Теория вероятности

Элементы комбинаторики

1. Перестановка – комбинация из m элементов, отличающихся только порядком их следования $P_m = m!$
2. Сочетание – комбинация из n элементов по m определенных элементов, отличающихся между собой составом $C_n^m = \frac{n!}{m! \cdot (n-m)!}$
3. Размещение – комбинация из n по m определенных элементов отличающихся и составом, и порядком следования $A_n^m = P_m \cdot C_n^m = \frac{n!}{(n-m)!}$

Частные случаи: $C_n^n = \frac{n!}{n! \cdot 0!} = 1 \qquad C_n^0 = \frac{n!}{0! \cdot n!} = 1$

Классическая вероятность: Вероятностью называется отношение числа равно возможных несовместимых элементарных исходов, благоприятствующих появлению данного события (m) к общему числу элементарных исходов (n). $P(A) = \frac{m}{n}$

Геометрической вероятностью: называется отношение меры области в которую должна попасть точка к мере области, в которую может попасть точка.

$$P(A) = \frac{\text{mes g}}{\text{mes G}}$$

Вероятность произведения двух событий $P(AB) = P(A) \cdot P(B/A) \equiv P(B) \cdot P(A/B)$

Вероятность полного события $P(A) = \sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)$

Формула Бейеса $P(B_i / A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum_{i=1}^{n} P(B_i) \cdot P(A/B_i)}$.

Формула Бернулли $P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}$, $\qquad q = 1 - p$ -

1 Локальная формула Лапласа $P_n(k) \approx \frac{1}{\sqrt{n \cdot p \cdot q}} \cdot \varphi(x)$, $\quad x = \frac{k - n \cdot p}{\sqrt{n \cdot p \cdot q}}$, $\quad \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

2 Интегральная формула Лапласа

$$\boxed{P_n(k_1, k_2) \approx \Phi(b) - \Phi(a) \ , \quad a = \frac{k_1 - n \cdot p}{\sqrt{n \cdot p \cdot q}}, \qquad b = \frac{k_2 - n \cdot p}{\sqrt{n \cdot p \cdot q}}}$$

3 **Формула Пуассона** $P(k) \approx \frac{(n \cdot p)^k}{k!} \cdot e^{-n \cdot p}$

Дискретная случайная величина

Вероятность попадания случайной величины в интервал $P(a \le x \le b) = F(b) - F(a)$

Функция распределения: $F(x_k) = \sum_{i=1}^{k-1} P_i$.

Мат. ожидание $M(x) = \sum_{i=1}^{n} x_i \cdot P_i$, для биноминального распределения $M(x) = n \cdot P$

Дисперсия: $\mathrm{D(X) = M(X^2) - M^2(X)}$, для биноминального распределения $D(x) = n \cdot p \cdot q$

Непрерывная случайная величина

Вероятность попадания случайной величины в интервал $\mathrm{P(a \le X < b) = \int_a^b f(x)dx}$

Свойство плотности $\int_{-\infty}^{\infty} f(x)dx = 1$. **Мат. ожидание** $M(x) = \int_a^b x \cdot f(x)dx$.

Дисперсия: $D(x) = \int_a^b x^2 \cdot f(x)dx - M^2(x)$.

Характеристики равномерно распределенной непрерывной случайной величины

$$f(x) = \begin{cases} 0, x < a \\ \dfrac{1}{b-a}, a \le x < b \\ 0, x > 0 \end{cases}. \mathrm{F(x)} = \begin{cases} \mathrm{0, x < a} \\ \dfrac{\mathrm{x-a}}{\mathrm{b-a}}, \mathrm{a \le x < b} \\ \mathrm{1, x \ge b} \end{cases}. \; M(x) = \frac{b+a}{2}. D(x) = \frac{(b-a)^2}{12}.$$

Характеристики показательного распределения непрерывной случайной величины

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda \cdot x}, x \ge 0 \\ 0, x < 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda \cdot x}, x \ge 0 \\ 0, x < 0 \end{cases}$$

$$M(x) = \frac{1}{\lambda}$$

$$D(x) = \frac{1}{\lambda^2}$$

$$\sigma(x) = \frac{1}{\lambda}$$

$$P(c \le x < d) = F(d) - F(c) = 1 - e^{-\lambda \cdot d} - 1 + e^{-\lambda \cdot c} = e^{-\lambda \cdot c} - e^{-\lambda \cdot d}$$

Характеристики нормального распределения непрерывной случайной величины

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-M)^2}{2\sigma^2}}, \quad x \in (-\infty; \infty). \qquad \mathrm{P(a \le X < b) = \Phi\left(\frac{b-M}{\sigma}\right) - \Phi\left(\frac{a-M}{\sigma}\right)}$$

$\mathrm{P(|X-M| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)}$; правило трех сигм: $\mathrm{P(|X-M| < 3o) = 2\Phi(3) \approx 1}$;

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Полигон частот или относительных частот – график $P(n_i)$ или $P(w_i)$

Частота n_i число появлений варианты на данном участке.

Относительная частота $w_i = \dfrac{n_i}{n}$, где n - объём выборки $n = \sum\limits_{i=1}^{m} n_i$, где m - число участков.

Эмпирическая функция распределения: $F_k = \sum\limits_{i=1}^{k} w_i$

Вариационный ряд:

Это выборка, элементы которой расположены в порядке возрастания их значений.

Гистограмма частот или относительных частот

Это ступенчатая фигура с высотой степени $\dfrac{n_i}{h}$ или $\dfrac{w_i}{h}$ и основанием $h = \dfrac{b-a}{m}$, где a, b - соответственно нижняя и верхняя границы отрезка, в котором расположены все варианты.

Точечные оценки:

Выборочное среднее = точечная оценка мат. ожидания: $\overline{x}_B = \dfrac{1}{n}\sum\limits_{i=1}^{m} n_i x_i$;

Выборочное квадратов $\overline{x}_B^2 = \dfrac{1}{n}\sum\limits_{i=1}^{m} n_i x_i^2$; Выборочная дисперсия: $D_B = \overline{\left(x^2\right)}_B - \left(\overline{x}_B\right)^2$.

Выборочное средне квадратическое отклонение: $\sigma_B = \sqrt{D_B}$

Точечная оценка дисперсии: $S^2 = \dfrac{n}{n-1} D_B$.

Точечная оценка средне квадратического отклонения $S = \sqrt{S^2}$

Табулирование выборки

Умножение или деление на константу: $Y = CX \to x_B = \dfrac{y_B}{C}$; $Y = \dfrac{X}{C} \to x_B = Cy_B$

Прибавление или вычитание константы: $Y = C + X \to x_B = y_B - C$; $Y = X - C \to x_B = y_B + C$

ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ДЛЯ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Мат. ожидания

При известном σ : $M \in \left(\overline{x}_B - \dfrac{t\sigma}{\sqrt{n}}; \overline{x}_B + \dfrac{t\sigma}{\sqrt{n}}\right)$. t - по таблице приложения 2 по известной функции Лапласа F(t)= ω /2, где ω – заданная функция надежности.

При неизвестном σ: $\to M \in \left(\overline{x}_B - \dfrac{St_v}{\sqrt{n}}; \overline{x}_B + \dfrac{St_v}{\sqrt{n}}\right)$.

где t_v – по таблице приложения 3 в зависимости от n и ω

Средне квадратическое отклонение

$\sigma \in \left(S(1-q); S(1+q)\right)$. где q – по таблице приложения 4 в зависимости от n и ω

ТЕОРЕТИЧЕСКИЕ ЧАСТОТЫ

Для нормального распределения

$n_i' = \frac{nh}{\sigma_B}\varphi(u_i)$. где $\varphi(u_i)$ по таблице приложения 1 в зависимости от $u_i = \frac{x_i^* - \overline{x}_B}{\sigma_B}$.

КРИТЕРИЙ СОГЛАСИЯ ПИРСОНА

$\chi^2 = \sum_{i=1}^{m}\frac{\left(\overline{n}'_i - n_i\right)^2}{\overline{n}'_i}$. Если $\chi^2 \le \chi^2_{кр}$, то гипотеза о нормальном распределении принимается, в противном случае – отвергается. $\chi^2_{кр}$ определяется в зависимости от уровня значимости: α и степени свободы: $K = m - 3$ по таблице приложения 5.

РАСЧЁТ ТЕОРЕТИЧЕСКИХ ЧАСТОТ

Дискретная случайная величина

Теоретическая частота вычисляется по формуле: $n_i = nP_i$

Где **n** – объём выборки, $\mathbf{P_i}$ – вероятность, вычисленная для соответствующего типового распределения.

ПУАССОНОВСКОЕ Е РАСПРЕДЕЛЕНИЕ

В терминах теории вероятности $P_n(k) = \frac{(np)^k}{k!}$. В терминах статистики $P_i = \frac{\left(\overline{x}_B\right)^i}{i!}e^{-x_B}$

Непрерывная случайная величина

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

В терминах теории вероятности: $P(a \le x < b) = \Phi(\frac{b-M}{\sigma}) - \Phi(\frac{a-M}{\sigma})$

В терминах статистики: $P_i = \frac{1}{\sigma_B}\phi(u_i)h$, где $\phi(u_i)$ из приложения 1, $u_i = \frac{x_i^* - \overline{x}_B}{\sigma_B}$,

Проверка гипотезы о соответствии характеристик выборки нормальному распределению.

1. <u>Выборочное среднее при известном средне квадратическом отклонении</u>

Критерий: $T_x = \frac{\overline{x_B} - M}{\sigma}\sqrt{n}$.

Критическое значение этого критерия определяется по функции Лапласа (Приложение 2) из условия: $\Phi(t) = \frac{\omega}{2}$, где ω - надежность (0,95-0,999), n - объём выборки.

Если $T_x < t$, то гипотеза о соответствии выборочное среднего нормальному распределению принимается, в противном случае – отвергается.

2. <u>Выборочное среднее при неизвестном средне квадратическом отклонении</u>

Критерий: $T_x = \frac{\overline{x_B} - M}{S}\sqrt{n}$.

Критическое значение этого критерия определяется для распределения Стьюдента в зависимости от степени свободы $k = n-1$ и уровня значимости $\alpha = 1-\omega$. (Приложение 6)

Если $T_x < t_s$, то гипотеза о соответствии выборочное среднего нормальному распределению принимается, в противном случае – отвергается.

<u>3. Исправленная дисперсия</u>

Критерий: $\chi^2 = \frac{(n-1)S^2}{D}$.

Критическое значение этого критерия определяется для распределения Стьюдента в зависимости от степени свободы $k = n-1$ и уровня значимости $\alpha = 1-\omega$. (Приложение 5)

Если $\chi^2 < \chi^2_{кр}$, то гипотеза о соответствии исправленной дисперсии нормальному распределению принимается, в противном случае – отвергается.

ЛИНЕЙНАЯ РЕГРЕССИЯ

1 часть Выборочное уравнение линейной регрессии. **g(x)=a$_0$+a$_1$x**

$$a_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \quad a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Относительна погрешность $\Delta_i = \frac{g(x_i) - y_i}{y_i} 100\%$.

Между двумя случайными величинами существует функциональная зависимость, если все погрешности Δ_i=0 , если $\Delta i \neq 0$, то линейная корреляционная зависимость,
если $\Delta i > 5\%$ - нелинейная корреляционная зависимость.

2 часть Корреляционная таблица и выборочный коэффициент корреляции.

Y_i / X_i	x_1	x_2	…	x_m	n_{y_i}
y_1	$n_{x_1 y_1}$	$n_{x_2 y_1}$	…	$n_{x_m y_1}$	n_{y_1}
y_2	$n_{x_1 y_2}$	$n_{x_2 y_2}$	…	$n_{x_m y_2}$	n_{y_2}
….	….	….	….	….	….
y_k	$n_{x_1 y_k}$	$n_{x_2 y_k}$	….	$n_{x_m y_k}$	n_{y_k}
n_{x_j}	n_{x_1}	n_{x_2}	….	n_{x_m}	n

Выборочный коэффициент корреляции: $r_B = \frac{S_{XY} - n\overline{x}_B \overline{y}_B}{n\sigma_{BX}\sigma_{BY}}$,

где $\overline{x}_B = \frac{1}{n}\sum_{i=1}^{m} n_{x_i} x_i$, $\overline{y}_B = \frac{1}{n}\sum_{j=1}^{k} n_{y_j} y_j$, $\left(\overline{x^2}\right)_B = \frac{1}{n}\sum_{i=1}^{m} n_{x_i} x_i^2$, $S_{xy} = \sum_{i=1}^{m}\sum_{j=1}^{k} n_{x_i y_j} x_i y_j$

$$D_{Bx} = \left(x^2\right)_B - \left(x_B\right)^2; D_{By} = \left(y^2\right)_B - \left(y_B\right)^2; \sigma_{Bx} = \sqrt{D_{Bx}}; \sigma_{By} = \sqrt{D_{By}}.$$

Выборочное уравнение линейной регрессии (проверка): $g(x) = \overline{y}_B + r_B \frac{\sigma_{By}}{\sigma_{Bx}}\left(x - \overline{x}_B\right)$.

Между двумя случайными величинами существует функциональная зависимость, если $r_B = 1$, в противном случае корреляционная зависимость.

3 часть Корреляционное отношение

$\eta = \frac{\sigma_{yx}}{\sigma_{By}}$, где $\sigma_{yx} = \sqrt{D_{yx}}$, $D_{yx} = 1/n \sum_{i=1}^{m} n_{xj} \left(g(x_i) - \overline{y}_B\right)^2$

Между двумя случайными величинами существует функциональная зависимость, если η=1,
между случайными величинами не существует никакой зависимости, если η=0,
между случайными величинами существует линейная зависимость, если $\eta = |r_B|$.