Домашнее задание по теории вероятностей №2

Олег Тарасов

2024-10-13

Задание 1.

Доказать, что из $Pr\{A\} = Pr\{A|B\}$ автоматически следует, что $Pr\{A\} = Pr\{A|\overline{B}\}$.

Преобразуем условную вероятность

$$Pr\{A|\overline{B}\} = \frac{Pr\{A\overline{B}\}}{Pr\{\overline{B}\}} = \frac{Pr\{A\} - Pr\{AB\}}{Pr\{\overline{B}\}} = ...$$

По условию $Pr\{A\} = Pr\{A|B\}$, поэтому выражение можно преобразовать далее как

$$... = \frac{Pr\{A|B\} - Pr\{AB\}}{Pr\{\overline{B}\}} = \frac{Pr\{AB\}/Pr\{B\} - Pr\{AB\}}{Pr\{\overline{B}\}}$$

Вынеся $Pr\{B\}$ в общий знаменатель, получаем

$$\frac{Pr\{AB\} - Pr\{AB\}Pr\{B\}}{Pr\{B\}Pr\{\overline{B}\}} = \frac{Pr\{AB\}\big(1 - Pr\{B\}\big)}{Pr\{B\}Pr\{\overline{B}\}} = \frac{Pr\{AB\}Pr\{\overline{B}\}}{Pr\{B\}Pr\{\overline{B}\}} = \frac{Pr\{AB\}}{Pr\{B\}} =$$

$= Pr\{A|B\}$, что по условию равняется $Pr\{A\}$.

QED

Задание 2.

Доказать, что из $RR = 1$ следует, что случайные события – независимы.

Отношение рисков наблюдать событие A при условии наступления либо не наступления события B равняется $RR = \frac{Pr\{A|B\}}{Pr\{A|\overline{B}\}} = 1$, следовательно, $Pr\{A|B\} = Pr\{A|\overline{B}\}$.

Представим условные вероятности в виде

$$Pr\{A|B\} = \frac{Pr\{AB\}}{Pr\{B\}}$$

$$Pr\{A|\overline{B}\} = \frac{Pr\{A\overline{B}\}}{Pr\{\overline{B}\}} = \frac{Pr\{A\} - Pr\{AB\}}{Pr\{\overline{B}\}} = \frac{Pr\{A\}}{Pr\{\overline{B}\}} - \frac{Pr\{AB\}}{Pr\{\overline{B}\}}$$

Приравняв эти два выражения и перенеся вычитаемое через знак равенства, получаем

$$\frac{Pr\{A\}}{Pr\{\overline{B}\}} = \frac{Pr\{AB\}}{Pr\{B\}} + \frac{Pr\{AB\}}{Pr\{\overline{B}\}}$$

Преобразуя правую часть уравнения, получаем

$$Pr\{AB\}\left(\frac{1}{Pr\{B\}} + \frac{1}{Pr\{\overline{B}\}}\right) = Pr\{AB\}\frac{Pr\{B\} + Pr\{\overline{B}\}}{Pr\{B\}Pr\{\overline{B}\}} = \frac{Pr\{AB\}}{Pr\{B\}Pr\{\overline{B}\}}$$

Поскольку $Pr\{B\} \neq 0$ и $Pr\{\overline{B}\} \neq 0$, то, приравняв два предыдущих выражения и домножив на $Pr\{B\}Pr\{\overline{B}\}$, получаем

$$Pr\{A\}Pr\{B\} = Pr\{AB\},$$

что соответствует определению независимых событий.

QED

Задание 3.

A. Найдите математическое ожидание и дисперсию числа циклов терапии при первичном выявлении и при рецидиве (отдельно).

При первичном выявлении заболевания математическое ожидание количества циклов химиотерапии $\mu_1 = 1*0,5+2*0,5 = 1,5$, а дисперсия $\sigma_1^2 = \big((1-1,5)^2*0,5+(2-1,5)^2*0,5\big) = 0,25$.

При рецидиве математическое ожидание количества циклов терапии $\mu_2 = 2*0,25+3*0,75 = 2,75$, а дисперсия $\sigma_2^2 = \big((2-2,75)^2*0,25+(3-2,75)^2*0,75\big) = 0,1875$.

B. Предположим, что мы изучаем только рецидивировавших пациентов.

- ***Постройте таблицу распределения общего числа циклов терапии у рецидивировавших пациентов («дебютных» + «рецидивных»).***

Если допустить, что выбор числа циклов при рецидиве не зависит от того, сколько циклов было в дебюте, то таблица имеет следующий вид:

Количество циклов	3	4	5
Вероятность	0,125	0,5	0,375

- ***Найдите математическое ожидание и дисперсию этой величины.***

Поскольку новая случайная величина является суммой двух исходных, и мы принимаем допущение об их независимости, то математическое ожидание и дисперсия этой величины равны соответственно сумме математических ожиданий и сумме дисперсий двух исходных величин: $\mu_3 = \mu_1 + \mu_2 = 4,25$, $\sigma_3^2 = \sigma_1^2 + \sigma_2^2 = 0,4375$.

Задание 4 *(добавлено 2024-10-20).*

Переделайте скрипт так, чтобы в нем можно было бы анализировать ошибку в оценке вероятности события в зависимости от истинной вероятности и объема выборки.

В скрипте истинная вероятность события `true_prob` задается явно в виде одного числа. Далее создаются выборки из n независимых пациентов, каждый из которых излечивается с вероятностью `true_prob`. В качестве выборочной оценки вероятности используется доля излечившихся пациентов. В качестве меры ошибки оценки вероятности я использовал дисперсию выборочных оценок по 1000 выборок. Скрипт позволяет визуализировать распределение выброчных оценок вероятности по отдельности в каждой серии экспериментов (с заданными числом пациентов и истинной вероятностью). Другие визуализации я не делал.

Полный скрипт выложен на github, ссылка добавлена в гугл-класс.

Какие закономерности вы можете вычислить, экспериментируя со скриптом?

Я проанализировал выборки размером от 3 до 1000 пациентов с увеличением приблизительно в 1,5 раза на каждом шаге и значения истинных вероятностей от 0,1 до 0,9 с шагом 0,1. С увеличением объема выборки ошибка монотонно уменьшается, чем дальше — тем медленнее, зависимость на вид похожа на $1/\sqrt{x}$. Зависимость ошибки от истинной вероятности выпуклая немонотонная и напоминает многочлен второй степени с отрицательным коэффициентом при старшей степени: при истинных вероятностях 0 и 1 ошибка при выбранном методе оценики, очевидно, будет нулевой (в скрипте я это не демонстрирую), а максимум достигается при истинной вероятности 0,5.