Дата последней сборки документа:
4 февраля 2024 г. в 09:55

Классические и продвинутые темы теории вероятностей и математической статистики

Подвойский А.О.

Здесь приводятся заметки по некоторым вопросам, касающимся машинного обучения, анализа данных, программирования на языках `Python`, `R` и прочим сопряженным вопросам так или иначе, затрагивающим работу с данными.

Содержание

1. Эмпирическая и теоретическая функции распределения

Построим по выборке $X_1, X_2, \ldots, X_n$ случайную ступенчатую функцию $\hat{F}_n(x)$, возрастающую скачками величины $1/n$ в точках $X_{(i)}$ (i-ая порядковая статистика). Эта функция называется *эмпирической функцией распределения*. Чтобы задать значения в точках разрывов, формально определим ее так, чтобы она была непрерывна справа

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_{(i)} \leqslant x\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \leqslant x\}}.$$

В отличие от эмпирической функции распределения выборки, интегральную функцию $F(x)$ распределения генеральной совокупности называют *теоретической функцией распределения.*

Различие между эмпирической и теоретической функциями распределения $F(x)$ состоит в том, что теоретическая функция определяет *вероятность* события $X_i \leqslant x$, а эмпирическая функция $\hat{F}_n(x)$ определяет *относительную частоту* этого события. Из теоремы Бернулли следует, что относительная частота события $X_i \leqslant x$, т.е. $\hat{F}_n(x)$ *стремится по вероятности* к вероятности $F(x)$ этого события, т.е. $\hat{F}_n(x) \xrightarrow{\mathbf{P}} F(x)$. Другими словами числа $\hat{F}_n(x)$ и $F(x)$ мало отличаются одно от другого [1, 191].

2. Неравенства Чебышева

Неравенство Маркова (еще называют неравенством Чебышева) дает грубую оценку вероятности события, состоящего в том, что неотрицательная случайная величина X с конечным математическим ожиданием $\mu = \mathbf{E}X$ превысит некоторую положительную детерминированную величину a

$$\mathbf{P}(|X| \geqslant a) \leqslant \frac{\mathbf{E}|X|}{a}, \quad a > 0.$$

Неравенство Чебышева (неравенством Чебышева-Бьенеме) дает грубую оценку вероятности события, состоящего в том, что случайная величина X отклонится от своего конечного среднего μ на величину небольшую a

$$\mathbf{P}\big(|\, X - \mu \,| \geqslant a\big) \leqslant \frac{\sigma^2}{a^2}, \quad a > 0,$$

где σ^2 – дисперсия X.

Другими словами неравенство Чебышева-Бьенеме дает грубую верхнюю оценку вероятности выброса центрированной случайной величины за положительный порог a.

В качестве следствия получим такназыаемое «правило трех сигм», которое означает, что *вероятность случайной величине отличаться от своего математического ожидания более чем на три среднеквадратических отклонения, мала.*

Разумеется, для каждого распределения величина этой вероятности своя. Можно получить верную для всех распределений с конечной дисперсией оценку сверху для вероятности случайной величине отличаться от своего математического ожидания более чем на три корня из дисперсии

$$\text{Если } \mathbf{E}X^2 < \infty, \text{то } \mathbf{P}(|X - \mathbf{E}X| \geqslant 3\sqrt{\mathbf{D}X}) \leqslant \frac{1}{9}.$$

Неравенство Высочанского-Петунина дает оценку вероятности события, состоящего в том, что неотрицательная случайная величина X с одномодальным распределением, конечными средним μ и дисперсией σ^2 не отклониться от своего среднего больше чем на $\lambda\sigma$

$$\mathbf{P}\big(|\, X - \mu \,| \geqslant \lambda\sigma\big) \leqslant \frac{4}{9\lambda^2}.$$

В приложениях математической статистики используется эвристическое правило $\lambda = 3$, что соответствует верхней границе вероятности $\dfrac{4}{81} \approx 0.04938$.

3. Доверительные интервалы

Доверительный интервал – интервал, покрывающий неизвестный скалярный параметр θ с заданной *доверительной вероятностью* $(1-\alpha)$

$$P(\hat{\theta}_1(X_1, X_2, \dots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \dots, X_n)) \geqslant 1-\alpha,$$

где $\hat{\theta}_{1,2}$ – нижняя и верхняя граница доверительного интервала (*случайные величины*), α – уровень значимости (она же вероятность ошибки первого рода).

Наиболее часто уровень значимости принимают равным 0.05 или 0.01. Если, например, принят уровень значимости равный 0.05, то означает, что в пяти случаях из ста мы рискуем допустить ошибку первого рода (отвергнуть правильную гипотезу) [1, 284].

Границы доверительного интервала являются *случайными величинами* – функциями от выборки (или другими словами границы доверительного интервала являются *статистиками*) – поэтому правильнее говорить не о вероятности попадания θ в доверительный интервал, а о вероятности того, что доверительный интервал покроет неизвестный параметр θ [1, стр. 216].

Смысл *надежности* (*доверительной вероятности*) $\gamma = 0.95$ заключается в том, что если произведено достаточно большое число выборок, то 95% из них определяет такие доверительные интервалы, в которых параметр действительно заключен. И лишь в 5% случаев он может выйти за границы доверительного интревала [1, стр. 219].

Надежность (доверительная вероятность) $\gamma = 0.95$ указывает, что если произведено достаточно большое число выборок, то 95% из них определяет такие доверительные интервалы, в которых параметр действительно заключен; лишь в 5% случаев он может выйти за границы доверительного интервала [1, стр. 219].

Интервалы в нормальной модели Допустим, что элементы выборки X_i распределены по закону $\mathcal{N}(\theta, \sigma^2)$, причем параметр масштаба σ известен, а параметр сдвига θ – нет. Эту модель часто применяют к данным, полученным при независимых измерениях некоторой величины θ с помощью прибора (или метода), имеющего известную среднюю погрешность (стандартную ошибку) σ.

Если случайная величина X распределена нормально $\mathcal{N}(\theta, \sigma^2)$, то выборочная средняя $\bar{X}$, найденная по независимым наблюдениям, также распределена нормально. Параметры распределения таковы [1]

$$\mathbf{E}(\bar{X}) = \theta, \sqrt{\mathbf{D}(\bar{X})} = \frac{\sigma}{\sqrt{n}} \quad \to \quad \bar{X} \sim \mathcal{N}(\theta, \sigma^2/n).$$

Для центрированной и нормированной случайной величины $\sqrt{n}(\bar{X}-\theta)/\sigma \sim \mathcal{N}(0,1)$ в качестве границ интервала с доверительной вероятности $1-\alpha$ можно взять

$$\hat{\theta}_1 = \bar{X} - \sigma/\sqrt{n}\, x_{1-\alpha/2}, \quad \hat{\theta}_2 = \bar{X} + \sigma/\sqrt{n}\, x_{1-\alpha/2}.$$

Таким образом, с вероятностью 0.95 истинное значение параметра сдвига θ находится в интервале $\bar{X} \pm 1.96\, \sigma/\sqrt{n} \approx \bar{X} \pm 2\, \sigma/\sqrt{n}$ (правило двух сигм) [2, 147].

На практике, если значение σ неизвестно, то его заменяют на *состоятельную оценку* $\hat{\sigma} = S$, где $S^2 = \dfrac{1}{2} \sum (X_i - \hat{X})^2$.

А вот если выборка *маленькая*, про ее параметры ничего неизвестно и объем выборки небольшой ($n \leqslant 30$), тогда вместо *нормального распределения* используют *распределение Стьюдента* (t-распределение).

Тогда доверительный интервал будет иметь вид

$$\left(\bar{x} - \frac{s}{\sqrt{n}} t_\alpha(n-1); \bar{x} + \frac{s}{\sqrt{n}} t_\alpha(n-1)\right),$$

где $t_\alpha(n-1)$ – это квантиль распределения Стьюдента уровня $1-\alpha/2$ с $n-1$ степенями свободы.

Замечание

Распределение Стьюдента стремиться к нормальному распределению при $n \to \infty$

Число степеней свободы зависит от того, сколько имеется связей между наблюдениями. Так как мы знаем среднее, то наблюдения связаны одним равенством и степеней свободы становится на одну меньше.

Оценка $\hat{\theta}$ параметра θ называется *состоятельной*, если для всех $\theta \in \Theta$ последовательность

$$\hat{\theta}_n = \hat{\theta}(X_1, \dots, X_n) \xrightarrow{\mathbf{P}} \theta, \quad n \to \infty.$$

Здесь $\xrightarrow{\mathbf{P}}$ обозначает *сходимость по вероятности*

$$\forall \varepsilon > 0, \ \mathbf{P}(|\hat{\theta} - \theta| > \varepsilon) \to 0, \quad n \to \infty.$$

Состоятельность оценки (а точнее – последовательности оценок $\{\hat{\theta}_n\}$) означает концентрацию вероятностной массы около истинного значения параметра θ с ростом размера выборки n [2, 75].

4. Сходимости

Из сходимости «почти наверное» следует сходимость «по вероятности». А из сходимости «по вероятности» следует сходимость «почти наверное».

Случайные величины $\xi_1, \xi_2, \dots, \xi_n$ сходятся при $n \to \infty$ к случайной величине ξ

- *Сходимость «почти наверное»* (или с вероятностью 1): $\xi_n \xrightarrow{\text{п.н.}} \xi$, если $\mathbf{P}\{\omega : \xi_n(\omega) \to \xi(\omega)\} = 1$,
- *в среднем квадратическом*: $\xi_n \xrightarrow{\text{с.к.}} \xi$, если $\mathbf{E}(\xi_n - \xi)^2 \to 0$,
- *по вероятности*: $\xi_n \xrightarrow{\mathbf{P}} \xi$, если $\forall \varepsilon > 0$ $\mathbf{P}(|\xi_n - \xi| > \varepsilon) \to 0$,
- *по распределению*: $\xi_n \xrightarrow{d} \xi$, если функция распределения $F_{\xi_n}(x)$ сходится к $F_\xi(x)$ в точках непрерывности последней.

5. Центральная предельная теорема

Пусть $X_1, \dots, X_n$ – независимые одинаково распределенные случайные величины. Положим $S_n = X_1 + X_2 + \dots + X_n$.

Если $0 < \sigma^2 = \mathbf{D}X_1 < \infty$, то

$$S_n^* = \frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} = \frac{S_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} Z, \quad n \to \infty,$$

где Z – стандартная нормальная случайная величина, $Z \sim \mathcal{N}(0,1)$.

По центральной предельной теореме среднее значение одинаково распределенных случайных величин стремится к нормальному распределению. Более того верна теорема.

Теорема: Если распределение генеральной совокупности имеет конченые математические ожидание и дисперсию, то при $n \to \infty$ основные выборочные характеристики (среднее, дисперсия, эмпирическая функция распределения) являются нормальными.

Пример Пусть случайные величины $Z_1, \ldots, Z_k$ распределены по закону $\mathcal{N}(0,1)$ и независимы. Тогда распределение случайной величины $R_k^2 = Z_1^2 + \cdots + Z_k^2$ называют распределением χ^2 с k степенями свободы (кратко $R_k^2 \sim \chi_k^2$).

Отметим, что каждое слагаемое имеет гамма-распределение с параметрами $\alpha = \lambda = 1/2$, т.е. $Z_i^2 \sim \Gamma(1/2, 1/2)$.

Поскольку R_k^2 – это сумма независимых и одинакового распределенных случайных величин Z_i^2, то согласно *центральной предельной теореме* имеет место *сходимость по распределению*

$$(R_k^2 - \mathbf{E}R_k^2)/\sqrt{\mathbf{D}R_k^2} = (R_k^2 - k)/\sqrt{2k} \xrightarrow{d} Z \sim \mathcal{N}(0,1), \quad k \to \infty.$$

Нормальное приближение является довольно точным уже при $k > 30$.

6. Фактический (достигаемый) уровень значимости

При проверке статистических гипотез в общем случае задается малое число α – вероятность, с которой мы можем позволить себе отвергнуть верную гипотезу (скажем, 0.05). Это число называют *уровнем значимости.*

Исходя из предположения, что гипотеза H верна, определяется *наименьшее* (самое крайнее левое) значение $x_{1-\alpha}$, удовлетворяющее условию

$$\mathbf{P}(T(X_1, \ldots, X_n) \geqslant x_{1-\alpha} \,|\, H) = \int\limits_{x_{1-\alpha}}^{+\infty} p_T(x)\, dx \leqslant \alpha.$$

Другими словами, вероятность события, состоящего в том, что статистика примет значение большее $(1-\alpha)$-квантиля (вероятность маловероятного события) должна быть не больше заранее заданного уровня значимости α.

Если функция распределения статистики T непрерывна, то $x_{1-\alpha}$ является, очевидно, ее $(1-\alpha)$-квантилью. Такое $x_{1-\alpha}$ называют *критическим значением*: гипотеза H отвергается, если

$$t_0 = T(x_1, \ldots, x_n) \geqslant x_{1-\alpha}$$

(произошло маловероятное событие), и принимается – в противном случае.

При этом величина

$$\alpha_0 = \mathbf{P}(T(X_1, \ldots, X_n) \geqslant t_0 \,|\, H) = \int\limits_{t_0}^{+\infty} p_T(x)\, dx$$

задает *фактический (достигаемый) уровень значимости.* Он равен вероятности того, что статистика T (измеряющая степень отклонения полученной реализации от наиболее типичной) за счет случайности примет значение t_0 или даже больше. Другими словами, фактический (достигаемый) уровень значимости оценивает вероятность того, что случайная величина $T(X_1, \dots, X_n)$ попадет в область $[\,t_0, +\infty\,)$, где t_0 – это значение статистики, найденное по выборке.

Фактический (достигаемый) уровень значимости[1] – наименьший уровень значимости, на котором проверяемая (нулевая) гипотеза принимается[2] [2, 161].

Фактический (достигаемый) уровень значимости – это вероятность получить значение статистики как в эксперименте или более экстремальное ее значение при условии справедливости нулевой гипотезы.

Подытожив сказанное выше, можно получить следующее правило: если фактический (достигаемый) уровень значимости α_0 меньше заранее заданного уровня значимости α, то говорят, что данные свидетельствуют против нулевой гипотезы H_0 в пользу альтернативной и у нас есть основания отвергнуть нулевую гипотезу

$$\boxed{\text{если} \quad \alpha_0 < \alpha \quad \text{тогда} \quad \cancel{H_0}}$$

Критическое значение $x_{1-\alpha}$ допускается интерпретировать как квантиль уровня $(1-\alpha)$ только для статистик с непрерывной функцией распределения

Вычисление фактического (достигаемого) уровня значимости нередко позволяет избежать категоричных (и при этом ошибочных) выводов, сделанных только на основе сравнения наблюдаемого значения статистики t_0 с критическим значением $x_{1-\alpha}$, найденным для формально заданного α.

7. Теоретические и выборочные квантили

Пусть $\alpha \in (0, 1)$. Для *непрерывной* функции распределения F *теоретической α-квантилью* x_α (или квантилью уровня α) называется решение уравнения $F(x_\alpha) = \alpha$, т.е. $x_\alpha = F^{-1}(\alpha)$.

Так же, как и в случае медианы ($\alpha = 1/2$) это решение может быть не единственным.

Оценить x_α можно с помощью порядковой статистики $X_{([\alpha n]+1)}$, где $[\cdot]$ – обозначает целую часть. Эту оценку называют *выборочной α-квантилью.*

8. Ошибки I и II рода

Пример рассмотрим модель $X_i \sim \mathcal{N}(\theta, \sigma^2)$, где дисперсия известна, а математическое ожидание нет. Для проверки гипотезы $H_0 : \theta = \theta_0$ можно применить критерий, основанный на статистике $T(X_1, \cdots, X_n) = \bar{X}$.

Если H_0 верна, то $\bar{X} \sim \mathcal{N}(\theta_0, \sigma^2/n)$. Найдем *критическое значение* t_α из условия

$$\alpha = \mathbf{P}_{\theta_0}(\bar{X} \geqslant t_\alpha).$$

[1] p-value
[2] Наверное, правильнее говорить *не отвергается*

Тогда (центрируем и нормируем случайную величину $\bar{X}$)

$$\alpha = \mathbf{P}\left(\frac{\sqrt{n}(\bar{X}-\theta_0)}{\sigma} \geqslant \frac{\sqrt{n}(t_\alpha-\theta_0)}{\sigma}\right) = 1-\Phi\left(\frac{\sqrt{n}(t_\alpha-\theta_0)}{\sigma}\right), \text{ так как } \frac{\sqrt{n}(\bar{X}-\theta_0)}{\sigma} \sim \mathcal{N}(0,1),$$

где $\Phi(x)$ – функция распределения закона $\mathcal{N}(0,1)$.

Из последнего соотношения получаем *критическое значение*

$$t_\alpha = \theta_0 + \sigma\, x_{1-\alpha}/\sqrt{n}.$$

Если значение выборочного среднего $\bar{x} \geqslant t_\alpha$, то гипотеза H_0 отвергается. Если нулевая гипотеза верна, то неравенство $\bar{X} \geqslant t_\alpha$ выполняется с вероятностью α. Отвергая в этом случае верную гипотезу H_0, мы совершаем *ошибку I рода*.

С другой стороны, может оказаться, что на самом деле верна не гипотеза H_0, а ее альтернатива $H_1 : \theta = \theta_1$. Если при этом случится, что $\bar{x} < t_\alpha$, то мы примем ошибочную гипотезу H_0 вместо H_1, тем самым допустив *ошибку II рода*.

Найдем вероятность β ошибки II рода для рассматриваемой модели. Когда верна альтернативная гипотеза, выборочное среднее распределено по закону $\mathcal{N}(\theta_1, \sigma^2/n)$, поэтому

$$\beta = \mathbf{P}_{\theta_1}(\bar{X} < t_\alpha) = \Phi\left(\frac{\sqrt{n}(t_\alpha-\theta_1)}{\sigma}\right) = \Phi\left(x_{1-\alpha} - \frac{\sqrt{n}(\theta_1-\theta_0)}{\sigma}\right).$$

Обобщить сказанное выше можно так

$$\alpha = \mathbf{P}(\,\mathrm{Rej}\, H_0^+\,), \text{ ошибка I рода},$$
$$\beta = \mathbf{P}(\,\mathrm{Rej}\, H_1^+\,) = \mathbf{P}(\,\neg\mathrm{Rej}\, H_0^-\,), \text{ ошибка II рода}.$$

Гипотеза H_0 заключается в том, что $\theta \in \Theta_0$, а альтернатива H_1 – в том, что $\theta \in \Theta_1$. Когда множество $\Theta_0(\Theta_1)$ состоит из единственной точки, гипотеза H_0 (альтернатива H_1) называется *простой*, иначе – *сложной*.

9. Оценка Ходжеса-Лемана

Оценка Ходжеса-Лемана (HL-оценка) – это оценка *параметра сдвига* случайной величины.

Оценка Ходжеса-Лемана параметра сдвига θ определяется в виде *медианы средних Уолша* $(X_i + X_j)/2, (1 \leqslant i \leqslant j \leqslant n)$, общее число которых равно $n(n+1)/2$, и записывается в виде [6]

$$HL = \mathrm{med}\,\frac{X_i + X_j}{2}, 1 \leqslant i \leqslant j \leqslant n, \tag{1}$$

где med – выборочная медиана, $(X_i + X_j)/2,\ (i \leqslant j)$ – средние Уолша, $\{X_i\}_{i=1}^n$ – последовательность независимых одинаково распределенных случайных величин.

В другой форме оценка Ходжеса-Лемана записывается в виде (по сути та же самая *медиана средних Уолша*) [7]

$$HL = \begin{cases} W_{(r+1)}, & N = 2\,r+1, \\ (W_{(r)} + W_{(r+1)})/2, & N = 2\,r, \end{cases}$$

где $W_{(1)}, \dots, W_{(N)}$, $N = n(n+1)/2$ – упорядоченные значения *средних Уолша* $(X_i+X_j)/, 1 \leqslant i \leqslant j \leqslant n$.

Оценка Ходжеса-Лемана определяется как *медиана средних Уолша*, т.е. медиана ряда $z_1 \leqslant z_2 \leqslant \ldots \leqslant z_{\frac{n(n+1)}{2}}$, где $z_k = \dfrac{x_i + x_j}{2}$, $(i < j)$. Следует отметить высокую устойчивость этой оценки к отклонениям от нормальности распределения и засоренности выборки аномальными наблюдениями [3, стр. 103].

Для нормальной модели ее *абсолютная эффективность* АЭ$(HL) = 0.955$, то есть она проигрывает оптимальному выборочному среднему $\bar{X}$ менее 5% в эффективности. Оценка Ходжеса-Лемана является B-робастной, и следовательно, защищена от наличия выбросов в выборке.

В сравнении с этими характеристиками оценки Ходжеса-Лемана, *выборочное среднее* $\bar{X}$, являясь оптимальной оценкой параметра сдвига θ *нормального* распределения, имеет абсолютную эффективность АЭ$(\bar{X}) = 1$, однако она теряет свойства оптимальности *даже при небольших отклонениях от нормального распределения* [6].

При использовании *статистики знаков* процедура Ходжеса-Лемана приводит к оценке $\hat{\theta}$ параметра θ в виде *выборочной медианы* $\hat{\theta} = \mathrm{med}(X_1, \dots, X_n)$ [5, стр. 153].

Использование *критерия знаковых рангов Уилкоксона* приводит к *оценке Ходжеса-Лемана* (медиана средних Уолша) [5, стр. 166, стр. 167] в виде (1).

А использование *одновыборочного критерия Стьюдента* в контексте процедуру Ходжеса-Лемана приводит к *выборочному среднему* $\bar{X}$ [5, стр. 167].

10. Критерий Холлендера-Прошана

В задачах теории надежности экспоненциальное распределение наработки на отказ $f(x) = \lambda e^{-\lambda x}$ характеризуется значением параметра $\lambda = const$, т.е. постоянством интенсивности отказов изделия во времени.

Отсюда следует, что вероятность безотказной работы изделия за время Δt определяется только промежутком времени Δt и не зависит от того, работало изделие раньше или нет.

Другими словами, *вероятность безотказной работы* нового изделия и изделия, проработавшего часть времени, должна быть одинакова. Проверка этого обстоятельства и является целью *критерия Холлендера-Прошана* [3, 295].

Статистикой Холлендера-Прошана является величина [2, 182]

$$T_n = \sum_{i>j>k} \psi(X_{(i)}, X_{(j)} + X_{(k)}),$$

где

$$\psi(a,b) = \begin{cases} 1, & \text{если } a > b, \\ 1/2, & \text{если } a = b, \\ 0, & \text{если } a < b. \end{cases}$$

Суммирование здесь производится по всем $n(n-1)(n-2)/6$ упорядоченным тройками (i,j,k), для которых $i > j > k$.

Для достаточно большой выборки можно воспользоваться нормальным приближением (на основании центральной предельной теоремы)

$$\frac{T_n - \mathbf{E}T_n}{\sqrt{\mathbf{D}T_n}} \xrightarrow{d} \xi \sim \mathcal{N}(0,1),$$

где

$$\mathbf{E}T_n = n(n-1)(n-2)/8,\ \mathbf{D}T_n = \frac{3}{2}n(n-1)(n-2)\left[\frac{5}{2592}(n-3)(n-4) + \frac{7}{432}(n-3) + \frac{1}{48}\right].$$

11. Ковариация и корреляция

Случайные величины X и Y называют независимыми, если их совместная плотность $f_{XY}(x,y)$ факторизуется, то есть для нее выполняется следующее равенство [4, стр. 29]

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

Факт *независимости* случайных величин X, Y и «силу» (или «тесноту») их связи описывают с помощью таких числовых характеристик, как **ковариация** $cov(X,Y)$ и **коэффициент корреляции** ρ_{XY}.

Ковариация (корреляционный момент) двух случайных величин X и Y определяется в виде

$$cov(X,Y) = \mathbf{E}[(X-\mathbf{E}X)(Y-\mathbf{E}Y)] = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y,$$

где $\mathbf{E}X$ и $\mathbf{E}Y$ – математические ожидания случайных величин X и Y, а $\mathbf{E}(XY)$ – математическое ожидание производения случайных величин X и Y, вычисляемое по формуле

$$\mathbf{E}(XY) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xyf_{XY}(x,y)dxdy.$$

Отсюда следует формула для математического ожидания от произведения в общем случае зависимых СВ

$$\mathbf{E}(XY) = \mathbf{E}X\,\mathbf{E}Y + cov(X,Y)$$

В частном случае независимых СВ

$$\mathbf{E}(XY) = \mathbf{E}X\,\mathbf{E}Y$$

Отметим, что ковариация случайной величины X с самой собой равна дисперсии, то есть $cov(X,X) = \mathbf{D}X$. Отметим также, что если СВ X и Y *независимые*, то ковариация $cov(X,Y) = 0$ [4, стр. 29]. Это следует из того, что для независмых СВ $f_{XY}(x,y) = f_X(x)f_Y(y)$. Тогда

$$cov(X,Y) = \mathbf{E}(XY) - \mathbf{E}X\mathbf{E}Y = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xyf_{XY}(x,y)dxdy - \mathbf{E}X\mathbf{E}Y = ...$$

$$... = \underbrace{\int\limits_{-\infty}^{+\infty} xf_X(x)dx}_{\mathbf{E}X}\underbrace{\int\limits_{-\infty}^{+\infty} yf_Y(y)dy}_{\mathbf{E}Y} - \mathbf{E}X\mathbf{E}Y = 0$$

Коэффициентом корреляции СВ X и Y называют число ρ_{XY}, которые вычисляют по формулы

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y},$$

и если $\rho_{XY} = 0$, то СВ X и Y называют *некорреллированными.*

Отметим, что понятия *независимости* и *некоррелированости* не являются тождественными [4, стр. 30]. Из независимости СВ X и Y, то есть из равенства $cov(X,Y) = 0$, следует некоррелированность, то есть $\rho_{X,Y} = 0$.

Обратное утверждение в общем случае неверно, исключение составляет, например, гауссовский случай, для которого из некоррелированности следует независимость.

Рассмотрим гауссовский случай, то есть предполагаем, что $L(X,Y) = NN(a_X, a_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Эта запись означает, что двумерная СВ $Z = (X,Y)$ имеет двумерное нормальное распределение вероятностей с плотностью [4, стр. 46]

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times \dots$$
$$\dots \times \exp\left(- \frac{1}{2(1-\rho^2)} \left[\left(\frac{x-a_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x-a_X}{\sigma_X} \right) \left(\frac{y-a_Y}{\sigma_Y} \right) + \left(\frac{x-a_Y}{\sigma_Y} \right)^2 \right] \right),$$

где $a_x = \mathbf{E}X$, $a_Y = \mathbf{E}Y$, $\sigma_X^2 = \mathbf{D}X$, $\sigma_Y^2 = \mathbf{D}Y$ и ρ – коээфициент корреляции.

Пусть СВ X и Y некоррелированы, т.е. $\rho_{X,Y} = 0$. Тогда выполняется равенство

$$f_{XY}(x,y) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-a_X}{\sigma_X} \right)^2 \right) \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y-a_Y}{\sigma_Y} \right)^2 \right) = f_X(x)\, f_Y(y),$$

из которого следует независимость СВ X и Y.

Список литературы

1. *Гмурман В.Е.* Теория вероятностей и математическая статистика. – М.: Высшая школа, 1972. – 368 с.
2. *Лагутин М.Б.* Наглядная математическая статистика. – М.: БИНОМ, 2009. – 472 с.
3. *Кобзарь А.И.* Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2012. – 816 с.
4. *Шуленин В.П.* Математическая статистика. Ч.1. Параметрическая статистика. – Томск. Изд-во НТЛ, 2012. – 540 с.
5. *Шуленин В.П.* Математческая статистика. Ч.2. Непараметрическая статистика. – Томск. Изд-во НТЛ, 2012. – 388 с.
6. *Шуленин В.П.* Свойства адаптивных оценок Ходжеса-Лемана в асимптотике и при конечных объемах выборки // Вестник Томского Государственного Университета. Управление, вычислительная техника и информатика. – 2010. – №2(11). – С. 96-112
7. *Шуленин, В.П.* Свойства робастности модифицированных оценок Ходжеса-Лемана // Известия высших учебных заведений. Физика. – 2020. – №4. – С. 40-54