Лекция 4.

Основные распределения математической статистики

Def. Случайная величина имеет нормальное распределение $\xi \in N(a, \sigma^2)$ с параметрами a и σ^2, если ее плотность имеет вид $f_\xi(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$
На практике нормальное распределение встречается чаще всего в силу ЦПТ

Def. Распределение $N(0, 1)$ с параметрами $a = 0, \sigma^2 = 1$ называется стандартным нормальным распределением. Его плотность равна $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. В дальнейшем такую случайную величину будем называть стандартной нормалью
Свойства

1. $a = E\xi \qquad \sigma^2 = D\xi$
2. Линейность: $\xi \in N(a, \sigma^2)$, то $\eta = b\xi + \gamma \in N(ab + \gamma, b^2\sigma^2)$
3. Стандартизация: Если $\xi \in N(a, \sigma^2)$, то $\eta = \dfrac{\xi - a}{\sigma} \in N(0, 1)$
4. Устойчивость относительно суммирования: если $\xi_1 \in N(a_1, \sigma_1^2)$, $\xi_2 \in N(a_2, \sigma_2^2)$, независимы то $\xi_1 + \xi_2 \in N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$

Распределение «хи-квадрат»

Def. Распределение «хи-квадрат» H_n со степенями свободы n называется распределение суммы квадратов независимых стандартных нормальных величин: $\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2$, где $X \in N(0, 1)$ и независимы
Свойства

1. $E\chi_n^2 = n$

   Так как $\forall i\ X_i \in N(0, 1)$, то $EX_i^2 = DX_i^2 + (EX_i)^2 = 1 \Longrightarrow E(X_i^2 + \dots X_n^2) = \sum_{i=1}^{n} EX_i^2 = n$

2. Устойчивость относительно суммирования: если $X \in H_n$, $Y \in H_m$, независимы, то $X + Y \in H_{n+m}$ (по определению)
3. $\dfrac{\chi_k^2}{k} \xrightarrow[k\to\infty]{p} 1$ (по Закону Больших Чисел)

Распределение Стьюдента

Def. Пусть $X_0, X_1, \ldots, X_k$ - независимые стандартные нормальные величины. Распределением Стьюдента T_k с k степенями свободы называется распределение случайной величины $t_k = \dfrac{X_0}{\sqrt{\frac{1}{k}(X_1^2 + \cdots + X_k^2)}} = \dfrac{X_0}{\sqrt{\frac{\chi_k^2}{k}}}$

<u>Свойства</u>

1. $Et_k = 0$ - в силу симметрии
2. $t_k \rightrightarrows N(0,1)$ (на практике при $k \geq 100$ распределение Стьюдента можно считать стандартным нормальным)

Распределение Фишера-Снедекера

Def. Распределением Фишера-Снедекера $F_{n,m}$ (другое название - F-распределение) со степенями свободы n и m называется распределение случайной величины $f_{n,m} = \dfrac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$, где χ_n^2 и χ_m^2 - независимые случайные величины с распределением «хи-квадрат»

<u>Свойства</u>

1. $Ef_{n,m} = \dfrac{m}{m-2}$ при $m > 2$
2. $f_{n,m} \xrightarrow[n,m\to\infty]{p} 1$

Математическое ожидание и дисперсия случайного вектора

Пусть $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ - случайный вектор, где случайная величина X_i - компонента (координата) случайного вектора

Def. Математическим ожидание случайного вектора называется вектор с координатами из математических ожиданий компонент: $E\vec{X} = \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}$

Def. Дисперсией случайного вектора (или матрицей ковариаций) случайного вектора $\vec{X}$ называется матрица $D\vec{X} = E(\vec{X} - E\vec{X})(\vec{X} - E\vec{X})^T$, состоящая из элементов $d_{ij} = \mathrm{cov}(X_i, X_j)$

Nota. На главной диагонали стоят дисперсии компонент: $d_{ii} = DX_i$

Nota. $D\vec{X}$ - симметричная положительно определенная матрица

<u>Свойства</u>

1. $E(A\vec{X}) = AE\vec{X}$
2. $E(\vec{X}+\vec{B}) = E\vec{X}+\vec{B}$, где $\vec{B}$ - вектор чисел
3. $D(A\vec{X}) = A\cdot D\vec{X}\cdot A^T$
4. $D(\vec{X}+\vec{B}) = D\vec{X}$

Многомерное нормальное распределение

Def. Пусть случайный вектор $\vec{\xi} = \begin{pmatrix}\xi_1\\ \vdots \\ \xi_n\end{pmatrix}$ имеет вектор средних $\vec{a} = E\vec{\xi}$, K - симметричная положительно определенная матрица. Вектор $\vec{\xi}$ имеет нормальное распределение в $\mathbb{R}^n$ с параметрами $\vec{a}$ и K, если его плотность $f_{\vec{\xi}}(\vec{X}) = \dfrac{1}{\left(\sqrt{2\pi}\right)^n\sqrt{\det K}}e^{-\frac{1}{2}(\vec{X}-\vec{a})^T K^{-1}(\vec{X}-\vec{a})}$

Свойства

1. Матрица $K = D\vec{\xi} = \left(\operatorname{cov}(\xi_i,\xi_j)\right)$ - матрица ковариаций
2. При $\vec{a}=\vec{0}$ и $K=E$ имеем вектор из независимых стандартных нормальных величин

> При $\vec{a} = \vec{0}$ и $K = E$: $f_{\vec{\xi}}(X_1,\dots,X_n) = \dfrac{1}{\left(\sqrt{2\pi}\right)^n}e^{-\frac{1}{2}\begin{pmatrix}X_1 & \dots & X_n\end{pmatrix}E\begin{pmatrix}X_1 & \dots & X_n\end{pmatrix}^T} =$
> $\dfrac{1}{\left(\sqrt{2\pi}\right)^n}e^{-\frac{1}{2}(X_1^2+\dots+X_n^2)} = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}X_1^2}+\dots+\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}X_n^2}$
> Так как плотность распалась на произведение плотностей стандартного нормального распределение, то все компоненты имеют стандартное нормальное распределение

Далее вектор из независимых стандартных нормальных величин для краткости будем называть стандартным нормальным вектором

3. $\sqsupset\vec{X}$ - стандартный нормальный вектор, B - невырожденная матрица, тогда вектор $\vec{Y} = B\vec{X}+\vec{a}$ имеет многомерное нормальное распределение с параметрами $\vec{a}$ и $K = BB^T$
4. $\sqsupset\vec{Y}\in N(\vec{a},K)$. Тогда вектор $\vec{X} = B^{-1}(\vec{Y}-\vec{a})$ - стандартный нормальный вектор, где $B = \sqrt{K}$
   Следствие. Эквивалентное определение: Многомерное нормальное распределение - это то, которое получается из стандартного нормального вектора при помощи невырожденного преобразования и сдвига
5. $\sqsupset\vec{X}$ - стандартный нормальный вектор, C - ортогональная матрица. Тогда $\vec{Y} = C\vec{X}$ - стандартный нормальный вектор

Так как C - ортогональная, то $C^T = C^{-1}$. Тогда по третьему свойству $K = CC^T = E$, а по второму свойству $\vec{Y}$ - стандартный нормальный вектор

6. $\sqsupset$ случайный вектор $\xi \in N(\vec{a}, K)$. Тогда его координаты независимы тогда и только тогда, когда они не коррелированы (то есть матрица коваріаций K диагональная)
   Следствие. Если плотность совместного распределения случайных величин ξ и η ненулевая, то они независимы тогда и только тогда, когда их коэффициент корреляции равен нулю

Многомерная центральная предельная теорема

Th. Среднее арифметическое независимых одинаково распределенных случайных векторов слабо сходится к многомерному нормальному распределению

Лемма Фишера

Пусть вектор $\vec{X}$ - стандартный нормальный вектор, C - ортогональная матрица, $\vec{Y} = C\vec{X}$. Тогда $\forall 1 \le k \le n-1$ случайная величина $T(\vec{X}) = \sum\limits_{i=1}^{n} X_i^2 - Y_1^2 - Y_2^2 - \ldots Y_k^2$ не зависит от $Y_1, Y_2, \ldots, Y_k$ и имеет распределение «хи-квадрат» со степенями свободы $n-k$

Так как C - ортогональное преобразование, то $\|\vec{X}\| = \|\vec{Y}\|$, то есть $\sum\limits_{i=1}^{n} X_i^2 = \sum\limits_{i=1}^{n} Y_i^2 \Longrightarrow$

$T(\vec{X}) = \sum\limits_{i=1}^{n} X_i^2 - Y_1^2 - Y_2^2 - \ldots Y_k^2 = Y_{k+1}^2 + \cdots + Y_n^2$

Согласно свойству 5 $Y_i \in N(0,1)$ и независимы, то по определению «хи-квадрат» $T(\vec{X}) \in H_{n-k}$ и не зависит от $Y_1, \ldots, Y_k$

Основная теорема

Эта теорема также известна как основное следствие леммы Фишера

Th. Пусть $(X_1, \dots, X_n)$ - выборка из нормального распределения $N(a, \sigma^2)$, $\overline{x}$ - выборочное среднее, S^2 - исправленная дисперсия.

Тогда справедливы следующие высказывания:

1. $\sqrt{n}\dfrac{\overline{x} - a}{\sigma} \in N(0, 1)$
2. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - a)^2}{\sigma^2} \in H_n$
3. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{x})^2}{\sigma^2} = \frac{nD^*}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \in H_{n-1}$
4. $\sqrt{n}\dfrac{\overline{x} - a}{S} \in T_{n-1}$
5. $\overline{x}$ и S^2 независимы

1. Так как $X_i \in N(a, \sigma^2)$, то $\displaystyle\sum_{i=1}^{n} X_i \in N(na, n\sigma^2) \Longrightarrow \overline{x} \in N\left(a, \frac{\sigma^2}{n}\right) \Longrightarrow \overline{x} - a \in N\left(0, \frac{\sigma^2}{n}\right) \Longrightarrow$ $\dfrac{\sqrt{n}}{\sigma}(\overline{x} - a) \in N(0, 1)$
2. Так как $X_i \in N(a, \sigma^2)$, то $\dfrac{X_i - a}{\sigma} \in N(0, 1)$ и $\displaystyle\sum_{i=1}^{n} \frac{(X_i - a)^2}{\sigma^2} \in H_n$ по определению
3. $\displaystyle\sum_{i=1}^{n} \frac{(X_i - \overline{x})^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - a}{\sigma} - \frac{\overline{x} - a}{\sigma}\right)^2 = \sum_{i=1}^{n} (z_i - \overline{z})^2$, где $z_i = \dfrac{X_i - a}{\sigma} \in N(0, 1)$, $\overline{z} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} z_i = \dfrac{\sum_{i=1}^{n} X_i - na}{\sigma} = \dfrac{\overline{x} - a}{\sigma}$

   Поэтому можно считать, что изначально $X_i \in N(0, 1)$

   $T(\vec{X}) = \displaystyle\sum_{i=1}^{n} (X_i - \overline{x})^2 = nD^* = n(\overline{x^2} - \overline{x}^2) = \sum_{i=1}^{n} X_i^2 - n\overline{x}^2 = \sum_{i=1}^{n} X_i^2 - Y_1^2$, где $Y_1 = \sqrt{n}\overline{x} = \dfrac{X_1}{\sqrt{n}} + \dots + \dfrac{X_n}{\sqrt{n}}$

   Строчка $\left(\dfrac{1}{\sqrt{n}}, \dots, \dfrac{1}{\sqrt{n}}\right)$ имеет длину 1, поэтому ее можно дополнить до ортогональной матрицы C, тогда Y_1 - первая компонента $\vec{Y} = C\vec{X}$, и согласно лемме Фишера $T(\vec{X}) = \displaystyle\sum_{i=1}^{n} (X_i - \overline{X})^2 \in H_{n-1}$
5. Согласно лемме Фишера $T(\vec{X}) = (n-1)S^2$ не зависит от $Y_1 = \sqrt{n}\overline{x} \Longrightarrow S^2$ и $\overline{x}$ - независимы
4. $\sqrt{n}\dfrac{\overline{x} - a}{S} = \sqrt{n}\dfrac{\overline{x} - a}{\sigma} \cdot \dfrac{1}{\sqrt{\frac{S^2(n-1)}{\sigma^2} \cdot \frac{1}{n-1}}} = \dfrac{\sqrt{n}\frac{\overline{x} - a}{\sigma}}{\frac{\chi^2_{n-1}}{n-1}}$

   Так как по пятому пункту числитель и знаменатель независимы, по определению получаем распределение Стьюдента