'_{t-1} \quad Y'_{t-2} \quad \cdots \quad Y'_{t-p} \quad z'_t)$ и его размер $1 \times [np + d + 1]$,

$\beta = \text{vec}(c, A_1, A_2, \cdots, A_p, D)$ и его размер $[n(np + d + 1)] \times 1$.

Неизвестные параметры модели: β – вектор коэффициентов и Σ – ковариационная матрица Y_t ⁸.

Данная модель предложена Кристофером Симсом как альтернатива системам одновременных уравнений. Особенностью VAR моделей считается, что все переменные являются зависимыми друг от друга в отличие от стандартной регрессионной модели, где всегда имеется независимая переменная. VAR является широко используемой моделью, преимущество которой заключается в простоте и гибкости управления данными.

⁷ Смотрите: Приложение 1 определения 10 и 11.

⁸ Смотрите: Приложение 2 определение 17.

Вектор ошибок в модели VAR должен удовлетворять следующим условиям:

1. $E[\varepsilon_t] = 0$ — средняя ошибка за период равна нулю, или, точнее, $E[\varepsilon_t|X_t] = 0$, то есть сумма ошибок при любом значении (X_t) равна нулю (*Strict exogeneity*). Другими словами, данные независимой переменной (X_t) за любой период, не дают никакой информации о средних значениях ошибок ($E[\varepsilon_t]$). Исходя из этого, берем математическое ожидание с обеих сторон уравнения (3) и получаем следующее отношение (обратите внимание, $X_t\beta$ константа, из-за того, что ε_t случайная величина, зависимая переменная величина Y_t будет производной случайной величиной, полученной посредством суммирования $X_t\beta$ к значениям зависимой переменной величины ε_t ⁹):

$$E[Y_t] = E[X_t\beta + \varepsilon_t] = X_t\beta + E[\varepsilon_t] = X_t\beta. \quad (4)$$

2. $E[\varepsilon_t\varepsilon_t'] = V[Y_t]$, то есть математическое ожидание матрицы $\varepsilon_t\varepsilon_t'$ будет ковариационной матрицей случайного вектора Y_t , поскольку если найдем вектор ошибок из уравнения (3):

$$\varepsilon_t = Y_t - X_t\beta.$$

Тогда выражение (4) согласно 1-свойству определения 17 приложения 2:

$$\begin{aligned} E[\varepsilon_t\varepsilon_t'] &= E[(Y_t - X_t\beta)(Y_t - X_t\beta)'] \\ &= E[(Y_t - E[Y_t])(Y_t - E[Y_t])'] = V[Y_t] = \Sigma. \end{aligned}$$

Напомним, что любая ковариационная матрица является симметричной и положительно - полуопределенной матрицей¹⁰ (*No perfect multicollinearity*).

3. $E[\varepsilon_t\varepsilon_{t-n}'] = 0$ — между ошибками корреляция отсутствует, то есть $Cov(\varepsilon_{i,t}, \varepsilon_{j,t-n}) = 0$ ¹¹ (*No autocorrelation*). Другими словами, информация

⁹Для подробной информации: <http://web.stanford.edu/class/polisci100a/regress2>. Section 3.

¹⁰Смотрите: Приложение 1 определение 16 и приложение 2 определение 17, 2-свойства.

¹¹Смотрите: Приложение 2 определение 14.

об ошибках одного наблюдения не даёт никакой информации о других ошибках произвольного наблюдения.

Согласно указанным выше условиям и центральной предельной теореме¹² вектор ошибок подчиняется закону¹³ многомерного нормального распределения со следующими параметрами:

$$\varepsilon_t \sim N(0, \Sigma).$$

III. Внедрение модели для Узбекистана и прогнозы

Описание данных

Для построения квартальной векторной модели авторегрессии были использованы 4 макроэкономических показателя в качестве эндогенной переменной: инфляция, реальный рост ВВП, средневзвешенная процентная ставка по межбанковским депозитам и изменение обменного курса сума к доллару США. Данные охватывают периоды с I квартала 2010 года по I квартал 2021 года. Источники данных: Государственный комитет Республики Узбекистан по статистике и Центральный банк Республики Узбекистан.

Оценка параметров VAR модели¹⁴

В данной работе все расчеты проведены с помощью программы MATLAB®. При разработке модели в качестве порядка VAR выбраны 5 лагов. Прежде чем оценить модель, данные по инфляции были очищены от сезонности с помощью стандартного метода X13-ARIMA-SEATS. Параметры модели оценены с помощью генератора выборки Гиббса (семплирование по Гиббсу). При этом в качестве априорного распределения коэффициентов модели использовалась априорное распределение Миннесоты. Общее количество повторов генератора выборки Гиббса составило 25000, при этом результаты

¹²Смотрите: Приложение 3.

¹³Смотрите: Приложение 2 определение 18.

¹⁴Методология оценки параметров модели рассмотрены в следующих разделах.

последних 2500 повторов использовались для формирования распределения параметров модели и извлечения прогнозов.

Результаты

При оценке VAR модели с применением Байесовского метода получена следующая прогнозная траектория для уровня инфляции на предстоящий год.

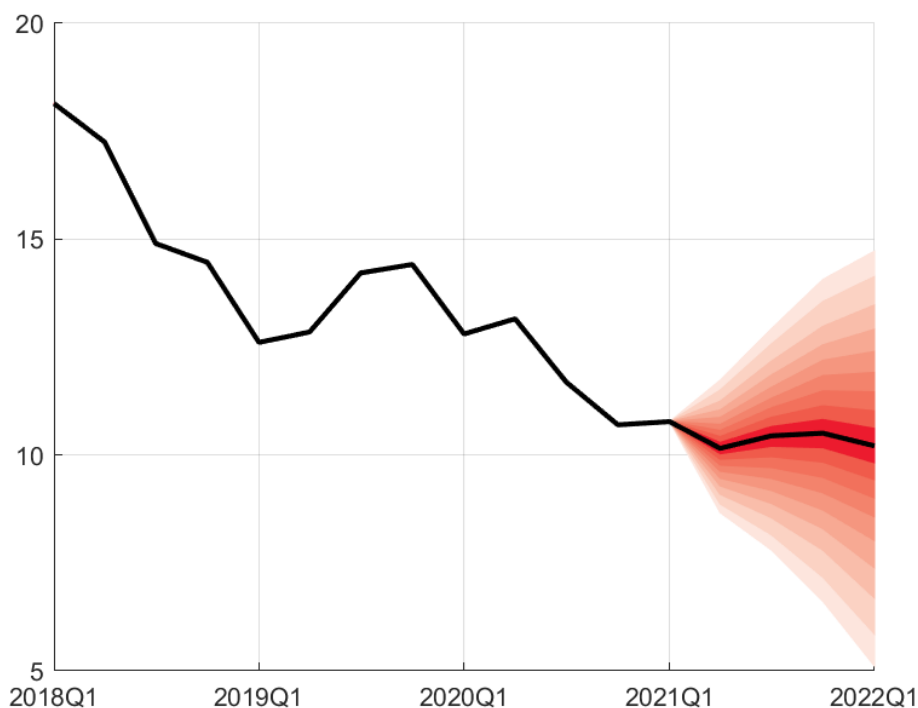


Рис. 1: Прогноз инфляции

При этом, уровень инфляции на конец 2021 года прогнозируется на уровне 9,7-10,6%.

Ниже приведены сравнительные результаты оценки параметров VAR модели с помощью метода наименьших квадратов (Multivariate Least Squares – MLS) и Байеса (BVAR) с точки зрения их соответствия (fitting) фактическим данным (Fakt) и показаны соответствующие прогнозы с применением этих двух методов (для BVAR взята медиана прогнозов).

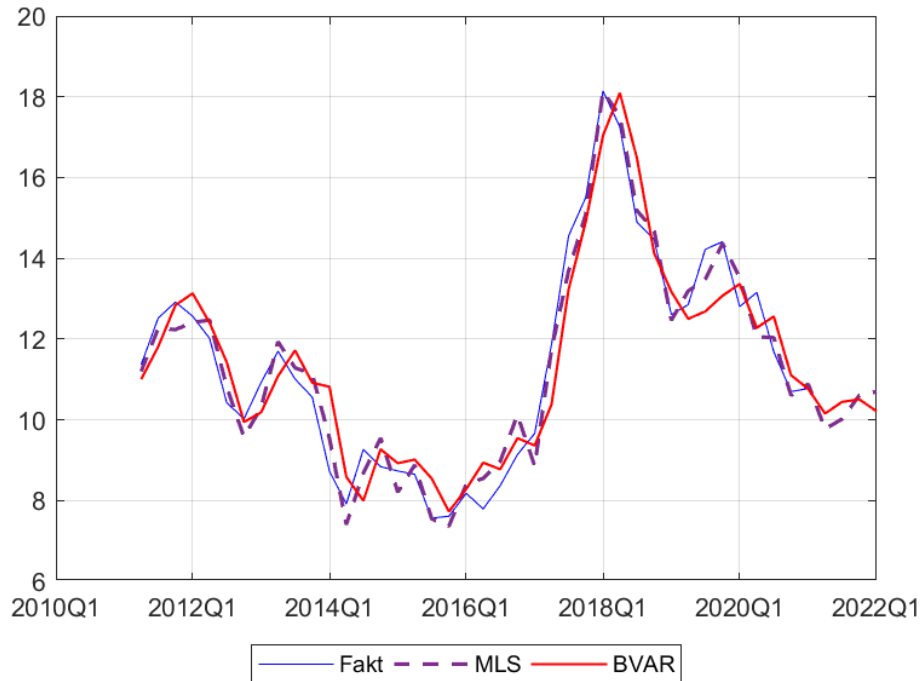


Рис. 2: Сравнение классического и Байесовского методов оценки

Перспективы совершенствования модели

В последующих этапах исследования при оценке параметров модели наряду с распределением Миннесоты планируется использовать более сложное и более совершенное нормальное-обратное априорное распределение Уишарта (Normal inverse Wishart Prior), поскольку последнее позволяет вывести апостериорное распределение ковариационной матрицы аналитическим способом в отличие от распределения Миннесоты, которое берет ковариационную матрицу как постоянную и с диагональной оценкой. Далее предусматривается увеличение числа эндогенных переменных, образование функций импульсных откликов и осуществление декомпозиции дисперсии.

IV. Вопрос оценки параметров модели

Параметром модели называется постоянный показатель, выражающий степень влияния определенного фактора на соответствующую зависимую переменную.

Классический метод оценки параметров модели VAR даёт возможность наиболее полного соответствия параметров к наблюдаемым данным благодаря большому количеству переменных, однако это может привести к излишней параметризации (“оверфиттинг”) модели, недостатком которой является возможное искажение получаемых прогнозов.

Это связано с тем, что, когда количество переменных относительно велико, а данные, рассматриваемые для оценки модели, относительно малы, особенно когда метод оценки (подобно методу наименьших квадратов) разработан так, чтобы быть максимально приближенным к данным, существует вероятность наибольшего влияния “шума” по сравнению с “сигналом” в формировании оценок параметров. Таким образом, даже если модель практически полностью согласуется с данными, точность ее прогноза резко снижается (Рисунок 3).

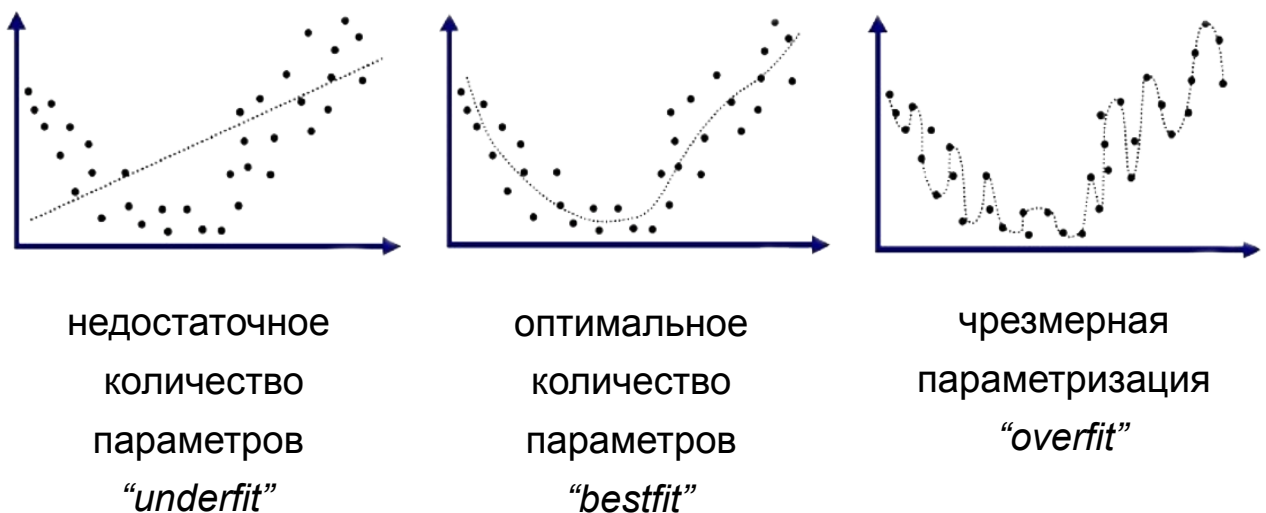


Рис. 3: Эволюция модели по мере увеличения параметров

В этом случае, такие модели рекомендуется оценивать, устанавливая некоторые ограничения, уменьшающие размер набора параметров. Следовательно, ключевая задача **найти как можно установить более надежные критерии ограничения**.

Применение байесовского принципа для оценки параметров VAR модели изначально был предложен Литтерманом (1980) как решение проблемы чрезмерной параметризации (overfitting), избегая жестких нулевых ограничений на коэффициенты. При установлении ограничений на параметры нельзя быть полностью уверенным в том, что некоторые коэффициенты равны нулю, и не рекомендуется игнорировать их возможный диапазон изменения. Точка зрения Байеса полностью соответствует этому подходу.

Вместо того, чтобы присваивать чрезмерно высокие веса значениям некоторых параметров, неопределенности касательно точных значений параметров модели можно рассматривать как **распределение вероятности**¹⁵ для вектора параметров. Степень неопределенности, сформированная этим распределением, затем может быть изменена информацией в наборе данных, если есть разница между исходными и обновленными данными. Пока исходная информация не очень расплывчата или не лишена информации оценка параметров будет меняться в части “сигнала”, а не “шума”, тем самым снижая риск чрезмерной параметризации.

V. Применение метода Байеса

Пусть случайный вектор Y задан некоторым законом распределения вероятностей. Нас интересует **вопрос оценки параметров модели** на основе имеющихся данных:

$$Y \sim f(\theta|Y),$$

¹⁵Смотрите: Приложение 2 определение 8.

где f – функция массы вероятности или плотности¹⁶, θ – параметры закона распределения, подлежащие к оценке, например, если в качестве f принять закон нормального распределения, тогда $\theta = \mu, \sigma^2$.

Следует отметить, что рассматриваемая нами задача является противоположностью обычных задач теории вероятностей.

Например, задачи классической теории вероятностей, в которых требуется найти распределение Y при заданных параметрах модели (θ):

– какова вероятность попадания “туза” из набора игральных карт?

– какова вероятность 6-разового попадания “герба” при подбрасывании монеты 8 раз?

Как утверждал Кинг (1998, 14), мы преследуем цели определения $P(Y|Модель)$ или $P(Информация|Модель)$. А по нашей терминологии $f(Y|\theta)$.

Однако, в рассматриваемой нами задаче у нас имеются данные и мы хотим определить модель, точнее **параметры модели**. Иными словами, мы хотим определить распределение неизвестных параметров, взяв данные за условия, то есть $P(Модель|Информация)$ или $f(\theta|Y)$. Это называется “обратной задачей вероятности”.

Для решения данной задачи можем воспользоваться **формулой Байеса**¹⁷:

$$P(\theta|Y) = \frac{P(\theta)P(Y|\theta)}{P(Y)},$$

где:

$P(\theta)$ – априорная вероятность (перед экспериментом),

$P(\theta|Y)$ – апостериорная вероятность (после эксперимента) и

$P(Y|\theta)$ – функция правдоподобия (likelihood).

Обратите внимание, что знаменатель формулы, то есть $P(Y)$ просто функция, которая зависит от данных. Учитывая, что $P(\theta|Y)$ и $P(Y|\theta)$ вычисляются для одних и тех же данных, можем опустить

¹⁶ *Смотрите*: Приложение 2 определения 9 и 10.

¹⁷ *Смотрите*: Приложение 2 определение 20.

знаменатель формулы (\propto – знак пропорциональности):

$$P(\theta|Y) \propto P(\theta)P(Y|\theta) \quad (5)$$

при этом $\frac{1}{P(Y)}$ называется коэффициентом пропорциональности.

Оценка параметров VAR модели, представленной в виде (3) методом Байеса состоит из 4 этапов:

1. **Формируется исходное предположение** об оцениваемых параметрах. Это исходное предположение обычно не исходит от заданных данных— Y_t и X_t и отражает информацию о β и Σ . Такие предположения могут быть сформированы на основе прошлого опыта исследователя или исследований на основе других данных (оценки аналогичных моделей). Главное, данные исходные предположения выражаются **в виде распределения вероятностей**. Например, исходное предположение матрицы коэффициентов β формируется в качестве распределения $p(\beta) \sim N(\beta_0, \Omega)$, при этом текущие предположения об элементах β суммируются в β_0 .
2. Собираются статистические данные по Y_t и X_t и записывается **функция правдоподобия** (likelihood function).
3. **Исходные предположения** о параметрах модели, сформированные на первом этапе **обновляются статистическими данными** с помощью функции правдоподобия, записанной на втором этапе. Другими словами, комбинируем априорное распределение параметров $p(\beta, \Sigma)$ и функцию правдоподобия $L(Y|\beta, \Sigma)$ с помощью формулы Байеса (5), и получаем **апостериорное распределение параметров** – $p(\beta, \Sigma|Y)$:

$$p(\beta, \Sigma|Y) \propto p(\beta, \Sigma)L(Y|\beta, \Sigma). \quad (6)$$

4. После нахождения апостериорного распределения $p(\beta, \Sigma|Y)$, будет возможным найти $p(\beta|Y)$ и $p(\Sigma|Y)$ посредством вычисления

β и Σ соответственно (при этом распределения $p(\beta|Y)$ и $p(\Sigma|Y)$ называются **маржинальными распределениями** на основе данных). Наконец, чтобы сформировать точечную оценку (point estimate) и измерения точности искомых параметров легко могут быть проанализированы соответствующие математические ожидания и дисперсии $p(\beta|Y)$ и $p(\Sigma|Y)$. Однако, на практике часто бывает очень сложным или даже невозможным аналитически вывести маргинальные распределения из $p(\beta, \Sigma|Y)$. Эта проблема может быть решена применением численного метода симуляции Монте Карло¹⁸.

Применение **каждого** вышеперечисленного этапа будет подробно рассмотрено в следующих разделах.

Поскольку формирование априорного распределения на 1-этапе требует **глубокого систематического подхода**, его определение отложим на следующие главы и сначала рассмотрим **задачу нахождения функции плотности вероятности данных (likelihood)**, то есть 2-этап. Она обычно обозначается как $L(Y|\theta)$.

Из уравнения (3) находим вектор ошибок:

$$\varepsilon_t = Y_t - X_t\beta.$$

Согласно с условиями, установленными для вектора ошибок VAR модели, вектор ε_t задается законом нормального распределения:

$$\varepsilon_t \sim N(0, \Sigma).$$

Функция правдоподобия для одного наблюдения за период t на основе вышеуказанных параметров будет выглядеть как:

$$L_t(Y|\beta, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(Y_t - X_t\beta)'\Sigma^{-1}(Y_t - X_t\beta)\right)$$

¹⁸Смотрите: Приложение 6.

Поскольку каждое наблюдение является независимым событием¹⁹ вероятность всего набора данных равна:

$$\begin{aligned}
 L &= L_1 \times L_2 \times L_3 \times \dots \times L_T = \prod_{t=1}^T L_t \\
 &= \prod_{t=1}^T \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (Y_t - X_t \beta)' \Sigma^{-1} (Y_t - X_t \beta) \right) \\
 &= \frac{1}{(2\pi)^{(Tn)/2} |\Sigma|^{T/2}} \exp \left(-\frac{1}{2} \sum_{t=1}^T (Y_t - X_t \beta)' \Sigma^{-1} (Y_t - X_t \beta) \right) \\
 &\propto |\Sigma|^{-T/2} \cdot \exp \left(-\frac{1}{2} \sum_{t=1}^T (Y_t - X_t \beta)' \Sigma^{-1} (Y_t - X_t \beta) \right). \tag{7}
 \end{aligned}$$

Так, функция правдоподобия на основе параметров модели:

$$L(Y|\beta, \Sigma) \propto |\Sigma|^{-T/2} \cdot \exp \left(-\frac{1}{2} \sum_{t=1}^T (Y_t - X_t \beta)' \Sigma^{-1} (Y_t - X_t \beta) \right). \tag{8}$$

VI. Определение апостериорного распределения

Чтобы упростить работу с уравнением (3), собираем все наблюдения за период для уравнения m и записываем следующим образом:

$$Y_m = X \beta_m + \varepsilon_m, \quad m = 1, \dots, n \tag{9}$$

где

Y_m и ε_m – векторы с размером $T \times 1$,
 β_m – векторы с размером $[np + d + 1] \times 1$,
 X – матрица с размером $T \times [np + d + 1]$:

¹⁹Смотрите: Приложение 2 определение 19.

$$Y_m = \begin{pmatrix} y_{m,p+1} \\ y_{m,p+2} \\ \vdots \\ y_{m,S} \end{pmatrix}, \quad \beta_m = \begin{pmatrix} c_m \\ a_{m,1}^1 \\ a_{m,2}^1 \\ \vdots \\ a_{m,n}^1 \\ \vdots \\ a_{m,1}^p \\ a_{m,2}^p \\ \vdots \\ a_{m,n}^p \\ b_{m,1} \\ \vdots \\ b_{m,d} \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ y_{1,p} & y_{1,p+1} & \cdots & y_{1,S-1} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,S-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,p} & y_{n,p+1} & \cdots & y_{n,S-1} \\ y_{1,p-1} & y_{1,p-2} & \cdots & y_{1,S-2} \\ y_{2,p-1} & y_{2,p-2} & \cdots & y_{2,S-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,p-1} & y_{n,p-2} & \cdots & y_{n,S-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1,1} & y_{1,2} & \cdots & y_{1,S-p} \\ y_{2,1} & y_{1,2} & \cdots & y_{1,S-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n,1} & y_{n,2} & \cdots & y_{n,S-p} \\ z_{1,p+1} & z_{1,p+2} & \cdots & z_{1,S} \\ \vdots & \vdots & \ddots & \vdots \\ z_{d,p+1} & z_{d,p+2} & \cdots & z_{d,S} \end{pmatrix}'.$$

Рассмотрим определение апостериорного распределения параметров **для трех случаев**. При этом ковариационная матрица Σ постоянна и диагональная матрица, представляется следующим образом

$$\text{var}[\varepsilon_i] = \sigma^2 \quad \text{— Homoscedasticity}$$

и

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \text{— No autocorrelation } (i \neq j)$$

соответственно, то есть

$$\Sigma = \sigma^2 I. \quad (10)$$

При оценке параметров сначала определяем апостериорное распределение β , предполагая, что σ^2 известно. Затем определяем

апостериорное распределение σ^2 , предполагая, что β известно, и наконец, рассмотрим общий случай, когда оба параметра неизвестны.

1-случай. Апостериорное распределение β , предполагая, что σ^2 известно. Рассмотрим вопрос оценки вектора β_m с размером $[np + d + 1] \times 1$ и состоящий из параметров уравнения (9) (предполагая, что известна дисперсия ошибки $\sigma_{m,m}^2 = \sigma^2$).

Сначала **формируем исходные взгляды на β** . Пусть его априорное распределение задается законом нормального распределения:

$$p(\beta_m) \sim N(\beta_0, \Omega) \propto |\Omega|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\beta_m - \beta_0)' \Omega^{-1} (\beta_m - \beta_0)\right) \quad (11)$$

где β_0 и Ω представляют априорное среднее и ковариационную матрицу β_m соответственно.

Исходя из предположения, что ошибки не связаны друг с другом [выражение (10)], мы получаем следующее для **функции правдоподобия**, как в выражении (8):

$$L(Y|\beta, \sigma^2) \propto (\sigma^2)^{-T/2} \cdot \exp\left(-\frac{1}{2\sigma^2}(Y_m - X\beta_m)'(Y_m - X\beta_m)\right). \quad (12)$$

Апостериорное распределение β определяется следующим образом:

$$p(\beta_m|Y) = p(\beta_m)L(Y|\beta, \sigma^2),$$

это пропорционально следующему выражению:

$$\begin{aligned} & (\sigma^2)^{-\frac{T}{2}} |\Omega|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[(\beta_m - \beta_0)' \Omega^{-1} (\beta_m - \beta_0) + \frac{1}{\sigma^2}(Y_m - X\beta_m)'(Y_m - X\beta_m)\right]\right\} \\ & \propto \exp\left\{-\frac{1}{2}\left[\left(\beta_m' \Omega^{-1} \beta_m - \beta_m' \Omega^{-1} \beta_0 - \beta_0' \Omega^{-1} \beta_m + \beta_0' \Omega^{-1} \beta_0\right) + \right. \right. \\ & \quad \left. \left. + \frac{1}{\sigma^2}\left(Y_m' Y_m - Y_m' X \beta_m - \beta_m' X' Y_m + \beta_m' X' X \beta_m\right)\right]\right\} \quad (13) \end{aligned}$$

Поскольку ковариационная матрица Ω в последнем выражении является симметричной (то есть $\Omega' = \Omega$) и в соответствии со свойствами обратной матрицы²⁰:

$$(\Omega^{-1})' = (\Omega')^{-1} = (\Omega)^{-1} = \Omega^{-1},$$

то есть обратная матрица симметричной матрицы также является симметричной.

Вычислим также размер произведения $\beta'_m \Omega^{-1} \beta_0$ ($k = np + d + 1$):

$$[\beta'_m]_{1 \times k} \cdot [\Omega^{-1}]_{k \times k} \cdot [\beta_0]_{k \times 1} = [c]_{1 \times 1}.$$

Следовательно, результатом произведения является матрица размерности 1×1 , или просто скалярная ($c = \text{conts}$). Учитывая скалярность транспонированной скалярной матрицы:

$$\beta'_m \Omega^{-1} \beta_0 = (\beta'_m \Omega^{-1} \beta_0)' = \beta'_0 \Omega^{-1} \beta_m. \quad (14)$$

Аналогично:

$$[\beta'_m]_{1 \times k} \cdot [X']_{k \times T} \cdot [Y_m]_{T \times 1} = [c]_{1 \times 1}.$$

И так,

$$\beta'_m X' Y_m = (\beta'_m X' Y_m)' = Y'_m X \beta_m. \quad (15)$$

Используя уравнения (14) и (15), можем записать выражение (13) следующим образом:

$$\exp \left\{ -\frac{1}{2} \left[(\beta'_m \Omega^{-1} \beta_m - 2\beta'_0 \Omega^{-1} \beta_m + \beta'_0 \Omega^{-1} \beta_0) + \frac{1}{\sigma^2} (Y'_m Y_m - 2Y'_m X \beta_m + \beta'_m X' X \beta_m) \right] \right\}, \quad (16)$$

Поскольку произведения

$$[Y'_m]_{1 \times T} \cdot [Y_m]_{T \times 1} = [c]_{1 \times 1} \text{ и } [\beta'_0]_{1 \times k} \cdot [\Omega^{-1}]_{k \times k} \cdot [\beta_0]_{k \times 1} = [c]_{1 \times 1}$$

²⁰ Смотрите: Приложение 1 определение 9

постоянны по отношению к β_m , выражение (16) пропорционально следующему:

$$\begin{aligned} & \exp \left\{ -\frac{1}{2} \left[\beta_m' \Omega^{-1} \beta_m - 2\beta_0' \Omega^{-1} \beta_m - 2 \cdot \frac{1}{\sigma^2} Y_m' X \beta_m + \frac{1}{\sigma^2} \beta_m' X' X \beta_m \right] \right\} \\ & = \exp \left\{ -\frac{1}{2} \left[\beta_m' \left(\Omega^{-1} + \frac{1}{\sigma^2} X' X \right) \beta_m - 2 \left(\Omega^{-1} \beta_0 + \frac{1}{\sigma^2} X' Y_m \right)' \beta_m \right] \right\}. \end{aligned} \quad (17)$$

Обратим внимание на следующее (здесь A и B векторы размерности $k \times 1$, C симметричная матрица размерности $k \times k$):

$$\begin{aligned} (A - B)' C^{-1} (A - B) &= A' C^{-1} A - B' C^{-1} A - A' C^{-1} B + B' C^{-1} B \\ &= A' C^{-1} A - 2B' C^{-1} A + B' C^{-1} B. \end{aligned}$$

Используя это, в выражении (17) введем следующие обозначения

$$\begin{aligned} A &= \beta_m, \\ C^{-1} &= \Omega^{-1} + \sigma^{-2} X' X, \\ B' C^{-1} &= \left(\Omega^{-1} \beta_0 + \sigma^{-2} X' Y_m \right)' = D' \end{aligned}$$

тогда в соответствии со свойствами обратной матрицы:

$$C = (C^{-1})^{-1} = (\Omega^{-1} + \sigma^{-2} X' X)^{-1}$$

и

$$\begin{aligned} B' C^{-1} = D' &\implies (B' C^{-1})' = (D')' \implies C^{-1} B = D \implies \\ \implies C C^{-1} B = C D &\implies B = C D = C \left(\Omega^{-1} \beta_0 + \sigma^{-2} X' Y_m \right). \end{aligned}$$

Следовательно, добавляя скалярное выражение $B' C^{-1} B$ к аргументу функции экспоненты в выражении (17) получаем следующее:

$$p(\beta_m | Y) \propto \exp \left\{ -\frac{1}{2} \left[\left(\beta_m - B \right)' C^{-1} \left(\beta_m - B \right) \right] \right\} \quad (18)$$

где

$$B = C \left(\Omega^{-1} \beta_0 + \sigma^{-2} X' Y_m \right) \quad \text{и} \quad C = \left(\Omega^{-1} + \sigma^{-2} X' X \right)^{-1}.$$

Обратите внимание, из выражения (18) следует, что апостериорным распределением параметров модели будет нормальное многомерное распределение с заданным математическим ожиданием B и ковариационной матрицей C , то есть:

$$p(\beta_m|Y) \sim N(B, C).$$

Иными словами, когда известны Ω^{-1} , β_0 и σ^{-2} , вектор B можем взять в качестве точечной оценки параметров.

Следует отметить, что если из формулы среднего апостериорного распределения параметров модели

$$B = (\Omega^{-1} + \sigma^{-2} X'X)^{-1} (\Omega^{-1}\beta_0 + \sigma^{-2} X'Y_m)$$

убрать априорные параметры (то есть β_0 и Ω^{-1}) (допустим, что априорное распределение имеет бесконечную дисперсию вокруг своего среднего значения: $\Omega^{-1} = 0$), апостериорное среднее преобразуется в вектор параметров оцененных методом наименьших квадратов (OLS):

$$B = (X'X)^{-1} X'Y_m.$$

В целом, когда апостериорное распределение $p(\theta|Y)$ и априорное распределение $p(\theta)$ принадлежат одному и тому же семейству распределения вероятности, они называются **сопряженными распределениями** (conjugate distributions), при этом априорное распределение называется **сопряженным априором** (conjugate prior) для функции правдоподобия $p(Y|\theta)$.

В результате вытекания выражения (18) из выражений (11) и (12) выше, стало известно, что семейство функций нормального распределения является самосопряженным (self-conjugate) распределением, когда в качестве функции правдоподобия берем нормальное распределение, то есть если функция априорного распределения и вероятности данных являются нормальными

распределениями, то апостериорное распределение также будет нормальным распределением.

2-случай. Апостериорное распределение σ^2 , предполагая, что известен β . Здесь мы также определим условное апостериорное распределение σ^2 в три этапа.

Сначала сформируем начальные представления σ^2 . Для получения сопряженных распределений берем распределение гаммы²¹ в качестве априорного распределения $\frac{1}{\sigma^2}$:

$$p\left(\frac{1}{\sigma^2}\right) \sim \Gamma\left(\frac{T_0}{2}, \frac{\theta_0}{2}\right) \propto \left[\frac{1}{\sigma^2}\right]^{\frac{T_0}{2}-1} \exp\left(-\frac{\theta_0}{2\sigma^2}\right).$$

В качестве **функции правдоподобия** используем выражение (12).

Вычислим апостериорное распределение σ^2 . При этом будем использовать функции априорного распределения и правдоподобия $\frac{1}{\sigma^2}$, сформированные на вышеуказанных этапах:

$$\begin{aligned} p\left(\frac{1}{\sigma^2} \middle| \beta, Y\right) &\propto p\left(\frac{1}{\sigma^2}\right) L(Y|\beta, \sigma^2) \\ &= \left[\frac{1}{\sigma^2}\right]^{\frac{T_0}{2}-1} \exp\left(-\frac{\theta_0}{2\sigma^2}\right) (\sigma^2)^{-\frac{T}{2}} \exp\left(-\frac{(Y_m - X\beta_m)'(Y_m - X\beta_m)}{2\sigma^2}\right) \\ &= \left[\frac{1}{\sigma^2}\right]^{\frac{T_0}{2}-1+\frac{T}{2}} \exp\left(-\frac{\theta_0 + (Y_m - X\beta_m)'(Y_m - X\beta_m)}{2\sigma^2}\right) \end{aligned}$$

Нетрудно заметить, что последнее выражение является распределением гаммы, задаваемое следующими параметрами:

$$\begin{aligned} T_1 &= \frac{T_0 + T}{2}, \\ \theta_1 &= \frac{\theta_0 + (Y_m - X\beta_m)'(Y_m - X\beta_m)}{2}. \end{aligned}$$

Из этого

$$\boxed{p\left(\frac{1}{\sigma^2} \middle| \beta, Y\right) \propto \Gamma(T_1, \theta_1).} \quad (19)$$

²¹ Смотрите: Приложение 4 определение 2 и 3-свойство.

3-случай. Апостериорное распределение σ^2 и β . Теперь рассмотрим вычисление их совместного апостериорного распределения, где и вектор коэффициентов β , и $1/\sigma^2$ неизвестны. Это также осуществим в три этапа.

Формирование исходных представлений. Используя формулу условной вероятности²² сформируем априорное распределение:

$$p\left(\frac{1}{\sigma^2}, \beta\right) = p\left(\frac{1}{\sigma^2}\right) \cdot p\left(\beta \middle| \frac{1}{\sigma^2}\right)$$

где

$$p\left(\frac{1}{\sigma^2}\right) \sim \Gamma\left(\frac{T_0}{2}, \frac{\theta_0}{2}\right) \propto \left[\frac{1}{\sigma^2}\right]^{\frac{T_0}{2}-1} \exp\left(-\frac{\theta_0}{2\sigma^2}\right),$$

$$p\left(\beta \middle| \frac{1}{\sigma^2}\right) \sim N(\beta_0, \sigma^2 \Omega) \propto |\sigma^2 \Omega|^{-1/2} \cdot \exp\left(-\frac{1}{2}(\beta_m - \beta_0)'(\sigma^2 \Omega)^{-1}(\beta_m - \beta_0)\right).$$

В качестве **функции правдоподобия** воспользуемся выражением (12).

Вычисление апостериорного распределения. Для этого, как обычно используем априорное распределение и функцию вероятности данных:

$$p\left(\frac{1}{\sigma^2}, \beta \middle| Y\right) \propto p\left(\frac{1}{\sigma^2}, \beta\right) \cdot L(Y|\beta, \sigma^2). \quad (20)$$

Данное распределение является совместной функцией распределения β и $1/\sigma^2$, и его форма намного сложнее, чем отдельные условные функции апостериорного распределения β и $1/\sigma^2$, приведенные в случаях 1 и 2. На следующем этапе следует отделить из выражения (20) необходимые нам функции условных апостериорных распределений β и $1/\sigma^2$.

²²Смотрите: Приложение 2 определение 20.

В целом, если функция $f(x, y)$ является совместной функцией плотности вероятности непрерывных случайных величин X и Y в пространстве $[a, b] \times [c, d]$, их отдельные функции плотности маргинального распределения определяются с помощью следующих интегралов²³:

$$f_X(x) = \int_a^b f(x, y) dy, \quad f_Y(y) = \int_c^d f(x, y) dx. \quad (21)$$

Аналогично, **теоретически**, используя формулы в (21) можно вычислить условные апостериорные распределения β и $1/\sigma^2$ из (20):

$$p(\beta|Y) = \int_0^{\infty} p\left(\frac{1}{\sigma^2}, \beta | Y\right) d\frac{1}{\sigma^2},$$

также,

$$p\left(\frac{1}{\sigma^2} | Y\right) = \int_0^{\infty} p\left(\frac{1}{\sigma^2}, \beta | Y\right) d\beta.$$

Однако, как отмечалось ранее, **на практике**, из-за сложного вида совместного апостериорного распределения параметров как в (20), нет возможности его определения аналитическим способом, а затем вычислить маргинальные распределения с помощью интегралов.

В свою очередь, исходя из трех случаев определения апостериорного распределения параметров можно сделать следующие выводы:

- относительно легче вычислять и работать с условными апостериорными распределениями, как в случаях 1 и 2;
- и наоборот, формирование маргинальных распределений из совместных апостериорных распределений, как в 3 случае, несколько затруднительно или невозможно в сложных моделях.

²³Для подробной информации: https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading7a.pdf, Sections 3.2 and 3.8.

VII. Генератор выборки Гиббса

Согласно нашему заключению, сделанному в предыдущей главе, имеет практическое значение определение апостериорного распределения параметров модели VAR, предполагая, что один из них известен. Для этого используется один из алгоритмов Марковской цепи Монте-Карло²⁴ – Генератор выборки Гиббса (Gibbs sampler).

Вспомним интересующий нас вопрос: нам необходимо апостериорное распределение параметров на основе данных, то есть, $p(\beta|Y)$ и $p(\Sigma|Y)$. Однако, как мы видели в 1 и 2 случаях выше, для определения одного из них необходимо знать значение другого: для определения $p(\beta|Y)$ необходимо значение σ^2 и для $p(\frac{1}{\sigma^2}|Y)$ нужен β .

Генератор выборки Гиббса является рекурсивным методом Монте-Карло, который постепенно образует необходимые нам распределения. Сначала начнем с произвольных значений $\beta^{(0)}$ и $\sigma^{2(0)}$ и сделаем выборку из их соответствующих функций плотности:

$$\begin{aligned} \sigma^2 &= \sigma^{2(0)} \\ \beta^{(0)} &\text{ из } p\left(\beta|Y, \sigma^{2(0)}\right) \\ \sigma^{2(1)} &\text{ из } p\left(\frac{1}{\sigma^2}\middle|Y, \beta^{(0)}\right) \\ \beta^{(1)} &\text{ из } p\left(\beta|Y, \sigma^{2(1)}\right) \\ \sigma^{2(2)} &\text{ из } p\left(\frac{1}{\sigma^2}\middle|Y, \beta^{(1)}\right) \\ &\vdots \\ \beta^{(k)} &\text{ из } p\left(\beta|Y, \sigma^{2(k-1)}\right) \\ \sigma^{2(k)} &\text{ из } p\left(\frac{1}{\sigma^2}\middle|Y, \beta^{(k)}\right). \end{aligned}$$

²⁴ Смотрите: Приложения 5 и 6.

Вектор $\mathbf{v}^{(k)} = (\beta^{(k)}, \sigma^{2(k)})$ образует Марковскую цепь, и при достаточно большом количестве итераций, полученные выборки примерно будут одинаковыми с реальным маргинальным апостериорным распределением. Другими словами, описанный выше процесс выборки повторяется M раз и сохраняется N последних (например, $N = 1000$) выборок.

Ниже рассмотрим алгоритм определения апостериорного распределения параметров уравнения (9) с помощью генератора выборки Гиббса. Это осуществляется в 4 шага.

Шаг 1. Определим априорные распределения и исходные значения.

Устанавливаем нормальное распределение для β по априорным параметрам Миннесоты:

$$p(\beta_m) \sim N(\beta_0, \Omega).$$

Иными словами, вводим априорное среднее – β_0 и априорную ковариационную матрицу – Ω для β_m .

Для $1/\sigma^2$ задаем априорное распределение гаммы по параметрам T_0 и θ_0 :

$$p\left(\frac{1}{\sigma^2}\right) \sim \Gamma\left(\frac{T_0}{2}, \frac{\theta_0}{2}\right).$$

Для запуска генератора выборки Гиббса потребуется исходное значение β_m или σ^2 . В данном примере в качестве исходного значения σ^2 берем его оценку по OLS, то есть

$$\sigma^2 = \hat{\sigma}_{OLS}^2 = \frac{\varepsilon' \varepsilon}{T - n}.$$

Следует отметить, что в линейных моделях, таких как VAR, когда количество итераций Гиббса достаточно велико, начальные значения параметров, с практической точки зрения, менее важны при формировании их конечных значений.

Шаг 2. Поскольку нам известно σ^2 , сформируем случайное число из апостериорного распределения β . Апостериорное распределение

β согласно выражению (18) будет нормальным распределением со следующими параметрами:

$$p(\beta_m|Y) \sim N(B, C),$$

где

$$B = C (\Omega^{-1}\beta_0 + \sigma^{-2}X'Y_m),$$

$$C = (\Omega^{-1} + \sigma^{-2}X'X)^{-1}.$$

Обратите внимание, у нас имеются значения всех компонентов, необходимых для расчета параметров B и C .

Теперь сформируем вектор случайных величин β_m из распределения $p(\beta_m|Y)$. Для этого воспользуемся следующим алгоритмом.

Алгоритм 1. Чтобы сгенерировать вектор размером $r \times 1$ (обозначим вектор как z) из распределения $N(\mu, \omega)$, с помощью компьютерной программы необходимо создать вектор z_0 размером $r \times 1$ из стандартного нормального распределения. Эти стандартные нормальные числа преобразуются в случайную величину z со средним значением μ и дисперсией ω с использованием следующей замены, обратной стандартизации случайных величин:

$$z = \mu + z_0 \cdot \omega^{1/2}.$$

Используя приведенный выше алгоритм, создадим стандартный нормальный вектор $\bar{\beta}$ размером $1 \times [np + d + 1]$ и тем самым получаем случайное число из апостериорного распределения β :

$$\beta_m^{(1)} = B + [\bar{\beta} \cdot C^{1/2}]'$$

где $\beta_m^{(1)}$ – означает вектор β_m , сформированный итерацией Гиббса 1.

Шаг 3. Поскольку $\beta^{(1)}$ известно, выводим σ^2 из апостериорного распределения $1/\sigma^2$. Апостериорное распределение $1/\sigma^2$ определяется

согласно выражению (19) распределением гаммы со следующими параметрами:

$$p\left(\frac{1}{\sigma^2} \middle| \beta, Y\right) \propto \Gamma\left(\frac{T_1}{2}, \frac{\theta_1}{2}\right).$$

где

$$T_1 = T_0 + T,$$

$$\theta_1 = \theta_0 + \left(Y_m - X\beta_m^{(1)}\right)' \left(Y_m - X\beta_m^{(1)}\right).$$

Обратите внимание, что апостериорный параметр θ_1 рассчитывается с использованием $\beta_m^{(1)}$, созданного на предыдущем шаге.

Теперь образуем случайную $1/\sigma^2$ из распределения $\Gamma\left(\frac{T_1}{2}, \frac{\theta_1}{2}\right)$. Для этого воспользуемся следующим алгоритмом.

Алгоритм 2. Чтобы сгенерировать число λ из распределения гаммы со степенью независимости $\frac{T}{2}$ и параметром $\frac{\theta}{2}$ сначала образуем T -ое количество случайных величин из стандартного нормального распределения и объединим их в вектор τ :

$$\tau = [X_1 \quad X_2 \quad \dots \quad X_T]'$$

В этом случае, согласно свойству распределения хи-квадрат²⁵:

$$\tau'\tau = X_1^2 + X_2^2 + \dots + X_T^2 \sim \Gamma\left(\frac{T}{2}, \frac{1}{2}\right)$$

Необходимое нам число λ с использованием свойства распределения гаммы²⁶:

$$\lambda = \frac{\tau'\tau}{\theta} \sim \Gamma\left(\frac{T}{2}, \frac{\theta}{2}\right).$$

Используя приведенный выше алгоритм, создадим число λ из распределения $\Gamma\left(\frac{T_1}{2}, \frac{\theta_1}{2}\right)$, тогда:

$$\sigma^{2(1)} = \frac{1}{\lambda}.$$

²⁵Смотрите: Приложение 4, 3-свойство.

²⁶Смотрите: Приложение 4, 2-свойство.

где $\sigma^{2(1)}$ это число σ^2 , образованное в итерации Гиббса 1.

Шаг 4. Повторяя шаги 2 и 3 k раз, получаем значения $\beta_m^{(1)}, \beta_m^{(2)}, \dots, \beta_m^{(k)}$ и $\sigma^{2(1)}, \sigma^{2(2)}, \dots, \sigma^{2(k)}$. Из значений β_m и σ^2 , сформированных на последних N итерациях составим эмпирическое распределение этих параметров. Данное эмпирическое распределение приблизительно соответствует реальному маргинальному апостериорному распределению параметров. Следовательно, мы можем использовать их среднее значение как точечную оценку β_m и σ^2 .

Следует также отметить, что первоначальные $k - N$ итераций не учитываются при вычислении эмпирического распределения и называются “выгорающими” (burn-in) итерациями. Эти итерации необходимы только для того, чтобы генератор выборки Гиббса начал сходимость (convergence) с реальными значениями параметров.

Наконец, как только мы определили все параметры модели VAR, мы сможем **сгенерировать прогнозные траектории** для рассматриваемых переменных, используя рекурсивный метод.

VIII. Априор Миннесоты

Первым шагом в оценке модели VAR методом Байеса является определение функции априорного распределения $p(\beta, \Sigma)$ для параметров модели. При этом, одним из наиболее широко используемых методов является метод, предложенный Литтерманом (1986), иногда называемый априорным методом Миннесоты.

Рассмотрим вопрос определения среднего значения априорного распределения β_0 и ковариационной матрицы Ω , необходимых для использования формулы (18).

Опираясь на три особенности макроэкономических временных рядов, Литтерман формирует свое исходное (априорное) распределение, аналогичное выражению (11):

- 1) трендовое изменение, характерное для макроэкономических

показателей;

2) недавние значения данных предоставляют больше информации о текущем значении, чем прошлые значения;

3) прошлые значения наблюдаемой переменной содержат больше информации о её текущем состоянии, чем прошлые значения других переменных.

Эти закономерности могут быть применены в оценке посредством сопоставления параметров со следующими распределениями вероятности:

1) среднее значение коэффициентов, установленных для всех лагов, кроме первого, равно нулю;

2) дисперсия коэффициентов обратно пропорциональна количеству лагов;

3) для коэффициентов переменной j в уравнении m устанавливается более низкая априорная дисперсия, чем для коэффициентов переменной m .

Вышеупомянутые требования могут быть представлены вектором (называемым гиперпараметром), состоящим из параметров. В частности, Литтермэн предполагает, что диагональные элементы ковариационной матрицы Ω являются случайными величинами следующего состава:

$$\begin{aligned} & \left(\frac{\lambda_1}{l^{\lambda_3}} \right)^2 \text{ если } m = j, \\ & \left(\frac{\sigma_m \lambda_1 \lambda_2}{\sigma_j l^{\lambda_3}} \right)^2 \text{ если } m \neq j, \\ & (\sigma_m \lambda_4)^2 \text{ для константы.} \end{aligned}$$

где в уравнении m зависимая переменная обозначена как m , а независимые переменные как j . Следовательно, если $m = j$, то имеется ввиду коэффициент перед лагом m . σ_m и σ_j – дисперсии ошибок авторегрессии, оцененные методом наименьших квадратов (OLS),

используя переменные в VAR. Отношение σ_m и σ_j в приведенной выше формуле контролирует вероятность нахождения переменных m и j в разных шкалах (единицах). l – порядок лагов ($l = 1, \dots, p$). λ – параметры, устанавливаемые исследователем для контроля априорной жесткости:

– λ_1 управляет жесткостью собственных лагов наблюдаемой зависимой переменной (через стандартное отклонение)

– λ_2 контролирует жесткость лагов других переменных в уравнении;

– λ_3 контролирует степень обнуления коэффициентов перед лагами, выше первого. По мере его роста коэффициенты перед верхними лагами приближаются к нулю;

– λ_4 контролирует степень неопределенности коэффициентов перед переменными.

Наконец, среднее априорного распределения определяется следующим образом:

$$\beta_0 = (0, \dots, 0, \lambda_0, 0, \dots, 0)',$$

где λ_0 – означает априорное среднее коэффициента перед первым лагом зависимой переменной в m -уравнении и располагается на месте m в векторе, а также предполагается, что они подвержены авторегрессии 1-порядка AR(1).

Установление значений для гиперпараметров, контролирующих исходных (априорных) значений является очень важным вопросом. Обычно, в литературе для параметров используются следующие значения²⁷:

$$\lambda_1 = 0.2$$

$$\lambda_2 = 0.5$$

$$\lambda_3 = 1 \text{ или } 2$$

$$\lambda_4 = 10^5.$$

²⁷Canova, Fabio, 2007, Methods for Applied Macroeconomic Research, Princeton University Press, Princeton.

Приложение 1. Основы матричной алгебры

Определение 1. Прямоугольная таблица чисел называется **матрицей**.

Размер матрицы записывается в виде $m \times n$, что означает, что матрица состоит из m строк и n столбцов. Обычно матрицы обозначаются заглавными буквами латинского алфавита.

$$\mathbf{A} = (a_{i,j}) = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}.$$

Здесь числа $a_{i,j}$ называются **элементами** матрицы.

Определение 2. Матрица, количество строк которой равно количеству столбцов, т.е. $m = n$, называется **квадратной матрицей** n -го порядка.

Определение 3. Квадратная матрица n -го порядка, элементы главной диагонали которой равны единице, а все остальные элементы равны нулю, называется **единичной матрицей** и обозначается как \mathbf{I} .

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Определение 4. Матрица, образованная заменой строк данной матрицы \mathbf{A} на столбцы, называется **транспонированной матрицей** и обозначается как \mathbf{A}^T или \mathbf{A}' , т.е.

$$\text{Если } \mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}, \text{ то } \mathbf{A}' = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{m,n} \end{pmatrix}.$$

Определение 5. Если квадратная матрица A равна ее транспонированной матрице, такая матрица называется **симметричной матрицей**, то есть $A^T = A$. Элементы симметричной матрицы расположены симметрично относительно главной диагонали, т.е. $a_{i,j} = a_{j,i}$.

Теперь определим **операции с матрицами**.

Определение 6. **Суммой матриц** A и B , каждая размером $m \times n$, называется матрица размера $m \times n$, образованная путем сложения соответствующих элементов строк и столбцов этих матриц, то есть $A + B = (a_{i,j} + b_{i,j})$ или:

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{pmatrix}.$$

Определение 7. **Скалярным произведением** матрицы A и произвольного действительного числа λ (обычно называемого скаляром), называется следующая матрица:

$$\lambda A = \lambda \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

Определение 8. **Произведением матриц** A размерности $m \times n$ и B размерности $n \times z$ называется матрица размерности $m \times z$ такая, что

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1z} \\ b_{21} & b_{22} & \cdots & b_{2z} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nz} \end{pmatrix}$$

$$\mathbf{C} = \mathbf{AB} = \left(a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj} \right) = \left(\sum_{k=1}^n a_{ik}b_{kj} \right)$$

Здесь $i = 1, 2, \dots, m$ и $j = 1, 2, \dots, z$. Или

$$\mathbf{C} = \begin{pmatrix} a_{11}b_{11} + \dots + a_{1n}b_{n1} & a_{11}b_{12} + \dots + a_{1n}b_{n2} & \dots & a_{11}b_{1z} + \dots + a_{1n}b_{nz} \\ a_{21}b_{11} + \dots + a_{2n}b_{n1} & a_{21}b_{12} + \dots + a_{2n}b_{n2} & \dots & a_{21}b_{1z} + \dots + a_{2n}b_{nz} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \dots + a_{mn}b_{n1} & a_{m1}b_{12} + \dots + a_{mn}b_{n2} & \dots & a_{m1}b_{1z} + \dots + a_{mn}b_{nz} \end{pmatrix}.$$

Схематично это можно описать следующим образом:

$$\begin{array}{ccc} \mathbf{A} & \mathbf{B} & \mathbf{A} \cdot \mathbf{B} \\ \text{строка } i \rightarrow \begin{pmatrix} a_{i1} & a_{i2} & a_{i3} & \dots & a_{in} \end{pmatrix} & \begin{pmatrix} b_{1j} \\ b_{2j} \\ b_{3j} \\ \vdots \\ b_{nj} \end{pmatrix} & = \begin{pmatrix} \sum_{k=1}^n a_{ik}b_{kj} \end{pmatrix} \\ & \uparrow & \uparrow \\ & \text{столбец } j & \text{элемент } (i, j) \end{array}$$

Основные свойства операций с матрицами:

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.
2. $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$.
3. $\mathbf{AB} \neq \mathbf{BA}$.
4. $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$, $(\mathbf{B} + \mathbf{C})\mathbf{D} = \mathbf{BD} + \mathbf{CD}$.
5. $c(\mathbf{AB}) = (c\mathbf{A})\mathbf{B}$, $(\mathbf{AB})c = \mathbf{A}(\mathbf{B}c)$; $c = const$.
6. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$.
7. $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$.
8. $(\mathbf{A}^T)^T = \mathbf{A}$.
9. $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.
10. $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

Определение 9. Если существует квадратная матрица \mathbf{B} для квадратной матрицы \mathbf{A} порядка n , такая что $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, то матрица \mathbf{B} называется **обратной матрицей** \mathbf{A} , а матрица \mathbf{A} называется **обратимой матрицей**. Матрица, обратная для матрицы \mathbf{A} , обозначается как \mathbf{A}^{-1} .

Обратная матрица имеет следующие свойства:

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
2. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
3. $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.
4. $(c\mathbf{A})^{-1} = c^{-1}\mathbf{A}^{-1}$, $c = const$, $c \neq 0$.

Определение 10. Произведением Кронекера $\mathbf{A} \otimes \mathbf{B}$ матриц \mathbf{A} размера $m \times n$ и \mathbf{B} размера $p \times q$ называется матрица размера $mp \times nq$, в которой

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}$$

Или:

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{pmatrix}.$$

Например,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \otimes \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} = \begin{pmatrix} 1 \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} & 2 \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} \\ 3 \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} & 4 \begin{pmatrix} 0 & 5 \\ 6 & 7 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \times 0 & 1 \times 5 & 2 \times 0 & 2 \times 5 \\ 1 \times 6 & 1 \times 7 & 2 \times 6 & 2 \times 7 \\ 3 \times 0 & 3 \times 5 & 4 \times 0 & 4 \times 5 \\ 3 \times 6 & 3 \times 7 & 4 \times 6 & 4 \times 7 \end{pmatrix} = \begin{pmatrix} 0 & 5 & 0 & 10 \\ 6 & 7 & 12 & 14 \\ 0 & 15 & 0 & 20 \\ 18 & 21 & 24 & 28 \end{pmatrix}.$$

Определение 11. Линейное преобразование, которое преобразует матрицу A размера $m \times n$ в вектор-столбец размера $mn \times 1$, называется **векторизацией** и обозначается как $\text{vec}(A)$. Другими словами, чтобы векторизовать матрицу, необходимо последовательно поместить каждый из ее столбцов в один столбец:

$$\text{vec}(A) = \left(a_{11} \ \dots \ a_{m1} \ a_{12} \ \dots \ a_{m2} \ \dots \ a_{1n} \ \dots \ a_{mn} \right)^T.$$

Например, для матрицы $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, вектором будет $\text{vec}(A) = \begin{pmatrix} a \\ c \\ b \\ d \end{pmatrix}$.

Определение 12. Векторы x_1, x_2, \dots, x_r называются **линейными независимыми векторами**, если из равенства $c_1x_1 + c_2x_2 + \dots + c_rx_r = 0$ следует, что $c_1 = c_2 = \dots = c_r = 0$.

Определение 13. Векторы x_1, x_2, \dots, x_r называются **линейно зависимыми векторами**, если равенство $c_1x_1 + c_2x_2 + \dots + c_rx_r = 0$ возможно для чисел c_1, c_2, \dots, c_r , которые одновременно не равны нулю. Другими словами, по крайней мере, один вектор в данном множестве векторов может быть представлен как линейная комбинация других векторов в этом множестве.

Например, векторы $x_1 = (1, 1, 1)$, $x_2 = (-1, 2, 1)$, $x_3 = (1, 4, 3)$ линейно зависимые, так как, $2x_1 + x_2 - x_3 = (0, 0, 0)$.

Определение 14. Максимальное количество линейных свободных столбцов матрицы A размера $m \times n$ называется **рангом** матрицы A и обозначается $\text{rang}(A)$.

Для произвольной матрицы \mathbf{A} размера $m \times n$ выполняется неравенство $\text{rang}(\mathbf{A}) \leq \min(m, n)$.

Определение 15. Квадратичной формой от n переменных называется сумма, каждое слагаемое является или квадратом одного из этих переменных, или произведением двух разных переменных.

Если заданы переменные x_1, x_2, \dots, x_n , обозначая коэффициент при x_i^2 через a_{ii} , а при произведении $x_i x_j$ и $x_j x_i$ ($i \neq j$) – через a_{ij} и a_{ji} , соответственно, то квадратичную форму Q можно представить в виде:

$$\begin{aligned} Q(x_1, x_2, \dots, x_n) &= a_{11}x_1^2 + a_{12}x_1x_2 + \dots + a_{1n}x_1x_n + \dots \\ &\quad + a_{n1}x_nx_1 + a_{n2}x_nx_2 + \dots + a_{nn}x_n^2 \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j. \end{aligned}$$

Где $x_i x_j = x_j x_i$ и $a_{ij} = a_{ji}$.

Например, квадратичная форма, состоящая из двух переменных x_1, x_2 будет иметь вид:

$$\begin{aligned} Q(x_1, x_2) &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2 \\ &= a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2 \\ &= ax_1^2 + 2bx_1x_2 + cx_2^2. \end{aligned}$$

Теперь рассмотрим вопрос выражения приведенной выше квадратичной формы в векторно-матричной форме. Дадим следующие определения:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

В этом случае:

$$Q = ax_1^2 + 2bx_1x_2 + cx_2^2 = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^T \mathbf{A} \mathbf{x}.$$

В общем, последнее уравнение также может быть записано для квадратичной формы, состоящей из произвольных n переменных. То

есть:

$$Q(x_1, x_2, \dots, x_n) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j.$$

Где, \mathbf{x} – построенный из переменной вектор размера $n \times 1$, \mathbf{A} – симметричная матрица размера $n \times n$, которая называется **матрицей квадратичной формы**.

Например, требуется написать матрицу следующей квадратичной формы:

$$Q(x_1, x_2, x_3) = 2x_1^2 - 5x_2^2 + 8x_3^2 + 4x_1x_2 - 2x_1x_3 + 6x_2x_3.$$

Где, $a_{11} = 2, a_{22} = -5, a_{33} = 8, a_{12} = a_{21} = 2, a_{13} = a_{31} = -1, a_{23} = a_{32} = 3$.
Значит,

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & -1 \\ 2 & -5 & 3 \\ -1 & 3 & 8 \end{pmatrix}.$$

Определение 16. Симметричная матрица \mathbf{A} :

а) называется **положительно определенной**, если $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ для любого $n \times 1$ размерного вектора \mathbf{x} , составленного из произвольных действительных чисел, отличных от 0;

б) называется **положительно полуопределенной**, если $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ для любого $n \times 1$ размерного вектора \mathbf{x} , составленного из произвольных действительных чисел.

Например. Единичная матрица $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ положительно определена (и, как таковая, также полуопределена). Так как для вектора \mathbf{x} состоящего из действительных элементов a и b :

$$\mathbf{x}^T \mathbf{I} \mathbf{x} = \begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = a^2 + b^2 \geq 0.$$

Симметричная действительная матрица $M = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$

положительно определена. Так как, для вектора \mathbf{x} , состоящего из элементов a, b и c отличных от 0:

$$\begin{aligned} \mathbf{x}^T M \mathbf{x} &= (\mathbf{x}^T M) \mathbf{x} = \left([2a - b] \quad [-a + 2b - c] \quad [-b + 2c] \right) \begin{pmatrix} a \\ b \\ c \end{pmatrix} \\ &= (2a - b)a + (-a + 2b - c)b + (-b + 2c)c \\ &= 2a^2 - 2ab + 2b^2 - 2bc + 2c^2 \\ &= a^2 + a^2 - 2ab + b^2 + b^2 - 2bc + c^2 + c^2 \\ &= a^2 + (a - b)^2 + (b - c)^2 + c^2 > 0. \end{aligned}$$

Свойства положительно определенных и положительно полуопределенных матриц:

1. Диагональные элементы положительно определенной матрицы также строго положительны. Диагональные элементы положительной полуопределенной матрицы неотрицательны.

2. Если матрица положительно определена, она обратима, и обратная матрица также положительно определена.

3. Для произвольной матрицы X произведения $X^T X$ и XX^T будут положительно полуопределенными.

4. Если для $m \times n$ размерной матрицы X верно $\text{rang}(X) = n \leq m$, то в данном случае $X^T X$ произведение положительно определено и значит будет обратной матрицей.

Приложение 2. Основы теории вероятности

Первоначальные понятия теории вероятностей – это такие понятия, как эксперимент, событие, элементарное событие, вероятность, и мы коротко остановимся на их описании.

Определение 1. Эксперимент заключается в выполнении ряда определенных условий, благодаря которым событие происходит. Предположим, мы подбросили монету 10 раз и посчитали сторону с выпавшим гербом. Это пример эксперимента.

Определение 2. Результат эксперимента называется **событием**. В опыте, подобном описанному выше, это явление, когда выпадает «орёл» или «решка» монеты.

Выделяют три вида событий: достоверные, невозможные и случайные.

Достоверное событие – это событие, которое неизбежно произойдет в результате эксперимента.

Невозможное событие – это событие, которое никогда не происходит в результате эксперимента.

Случайное событие – это событие, которое может произойти или не произойти в результате эксперимента.

Например, был брошен игровой кубик. В таком случае:

- «очко не больше 6» достоверное событие;
- «очко равно 10» невозможное событие;
- «очко положительное число» случайно событие.

Определение 3. **Элементарное событие** - это любой результат эксперимента. Например, пусть эксперимент состоит в том, чтобы дважды бросить монету. При этом элементарные события таковы: «орёл-орёл», «орёл-решка», «решка-орёл», «решка-решка».

Определение 4. Вероятность – это число, которое количественно определяет возможность возникновения случайного события. Вероятность элементарного события A определяется как $P(A)$, где уместно соотношение $0 \leq P(A) \leq 1$.

Определение 5. Переменная, которая случайным образом принимает одно значение из набора значений экспериментального результата, называется **случайной величиной**. Например, когда монета подбрасывается 10 раз, количество выпадений стороны герба является случайной величиной, а ее набор значений состоит из целых чисел от 0 до 10.

Случайные величины обычно обозначаются заглавными буквами латинского алфавита, такими как X , Y , Z , ... а их допустимые значения обозначаются соответствующими строчными буквами x, y, z, \dots . Например, в приведенном выше примере обозначим через X количество выпадений герба, когда монета подбрасывается 10 раз. В этом случае X не зависит от какого-либо определенного значения, но мы знаем, что X принимает значение из набора $\{0, 1, 2, \dots, 10\}$. Результатом вышеупомянутого эксперимента может быть, например, $x = 6$.

Определение 6. Дискретная случайная величина – это случайная величина, множество значений которой не более чем счётно (то есть конечно или счётно). Для информации, счётное множество – бесконечное множество, элементы которого возможно пронумеровать натуральными числами.

Примеры:

1. Количество очков, выпавших на игровом кубике, представляет собой дискретную случайную величину, набор значений которой равен $\{1, 2, 3, 4, 5, 6\}$.

2. Количество дефектных продуктов в данной партии $\{4, 5, 2, 3, \dots\}$ дискретная случайная величина, которая принимает все значения от

нуля до общего количества продуктов в партии.

3. Количество выстрелов до первого попадания в цель $\{1, 5, 3, \dots\}$ случайная величина, принимающая значения из набора всех натуральных чисел.

Определение 7. Случайная величина, которая может принимать все значения в конечном или бесконечном интервале, называется **непрерывной случайной величиной**.

Примеры:

1. Дальность артиллерийского снаряда (2.5–3 км) – это непрерывная случайная величина, принимающая значения в определенном диапазоне положительных чисел.

2. Прирост массы скота за месяц (-0.5 кг, 5.3 кг, ...) – это непрерывная случайная величина, принимающая значения в определенном диапазоне чисел.

Определение 8. Закон распределения случайной величины - это совокупность всех ее значений и соответствующих им вероятностей.

Определение 9. Закон распределения дискретной случайной величины называется **функцией массы вероятности** (probability mass function). Причина такого обозначения в том, что каждое значение, которое может принять случайная величина, имеет определенную массу (вероятность). Сумма всех вероятностей равна 1.

Функцию массы вероятности можно записать следующей формулой:

$$f(x_j) = p_j, \quad j = 1, 2, \dots, k,$$

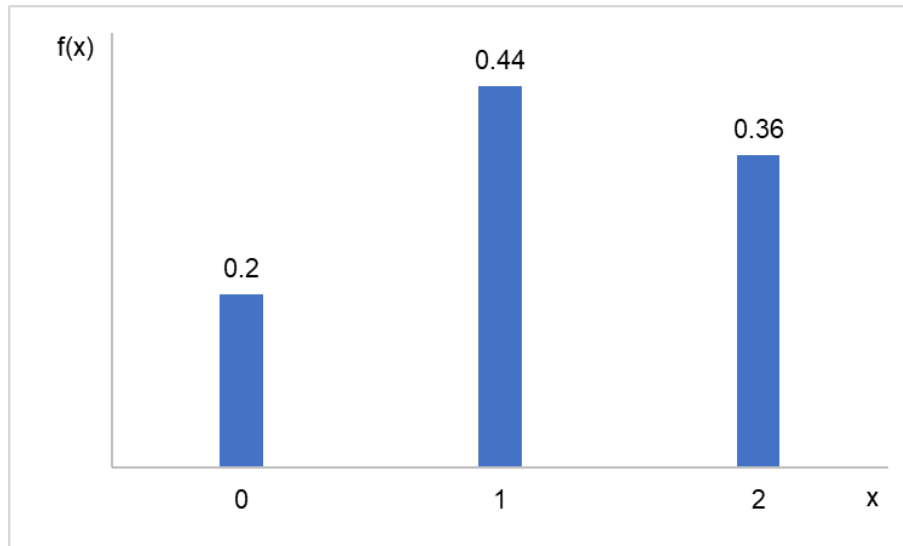
где $f(x)$ — означает вероятность, когда случайная величина X принимает какое-либо действительное значение x .

Например, пусть X будет количеством мячей, заброшенных в корзину баскетболистом за 2 попытки. В этом случае случайная

величина X может принимать значения $\{0, 1, 2\}$. Предположим, что функция массы вероятности X задается следующим образом:

$$f(0) = 0.20, \quad f(1) = 0.44, \quad f(2) = 0.36.$$

Графически это можно проиллюстрировать следующим образом:

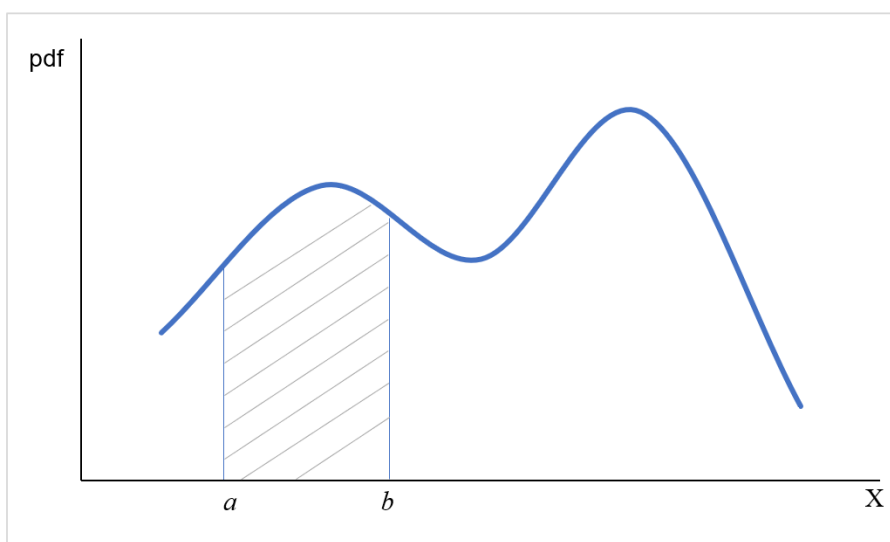


Определение 10. Закон распределения непрерывной случайной величины называется **функцией плотности вероятности** (probability density function) и записывается следующим образом:

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Из курса математического анализа известно, что определенный интеграл функции вычисляет поверхность, образованную между графиком функции и осью Ox . Следовательно, вероятность того, что непрерывная случайная величина X примет значения между точками a и b , также равна значению поверхности в интервалах a и b под функцией плотности вероятности. Вся поверхность под функцией плотности вероятности должна быть равна 1, то есть:

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$



На этом этапе давайте глубже рассмотрим термин плотность. Из физики известно:

$$\text{плотность} = \frac{\text{масса}}{\text{объем}}$$

или сколько массы (кг) соответствует на 1 единицу объема (м^3). Например, плотность воды равна 997 кг/м^3 .

Также, найдем плотность населения для территории:

$$\text{плотность населения} = \frac{\text{численность населения}}{\text{площадь территории}}.$$

Например, средняя плотность населения в Узбекистане 74.1 чел/км^2 .

Аналогично, в теории вероятности, если мы посмотрим на вероятность как на массу (как в дискретном случае), плотность указывает, сколько массы находится в интервале. Рассмотрим это с другой стороны:

Обозначим через δ небольшое положительное изменение в начале рассматриваемого интервала x , и вычислим вероятность в этом интервале:

$$P(x \leq X \leq x + \delta) = \int_x^{x+\delta} f(x) dx \approx f(x) \cdot \delta.$$

Следовательно, при достаточно малых значениях δ , вероятность (поверхность под графиком функции) приблизительно равна произведению δ (ширины прямоугольника) и значения функции плотности (длины прямоугольника). Из последнего уравнения получаем:

$$f(x) \approx \frac{P(x \leq X \leq x + \delta)}{\delta}.$$

Отсюда следует, что плотность означает вероятность (массу), соответствующую 1 единице длины отрезка. Плотность не обязательно должна быть меньше 1, потому что она не является вероятностью.

Определение 11. Математическим ожиданием случайной величины X называется взвешенное среднее по вероятностям ее возможных значений и обозначается как $E(X)$ или μ .

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \dots + x_k f(x_k) = \sum_{j=1}^k x_j f(x_j).$$

Определение 12. Дисперсией случайной величины X называется математическое ожидание квадрата ее отклонения от математического ожидания и обозначается как $Var(X)$ или σ_X^2 или σ^2 :

$$Var(X) = E[(X - \mu)^2].$$

Дисперсия показывает, насколько далека случайная величина от своего математического ожидания.

Определение 13. Стандартным отклонением случайной величины X называется положительный квадратный корень, полученный из его дисперсии, и обозначается $sd(X)$ или σ_X или σ :

$$\sigma_X = +\sqrt{Var(X)}.$$

Определение 14. Ковариация случайных величин X и Y – это математическое ожидание произведения отклонений этих случайных

величин от соответствующих средних значений и обозначается $Cov(X, Y)$ или σ_{XY} :

$$Cov[X, Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y].$$

Ковариация измеряет линейную связь между двумя случайными величинами. Положительная ковариация означает, что две случайные величины изменяются в одном направлении, а отрицательная ковариация означает, что они изменяются в противоположных направлениях. Ковариация 0 означает, что случайные величины независимы.

Тот факт, что абсолютная величина ковариации зависит от масштаба (единиц) рассматриваемых случайных величин, усложняет ее интерпретацию. Например, мы хотим изучить взаимосвязь между уровнем образования и годовым доходом. Мы можем получить образование как X , а доход как Y . Абсолютное значение коэффициента ковариации варьируется в зависимости от того, измеряем ли мы образование годами или месяцами, а доход - в сумах или других валютах. Очевидно, совершенно не имеет значения, в каких единицах мы измеряем эти случайные величины с точки зрения степени зависимости между ними.

В этой связи, степень линейной взаимосвязи между случайными величинами (коэффициент корреляции) находится путем деления ковариации случайных величин на их соответствующие стандартные отклонения (т.е. устранение проблемы различных единиц).

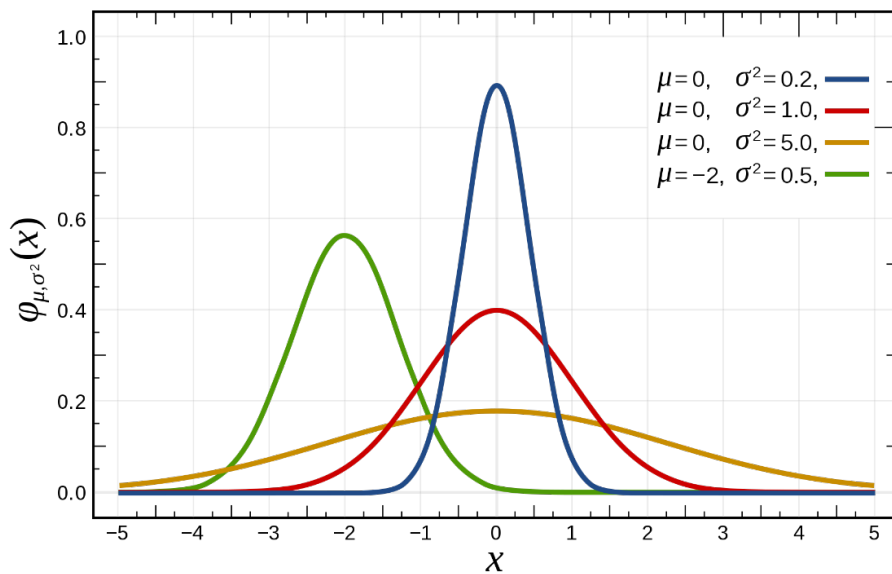
Определение 15. Коэффициенты корреляции случайных величин X и Y определяются следующими отношениями и обозначаются как $Corr(X, Y)$ или ρ_{XY} :

$$Corr(X, Y) = \frac{Cov(X, Y)}{sd(X) \cdot sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \in [-1; 1].$$

Определение 16. Непрерывная случайная величина с функцией плотности вероятности следующего вида называется **нормальной случайной величиной**:

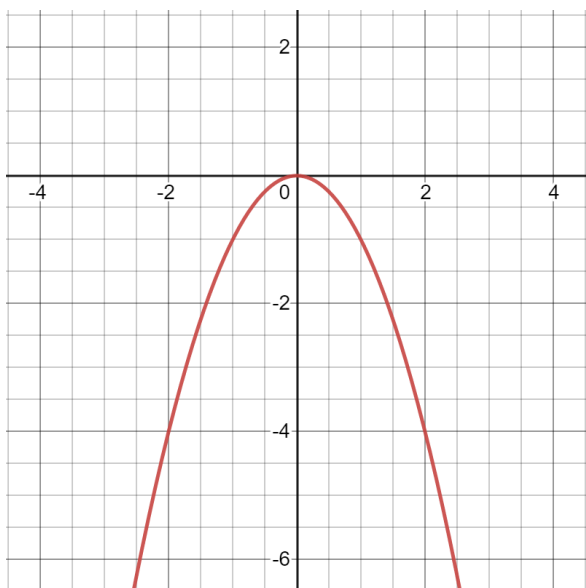
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty.$$

где μ – математическое ожидание случайной величины, σ^2 – ее дисперсия, которые называются **параметрами** закона нормального распределения. Обычно, если случайная величина X задана по закону нормального распределения со средним μ и дисперсией σ^2 обозначаются $X \sim Normal(\mu, \sigma^2)$ или короче $X \sim N(\mu, \sigma^2)$. Закон нормального распределения иногда также называют **законом распределения Гаусса** в честь известного математика К.Ф. Гаусса. Его график выглядит следующим образом:

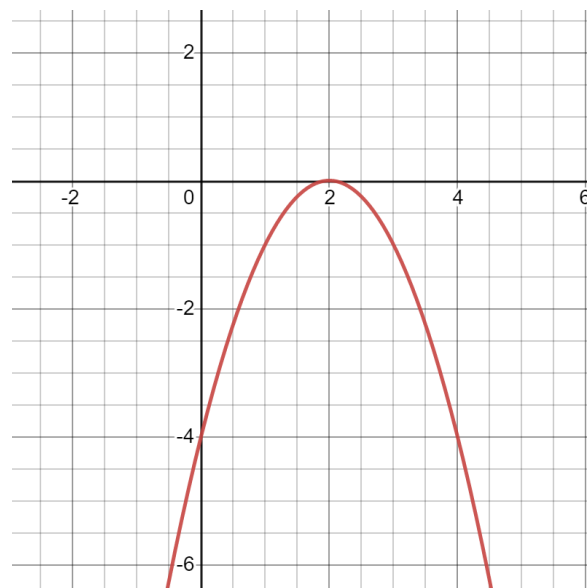


Хотя закон нормального распределения на первый взгляд может показаться сложным, его можно объяснить шаг за шагом более простым способом.

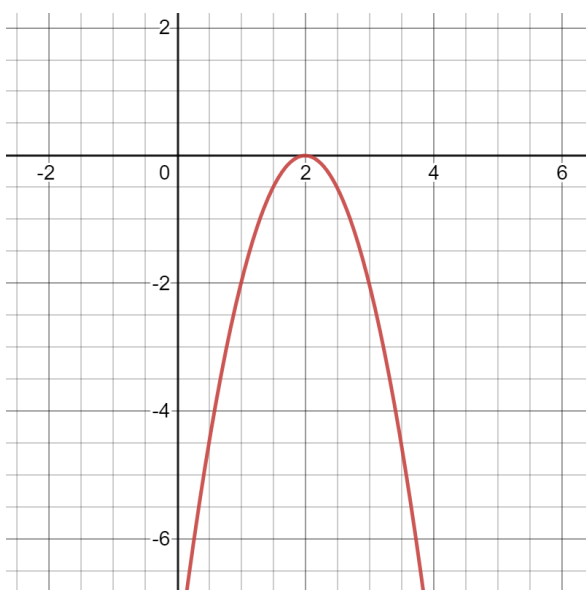
Известно, что график функции $y = x^2$ состоит из параболы. Если $y = -x^2$, ветви параболы направлены вниз. Если $y = -(x - \mu)^2$, центр параболы смещен на μ . Если $y = \frac{-(x-\mu)^2}{2\sigma^2}$, ширина ветвей параболы изменена на σ^2 . $y = \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$ обеспечивает колоколообразного графика (bell-shaped) и постоянную положительность вероятности.



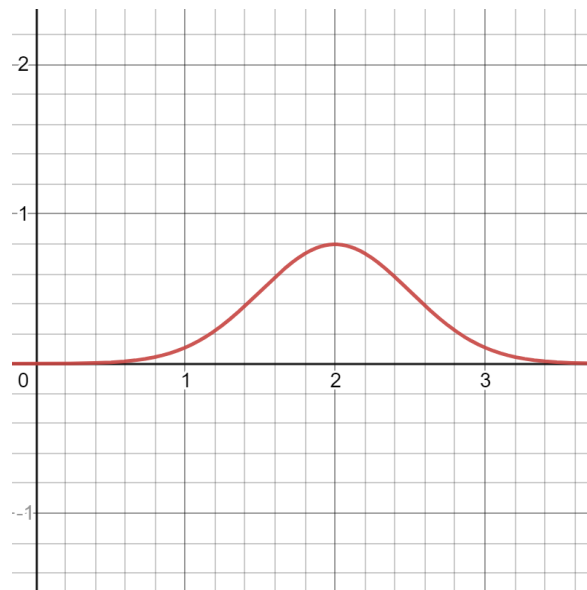
(a) $y = -x^2$



(b) $y = -(x - \mu)^2, \mu = 2$



(c) $y = \frac{-(x-\mu)^2}{2\sigma^2}, \sigma^2 = 0.25$



(d) $y = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$

Учитывая, что значение интеграла, называемого интегралом

Гаусса, равно $\sqrt{\pi}$:

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}, \quad (22)$$

также, учитывая, что значение интеграла представляет вероятность и должно быть равно 1 (мы преобразуем выражение под интегралом произведение в 1, умножив его на $\frac{1}{\sigma\sqrt{2\pi}}$ что называется нормализующим множителем):

$$\frac{1}{\sqrt{2\sigma^2} \cdot \sqrt{\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) dx = 1$$

и как в предыщем определении, мы получим функцию плотности вероятности в виде $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)$.

Закон нормального распределения важен в статистике и эконометрике. Предположение, что случайные величины заданы по закону нормального распределения, упрощает вычисления вероятностей. Также замечено, что некоторые случайные величины в жизни подчиняются закону нормального распределения. Например, рост и вес человека, результаты тестов, некоторые физические процессы и т.д.

Теперь приступим к изучению **закона нормального распределения многих переменных (ЗНМП)**. При этом мы широко используем понятия матричной алгебры (Приложение 1).

Концепция ковариационной матрицы очень важна для понимания ЗНМП. Это позволяет нам кратко сделать вывод о ковариации между всеми парами переменных.

Определение 17. Матрица размером $n \times n$ у которой элемент (i, j) равен $Cov(X_i, X_j)$ называется **ковариационной матрицей** случайного вектора $X = [X_1 \cdots X_n]^T$ и обычно обозначается $V(X)$ или Σ :

$$\Sigma = \begin{pmatrix} Cov[X_1, X_1] & Cov[X_1, X_2] & \cdots & Cov[X_1, X_n] \\ Cov[X_2, X_1] & Cov[X_2, X_2] & \cdots & Cov[X_2, X_n] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[X_n, X_1] & Cov[X_n, X_2] & \cdots & Cov[X_n, X_n] \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Здесь по диагонали ковариационной матрицы расположена дисперсия каждой случайной величины. Так как, $Cov[X_i, X_j] = Cov[X_j, X_i]$, ковариационная матрица является симметричной матрицей.

1-свойство. Ковариационная матрица произвольного случайного вектора X со средним μ может быть вычислена следующим образом:

$$\Sigma = E[(X - \mu)(X - \mu)^T] = E[XX^T] - \mu\mu^T.$$

Доказательство.

$$\begin{aligned} \Sigma &= \begin{pmatrix} Cov[X_1, X_1] & \cdots & Cov[X_1, X_n] \\ \vdots & \ddots & \vdots \\ Cov[X_n, X_1] & \cdots & Cov[X_n, X_n] \end{pmatrix} \\ &= \begin{pmatrix} E[(X_1 - \mu_1)^2] & \cdots & E[(X_1 - \mu_1)(X_n - \mu_n)] \\ \vdots & \ddots & \vdots \\ E[(X_n - \mu_n)(X_1 - \mu_1)] & \cdots & E[(X_n - \mu_n)^2] \end{pmatrix} \\ &= E \begin{pmatrix} (X_1 - \mu_1)^2 & \cdots & (X_1 - \mu_1)(X_n - \mu_n) \\ \vdots & \ddots & \vdots \\ (X_n - \mu_n)(X_1 - \mu_1) & \cdots & (X_n - \mu_n)^2 \end{pmatrix} \\ &= E \left(\begin{pmatrix} X_1 - \mu_1 \\ \vdots \\ X_n - \mu_n \end{pmatrix} \begin{pmatrix} X_1 - \mu_1 & \cdots & X_n - \mu_n \end{pmatrix} \right) \\ &= E[(X - \mu)(X - \mu)^T]. \end{aligned}$$

Для вывода последнего уравнения использовалось следующее

соотношение для произвольного вектора порядка n :

$$zz^T = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \begin{pmatrix} z_1 & z_2 & \cdots & z_n \end{pmatrix} = \begin{pmatrix} z_1 z_1 & z_1 z_2 & \cdots & z_1 z_n \\ z_2 z_1 & z_2 z_2 & \cdots & z_2 z_n \\ \vdots & \vdots & \ddots & \vdots \\ z_n z_1 & z_n z_2 & \cdots & z_n z_n \end{pmatrix}.$$

2-свойство. Ковариационная матрица произвольного случайного вектора X является положительной полуопределенной матрицей.

Доказательство. Как и в доказательстве вышеупомянутого свойства, $z = X - \mu$, и произведение zz^T для произвольной матрицы z всегда является положительным полуопределенным в соответствии со свойствами положительных полуопределенных матриц.

Определение 18. Если функция плотности вероятности случайного вектора $X = [X_1 \cdots X_n]^T$, состоящего из n случайных величин, со средней μ и ковариационной матрицей Σ имеет следующий вид, то говорят что данный случайный вектор X задан по **закону многомерного нормального распределения** и записывается как $X \sim N(\mu, \Sigma)$:

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right),$$

здесь μ – n -порядковый вектор и Σ положительная полуопределенная симметричная матрица размером $n \times n$, $|\Sigma|$ ее детерминант.

Согласно 2-свойству 17-определения, ковариационная матрица произвольного случайного вектора X будет положительной полуопределенной. Но для того чтобы иметь как в вышеуказанном определении Σ^{-1} , Σ должна быть обратимой, и, значит, ранг матрицы должен быть равен ее порядку (т.е. матрица полного ранга [full rank]). Поскольку необходимо, чтобы любая полноранговая симметричная матрица была положительно определена, Σ также симметрично положительно определена и, следовательно, является обратимой.

Тогда как в функции плотности закона нормального распределения, выражение $-\frac{1}{2\sigma^2}(x - \mu)^2$ в степени числа e является квадратичной функцией переменной x и ветви параболы направлены вниз, а выражение функции плотности ЗНМП $-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)$ квадратичная форма векторов переменной x . Так как, Σ положительно определена, обратная Σ^{-1} тоже положительно определена, или для произвольного, отличного от 0 вектора z уместно соотношение $z^T \Sigma^{-1} z > 0$. Из этого, для произвольных векторов $x \neq \mu$ вытекает нижеследующее неравенство:

$$(x - \mu)^T \Sigma^{-1}(x - \mu) > 0$$

$$-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) < 0.$$

Как и в случае одной переменной, график окончательного выражения можно представить себе как квадратичную поверхность в форме колокола, направленную вниз.

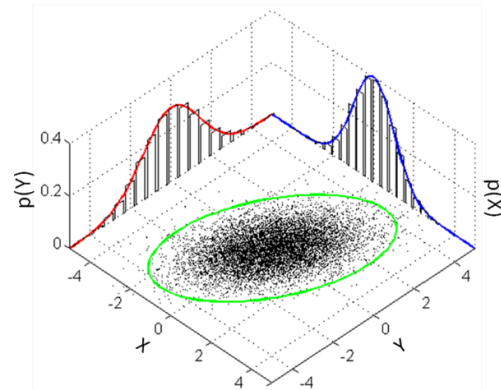
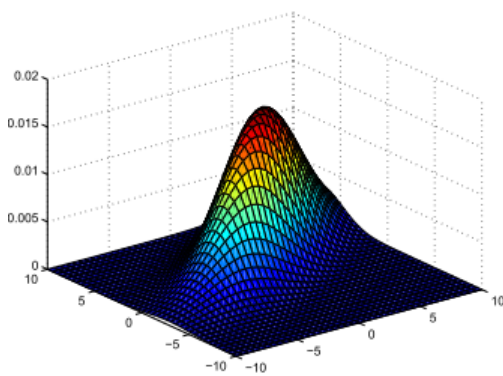


График функции плотности двух случайных величин

Кроме того, так же, как нормализующий коэффициент $\frac{1}{\sigma\sqrt{2\pi}}$ в случае одной переменной гарантирует, что значение определенного интеграла равно 1, $\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}$ перед функцией плотности ЗПНМ также гарантирует, что значение n -го интеграла равно 1:

$$\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) dx_1 dx_2 \cdots dx_n = 1.$$

Чтобы лучше понять, что такое ЗНМП, давайте рассмотрим более простой случай, то есть ситуацию, в которой количество случайных величин равно двум ($n = 2$), а ковариационная матрица Σ является диагональной матрицей.

Введем следующие обозначения:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}.$$

В этом случае, функция плотности ЗНМП будет следующего вида:

$$\begin{aligned} f(x; \mu, \Sigma) &= \frac{1}{2\pi \begin{vmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{vmatrix}^{1/2}} \exp \left[-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right] \\ &= \frac{1}{2\pi (\sigma_1^2 \cdot \sigma_2^2)^{1/2}} \exp \left[-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}^T \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix} \right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{1}{2} \begin{pmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1^2} (x_1 - \mu_1) \\ \frac{1}{\sigma_2^2} (x_2 - \mu_2) \end{pmatrix} \right] \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{1}{2\sigma_1^2} (x_1 - \mu_1)^2 - \frac{1}{2\sigma_2^2} (x_2 - \mu_2)^2 \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[-\frac{1}{2\sigma_1^2} (x_1 - \mu_1)^2 \right] \cdot \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left[-\frac{1}{2\sigma_2^2} (x_2 - \mu_2)^2 \right]. \end{aligned}$$

Из последнего уравнения видно, что функция плотности нормального распределения двух переменных равна **произведению функций плотности нормального распределения двух независимых переменных**.

Аналогичным образом можно показать, что функция плотности n переменных ЗНМП равна произведению функций плотности нормального распределения n независимых переменных.

Ясно, что приведенные выше выводы основаны на предположении, что Σ - диагональная матрица. Однако даже если Σ не диагональная, на

графике функций плотности не происходит значительного изменения²⁸.

Определение 19. Если вероятность наступления каждого из двух событий A и B не зависит от того, происходит последнее или нет, эти события называются **независимые событиями**. В противном случае говорят, что эти события взаимосвязаны.

Вероятность того, что независимые события происходят вместе, равна произведению вероятностей каждого из этих событий:

$$P(AB) = P(A) \cdot P(B).$$

Определение 20. Пусть A и B - взаимозависимые события. Когда очевидно, что событие B произошло, вероятность того, что событие A произойдет, называется условной вероятностью и обозначается $P(A|B)$:

$$P(A|B) = \frac{P(AB)}{P(B)}, \quad \text{бунда } P(B) \neq 0.$$

Аналогично, рассчитаем $P(B|A)$:

$$P(B|A) = \frac{P(BA)}{P(A)}.$$

Учитывая что, $P(BA) = P(AB)$, из последнего уравнения:

$$P(AB) = P(A) \cdot P(B|A).$$

Значит:

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B)}.$$

Последняя найденная формула называется **формулой Байеса**.

²⁸Для подробной информации: Chuong B. Do. The Multivariate Gaussian Distribution. Stanford University, 2008.

Приложение 3. Теоремы пределов

В данном приложении рассматривается одна из важных теорем теории вероятностей – **теорема центрального предела**. Для этого сначала рассмотрим следующие важные неравенства и закон больших чисел.

Неравенство Маркова. X – неотрицательная случайная величина ($X \geq 0$) и $a > 0$. В этом случае, вероятность того, что X равно или больше a , равна или меньше отношения математического ожидания X к a :

$$P(X \geq a) \leq \frac{E(X)}{a}.$$

Можно показать, что это неравенство всегда верно. Согласно определению математического ожидания:

$$E[X] = \sum_x xf(x) \geq \sum_{x \geq a} xf(x) \geq \sum_{x \geq a} af(x) = aP(X \geq a).$$

Значит, $E[X] \geq aP(X \geq a)$. Из этого $P(X \geq a) \leq \frac{E(X)}{a}$.

Неравенство Маркова связывает вероятность с математическим ожиданием. Согласно этому неравенству, если математическое ожидание случайной величины мало, то вероятность того, что X примет большие значения, также мала.

Неравенство Чебышева. Подставив в неравенстве Маркова $(X - \mu)^2$ вместо X , a^2 вместо a , получим следующее неравенство:

$$E[(X - \mu)^2] \geq P((X - \mu)^2 \geq a^2) \cdot a^2 = P(|X - \mu| \geq a) \cdot a^2.$$

Согласно определению дисперсии $E[(X - \mu)^2] = Var(X)$:

$$Var(X) \geq P(|X - \mu| \geq a) \cdot a^2. \quad (23)$$

Последнее неравенство связывает дисперсию случайной величины с вероятностью того, что она далека от своего среднего значения. Если дисперсия мала, то вероятность отклонения от

среднего также мала. Для более подробного пояснения запишем неравенство (23) следующим образом:

$$\sigma_X^2 \geq P(|X - \mu| \geq a) \cdot a^2 \implies P(|X - \mu| \geq a) \leq \frac{\sigma_X^2}{a^2}. \quad (24)$$

Из вышеуказанного неравенства видно, если a крупное число, $\frac{\sigma_X^2}{a^2}$ будет маленьким числом.

Кроме того, если дисперсия является небольшим числом, вероятность того, что случайная величина далека от своего среднего значения, также мала. По мере того, как мы удаляемся от среднего, то есть по мере увеличения a , эта вероятность уменьшается. Неравенство (24) называется **неравенством Чебышева**.

В неравенство (24) введем обозначение $a = k\sigma$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Так мы оценим вероятность того, что событие будет далеко от среднего в размере квадратичного отклонения k . Например, последнее неравенство показывает нам, какая часть группы, которая участвовала в экзамене, была далека от среднего значения результатов теста на 3 стандартных отклонения. Это может быть не более 1/9 группы.

Определение приближения в вероятности. Если для произвольного положительного числа ε верно:

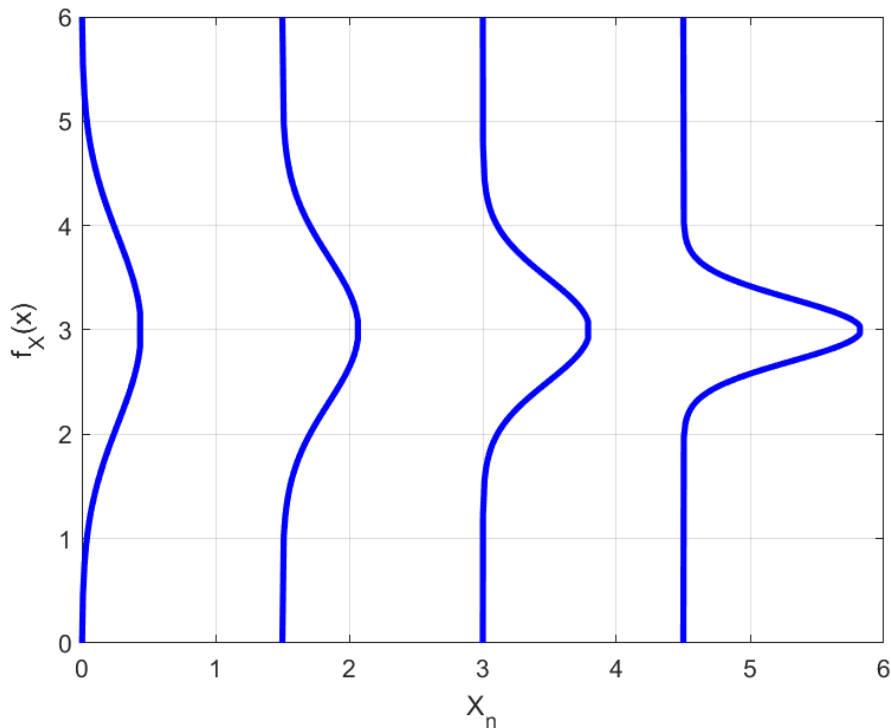
$$\lim_{n \rightarrow \infty} P(|X_n - a| \geq \varepsilon) = 0$$

тогда говорится последовательность случайной величины X_n **приближается по вероятности** к числу a .

То есть (почти все) функции массы вероятности или плотности случайной последовательности величин X_n через некоторое время концентрируются (накапливаются) вокруг числа a .

Другими словами, функция плотности члена X_i последовательности случайной величины X_n , собирается на отрезке

$[a - \varepsilon_i; a + \varepsilon_i]$. Но за пределами диапазона также есть вероятность (поверхность под функцией плотности). Эти вероятности уменьшаются с увеличением n и стремятся к 0. Мы можем увидеть это на примере последовательности нормальных случайных величин в следующем условном графике.



Из графика видно что, соответствующие функции плотности $X_i - f_X(x)$, по мере увеличения n , концентрируются вокруг точки $a = 3$, и в этом случае, например, когда $\varepsilon = 1$, вероятность принятия значений за пределами интервала $[2; 4]$ также приближается к 0.

Закон больших чисел. Пусть будет задано n свободных, имеющих одинаковые средние μ и дисперсии σ^2 случайных величин X_1, X_2, \dots, X_n . Среднее данных случайных величин:

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

где M_n считается средним выборки (sample mean) само тоже является случайной величиной. Математическое ожидание и дисперсия данной

случайной величины:

$$E[M_n] = \frac{E[X_1] + E[X_2] + \dots + E[X_n]}{n} = \frac{\mu n}{n} = \mu,$$
$$Var(M_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \quad \text{так как} \quad Var(aX + b) = a^2 Var(X).$$

Согласно неравенству Чебышева:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{Var(M_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}.$$

Если перейдем к пределу в последнем уравнении, для произвольной $\varepsilon > 0$:

$$P(|M_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0.$$

Из этого видно, согласно определению приближения к вероятности, случайная величина M_n (средняя выборки) приближается к числу μ (средней генеральной совокупности) и это называется **законом больших чисел**. Другими словами, по мере увеличения размера набора выборки его среднее значение будет более хорошей оценкой общего среднего значения совокупности.

Обозначим через S_n набор n свободных, имеющих одинаковую дисперсию σ^2 случайных величин X_1, X_2, \dots, X_n :

$$S_n = X_1 + X_2 + \dots + X_n, \quad Var(S_n) = n\sigma^2.$$

Чтобы не изменять его дисперсию даже при добавлении к этой сумме других случайных величин, разделим его на \sqrt{n} . В этом случае:

$$Var\left(\frac{S_n}{\sqrt{n}}\right) = \frac{Var(S_n)}{(\sqrt{n})^2} = \frac{n\sigma^2}{n} = \sigma^2.$$

“Стандартизируем” S_n следующим методом:

$$Z_n = \frac{S_n - E[S_n]}{\sigma_{S_n}} = \frac{S_n - nE[X]}{\sqrt{n}\sigma}$$

при этом, среднее полученной случайной величины Z_n будет равно 0, а дисперсия 1:

$$E[Z_n] = 0, \quad Var(Z_n) = 1.$$

Теорема центрального предела. Для произвольного числа c и стандартной нормальной случайной величины Z_n уместно следующее соотношение:

$$P(Z_n \leq c) \rightarrow P(Z \leq c)$$

где $Z \sim N(0, 1)$, $P(Z \leq c)$ – функция стандартного нормального кумулятивного распределения (cumulative distribution function) – $\Phi(c)$:

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

Смысл этой теоремы в том, что **распределение суммы большого числа свободных случайных величин близко к нормальному распределению**, когда выполняются очень общие условия. Другими словами, когда у нас есть какое-либо «шумное» событие, а «шум», который мы наблюдаем, состоит из множества случайных событий, которые не зависят друг от друга, общий эффект, который мы наблюдаем, может быть охарактеризован нормальным случайным распределением. Например, мы можем предположить, что движения цен могут быть связаны с небольшими решениями и другими небольшими событиями многих различных экономических агентов, вовлеченных в рынок. Следовательно, *распределение товарных цен может быть хорошо описано нормальными случайными величинами.*

Приложение 4. Некоторые важные распределения вероятности.

Определение 1. Для произвольного $\alpha > 0$ заданная функция, использующая следующую интегральную формулу, называется **Гамма-функцией** и обозначается как $\Gamma(\alpha)$:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx. \quad (25)$$

Рассмотрим следующее важное свойство гамма-функции. Известно, что при $\alpha = 1$:

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = -e^{-x} \Big|_0^{\infty} = -0 + 1 = 1.$$

При $\alpha > 1$ через интегрирование по частям:

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{\infty} x^{\alpha-1} e^{-x} dx = \int_0^{\infty} x^{\alpha-1} d(-e^{-x}) \\ &= x^{\alpha-1} (-e^{-x}) \Big|_0^{\infty} - \int_0^{\infty} (-e^{-x}) (\alpha-1)x^{\alpha-2} dx \\ &= (\alpha-1) \int_0^{\infty} x^{(\alpha-1)-1} e^{-x} dx = \boxed{(\alpha-1)\Gamma(\alpha-1)}. \end{aligned}$$

Используя вышеизложенное, мы получаем следующее для значений гамма-функции в натуральных числах:

$$\Gamma(2) = 1 \cdot 1, \Gamma(3) = 2 \cdot 1, \Gamma(4) = 3 \cdot 2 \cdot 1, \Gamma(5) = 4 \cdot 3 \cdot 2 \cdot 1$$

и согласно индукции

$$\boxed{\Gamma(n) = (n-1)!}$$

Следовательно, гамма-функция будет расширением факториальной функции до положительных действительных чисел.

В частности, при $n = 1/2$ путем подстановки $x = u^2$ и используя значение интеграла Гаусса из выражения (22), получаем следующее:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} \frac{e^{-x}}{\sqrt{x}} dx = 2 \int_0^{\infty} e^{-u^2} du = 2 \cdot \frac{\sqrt{\pi}}{2} = \sqrt{\pi}.$$

Используя это, мы получим следующее, используя рекуррентную формулу:

$$0.5! = \Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2} \approx 0.8862.$$

Теперь разделим обе стороны равенства (25) на $\Gamma(\alpha)$ и введем обозначение $x = \beta y$ ($\beta > 0$):

$$1 = \int_0^{\infty} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} dx = \int_0^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} dy.$$

В этом случае функция

$$f(x|\alpha, \beta) = \begin{cases} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (26)$$

будет функцией плотности вероятности, потому что она неотрицательна и её интеграл равен 1.

Определение 2. Распределение с функцией плотности вероятности $f(x|\alpha, \beta)$ как и в уравнении (26), называется **гамма-распределением**, заданным параметрами α и β , и обозначается как $\Gamma(\alpha, \beta)$.

1-свойство. Если последовательность случайных величин X_1, X_2, \dots, X_n задана следующими гамма распределениями,

$$X_1 \sim \Gamma(\alpha_1, \beta), X_2 \sim \Gamma(\alpha_2, \beta), \dots, X_n \sim \Gamma(\alpha_n, \beta)$$

то их сумма $X_1 + X_2 + \dots + X_n$ имеет распределение вида $\Gamma(\alpha_1 + \alpha_2 + \dots + \alpha_n, \beta)$ ²⁹.

2-свойство. Для случайной величины $X \sim \Gamma(\alpha, \beta)$ и отличного от 0 действительного числа с уместно следующее соотношение³⁰:

$$cX \sim \Gamma\left(\alpha, \frac{\beta}{c}\right).$$

Определение 3. Если заданы свободные, имеющие одинаковые средние и дисперсию n стандартных нормальных случайных величин X_1, X_2, \dots, X_n ,

$$X_1^2 + X_2^2 + \dots + X_n^2$$

заданное со степенью свободы суммы n называется **распределение χ_n^2** (распределение хи-квадрат).

Теорема. $n\sigma^2$ дисперсия, свободных, с одинаковым средним и дисперсией σ^2 n случайных величин X_1, X_2, \dots, X_n подчиняется распределению χ_{n-1}^2 заданному со степенью свободы $(n - 1)$ ³¹, или

$$n\sigma^2 = n \cdot (E[X^2] - E[X]^2) \sim \chi_{n-1}^2.$$

3-свойство. Распределение χ_n^2 является гамма распределением с параметрами $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$. Правильность этого мы покажем следующим образом.

Если дано нормальная стандартная случайная величина $X \sim N(0, 1)$, найдем распределение случайной величины X^2 . Кумулятивная функция распределения X^2 выглядит следующим

²⁹Смотрите для доказательства: <https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture6.pdf>

³⁰Смотрите для доказательства: https://sites.stat.washington.edu/thompson/S341_10/Notes/week4.pdf, 10.4-(ii)

³¹Смотрите для доказательства: <https://ocw.mit.edu/courses/mathematics/18-443-statistics-for-applications-fall-2006/lecture-notes/lecture5.pdf>

образом:

$$\mathbf{P}(X^2 \leq x) = \mathbf{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Воспользовавшись тем что функция плотности случайной величины равна производной кумулятивной функции ее распределения, найдем функцию плотности X^2 :

$$\begin{aligned} f_{X^2}(x) &= \frac{d}{dx} \mathbf{P}(X^2 \leq x) = \frac{d}{dx} \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{x})^2}{2}} (\sqrt{x})' - \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{x})^2}{2}} (-\sqrt{x})' \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} e^{-\frac{x}{2}} = \boxed{\frac{1}{\sqrt{2\pi}} x^{\frac{1}{2}-1} e^{-\frac{x}{2}}}. \end{aligned}$$

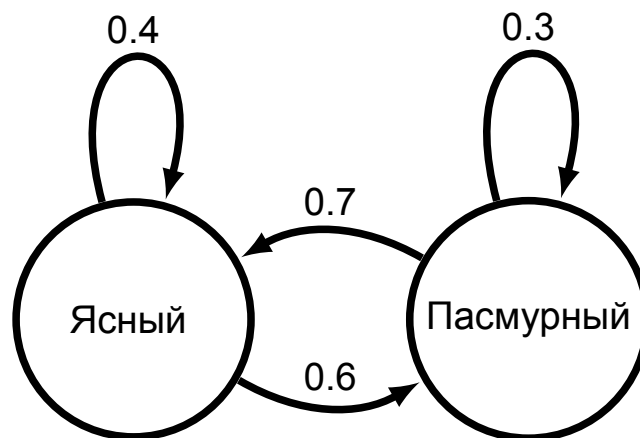
Можно увидеть что, последнее выражение является функцией плотности гамма распределения с параметрами $\Gamma(\frac{1}{2}, \frac{1}{2})$, или $X^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$. Согласно вышеприведенного 1-свойства

$$\boxed{X_1^2 + X_2^2 + \dots + X_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)}.$$

Приложение 5. Цепь Маркова

Цепь Маркова – это математическая система, которая переходит из одного состояния в другое в соответствии с определенными вероятностными правилами.

Мы объясним цепь Маркова на следующем простом примере. В любой день может быть одна из двух ситуаций: ясный день или пасмурный день. Предположим, сегодняшняя погода зависит только от вчерашней погоды. Также переход от одного случая к другому должен полностью контролироваться следующими четырьмя вероятностями:



Например, если сегодня ясный день, вероятность того, что завтра будет пасмурный, составляет 0,6 согласно диаграмме выше. Точно так же вероятность наступления солнечного дня после пасмурного дня или снова пасмурного дня составляет 0,7 и 0,3 соответственно, согласно диаграмме. **Набор состояний** {“Ясный”, “Пасмурный”}, который состоит из общего числа случаев, которые могут произойти в этой цепи Маркова, называется **текущее состояние** может принимать значение “Ясный” или “Пасмурный”.

Обозначим состояние в момент времени t через X_t . Цепь Маркова будет следующим процессом:

$$X_0, X_1, X_2, \dots, X_t, \dots$$

На практике цепь Маркова может состоять из многих случаев. Кроме того, наиболее важной особенностью цепи Маркова является то, что в ней **произвольное состояние считается зависящим только от предыдущего состояния**, и это называется свойством Маркова, т.е.

$$P(X_{t+1}|X_t, X_{t-1}, \dots, X_1, X_0) = P(X_{t+1}|X_t).$$

Матрица, состоящая из вероятностей перехода из состояния в состояние, называется **матрицей перехода**:

	Ясный	Пасмурный
Ясный	0.4	0.6
Пасмурный	0.7	0.3

Например, чтобы узнать вероятность, будет ли завтра “Ясный” день, если сегодня солнечный “Пасмурный”, мы смотрим на число на пересечении строки “Пасмурный” и столбца “Ясный” матрицы.

Предположим, сегодня (день 0) погода “Ясная”. Мы рассматриваем вероятность “Ясной” или “Пасмурной” погоды завтра как средневзвешенное значение вероятности текущей ситуации и перехода на следующий день соответственно:

$$\begin{aligned} P(\text{Ясный}_1|\text{Ясный}_0) &= P(\text{Ясный}_0) \cdot P(\text{Ясный}|\text{Ясный}) + \\ &+ P(\text{Пасмурный}_0) \cdot P(\text{Ясный}|\text{Пасмурный}) \\ &= 1 \cdot 0.4 + 0 \cdot 0.7 = 0.4, \end{aligned}$$

$$\begin{aligned} P(\text{Пасмурный}_1|\text{Пасмурный}_0) &= P(\text{Ясный}_0) \cdot P(\text{Пасмурный}|\text{Ясный}) + \\ &+ P(\text{Пасмурный}_0) \cdot P(\text{Пасмурный}|\text{Пасмурный}) \\ &= 1 \cdot 0.6 + 0 \cdot 0.3 = 0.6. \end{aligned}$$

Используя вышеизложенное, можно записать вероятность погоды на каждый день в форме таблицы:

	X_0	X_1	X_2	...	
$P(\text{Ясный})$	1	0.4	0.58	...	τ_1
$P(\text{Пасмурный})$	0	0.6	0.42	...	τ_2

Если аналогичным образом рассчитаем вероятность 2-дневной погоды:

$$P(\text{Ясный}_2 | \text{Ясный}_1) = 0.4 \cdot 0.4 + 0.6 \cdot 0.7 = 0.58,$$

$$P(\text{Пасмурный}_2 | \text{Пасмурный}_1) = 0.4 \cdot 0.6 + 0.6 \cdot 0.3 = 0.42.$$

Возникает естественный вопрос, если продолжим рассчитывать вероятности погоды для следующих дней, приблизятся ли данные вероятности к определенному числу, или постоянному состоянию – τ_1 и τ_2 (steady state). Если это так, через определенное время вероятности перехода из одного состояния в другое станут неизменными и будут равны τ_1 и τ_2 соответственно, или

$$\begin{cases} \tau_1 \cdot P(\text{Ясный} | \text{Ясный}) + \tau_2 \cdot P(\text{Ясный} | \text{Пасмурный}) = \tau_1 \\ \tau_1 \cdot P(\text{Пасмурный} | \text{Ясный}) + \tau_2 \cdot P(\text{Пасмурный} | \text{Пасмурный}) = \tau_2 \end{cases} \implies$$

$$\begin{cases} 0.4\tau_1 + 0.7\tau_2 = \tau_1 \\ 0.6\tau_1 + 0.3\tau_2 = \tau_2 \end{cases} \implies \begin{cases} 0.6\tau_1 = 0.7\tau_2 \\ 0.6\tau_1 = 0.7\tau_2 \end{cases}$$

Поэтому мы берем из нее одно уравнение и пользуясь тем фактом, что сумма τ_1 и τ_2 должна быть равна 1 согласно свойству вероятности, мы создаем следующую систему уравнений:

$$\begin{cases} 0.6\tau_1 = 0.7\tau_2 \\ \tau_1 + \tau_2 = 1 \end{cases} \implies \begin{cases} \tau_1 = \frac{7}{13} \\ \tau_2 = \frac{6}{13} \end{cases}$$

Следовательно, цепь Маркова вне зависимости от начального распределения состояний (в нашем примере $X_0 = \{P(\text{Ясный}) = 1, P(\text{Пасмурный}) = 0\}$), через определенное время достигает постоянного распределения состояний (stationary distribution) — $X = \{\tau_1, \tau_2\}$ и остается в нем на протяжении оставшегося периода. Следует отметить, что не все цепи Маркова также могут иметь постоянное распределение.

Приложение 6. Метод симуляций Монте Карло

Метод Монте-Карло – это широкий класс вычислительных алгоритмов, основанных на повторной случайной выборке для получения результатов в числовой форме. Основная идея метода – использование случайности при решении задач, которые сложно или невозможно решить другими методами. Проиллюстрируем применение этой идеи на следующем простом примере.

Задача. *Какова вероятность того, что стрела, выпущенная в квадратную цель с длиной стороны 2, коснется круга, нарисованного внутри квадрата, когда стрела точно попадет в цель?*

A – событие попадание стрелы в цель. Согласно геометрическому определению вероятности, запрашиваемая вероятность определяется как отношение площади круга к площади квадрата (радиус круга равен половине длины стороны квадрата $R = a/2 = 1$):

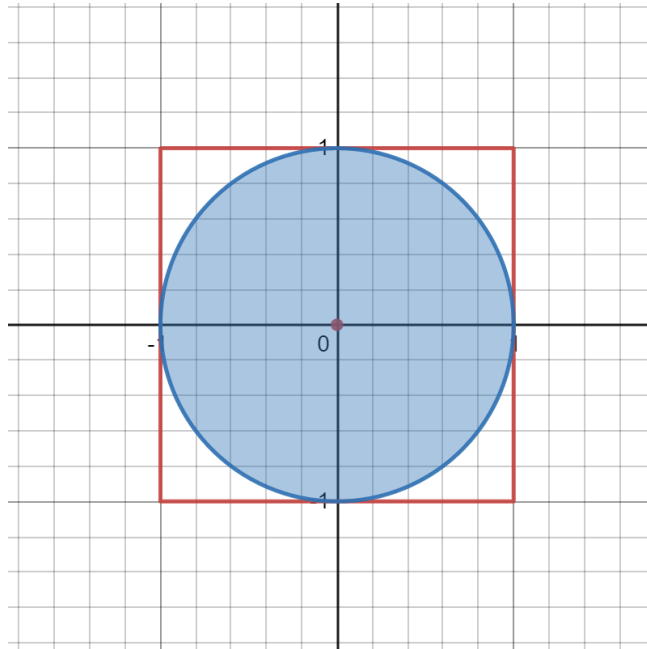
$$P(A) = \frac{\text{площадь круга}}{\text{площадь квадрата}} = \frac{\pi R^2}{a^2} = \frac{\pi}{4}.$$

Этот метод представляет собой аналитический способ решения данной проблемы, и для его решения требуются знания о вычислении площадей форм, как указано выше. Предположим, мы этого не осознаем и хотим решить проблему только путем проведения большого количества повторных экспериментов. Это в точности метод Монте-Карло. Для этого мы создаем случайную точку внутри квадрата с помощью компьютерной программы. Если образовавшаяся точка находится в круге, мы принимаем ее, в противном случае мы ее отвергаем. Данный эксперимент повторяем много раз, и согласно закону больших чисел³² получаем результат очень близкий действительной вероятности.

Подробное описание алгоритма (на примере языка программирования Python) приводится ниже, в нем для проверки

³²Смотрите: Приложение 3.

нахождения точки внутри круга используется неравенство – $(x - x_0)^2 + (y - y_0)^2 \leq R^2$, также за центр круга принята точка центра системы координат $(x_0, y_0) = (0, 0)$:



```
# Загружаем необходимую функцию для генерации
# равномерно распределенной случайной величины
from random import uniform

# Число экспериментов
N = 10000000

# Счётчик для подсчёта точек, попавших в окружность
circ = 0

# Повторение эксперимента N-раз
for i in range(1, N+1):
    # Создаём произвольную точку внутри квадрата
    x = uniform(-1, 1)
    y = uniform(-1, 1)
```

```
# Если точка находится внутри круга,  
# увеличиваем счётчик на 1  
if x**2 + y**2 <= 1: #Неравенство круга  
    circ += 1
```

Результаты эксперимента показывают, что около 78,5% точек внутри квадрата, находятся внутри окружности и это практически одинаково с вероятностью, рассчитанной аналитическим путем:

$$P_1 = \frac{circ}{N} = 0.7853951, \quad P_2 = \frac{\pi}{4} = 0.7853981, \quad \frac{P_1}{P_2} = 0.999996.$$

Список литературы

1. Blake A. and Mumtaz H., *Applied Bayesian Econometrics for Central Bankers*, Center for Central Banking Studies, Bank of England.
2. Ciccarelli M. and Rebucci A., 2003, *Bayesian VARs: A Survey of the Recent Literature with an Application to the European Monetary System*, International Monetary Fund.
3. Carriero A., 2018, *Bayesian VARs*.
4. Kadiyala K.R. and Karlsson S., 1997, *Numerical Methods for Estimation and Inference in Bayesian VAR-models*, Journal of Applied Econometrics, Vol. 12, pp. 99-132.
5. Greene H. W., 2002, *Econometric Analysis, Fifth Edition*, New York University.
6. Wooldridge J.M., 2016, *Introductory Econometrics, Sixth Edition*, Michigan State University.
7. Bertsekas D.P. and Tsitsiklis J.N., 2008, *Introduction to Probability, Second Edition*, Massachusetts Institute of Technology.
8. Rasulov A.S., Raimova G.M. va Sarimsakova X.K., 2006, *Ehtimollar nazariyasi va matematik statistika*, O'zbekiston faylasuflari milliy jamiyati nashriyoti.
9. Mamurov E. va Adirov T., 2005, *Ehtimollar nazariyasi va matematik statistika*, Toshkent moliya instituti.
10. Konstantopoulos T, *One Hundred Solved Exercises for the Subject: Stochastic Processes I*, www.docplayer.net.
11. Fewster R., 2013, *Markov Chains*, The State University of Auckland, New Zealand.

© Центральный банк Республики Узбекистан, 2021

Подготовлено Департаментом денежно-кредитной политики

Предложения и замечания можно направить по адресу:

E-mail: achilov@cbu.uz

Тел.: (+998) 71 212-60-22