**EST WS2024/2025**  
**Краткое изложение**  
**24 января 2025 года**

**Содержание**

1. **Основы**  
   1.1 Основные концепции  
   1.1.1 Операции над множествами  
   1.1.2 Определение  
   1.2 Комбинаторика
2. **Случай**  
   2.1 Случайная переменная  
   2.1.1 Определение  
   2.2 Функция распределения  
   2.2.1 Дискретный случай  
   2.2.2 Непрерывный случай
3. **Оценивание**  
   3.1 Оценка максимально правдоподобия  
   3.1.1 Пример  
   3.1.2 Предположение i.i.d  
   3.1.3 Неравенство Маркова  
   3.2 Линейная регрессия
4. **Тестирование гипотез**  
   4.1 Гипотезы  
   4.2 Распределение при нулевой гипотезе  
   4.2.1 Пермутационные тесты
5. **Каузальность**  
   5.1 Принцип Райхенбаха  
   5.1.1 Парадокс Симпсона  
   5.2 Подходы  
   5.2.1 Модель потенциальных исходов  
   5.2.2 Каузальные структурные уравнения

**1. Основы**

**1.1 Основные концепции**

Чтобы корректно сформулировать статистическую задачу, нам нужны:

* Выборочное (или «стихийное») пространство Ω.
* Элементарные исходы ωi, лежащие в Ω.
* События A, которые являются подмножествами Ω (A ⊆ Ω).

**1.1.1 Операции над множествами**

* Дополнение к A (часто пишут A^c): это множество всех ω из Ω, которые не лежат в A.
* Объединение A ∪ B: это множество всех ω, которые принадлежат хотя бы одному из множеств A или B.
* Пересечение A ∩ B: это множество всех ω, которые одновременно лежат и в A, и в B.
* Разность A – B: это множество всех ω, которые лежат в A, но не лежат в B.
* Если все элементы из A входят в B, то пишут A ⊆ B (A — подмножество B).
* |A| обозначает число (мощность) элементов в A.
* Два события называются дизъюнктными (несовместимыми), если их пересечение пусто (A ∩ B = ∅).

**1.1.2 Определение**

Пусть P — функция, которая ставит в соответствие каждому событию A некоторое неотрицательное число P(A). Называют P вероятностной мерой, если выполняются аксиомы Колмогорова:

1. P(A) ≥ 0 для любого события A.
2. P(Ω) = 1.
3. Если A1, A2, … попарно несовместны (дизъюнктны), то  
   P(A1 ∪ A2 ∪ …) = P(A1) + P(A2) + …

**Некоторые следствия**:

* P(∅) = 0.
* Если A лежит в B (A ⊂ B), то P(A) ≤ P(B).
* P(A^c) = 1 – P(A), где A^c — дополнение к A.
* Если A и B дизъюнктны, то P(A ∪ B) = P(A) + P(B). В общем случае:  
  P(A ∪ B) = P(A) + P(B) – P(A ∩ B).
* **Условная вероятность**. Вероятность события A при условии B:  
  P(A | B) = P(A ∩ B) / P(B),  
  при P(B) > 0.  
  Аналогично P(B | A) = P(A ∩ B) / P(A).

Из этого следует:  
P(A ∩ B) = P(A | B)·P(B) = P(B | A)·P(A).

* **Формула Байеса**:  
  P(A | B) = [P(B | A) · P(A)] / P(B).

**Пример**. Сравнивают экспресс-тест (ST) на COVID с PCR-тестом. Пусть имеются данные:

|  | **PCR+** | **PCR-** |
| --- | --- | --- |
| ST+ | 99 | 50 |
| ST- | 1 | 9850 |

* Чувствительность: P(ST+ | PCR+) = 99/100 = 0.99.
* Специфичность: P(ST– | PCR–) = 9850/9900 ≈ 0.995.
* Нужно найти P(PCR+ | ST+). По формуле Байеса это примерно 0.664, то есть 66,4%.

**1.2 Комбинаторика**

1. Сколькими способами можно выбрать k объектов из n с возвращением (разрешаются повторения)? Ответ: n^k.
2. Сколькими способами можно упорядочить n объектов (то есть переставить все n штук)? Ответ: n! (факториал).
3. Сколькими способами можно выбрать k объектов из n без возвращения и не учитывая порядок? Это сочетания: C(n, k) = n! / (k!(n – k)!).

**2. Случай**

**2.1 Случайная переменная**

**2.1.1 Определение**

Случайная переменная X — это функция из Ω в R, то есть X(ω) даёт вещественное число при каждом элементарном исходе ω ∈ Ω. Для описания поведения X используют функцию распределения F\_X(x) = P(X ≤ x).

**2.2 Функция распределения**

Случайная переменная может быть дискретной или непрерывной.

**2.2.1 Дискретный случай**

* X называется дискретной, если она может принимать счётное множество значений.
* Определяют функцию вероятностей fX(x) = P(X = x), где сумма по всем x равна 1.

Примеры дискретных распределений:

1. **Точечная масса** δ(a): P(X = a) = 1.
2. **Равномерное дискретное** на k точках. Вероятность каждой точки равна 1/k.
3. **Бернулли(p)**: X ∈ {0,1}, P(X=1) = p, P(X=0) = 1 – p.
4. **Биномиальное(n, p)**: число успехов при n независимых испытаниях Бернулли(p). Формула:  
   P(Y=k) = C(n, k) · p^k · (1 – p)^(n – k).
5. **Геометрическое(p)**: количество испытаний Бернулли(p) до первого успеха. Формула: P(X=k) = (1 – p)^(k – 1) · p.
6. **Пуассон(λ)**: X ∈ {0,1,2,…},  
   P(X=k) = (λ^k / k!) · e^(-λ).

**2.2.2 Непрерывный случай**

* В непрерывном случае вероятность отдельной точки равна нулю, поэтому рассматривают интервалы.
* Вводят плотность fX(x) (неотрицательную, с интегралом равным 1). Тогда  
  P(a < X < b) = ∫ от a до b fX(x) dx.
* Функция распределения: F\_X(x) = ∫ от –∞ до x fX(t) dt.
* Математическое ожидание (среднее) E(X):
  + для дискретного X: сумма по x (x · fX(x));
  + для непрерывного X: интеграл (x · fX(x)) dx.
* Дисперсия Var(X) = E[(X – μ)²]. Часто используют формулу Var(X) = E(X²) – (E(X))².

Примеры непрерывных распределений:

1. **Нормальное (гауссово)**. Обозначают X ~ N(μ, σ²). Плотность:  
   (1 / (sqrt(2πσ²))) \* exp(– (x – μ)² / (2σ²)).  
   Если X ~ N(μ, σ²), то Z = (X – μ) / σ имеет стандартное нормальное распределение N(0,1).

Центральная предельная теорема: сумма (или среднее) большого числа независимых одинаково распределённых величин приблизительно имеет нормальное распределение.

1. **Экспоненциальное(λ)**: X ≥ 0, плотность fX(x) = λ \* exp(–λx).  
   E(X) = 1/λ, Var(X) = 1/λ².
2. **Многомерное нормальное** N(μ, Σ). Часто применяется в задачах с вектором случайных переменных X = (X1, …, Xk). Матрица Σ называется ковариационной.

**3. Оценивание**

Как оценить неизвестный параметр φ распределения по выборке S = {x1, …, xm}? В байесовском подходе P(φ | S) пропорционально P(S | φ)\*P(φ). Если P(φ) считать постоянным (неинформативный приор), то остаётся максимум правдоподобия.

**3.1 Оценка максимально правдоподобия**

Пусть x1, …, xm являются независимыми и одинаково распределёнными (i.i.d) из fX(x; φ). Тогда

φ^ = argmax по φ произведения i=1..m fX(xi; φ).

**3.1.1 Пример**

X ~ N(μ, σ²). По выборке x1, …, xm оценки максимум правдоподобия дают:  
μ^ = (1/m) \* сумма(xi),  
σ^2^ = (1/m) \* сумма((xi – μ^)²).

(При практических расчётах часто используют деление на (m – 1) для несмещённой оценки дисперсии.)

**3.1.2 Предположение i.i.d.**

«i.i.d.» значит «independent and identically distributed» (независимые и одинаково распределённые). Это упрощает вычисления, так как совместная плотность выборки раскладывается в произведение.

**3.1.3 Неравенство Маркова**

Если случайная величина Z ≥ 0, то:  
P(Z ≥ ε) ≤ E(Z) / ε.

Это даёт верхнюю оценку вероятности того, что оценка отклонится от своей средней более чем на ε.

**3.2 Линейная регрессия**

Модель: Y = β0 + β1·X + ε, где ε ~ N(0, σ²). Имеем выборку (x1, y1), …, (xm, ym). Требуется оценить β0, β1. Метод наименьших квадратов (эквивалент максимуму правдоподобия при нормальном шуме):

минимизировать сумму по i=1..m [yi – (β0 + β1·xi)]².

Решение даёт привычные формулы для прямой регрессии.

**4. Тестирование гипотез**

**4.1 Гипотезы**

Пусть есть две выборки S\_X = {X1,…,Xm} и S\_Y = {Y1,…,Yn}, хотим проверить, равны ли, например, их средние. Пусть тестовая статистика T(S\_X, S\_Y) = |μ^X – μ^Y|.

* Нулевая гипотеза H0: «Нет различий», T=0.
* Альтернативная гипотеза H1: «Различие есть», T≠0 (двусторонний тест).

Если наблюдаемый T равен c, мы хотим понять, насколько это «необычно», исходя из H0. Принимается или отвергается H0, опираясь на p-значение (p-value). Если p ≤ α, отвергаем H0.

Ошибки:

* Ошибка I рода: отвергнуть H0, когда она верна (уровень α).
* Ошибка II рода: принять H0, когда она неверна (1 – β — мощность теста).

**4.2 Распределение при нулевой гипотезе**

Требуется оценить P(T ≥ c | H0). Если есть классические формулы (например, t-тест), то используют их. Иначе можно воспользоваться пермутационным методом.

**4.2.1 Пермутационные тесты**

Идея: если при H0 выборки S\_X и S\_Y «равноценны», то можно случайно «перемешивать ярлыки».

1. Объединяем данные S\_X и S\_Y в одно множество S\_Z,
2. Считаем исходную статистику T = c,
3. Многократно перемешиваем (переставляем) элементы S\_Z, каждый раз разбивая на две группы того же размера (m и n),
4. Каждый раз считаем новое T', получая выборку (T'\_1, …, T'\_k). Это «распределение T при H0»,
5. Оцениваем p-value как долю тех T'\_i, что не меньше c. Если p достаточно мало (≤ α), то отвергаем H0.

(Так же работают двувыборочный t-тест, ANOVA и т. д.)

**5. Каузальность**

Основной вопрос: как статистические зависимости связаны с причинно-следственными связями и влиянием интервенций?

**5.1 Принцип Райхенбаха**

Если X и Y статистически зависимы, то одна из трёх причинно-следственных схем должна иметь место:

1. X → Y,
2. Y → X,
3. Существует третья переменная Z, общая причина: X ← Z → Y.

**5.1.1 Парадокс Симпсона**

Пример: сравнение двух методов удаления почечных камней (большие и маленькие). В каждом из двух «подразделов» метод B даёт лучший процент успеха. Но если объединить оба подраздела, оказывается, что метод A даёт в сумме лучше. Причина в том, что метод A чаще использовался для мелких камней, а метод B — для крупных. «Размер камня» влияет и на выбор метода, и на успех операции.

**5.2 Подходы**

**5.2.1 Модель потенциальных исходов (Potential Outcomes Framework)**

* Имеется множество «единиц» (например, пациентов),
* Функция назначения S(u): куда попал пациент (лечение t или контроль c),
* Исход (outcome) Y(u, t) или Y(u, c).

Для одной и той же единицы мы не можем наблюдать оба исхода одновременно. Поэтому оценивают средний эффект E(Y(t) – Y(c)). В рандомизированном эксперименте выбор в группы t и c считается независимым от исхода.

**5.2.2 Каузальные структурные уравнения**

* Набор эндогенных переменных X1,…, XN,
* Набор экзогенных шумов ε1,…, εN,
* Уравнения вида Xi = fi(родители(Xi), εi) без циклов (DAG).

Пример:

* X2 = ε2,
* X1 = X2 + ε1,
* X3 = X1·X2 + ε3.

Тогда совместное распределение можно факторизовать, а при интервенции do{X2 = x'} X2 перестаёт зависеть от ε2 и фиксируется в x'. Анализируют, как это влияет на X1 и X3.