A χ^2-eloszlás

Szükségünk van az alábbi tényekre a kalkulusból. Az **Euler-féle bétaintegrál** $B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$, ahol α és β pozitív valós számok. A **gammafüggvény** $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$, $x > 0$. Nevezetes összefüggés $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, továbbá ismeretes, hogy $\Gamma(x+1) = x\Gamma(x)$, $\Gamma(n+1) = n!$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, $\Gamma(n+\frac{1}{2}) = \frac{1\cdot 3\cdots(2n-1)}{2^n}\sqrt{\pi}$. Legyenek $\xi_1,\ldots,\xi_n$ független, standard normális eloszlású valószínűségi változók, ekkor a $\chi^2 = \xi_1^2+\cdots+\xi_n^2$ valószínűségi változót n **szabadsági fokú** χ^2**-eloszlású**nak nevezzük. Más szemszögből: Legyenek $\xi_1,\xi_2,\ldots,\xi_n$ teljesen független valószínűségi változók azonos normális eloszlásúak m várható értékkel és σ szórással. Ekkor:

Az (m,σ)-paraméterű normális populációból származó független minta esetén a $\sum_{i=1}^n \frac{(\xi_i-m)^2}{\sigma^2}$ valószínűségi változó n szabadsági fokú χ^2-eloszlású.

Megjegyezzük, hogy ebben az állításban a várható érték és a szórás ismert. Ha az m várható érték ismert, a szórásnégyzetet becsülhetjük a $\frac{1}{n}\sum_{i=1}^n(\xi_i-m)^2$ értékkel, és alkalmazhatjuk az állítást. Vegyük észre, hogy a szummában n független négyzet szerepel.

Ha $\xi_1,\ldots,\xi_n$ független, standard normális eloszlású valószínűségi változók, akkor $\sum_{i=1}^n(\xi_i-\overline{\xi})^2$ $n-1$ szabadsági fokú χ^2-eloszlású.

Hiszen,

$$\sum_{i=1}^n \xi_i^2 = \sum_{i=1}^n(\xi_i-\overline{\xi}+\overline{\xi})^2 =$$

$$\sum_{i=1}^n(\xi_i-\overline{\xi})^2 - 2\overline{\xi}\sum_{i=1}^n(\xi_i-\overline{\xi}) + \sum_{i=1}^n\overline{\xi}^2 =$$

$$\sum_{i=1}^n(\xi_i-\overline{\xi})^2 + n\overline{\xi}^2,$$

és mivel ismeretes, hogy $\sum_{i=1}^n(\xi_i-\overline{\xi})^2$ és $n\overline{\xi}^2$ független valószínűségi változók:

$$M_{\sum_{i=1}^n \xi_i^2}(t) = M_{\sum_{i=1}^n(\xi_i-\overline{\xi})^2}(t)M_{n\overline{\xi}^2}(t)$$

Mint ismeretes, a χ^2-eloszlás generátorfüggvénye $M_{\chi^2}(t) = \frac{1}{(1-2t)^{n/2}}$, és azt, hogy a mintaátlag $(0,1/\sqrt{n})$-paraméterű normális eloszlású,

$$\frac{1}{(1-2t)^{n/2}} = M_{\sum_{i=1}^n(\xi_i-\overline{\xi})^2}(t)\frac{1}{(1-2t)^{1/2}},$$

vagyis

$$M_{\sum_{i=1}^n(\xi_i-\bar{\xi})^2}(t) = \frac{1}{(1-2t)^{(n-1)/2}},$$

éppen a $n-1$ szabadsági fokú χ^2-eloszlás generátorfüggvénye. Ezzel a bizonyítás kész.

Vegyük észre, hogy a szummában $n-1$ független négyzet szerepel, mivel a $\sum_{i=1}^n(\xi_i-\bar{\xi}) = 0$ összefüggésből az egyik eltérés a többiből meghatározható. Egy következmény:

Normális populáció esetén, ha S_n^{*2} a korrigált tapasztalati szórásnégyzet, akkor $\frac{(n-1)S_n^{*2}}{\sigma^2}$ $n-1$ szabadsági fokú χ^2-eloszlású.

Állítjuk, hogy:

A χ^2 valószínűségi változó sűrűségfüggvénye

$$f_n(x) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)} \text{ ha } x \geq 0$$

és $f_n(x) = 0$ egyébként.

Hiszen, ismeretes, hogy ha a ψ függvény monoton, differenciálható, nemnulla, akkor az $\eta = \psi(\xi)$ valószínűségi változó sűrűségfüggvénye $g(y) = \frac{f(\psi^{-1}(y))}{\psi'(\psi^{-1}(y))}$ ha $y \in (\inf\psi, \sup\psi)$ és 0 egyébként. Innen ξ_1^2 sűrűségfüggvénye $2\frac{\varphi(\sqrt{y})}{2\sqrt{y}} = \frac{1}{\sqrt{y}\sqrt{2\pi}}e^{-y/2}$ ha $y > 0$, 0 egyébként. Másrészről $f_1(y) = \frac{y^{-1/2}e^{-y/2}}{2^{1/2}\Gamma(1/2)} = \frac{1}{\sqrt{2\pi y}}e^{-y/2}$ ha $y > 0$, 0 egyébként. $n = 1$-re igaz állításunk. Indukcióval folytatjuk. A konvolúció sűrűségfüggvényéről tudjuk, hogy $h(z) = \int_{-\infty}^{\infty} f(z-x)g(x)dx$. Innen, ha $x < z$,

$$f_{n+1}(z) = \int_0^z \frac{1}{\sqrt{2\pi(z-x)}}e^{-(z-x)/2}\frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}dx =$$

$$\frac{e^{-z/2}}{\sqrt{2\pi}2^{n/2}\Gamma(n/2)}\int_0^z \frac{x^{n/2-1}}{\sqrt{z-x}}dx.$$

Legyen $x = zt$, $z > 0$, ekkor

$$\int_0^z \frac{x^{n/2-1}}{\sqrt{z-x}}dx = \int_0^1 \frac{z^{n/2-1}t^{n/2-1}}{\sqrt{z-zt}}\frac{dx}{dt}dt =$$

$$z^{(n+1)/2-1}\int_0^1 t^{n/2-1}(1-t)^{-1/2}dt = z^{(n+1)/2-1}B(n/2, 1/2),$$

ebből következik, hogy

$$f_{n+1}(z) = \frac{z^{(n+1)/2-1}e^{-z/2}}{2^{(n+1)/2}\Gamma(1/2)\Gamma(n/2)} \frac{\Gamma(1/2)\Gamma(n/2)}{\Gamma((n+1)/2)} =$$

$= \frac{z^{(n+1)/2-1}e^{-z/2}}{2^{(n+1)/2}\Gamma((n+1)/2)}$ ha $z > 0$, illetve 0 egyébként. Ezzel beláttuk az állításunkat.

Egyszerű számolás mutatja, hogy

A χ-eloszlás *sűrűségfüggvénye* $\frac{2x^{n-1}e^{-x^2/2}}{2^{n/2}\Gamma(n/2)}$ *ha* $x > 0$*, illetve 0 egyébként.*

Számítsuk ki a χ^2-eloszlás várható értékét és szórásnégyzetét! (A várható érték nyilván azonnal adódna mint a várható értékek összege.) A negyedik momentum, parciális integrálással,

$$E\xi^4 = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^4 e^{-x^2/2} dx = \frac{-1}{\sqrt{2\pi}} \int_0^\infty x^3 \left(-xe^{-x^2/2}\right) dx =$$

$$\frac{-1}{\sqrt{2\pi}} \left(\left[x^3 e^{-x^2/2} \right]_0^\infty - 3 \int_0^\infty x^2 e^{-x^2/2} dx \right) =$$

$$3 \int_0^\infty x^2 \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 3E\xi^2 = 3.$$

Parciális integrálással,

$$E\chi_n^2 = \int_{-\infty}^\infty x f_n(x) dx = \int_0^\infty \frac{x^{n/2}e^{-x/2}}{2^{n/2}\Gamma(n/2)} dx =$$

$$\frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty x^{n/2} e^{-x/2} dx$$

$$= \frac{1}{2^{n/2}\Gamma(n/2)} \left(\left[(-2)x^{n/2}e^{-x/2} \right]_0^\infty + 2\frac{n}{2} \int_0^\infty x^{n/2-1} e^{-x/2} dx \right)$$

$= n \int_0^\infty f_n(x)dx = n$. Nyilván $D^2\xi_i^2 = E\xi_i^4 - (E\xi_i^2)^2 = 3 - 1 = 2$, ahonnan $D^2\chi_n^2 = D^2\xi_1^2 + \cdots + D^2\xi_n^2 = 2n$.

Kaptuk, hogy

Az n szabadsági fokú χ^2*-eloszlás várható értéke n, és szórásnégyzete* $2n$*.*

Egyszerű számolás mutatja, hogy

A χ*-eloszlás várható értéke* $\frac{\sqrt{2}\Gamma((n+1)/2)}{\Gamma(n/2)}$.