Лекция 16

Условная дисперсия

Def. Условной дисперсией случайной величины ξ относительно случайной величины η называется случайная величина $D(\xi|\eta) = E((\xi - E(\xi|\eta))^2|\eta)$

Nota. То есть дисперсия соответствующего условного распределения

Свойства

1. $D(\xi|\eta) = E(\xi^2|\eta) - E^2(\xi|\eta)$
2. Закон полной дисперсии

   Th. $D\xi = E(D(\xi|\eta)) + D(E(\xi|\eta))$

   Из первого свойства $E(\xi^2|\eta) = D(\xi|\eta) + E^2(\xi|\eta)$
   $D\xi = E\xi^2 - (E\xi)^2 = E(E\xi^2|\eta) - E^2(E(\xi|\eta)) = E(D(\xi|\eta) + E^2(\xi|\eta)) - E^2(E(|\eta)) = E(D(\xi|\eta)) + E(E^2(\xi|\eta)) - E^2(E(\xi|\eta)) = E(D(\xi|\eta)) + D(E(\xi|\eta))$

   Следствие и смысл:

   - Если ξ и η независимы (некоррелированы), то $D(E(\xi|\eta)) = D(E\xi) = 0$ и $D\xi = E(D(\xi|\eta))$
   - Если имеется функциональная зависимость (то есть $\xi = g(\eta)$), то $D(E(\xi|\eta)) = D(E(g(\eta)|\eta)) = D(g(\eta)) = D\xi$
   - Таким образом по величине $R^2 = \dfrac{D(E(\xi|\eta))}{D\xi}$ $(0 \le R^2 \le 1)$ можно судить о силе корреляционной зависимости. Такая величина называется корреляционным отношением

Энтропия

Пусть ξ - результат эксперимента с исходами $A_1, A_2, \dots, A_N$, вероятности которых $p_1, p_2, \dots, p_N$

Def. Энтропией эксперимента называется величина $H(\xi) = -\sum\limits_{i=1}^{N} p_i \cdot \log_2 p_i$

Свойства энтропии:

1. Очевидно, что $H(\xi) \ge 0$, так как $p \ge 0$, а $\log_2 p_i \le 0$
2. $H(\xi) = 0 \iff \exists i$, такой что $p_i = 1, p_j = 0 \forall j \ne i$ - то есть эксперимент заканчивается всегда одним исходом, нет неопределенности
3. Максимум $H(\xi) = \log_2 N = H_0$ достигается при $p_1 = p_2 = \dots = \dfrac{1}{N}$ - то есть когда все вероятности одинаковы, ни одному исходу нельзя отдать предпочтение, и результат эксперимента получается максимально неопределенным

Рассмотрим $\varphi(x) = x\log_2 x$. Так как $\varphi''(x) = \dfrac{1}{x\ln 2} > 0$ при $x > 0$, следовательно $\varphi(x)$ выпукла вниз

Рассмотрим случайную величину η

η	p_1	p_2	$\dots$	p_n
p	$\frac{1}{N}$	$\frac{1}{N}$	$\dots$	$\frac{1}{N}$

По неравенству Йенсена $\varphi(E\eta) = \varphi(\sum_{i=1}^{N}\frac{p_i}{N}) = \varphi(\frac{1}{N}\sum_{i=1}^{N}p_i) = \varphi(\frac{1}{N}) = \dfrac{\log_2\frac{1}{N}}{N} \le E(\varphi(\eta)) = \dfrac{1}{N}\sum_{i=1}^{N}p_i\log_2 p_i = -\dfrac{1}{N}H(\eta)$

Получаем $\dfrac{\log_2\frac{1}{N}}{N} \le -\dfrac{1}{N}H(\eta)$, то есть $H \le \log_2 N$

Следствие: Энтропию можно рассматривать как меру неопределенности эксперимента

Ex. $\xi \in B_p$

ξ	0	1
p	$1-p$	p

$H(\xi) = -(1-p)\log_2(1-p) - p\log_2 p$

Ex. 1. Психолог Р. Хайман проводил такой эксперимент: перед человеком загорались с некоторой частотой лампочки, замерялась время реакции на загоревшуюся лампочку. Если лампочки загорались с одинаковой частотой, то энтропия была пропорциональна H_0

Ex. 2. Также с помощью энтропии определен второй закон термодинамики

Ex. 3. Теория кодирования информации

Если алфавит сообщения состоит из N символов, то каждому символу присваиваем последовательность одинаковой длины из 0 и 1, причем ее длина будет $\lceil\log_2 N\rceil$

Для передачи n символов потребуется последовательность длиной $n\lceil\log_2 N\rceil$

Цель: сократить длину последовательности

Для больших по объему сообщений можно заметно уменьшить эту величину, используя, что разные символы встречаются с разными частотами.

Если $p_1, p_2, \dots, p_N$ - эти частоты, то в сообщении длиной N i-ый символ появляется $v_i \approx np_i$ раз

Def. Сообщение длины N называется типичным с параметрами n и δ, если $|v_i - np_i| < \delta\ \forall 1 \le i \le N$

Пусть $M_{n,\delta}$ - число таких сообщений

Th. (частный случай теоремы Макмиллана)
$\frac{1}{n}\log_2 M_{n,\delta} \xrightarrow[n\to\infty]{} H = -\sum_{i=1}^{N} p_i \log_2 p_i$

Следствие: существует $\varepsilon > 0 \mid \frac{1}{n}\log_2 M_{n,\delta} < H + \varepsilon$ (или $M_{n,\delta} < 2^{n(H+\varepsilon)}$)
Если можно занумеровать эти типичные сообщения, то для них потребуется число символов $\log_2 2^{n(H+\varepsilon)} = n \cdot (H + \varepsilon)$
И поэтому с вероятностью приблизительно 1 можно сократить длины сообщение с коэффициентом сжатия $\gamma \approx \frac{nH}{nH_0} = \frac{H}{H_0}$, где $H_0 = \log_2 N$
Если все символы встречаются независимо, то дальнейшее сжатие невозможно, но так как буквы встречаются в определенных сочетаниях, то можно сжать информации дальше, используя этот факт
Пусть γ_∞ - коэффициент итогового сжатия
В русском языке $\gamma \approx 0.87$. Если считать слова символами нашего алфавита, то получится $\gamma_\infty \approx 0.24$ для литературного языка и $\gamma_\infty \approx 0.17$ для делового языка

Def. $1 - \gamma_\infty$ называют коэффициентом избыточности языка

Энтропия при непрерывном распределении

Def. Пусть ξ абсолютно непрерывная случайная величина с плотностью $f(x)$ и носителем $A = \{x \mid f(x) > 0\}$. Энтропией $H(\xi)$ называется величина $-\int_A f(x)\log_2 f(x)dx$

Th. Следующие распределения имеют наибольшую энтропию:

1. Если $A = [0, 1]$, то $U(0, 1)$
2. Если $A = [0, \infty)$ и $E\xi = 1$, то показательное E_1
3. Если $A = \mathbb{R}$ и $E\xi = 0$, а $D\xi = 1$, то $N(0, 1)$