Глава 11

Основни статистически разпределения

11.1 Квантили на разпределенията U, χ^2, T и F

В математическата статистика често се използват разпределения, които са свързани с нормалното разпределение. Това са разпределението χ^2 („хи-квадрат”) на К. Пирсън[1], $T-$разпределението на Стюдънт[2] и разпределението F на Фишер[3] (или Фишер – Снедекор[4]). В тази връзка ще използваме следните означения: U – случайна величина, имаща стандартно нормално разпределение $N(0,1)$; $\chi^2(k)$ – случайна величина, имаща разпределение χ^2 с k степени на свобода; $T(k)$ – случайна величина, имаща разпределение на Стюдънт с k степени на свобода; $F(k_1, k_2)$ – случайна величина, имаща разпределение на Фишер с k_1 и k_2 степени на свобода. Съответните функции на разпределение на тези случайни величини се означават с $U(x), \chi^2(x;k), T(x;k), F(x;k_1,k_2)$. Нека $R(x)$ е една от тези функции на разпределение и $0 < p < 1$.

Определение 11.1. Квантил от ред p на разпределението R *се нарича числото x_p, за което*

$$R(x_p) = p. \tag{11.1}$$

[1] Пѝрсън (Pearson) Карл (1857 – 1936) – английски математик, биолог, философ

[2] Стю̀дънт (Student) – псевдоним на William S. Gosset (1876 – 1937) – английски математик и статистик

[3] Фѝшер (Fisher) Роналд (1890 – 1962) – английски статистик и генетик

[4] Снедекор (Snedecor) Джордж (1881 – 1974) – американски математик и статистик

Квантилите от ред p на горните разпределения се означават съответно с $u_p, \chi^2_p(k), t_p(k), F_p(k_1, k_2)$. Тези квантили са дадени в Таблиците $4-9$, стр. $211-216$.

Понеже плътността на нормалното разпределение U е симетрична относно нулата, то

$$u_p = -u_{1-p}\,. \tag{11.2}$$

Пример 11.1. $u_{0.1} = -u_{0.9} = -1.282$ (*Таблица* 4).

Ще дадем определения и някои основни свойства на разпределенията χ^2, на Стюдънт и на Фишер.

11.2 Разпределение $\chi^2(k)$

Определение 11.2. *Случайната величина $\chi^2(k)$ е равна на сумата на квадратите на k независими случайни величини U_i, $i = 1, \ldots, k$, всяка от които има стандартно нормално разпределение $N(0,1)$, т.е.*

$$\chi^2(k) = U_1^2 + U_2^2 + \cdots + U_k^2. \tag{11.3}$$

Разпределението на тази случайна величина се нарича **разпределение χ^2 с k степени на свобода**.

Математическото очакване и дисперсията на $\chi^2(k)$ са равни на

$$E(\chi^2(k)) = k, \quad D(\chi^2(k)) = 2k.$$

При големи стойности на k ($k > 30$) разпределението $\chi^2(k)$ с достатъчна за практиката точност се апроксимира с нормално разпределение $N(k, \sqrt{2k})$. За пресмятане на квантилите $\chi^2_p(k)$ при $k > 30$ може да се използва формулата

$$\chi^2_p(k) \approx \frac{1}{2}[u_p + \sqrt{2k-1}]^2. \tag{11.4}$$

При малки стойности на p може да се използва по-точната формула

$$\chi^2_p(k) \approx k\left[1 - \frac{2}{9k} + u_p\sqrt{\frac{2}{9k}}\right]^3. \tag{11.5}$$

Пример 11.2. *Да се намерят квантилите* $\chi^2_{0.01}(10)$, $\chi^2_{0.95}(100)$, $\chi^2_{0.01}(100)$.

От таблицата за квантилите на разпределението χ^2 (*Таблица* 5) *намираме, че* $\chi^2_{0.01}(10) = 2.56$.

За пресмятането на $\chi^2_{0.95}(100)$ *ще използваме формула* (11.4). *Понеже* $u_{0.95} = 1.645$ (*Таблица* 4), *то*

$$\chi^2_{0.95}(100) \approx \frac{1}{2}[1.645 + \sqrt{2 \cdot 100 - 1}]^2 \approx 124.06.$$

Квантила $\chi^2_{0.01}(100)$ *ще пресметнем по формула* (11.5). *Понеже* $u_{0.01} = -u_{0.99} = -2.326$, *то*

$$\chi^2_{0.01}(100) \approx 100\left[1 - \frac{2}{9 \cdot 100} + (-2.326)\sqrt{\frac{2}{9 \cdot 100}}\right]^3 \approx 70.05.$$

11.3 Разпределение на Стюдънт $T(k)$

Определение 11.3. *Случайната величина* $T(k)$ *е отношение на две независими случайни величини* U *и* $\sqrt{\chi^2(k)/k}$, *т.е.*

$$T(k) = \frac{U}{\sqrt{\chi^2(k)/k}}\,\cdot$$

Разпределението на тази случайна величина се нарича **разпределение на Стюдънт с k степени на свобода**.

Математическото очакване и дисперсията на $T(k)$ са равни на

$$E(T(k)) = 0, \quad D(T(k)) = \frac{k}{k-2}\,\cdot$$

Понеже плътността на разпределението на Стюдънт е симетрична относно нулата, то

$$t_p(k) = -t_{1-p}(k). \tag{11.6}$$

При $k > 30$ за пресмятане на квантилите $t_p(k)$ може да се използва формулата

$$t_p(k) \approx u_p\,. \tag{11.7}$$

Пример 11.3. *Да се намерят квантилите* $t_{0.05}(8)$ *и* $t_{0.90}(40)$.

От Таблица 6 *намираме, че* $t_{0.95}(8) = 1.86$. *Като отчетем формула* (11.6), *получаваме,че* $t_{0.05}(8) = -1.86$. *Квантила* $t_{0.90}(40)$ *определяме по формула* (11.7): $t_{0.90}(40) \approx u_{0.90} \approx 1.28$. *За сравнение,* $t_{0.90}(40) = 1.303$.

11.4 Разпределение на Фишер $F(k_1, k_2)$

Определение 11.4. *Случайната величина* $F(k_1, k_2)$ *е равна на отношението на две независими случайни величини* $\chi^2(k_1)/k_1$ *и* $\chi^2(k_2)/k_2$, *т.е.*

$$F(k_1, k_2) = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}\,. \tag{11.8}$$

Разпределението на тази случайна величина се нарича **разпределение на Фишер с k_1 и k_2 степени на свобода**.

Математическото очакване на $F(k_1, k_2)$ е равно на

$$E(F(k_1, k_2)) = \frac{k_2}{k_2 - 2},\ k_2 > 2.$$

Квантилите на разпределението на Фишер от редовете p и $1-p$ са свързани с формулата

$$F_p(k_1, k_2) = \frac{1}{F_{1-p}(k_2, k_1)}\,. \tag{11.9}$$

Между случайните величини, имащи разпределения U, $T(k)$, $\chi^2(k)$ и $F(k_1, k_2)$, са в сила следните зависимости:

$$T^2(k) = F(1, k), \tag{11.10}$$

$$F(k, \infty) = \frac{\chi^2(k)}{k}, \tag{11.11}$$

$$\chi^2(1) = U^2. \tag{11.12}$$

Пример 11.4. *Да се намерят квантилите* $F_{0.01}(3, 5)$, $F_{0.9}(4, 100)$ *и* $F_{0.05}(1, 80)$.

Като отчетем формула (11.9) *и Таблица* 9, *получаваме*

$$F_{0.01}(3, 5) = \frac{1}{F_{0.99}(5, 3)} = \frac{1}{28.24} \approx 0.035.$$

Като отчетем формула (11.11) *и Таблица* 5, *получаваме*

$$F_{0.9}(4, 100) \approx \frac{\chi^2_{0.9}(4)}{4} = \frac{7.78}{4} = 1.945.$$

Накрая, като отчетем формулите (11.11) *и* (11.12) *и Таблица* 4, *намираме*

$$F_{0.05}(1, 80) \approx \chi^2_{0.05}(1) = u^2_{0.05} = (-u_{0.95})^2 = 1.645^2 \approx 2.706.$$

11.5 Основни статистики, свързани с нормално разпределена генерална съвкупност

Нека статистическата променлива X има нормално разпределение $N(m, \sigma)$ и при нейното статистическо проучване е получена извадката с обем n:

$$x_1, x_2, \ldots, x_n. \tag{11.13}$$

Тези стойности на извадката може да се разглеждат като конкретни реализации на стойностите на n случайни величини $X_1, X_2, \ldots, X_n$, които са независими и имат нормално разпределение $N(m, \sigma)$. Затова всяка статистика (функция на $x_1, x_2, \ldots, x_n$) е случайна величина, която има някакво разпределение.

Разпределенията на основните статистики, получени на базата на извадката (11.13) на нормално разпределена $N(m, \sigma)$ генерална съвкупност, може да се получат от следните факти:

1. Статистическото средно

   $$\overline{x} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

   има нормално разпределение $N(m, \frac{\sigma}{\sqrt{n}})$. Тогава статистиката

   $$U = \frac{\overline{x} - m}{\sigma/\sqrt{n}}$$

   има стандартно нормално разпределение $N(0, 1)$.

2. Статистическата дисперсия

   $$D_0 = s_0^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - m)^2$$

   има разпределение, което е свързано с разпределението χ^2 с n степени на свобода чрез съотношението

   $$s_0^2 = \frac{\sigma^2}{n}\chi^2(n).$$

3. Неизместената статистическата дисперсия

$$\overline{D} = \overline{s}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{x})^2$$

има разпределение, което е свързано с разпределението χ^2 с $(n-1)$ степени на свобода чрез съотношението

$$\overline{s}^2 = \frac{\sigma^2}{n-1}\chi^2(n-1).$$

4. Статистиката $\dfrac{\overline{x}-m}{\overline{s}/\sqrt{n}}$ е случайна величина, имаща разпределение на Стюдънт с $(n-1)$ степени на свобода, т.е.

$$\frac{\overline{x}-m}{\overline{s}/\sqrt{n}} = T(n-1).$$

5. Нека $\overline{s}_1^2$ и $\overline{s}_2^2$ са неизместени статистически дисперсии, изчислени по независими извадки с обеми n_1 и n_2 от две нормално разпределени генерални съвкупности с дисперсии съответно σ_1^2 и σ_2^2. Тогава отношението $\dfrac{\overline{s}_1^2/\sigma_1^2}{\overline{s}_2^2/\sigma_2^2}$ има разпределение на Фишер с (n_1-1) и (n_2-1) степени на свобода, т.е.

$$\frac{\overline{s}_1^2/\sigma_1^2}{\overline{s}_2^2/\sigma_2^2} = F(n_1-1, n_2-1).$$