Глава 13

Свойства на нормалните извадки, χ^2 и t разпределения

Ще разгледаме някои важни гранични резултати, свързани с извадки, направени от нормални съвкупности, т.е. предполагаме, че неизвестния признак, върху който правим наблюденията има нормално разпределение.

Теорема 13.1 *Нека* $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$, *и* $s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \bar{X})^2$. *Тогава:*
a) $\bar{X}$ *и* s^2 *са независими сл.в.,*
b) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$,
c) $(n-1)\dfrac{s^2}{\sigma^2} \sim \chi^2(n-1)$.

Доказателство: а) Б.О.О. ще смятаме, че $\mu = 0$ и $\sigma^2 = 1$. Тогава

$$
\begin{aligned}
s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \frac{1}{n-1}\left((X_1 - \bar{X})^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) \\
&= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} X_i - \sum_{i=2}^{n} X_i - n\bar{X} + (n-1)\bar{X}\right)^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) \\
&= \frac{1}{n-1}\left(\left(\sum_{i=1}^{n}(X_i - \bar{X}) - \sum_{i=2}^{n}(X_i - \bar{X})\right)^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) \\
&= \frac{1}{n-1}\left(\left(0 - \sum_{i=2}^{n}(X_i - \bar{X})\right)^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right) \\
&= \frac{1}{n-1}\left(\left(\sum_{i=2}^{n}(X_i - \bar{X})\right)^2 + \sum_{i=2}^{n}(X_i - \bar{X})^2\right),
\end{aligned}
$$

което показва, че s^2 е функция само на $(X_2 - \bar{X}, X_3 - \bar{X}, \dots, X_n - \bar{X})$.

Съвместната плътност на $X_1, X_2, \dots, X_n$ е $f(x_1, x_2, \dots, x_n) = \dfrac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum\limits_{i=1}^{n} x_i^2}$, $x_i \in (-\infty, \infty)$. Ще направим следната трансформация:

$y_1 = \bar{x}$,
$y_2 = x_2 - \bar{x}$,
...
$y_n = x_n - \bar{x}$.

Тогава обратната трансформация е:

$x_1 = y_1 - \sum_{i=2}^{n} y_i$,
$x_2 = y_2 + y_1$,
...
$x_n = y_n - y_1$.

Якобианът $|J| = \frac{1}{n}$ и съответно съвместната плътност на новите променливи има вида:

$$f(y_1, y_2, \dots, y_n) = \frac{n}{(2\pi)^{\frac{n}{2}}} e^{-\left[\frac{1}{2}\left(y_1 - \sum\limits_{i=2}^{n} y_i\right)^2 + \sum\limits_{i=2}^{n}(y_i + y_1)^2\right]}$$

$$= \left[\left(\frac{n}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{ny_1^2}{2}}\right] \left[\frac{n^{\frac{1}{2}}}{(2\pi)^{\frac{n-1}{2}}} e^{-\frac{1}{2}\left[\sum\limits_{i=2}^{n} y_i^2 + \left(\sum\limits_{i=2}^{n} y_i\right)^2\right]}\right].$$

Понеже съвместната плътност се разлага на произведение на две плътности, това показва, че Y_1 е независимо от $Y_2, \dots, Y_n$, следователно и $\bar{X}$ е независимо от s^2.

b) Ще използваме пораждаща моментите функция. За нормалното разпределение знаем, че тя има вида: $M_{X_i}(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$. Тогава за п.м.ф. на $\bar{X}$ имаме:

$$M_{\bar{X}}(t) = \mathbf{E}e^{t\bar{X}} = \mathbf{E}e^{t\frac{X_1 + \ldots + X_n}{n}} = \mathbf{E}e^{\frac{t}{n}Y} = M_Y\left(\frac{t}{n}\right) = \left[M_{X_i}\left(\frac{t}{n}\right)\right]^n$$

$$= \left[\exp\left(\mu\frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}\right)\right]^n = \exp\left(n\left(\mu\frac{t}{n} + \frac{\sigma^2 t^2}{2n^2}\right)\right) = \exp\left(\mu t + \frac{\frac{\sigma^2}{n}t^2}{2}\right),$$

където $Y = \sum\limits_{i=1}^{n} X_i$ и $X_1, \dots, X_n$ са независими. Оттук следва, че $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Преди да докажем последното твърдение в теоремата, да напомним следните факти за χ-квадрат разпределението (виж Твърдение 4.7 и Твърдение 4.8):

1) Ако $Z \sim N(0,1)$, то $Z^2 \sim \chi^2(1)$.

2) Ако $X_1, \dots, X_n$ са независими и $X_i \sim \chi^2(p_i)$, то $\sum\limits_{i=1}^{n} X_i \sim \chi^2(\sum\limits_{i=1}^{n} p_i)$.

Да се върнем към доказателството на теоремата.

c) Нека $\bar{X}_k = \dfrac{1}{k}\sum\limits_{i=1}^{k} X_i$ и $s_k^2 = \dfrac{1}{k-1}\sum\limits_{i=1}^{k}(X_i - \bar{X}_k)^2$. Тогава имаме следната рекурентна зависимост (докажете!):

$$(n-1)s_n^2 = (n-2)s_{n-1}^2 + \frac{n-1}{n}(X_n - \bar{X}_{n-1})^2.$$

Ще използваме индукция по k:

$$k = 2 : 0.S_1^2 = 0 \Rightarrow s_2^2 = \frac{1}{2}(X_2 - X_1)^2, \quad \frac{X_2 - X_1}{\sqrt{2}} \sim N(0,1) \Rightarrow s_2^2 \sim \chi^2(1)$$

Нека е изпълнено за $k : (k-1)s_k^2 \sim \chi^2(k-1)$, тогава за $k+1$:

$$ks_{k+1}^2 = (k-1)s_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2.$$

Първото събираемо $(k-1)s_k^2 \sim \chi^2(k-1)$ от индукционното предположение, а за второто имаме $X_{k+1} - \bar{X}_k \sim N(0, \frac{k+1}{k})$. Освен това $(X_{k+1} - \bar{X}_k)^2$ и s_k^2 са независими, защото $(X_{k+1}, \bar{X}_k)$ не зависи от s_k^2. Тогава $\frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2 \sim \chi^2(1)$ и $ks_{k+1}^2 \sim \chi^2(k)$.

Какво можем да кажем за разпределението на $\bar{X}$, когато σ^2 не е известно? Нека разгледаме представянето:

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{s^2/\sigma^2}}.$$

Знаем, че $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$, а $\sqrt{s^2/\sigma^2} \sim \sqrt{\chi^2(n-1)/(n-1)}$, така че задачата се свежда до намиране разпредлението на сл.в. $\frac{U}{\sqrt{V/p}}$, където $U \sim N(0,1)$ и $V \sim \chi^2(p)$. Съвместната плътност на U и V е:

$$f_{U,V}(u,v) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{1}{\Gamma(\frac{p}{2}) 2^{\frac{p}{2}}} v^{\frac{p}{2}-1} e^{-\frac{v}{2}}, \quad u \in (-\infty, \infty), \quad v > 0.$$

Правим трансформацията: $t = \frac{u}{\sqrt{v/p}}$, $w = v$. Якобианът $|J| = \left(\frac{w}{p}\right)^{\frac{1}{2}}$, а за маргиналната плътност на T получаваме:

$$\begin{aligned} f_T(t) &= \int_0^\infty f_{U,V}\left(t\left(\frac{w}{p}\right)^{\frac{1}{2}}, w\right)\left(\frac{w}{p}\right)^{\frac{1}{2}} dw \\ &= \frac{1}{\sqrt{2\pi}\Gamma(\frac{p}{2})2^{\frac{p}{2}}} \int_0^\infty e^{-\frac{t^2 w}{2p}} w^{\frac{p}{2}-1} e^{-\frac{w}{2}} \left(\frac{w}{p}\right)^{\frac{1}{2}} dw \\ &= \frac{1}{\sqrt{2\pi}\Gamma(\frac{p}{2})2^{\frac{p}{2}}\sqrt{p}} \int_0^\infty e^{-\frac{1}{2}\left(1+\frac{t^2}{p}\right)w} w^{\frac{p+1}{2}-1} dw \end{aligned}$$

Но под интеграла всъщност имаме плътността (ненормирана) на $\Gamma\left(\frac{p+1}{2}, \frac{2}{1+\frac{t^2}{p}}\right)$, следователно:

$$\begin{aligned} f_T(t) &= \frac{1}{\sqrt{2\pi}\Gamma(\frac{p}{2})2^{\frac{p}{2}}\sqrt{p}} \Gamma\left(\frac{p+1}{2}\right)\left[\frac{2}{1+\frac{t^2}{p}}\right]^{\frac{p+1}{2}} \\ &= \frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{p}{2}\right)} \frac{1}{\sqrt{p\pi}\left(1+\frac{t^2}{p}\right)^{\frac{p+1}{2}}}, \quad t \in (-\infty, \infty) \end{aligned}$$

Разпределението на T наричаме t-разпределение или разпределение на Стюдънт с p степени на свобода (бележим T_p).