Föreл. 14
161213

χ^2-test

I de hypotestest vi hittills studerat har hypoteserna gällt värdena på okända parametrar i en känd fördelning, exvis $Bin(n,p)$ eller $N(\mu,\sigma)$. Detta kallas parametrisk inferens. Vi ska idag studera några metoder för att undersöka vilken fördelning som ett givet stickprov kommer från. Detta kallas icke-parametrisk inferens. Ett exempel på nollhypotes och mothypotes skulle kunna vara H_0: $X \in Po(\lambda)$ för något $\lambda > 0$ och H_1: X är inte Poissonfördelad. Vi kommer att testa denna typ av hypoteser med så kallade χ^2-test.

Antag att ett försök har r möjliga utfall $A_1, \dots, A_r$ med sannolikheterna $\mathbb{P}(A_1), \dots, \mathbb{P}(A_r)$. Antag att försöket upprepas n gånger och att $x_1, \dots, x_r$ är de absoluta frekvenserna för utfallen $A_1, \dots, A_r$ ($\sum_{j=1}^r x_j = n$). Låt $p_1, \dots, p_r$ vara givna sannolikheter som uppfyller $\sum_{j=1}^r p_j = 1$. Vi vill testa nollhypotesen

$$H_0: \mathbb{P}(A_j) = p_j \qquad \text{för alla } j \in \{1, \dots, r\}$$

mot mothypotesen

$$H_1: \mathbb{P}(A_j) \neq p_j \qquad \text{för minst ett } j \in \{1, \dots, r\}.$$

Vi testar alltså om utfallen av försöket har den fördelning som ges av $p_1, \dots, p_r$ (goodness-of-fit-test). För att testa H_0 bildar vi testvariabeln

$$Q_{obs} = \sum_{j=1}^{r} \frac{(x_j - np_j)^2}{np_j}$$

Om H_0 är sann, så är np_j det förväntade värdet på den absoluta frekvensen x_j, så ett lågt värde på Q_{obs} indikerar en god överensstämmelse mellan $x_1, \dots, x_r$ och den givna fördelningen $p_1, \dots, p_r$, medan ett högt värde på Q_{obs} indikerar att H_0 bör förkastas. Det kan bevisas att om H_0 är sann, så är den stickprovsvariabel Q, som Q_{obs} är ett utfall av, approximativt $\chi^2(r-1)$-fördelad för stora n. Vi får följande test som för stora n har felrisk α

– Förkasta H_0 om $Q_{obs} > \chi^2_\alpha(r-1)$.

Approximationen som testet baseras på är god om $np_j \geq 5$ för alla $j \in \{1, \dots, r\}$.

Ex: En sexsidig tärning kastas 96 gånger. Utfallet blir 15 ettor, 7 tvåor, 9 treor, 20 fyror, 26 femmor och 19 sexor. Vi vill pröva hypotesen att tärningen är symmetrisk och nollhypotesen blir därmed H_0: $\mathbb{P}(A_j) = p_j = 1/6$ för alla $j \in \{1, \dots, 6\}$. Eftersom $np_j = 96 \cdot 1/6 = 16$ för alla j, så är approximationen som testet bygger på god. Testvariabeln blir

$$Q_{obs} = \frac{(15-16)^2}{16} + \frac{(7-16)^2}{16} + \frac{(9-16)^2}{16} + \frac{(20-16)^2}{16} + \frac{(26-16)^2}{16} + \frac{(19-16)^2}{16} = 16{,}6$$

och antalet frihetsgrader är $r-1 = 5$. Eftersom $\chi^2_{0,05}(5) = 11{,}1$, $\chi^2_{0,01}(5) = 15{,}1$ och $\chi^2_{0,001}(5) = 20{,}5$, så kan vi förkasta H_0 med felriskerna $0{,}05$ och $0{,}01$ men inte med felrisken $0{,}001$. Vi kan säga att tärningen är asymmetrisk och att resultatet är signifikant**.

En liknande metod kan användas för att testa om ett stickprov kommer från en fördelning med en okänd parameter $\theta = (\theta_1, \dots, \theta_s)$, ex.vis $Po(\mu)$. Nollhypotesen blir då H_0: $\mathbb{P}(A_j) = p_j(\theta)$ för alla $j \in \{1, \dots, r\}$. För att testa H_0, skattar vi först $\theta^*_{obs} = \theta^*(x_1, \dots, x_r)$ från stickprovet under antagandet att H_0 är sann. Sedan sätter vi in θ^*_{obs} istället för θ i nollhypotesen och beräknar testvariabeln $Q_{obs} = \sum_{j=1}^r (x_j - np_j(\theta^*_{obs}))/np_j(\theta^*_{obs})$. Om H_0 är sann, så är Q_{obs} approximativt $\chi^2(r-s-1)$-fördelad för stora n. Vi ser att antalet frihetsgrader minskar med antalet skattade parametrar s. Läs själva exempel 13.18 i boken där det testas om ett givet data är $Po(\mu)$-fördelat för något $\mu > 0$.

Nästa typ av test som vi ska undersöka kallas homogenitetstest. Sådana test används för att undersöka om ett antal försöksserier kommer från samma fördelning (är homogena). Antag att vi har s försöksserier om n_i försök och låt x_{ij} beteckna antalet gånger som utfall A_j uppkommer i försöksserie i. Vi vill testa nollhypotesen

H_0: $\mathbb{P}(A_j) = p_j$ för alla $j \in \{1, \dots, r\}$ och alla serier.

Serie \ Utfall	A_1	A_2	$\cdots$	A_r	# försök
1	x_{11}	x_{12}	$\cdots$	x_{1r}	n_1
2	x_{21}	x_{22}	$\cdots$	x_{2r}	n_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	
s	x_{s1}	x_{s2}	$\cdots$	x_{sr}	n_s
# utfall	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot r}$	n

Kontingenstabell

Observera att sannolikheterna p_j är ospecificerade. För att testa nollhypotesen använder vi testvariabeln

$$Q_{obs} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - n_i p_j^*)^2}{n_i p_j^*} \quad \text{där} \quad p_j^* = \frac{1}{n} \sum_{i=1}^{s} x_{ij} = \frac{x_{\cdot j}}{n}$$

Notera att p_j^* är den bästa möjliga skattningen av p_j givet att försöksserierna är homogena (dvs om H_0 är sann). Under detta antagande är Q_{obs} approximativt $\chi^2((r-1)(s-1))$-fördelad för stora n.

Ex: För att utvärdera en ny undervisningsmetod delas studenterna på en kurs upp i två grupper. Ena gruppen undervisas med en traditionell undervisningsmetod och den andra gruppen med en ny undervisningsmetod. Resultaten för de båda grupperna redovisas i kontingenstabellen. Vi vill testa nollhypotesen att undervisningsmetoderna är likvärdiga och får testvariabeln

Metod \ Betyg	VG	G	U	Totalt
Trad	22	46	32	100
Ny	32	52	18	102
Totalt	54	98	50	202

$$Q_{obs} = \frac{(22 - 100 \cdot 54/202)^2}{100 \cdot 54/202} + \frac{(46 - 100 \cdot 98/202)^2}{100 \cdot 98/202} + \dots + \frac{(18 - 102 \cdot 50/202)^2}{102 \cdot 50/202} = 5{,}83$$

Vi förkastar H_0 på 5%-nivån om $Q_{obs} > \chi^2_{0,05}((3-1)(2-1)) = \chi^2_{0,05}(2) = 5{,}99$, så enligt testet finns ingen signifikant skillnad mellan metoderna på 5%-nivån. Notera att testet är applicerbart ty $n_i p_j^* \geq 5$ för alla i, j.

Vi kan också använda χ^2-test för att undersöka om två egenskaper A och B är oberoende. Egenskaperna A och B skulle exempelvis kunna vara hårfärg och ögonfärg hos människor. Egenskapen A har utfallen $A_1, \dots, A_r$ och egenskapen B har utfallen $B_1, \dots, B_s$, så totalt finns det rs olika utfall $B_i \cap A_j$ med kända sannolikheter p_{ij}. Marginalsannolikheterna för varje utfall A_j och B_i ges av

$$\mathbb{P}(A_j) = \sum_{i=1}^{s} \mathbb{P}(B_i \cap A_j) = \sum_{i=1}^{s} p_{ij} = p_{\cdot j} \quad \text{respektive} \quad \mathbb{P}(B_i) = \sum_{j=1}^{r} p_{ij} = p_{i \cdot}$$

Vi vill testa nollhypotesen att egenskaperna A och B är oberoende, dvs att $p_{ij} = \mathbb{P}(B_i \cap A_j) = \mathbb{P}(B_i)\mathbb{P}(A_j) = p_{i\cdot} p_{\cdot j}$ för alla i, j. Vi utför n oberoende försök och låter x_{ij} beteckna de absoluta frekvenserna för utfallen $B_i \cap A_j$. Marginalsannolikheterna $p_{\cdot j}$ och $p_{i \cdot}$ för utfallen A_j och B_i kan skattas med $p_{\cdot j}^* = \frac{1}{n} \sum_{i=1}^{s} x_{ij}$ respektive $p_{i \cdot}^* = \frac{1}{n} \sum_{j=1}^{r} x_{ij}$. Vi bildar testvariabeln

$$Q_{obs} = \sum_{i=1}^{s} \sum_{j=1}^{r} \frac{(x_{ij} - n p_{i\cdot}^* p_{\cdot j}^*)^2}{n p_{i\cdot}^* p_{\cdot j}^*},$$

som är approximativt $\chi^2((r-1)(s-1))$-fördelad för stora n om H_0 är sann. Notera att datat i ett oberoendetest kan skrivas i en kontingenstabell på samma sätt som vid homogenitetstest. Skillnaden är att i oberoendetest, så är radmarginalerna slumpmässiga, medan de är givna på förhand i ett homogenitetstest. Räkningarna blir identiska, men slutsatserna olika i de båda testen.