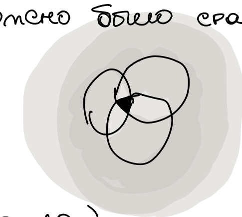
Ω - пр-во элементарных событий
(дискретно / конечно / сч. / беск. / непрерывно)

$\mathcal{A}$ – семейство подмн-в Ω (алгебра)
замкнуто относит. операции объедин

Булева алгебра

P – измеримая ф-я (вероятностная мера)

$f: \mathbb{R}^1 \to \mathbb{R}^1$
$P: \mathcal{A} \to \mathbb{R}^1$, чтобы можно было сравнить

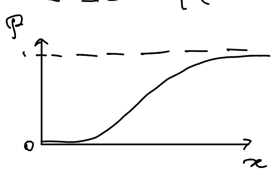


1. $P(\varnothing)=\varnothing$
2. $P(a) \geq 0$
3. $P(\Omega)=1$
4. $a_1, a_2 \in \mathcal{A} \Rightarrow P(a_1)+P(a_2) \geq P(a_1 \cup a_2)$

Основные задачи:

$\exists\ \Omega = \{\omega_1, \dots, \omega_n\} \Rightarrow P := (p_1, p_2, \dots, p_n)$

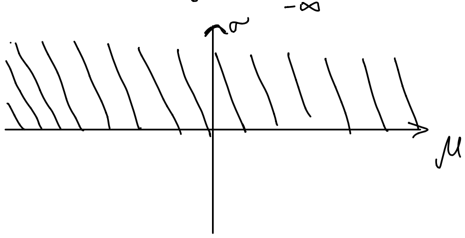
$\exists\ \Omega = \{(-\infty; x_1]; \dots\}$



Пространство изм. ф-й   Гауссовская модель.

cdf

$$F_\xi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$



$P(a)$

Плотность распр-я $\exists\ \frac{dF(x)}{dx} =: f(x)$ – p.d.f

Решение задач мат. статистики:

1) Оценка $\hat{P}$
2) Построение решающих правил

$F_\xi(x, \vec{\Theta})$

↑вектор параметров

имеем выборку (набор данных): $\{s_1, \dots, s_n\} \to \hat{\vec{\Theta}}$

↳оценка $\ne \vec{\Theta}$

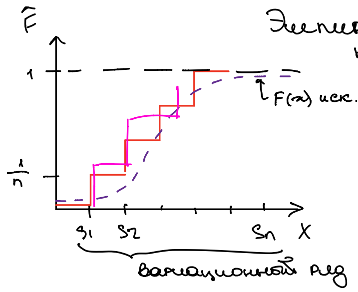
функции от выборки (статистики)

Св-ва оценок:

1. несмещённость: $M(\hat{\vec{\Theta}}) = \vec{\Theta}$ — интегральный оператор свёртки $\int_{sup} x\, d(F(x))$
2. состоятельность
3. Эффективность: обладает минимальной дисп.

Усреднение в МО — метод моментов

при большом кол-ве наблюдений оценки распр-на нормально



Эмпирич. ф-я распр-я непараметрич.

— эл-т последовательности

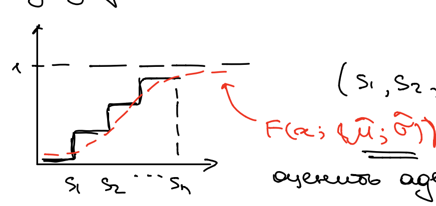
Статическая гипотеза

нулевая гипотеза H_0

альтернативная гипотеза H_a

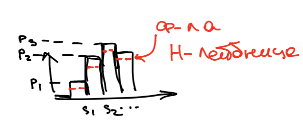
Результат проверки гипотезы — случайная величина.

Критерий ∀ статист. гипотезы в основе своей наз-ся статистике из заданного распр-я



$(s_1, s_2, \dots, s_n) \to \hat{\mu}, \hat{\sigma}$

оценить адекватность этой статистики



① ∃ Носитель меры ξ на подмн-ве, где задана выборка.

Потом строим непараметрич. оценку пл-ти распр-я (гистограмма)

Вектора: m интервалов → вектор из m эл-тов $(P_1, P_2, \dots P_m)$

② строим параметрич. $(\hat{P}_1, \hat{P}_2, \dots, \hat{P}_m)$
(построен на основе выборки) (случ. вел-ны)

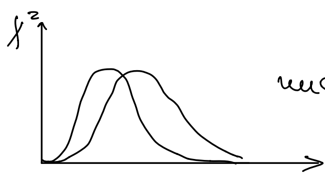
Оцениваем меру близости двух векторов (расстояние)
=> исп. случ. меру.
Строим оценку новой статистики с заданным распр-м

χ^2-критерий

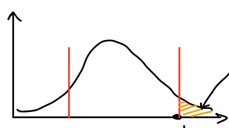
$$\frac{\sum_{i=1}^{m} (\hat{P}_i - P_i)^2}{m}$$

$\hat{P}_i \sim N(P_i, ?)$ потенциальн.
└ чтобы центрир. получить

?

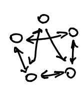


число степеней свободы



Как выбрать α?

Эконометрика

 → $O \in \mathbb{R}^m$

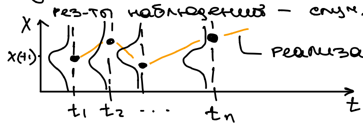
$o = (X_1, X_2, \dots, X_m)$ — $\boxed{\text{проекция системы в мн. подпр-во.}}$
┌ связи при их участников
└ кол-во участников рынка
└ объемы сделок
└ валюта

Если хотим рассматривать динамич. систему: $o = (X_1(t), X_2(t), \dots, X_m(t))$ —
— фазовое пр-во.

Временной ряд —

упорядоченная послед-ть наблюдений
$X(t_1), X(t_2), \dots, X(t_n)$

Случайная ф-я (арг. не случ.)



Сможем определить эволюцию модели т. и т.п. или сможем задать закон распр-я

Классификация случайных ф-й.

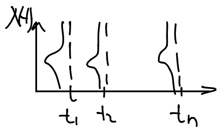
Ключевой признак

структура (топология) обл. опр-я с.ф. : t – дискр. зн-я => сл-я послед-ть.

структура (топология) обл. значений с.ф.: $X(t)$ – дискр. знач.

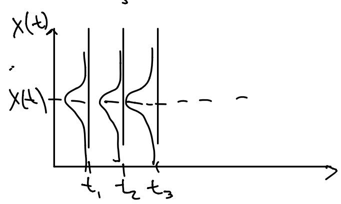
Классификация на основе инвариант

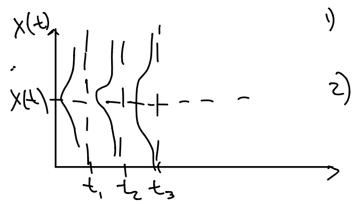
1. инвариантность сл. меры



в сечении всегда одно и то же распр.

2. инвариантность двух основных хар-к сл. меры моментной ф-ии, ковариационной ф-ии.



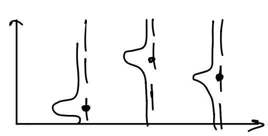


1) $M[X(t)] = const$

2) $K(t_1, t_2) = K(t_2 - t_1) = K(\tau)$

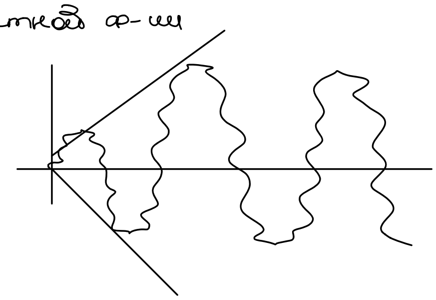
Нестационарные случ. п.

1. Нестационарные в смысле первой моментной ф-ии



мат. ожидание изм.
остальные моменты сохр.

2. Нестационарные в смысле $2^{ой}$ моментной ф-ии
Пр: Броуновское движение



Мат. ожидание сохр. инвариантой
Дисперсия изменяется

JSL
C/C++ / Java
Прилож., генерируются данные
+ считыватель файла
SVD single value
регуляризация решения

Полиномы до $3^{го}$ порядка

4 лаб. работы
1 ~ 2 недели

Детерминистические модели
(экзогенные /...)

$\exists$ случ. ф-я $V(t) \equiv \varnothing$

$\exists$ $W(t) = I\,(t = \tau : F_\tau(x))$
индик. ф-я — непрерывное распр-е

стохастич. экв.
(не совпад на $\leq$ сч.)
с лебеговой мерой $= 0$

Стохастические модели

1) операторы для отобр-е

2) связать нестационарность только с одной хар-ой процесса
(диспер $=$ const / мат. ожид. $=$ const)

$\exists$ классификации нестацион. процессов на основе их структуры.

I аддитивные модели: ордината = сумма ординат простых процессов.

$$X(t) = g(t) + h(t)$$

II мультипликативные модели: ордината = произв. орд простых пр.

$$X(t) = g(t) * h(t)$$

III мультипликат.-аддитивные модели: композиции

$$X(t) = g(t) + h(t) f(t)$$

IV квазипериодические модели: $F_x(x, t + dt) = F_x(x, t) + \delta(t)$ — б.м.в.

$\mathfrak{F}$ Простейший класс аддитивных моделей:

линейные регрессионные модели

$$X(t) = g(t) + \mathcal{E}(t)$$
нестац. случ. процесс

$\exists$ $g(t)$ — первомоментная ф-я процесса
поведение н.с.п. в среднем

типо $M[X(t)] = M[g(t)]$

$\exists$ $g(t)$ — детерминированная ф-я

$\mathcal{E}(t)$ — стохастическое слагаемое;
— случ. ф-я

$M[\mathcal{E}(t)] = \varnothing$

$D[\mathcal{E}(t)] = \sigma^2 < \infty$
физич. смысл.

Больше никаких предполож. о распр-ии $F_\mathcal{E}(e, t)$

Потребуем $g(t) \in L_2$ $\varepsilon(t) \in L_2$

В силу предположения об инерцион-ом хар-ре изм-я экономич. системы в среднем $g \in C^{(0)}[t_1; t_n]$

Решаем некорректно пост. задачу (у нас есть сумма, а хотим найти слагаемое)

Нужно $\hat{g}(t)$ и $\hat{\varepsilon}(t)$ $\varepsilon \sim L(0, \underbrace{\hat{\sigma}}_{\text{ищем}})$

Решим в параметрич. постановке.

Выберем базис в $C^{(0)}[t_1, t_n]$ из полиномов и построим $\hat{g}(t)$, как их композицию.

$$\underbrace{\hat{g}(t)}_{\substack{\text{объясняемая}\\ \text{переменная}}} = \beta_0 \underbrace{t^0 + \beta_1 t^1 + \ldots + \beta_k t^k}_{\substack{\text{переменная регрессов}\\ \text{(объясняющая переменная)}}}$$

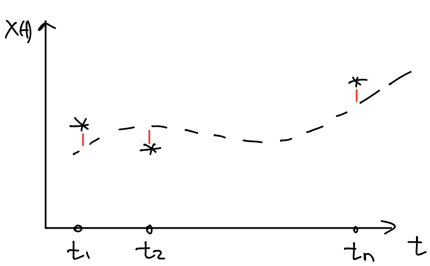
$$\overbrace{g(t)}^{\text{тренд}} = \vec{\beta}^T \cdot \vec{t} \leftarrow \text{ск. пр-е}$$

$$\vec{\beta} = \begin{pmatrix} \beta_0 \\ \cdots \\ \beta_k \end{pmatrix} \quad \vec{t} = \begin{pmatrix} t^0 \\ \cdots \\ t^k \end{pmatrix} \qquad Z = \begin{pmatrix} 1 & 1 & & 1 \\ t_1 & t_2 & & t_n \\ \cdots & \cdots & \cdots & \cdots \\ t_1^k & t_2^k & & t_n^k \end{pmatrix}$$

$k < n$

$$\vec{X}(t) = \vec{Z}^T \vec{\beta} + \vec{\varepsilon}(t) \qquad \vec{\varepsilon}(t) = \begin{pmatrix} \varepsilon(t_1) \\ \cdots \\ \varepsilon(t_n) \end{pmatrix}$$

слагаемые независимы $\Rightarrow$ можем идти посл-но



$L_2 \to l_2$

Расстояние функциональная $S(\hat{\beta}) := \sum_{i=1}^{n} (X(t_i) - \hat{g}(t_i))^2$

Параболоид в k-мерном евклидовом пр-ве.
имеем задачу выпуклого программирования.

Решение 1-й задачи: (усл. Каша)

$$\left\{\frac{\partial S}{\partial \beta} = \emptyset\right\} \leftarrow \text{условие } \exists \text{ глобального оптимума}$$

$$\frac{\partial}{\partial \beta_0}\left(X(t_1) - \beta_0 t_1^0 - \beta_1 t_1^1\right)^2 + \left(X(t_2) - \beta_0 t_2^0 - \beta_1 t_2^1\right)^2 \neq 0$$

$$\sum_{i=1}^{n} \beta_0 - \sum_{i=1}^{n} X(t_i) + \sum_{i=1}^{n} \beta_1 t_i = \emptyset \Rightarrow \hat{\beta}_0 = \frac{\sum_{i=1}^{n} X(t_i) - \sum_{i=1}^{n} \beta_1 t_i}{n} =$$

$$= \overline{X}(t) - \beta_1 \overline{t}$$

$$X_t \cdot z_t - \beta_0 \sum z_t$$

$$\frac{\partial}{\partial \beta_1} S(\beta) = \frac{\partial}{\partial \beta_1} \sum_{i=1}^{n} \left(X(t_i) - \beta_0 - \beta_1 \cdot t_i\right)^2 =$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X(t_i) \cdot t_i - \beta_0 \sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} t_i^2} \qquad \frac{\partial}{\partial \vec{\beta}} = \begin{pmatrix} \frac{\partial}{\partial \beta_0} \\ \cdots \\ \frac{\partial}{\partial \beta_{k-1}} \end{pmatrix}$$

$$S(\vec{\beta}) = (\vec{X} - z^T\vec{\beta})^T(\vec{X} - z^T\vec{\beta}) \Rightarrow$$

$$\Rightarrow \frac{\partial}{\partial \vec{\beta}} S(\vec{\beta}) = \frac{\partial}{\partial \vec{\beta}}(\vec{X} - z^T\vec{\beta})^T[\vec{X} - z^T\vec{\beta}] + [\vec{X} - z^T\vec{\beta}]^T \frac{\partial}{\partial \vec{\beta}}(\vec{X} - z^T\vec{\beta}) =$$

$$= (\vec{\emptyset} - z^T\vec{1})^T[\vec{X} - z^T\vec{\beta}] + [\vec{X} - z^T\vec{\beta}]^T(\vec{\emptyset} - z^T\vec{1}) =$$

$$\vec{1} := \frac{\partial \vec{\beta}}{\partial \vec{\beta}} \qquad = -\vec{1}^T z[\vec{X} - z^T\vec{\beta}] + [\vec{X}^T - \vec{\beta}^T z](-z^T\vec{1}) =$$

$$= -\vec{1}^T z \vec{X} + \vec{1}^T z z^T \vec{\beta} - \vec{X}^T z^T \vec{1} + \vec{\beta}^T z z^T \vec{1} =$$

$$= -\vec{1}^T z \vec{X} + \vec{1}^T \underbrace{z z^T}_{\text{симм. кв. м. ранга } k \text{ положит. опр.}} \vec{\beta} - (z X)^T \vec{1} + \vec{1}^T z^T z \vec{\beta} =$$

$$= -\vec{1}^T z \vec{X} + \vec{1}^T z z^T \vec{\beta} - \vec{1}^T z \vec{X} + \vec{1}^T z^T z \vec{\beta} =$$

$$= -2\left[\vec{1}^T z \vec{X}\right] + 2\left[\vec{1}^T z z^T \vec{\beta}\right] = -\vec{1}^T z \vec{X} + \vec{1}^T z z^T \vec{\beta} = \vec{0} \Rightarrow$$

$$\Rightarrow z z^T \vec{\beta} = z \vec{X} \qquad \text{Нормальная система уравнений}$$

Становится плохо обусловл. => при малых возмущениях будут появляться большие ошибки.

Псевдообратная матрица

Получим точечную оценку $\hat{\vec{\beta}}$: $M[\hat{\beta}] = \beta$

Оценить $\tilde{\sigma}$ - ?

$X(t) = g(t) + \varepsilon(t)$

Стохастическая добавка

$\varepsilon(t) := X(t) - g(t)$ тоже случ. вел-на

Из мат. стат.

Лаб №1 Сгенерировать реализации с.п. (в.р.) объёмом 10^3

$g(t) :=$ полинома степ. от 0 до 3-й.

$\varepsilon(t) :=$ выборка из $N(0; \sigma^2)$

σ - выбираем исходя из оценки вариации $g(t)$ на $[t_1; t_n]$

2) Вычисляем оценку $\hat{\vec{\beta}}$!
(посл-но получить для ст-ва 1, 2, ...
как оценка даёт лучшее приближ.

3) $\hat{\sigma}$ и анализ адекватности регрессионной модели.

$Z^T \vec{\beta} = X$

$A\beta = Z \cdot X$

ZZ^T

$I\hat{\beta} = A^{-1} Z \cdot X$

единич. матрица

$\exists\, L := A^{-1} Z$

response vector

$D[\hat{\beta}] = D[L \cdot X] = L \cdot D[X] \cdot L^T = L\sigma^2 I L^T = \sigma^2 L \cdot L^T = \sigma^2 A^{-1} Z \cdot (A^{-1} Z)^T =$

диаг. матрица единиц т.к. X – независ

$= \sigma^2 \cdot A^{-1} \cdot \underbrace{Z \cdot Z^T}_{A \cdot =} (A^{-1})^T = \sigma^2 I (A^{-1})^T =$

$= \sigma^2 \cdot I \cdot (A^{-1})^T$

$cov(\hat{\beta}_i, \hat{\beta}_j) = \sigma^2 a^{-1}_{ij}$

Оцениваем $\hat{\sigma}$

$$M[S(\vec{\beta})] = M\Big[\sum_i^n \varepsilon_i^2\Big] = \sum_i^n M[\varepsilon_i] = n\sigma^2$$

наем. $\hat{\varepsilon}$!

$\vec{\beta} = \hat{\beta} - d\beta$ — уклонение оценок от точных β

$d\beta = \hat{\beta} - \beta$

$$M[S(\vec{\beta})] = M\Big[\big(X - Z^T(\hat{\beta} - d\beta)\big)^T\big(X - Z^T(\hat{\beta} - d\beta)\big)\Big] = M\big[(X - Z^T\hat{\beta} + Z^T d\beta)^T(X - Z^T\hat{\beta} + Z^T d\beta)\big] =$$

$$= M\big[(X - Z^T\hat{\beta})^T(X - Z^T\hat{\beta}) + (X - Z^T\hat{\beta})^T Z^T d\beta + d\beta^T Z\cdot(X - Z^T\hat{\beta}) + d\beta^T ZZ^T d\beta\big]$$

1) $M[(X - Z^T\hat{\beta})^T(X - Z^T\hat{\beta})]$ — это мы должны найти.

2) и 3) $M[(X - Z^T\hat{\beta})^T Z^T d\beta]$ и $M[d\beta^T\cdot Z\cdot(X - Z^T\hat{\beta})]$ найти самостоятельно.

4) $M[d\beta^T A\cdot d\beta]$ $\quad d\beta = \hat{\beta} - \beta$

$$M[(\hat{\beta} - \beta)^T \underbrace{A(\hat{\beta} - \beta)}_{\sum_j a_{ij}(\hat{\beta}_j - \beta_j)}] = M\Big[\sum_{ij} a_{ij}(\hat{\beta}_j - \beta_j)(\hat{\beta}_i - \beta_i)\Big] = \sum_{i=j=1}^{k} M[a_{ij}(\hat{\beta}_j - \beta_j)(\hat{\beta}_i - \beta_i)] =$$

$$\sum_i\sum_j a_{ij}(\hat{\beta}_j - \beta_j)(\hat{\beta}_i - \beta_i)$$

$$= \sum_{i=j=1}^{k} a_{ij} M[(\hat{\beta}_j - \beta_j)(\hat{\beta}_i - \beta_i)] =$$

$$= \sum_{i=j=1}^{k} a_{ij}\,\mathrm{cov}(\hat{\beta}_i, \hat{\beta}_j) = \sum_{i=j=1}^{k} a_{ij}\sigma^2 a^{-1}_{ij} =$$

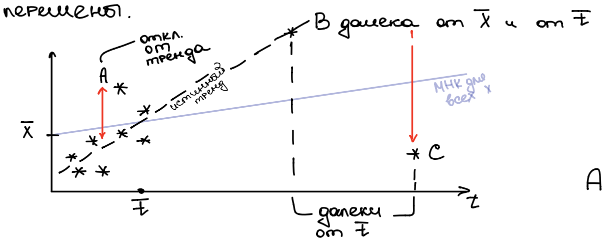
$$= \sigma^2 \sum_{i=j=1}^{k} a_{ij} a^{-1}_{ij} = \sigma^2\,\mathrm{tr}(AA^{-1}) = k\sigma^2$$

$$\underset{n\sigma^2}{\underset{\parallel}{M[S(\vec{\beta})]}} = M[(X - Z^T\beta)^T(X - Z^T\beta)] + \varnothing + \varnothing + k\sigma^2$$

$$M[(X - Z^T\beta)^T(X - Z^T\beta)] = n\sigma^2 - k\sigma^2 = (n-k)\sigma^2$$

отсюда получить $n\sigma^2$, а образовалась недоимка.
это цена за оценку пар-в $\hat{\beta}$, которых k штук.

Оценки в метрике L_2 неустойчивы к аномальным зн-ям объясняемой переменой.



Точки рычага: A, B, C
(Leverage points)

A, B — good leverage points
C — bad leverage point
(нужно будет менять Z)

ОМНК

Мы не знаем распр. ε ⇒ нет интервальных оценок

Нормальная регрессионная модель

ε_i – независ., с.в., одинак. р. $\forall \varepsilon_i \sim N(\varnothing, \sigma)$

$\forall (X(t_i) - g(t_i)) \sim N(\varnothing, \sigma)$

$X(t_i) \sim N(g(t_i), \sigma)$

Можем получить оценки методом максимального правдоподобия.

$$L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(X(t_i)-g(t_i))^2}{2\sigma^2}\right)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{\left(-\frac{\sum_{i=1}^{n}(X(t_i)-g(t_i))^2}{2\sigma^2}\right)} =$$

случ. откл.

$$= \ln L = -\frac{\sum_{i=1}^{n}(X(t_i)-g(t_i))^2}{2\sigma^2} - \frac{n}{2}\cdot \ln 2\pi\sigma^2$$

построение нормальных ур-й

$$\frac{\partial}{\partial\beta}\ln L(\varepsilon_1, \dots, \varepsilon_n) = \frac{\partial}{\partial\beta}\left(-\frac{\sum_{i=1}^{n}(X(t_i)-g(t_i))^2}{2\sigma^2} - \frac{n}{2}\cdot \cancel{\ln 2\pi\sigma^2}^{\,0}\right) = -\frac{1}{2\sigma^2}\frac{\partial}{\partial\beta}S(\beta) = \varnothing$$

МП оценка $\hat{\beta}$ с норм. распр. совпадает с МНК-оценкой.

$$\frac{\partial}{\partial\sigma}\ln L = \frac{\partial}{\partial\beta}\left(-\frac{S(\hat{\beta})}{2\sigma^2} - \frac{n}{2}\ln 2\pi\sigma^2\right) = \varnothing$$

МП-оценка $\hat{\sigma}^2 = \frac{S(\hat{\beta})}{n}$ ← МП-оценка смещённая

$$M[(X - z^T\hat{\beta})^T z^T d\beta] = M[(x - z^T\hat{\beta})^T z^T(\hat{\beta} - \beta)] = M[(x - \overbrace{z^T\hat{\beta}})(z^T\hat{\beta} - z^T\beta)] =$$

$$M[xz^T\hat{\beta} - xz^T\beta - z^T\hat{\beta}z^T\hat{\beta} - z^T\hat{\beta}z^T\beta]$$

Сгенерировать выборку double 10^7 и записать в файл.

$X = P + \varepsilon$