

Föreläsning 14

Thomas Önskog

11/12 2017

χ^2 -test

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

Det första test som vi undersöker är **test av given fördelning**. 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 $A_1, \dots, A_r$ ($\sum_{i=1}^r x_i = n$). Låt $p_1, \dots, p_r$ vara givna sannolikheter som uppfyller $\sum_{i=1}^r p_i = 1$. Vi vill testa nollhypotesen

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

mot mothypotesen

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

Vi testar alltså om utfallen av försöket har den fördelning som ges av $p_1, \dots, p_r$. För att testa H_0 bildar vi testvariabeln

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

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

$$\underline{\text{Förkasta } H_0 \text{ om } Q_{\text{obs}} > \chi_{\alpha}^2(r-1)}.$$

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

Exempel. En sexsidig tärning kastas 96 gånger. Utfallet blir 15 ettor, 7 tvåor, 9 treor, 20 fyror, 26 femmor och 19 sexor. Låt A_j beteckna händelsen att j prickar kommer upp vid ett kast med tärningen. Vi vill prova hypotesen att tärningen är symmetrisk och nollhypotesen blir därmed

$$H_0 : \mathbb{P}(A_j) = p_j = 1/6, \quad \text{för alla } j \in \{1, \dots, r\}.$$

Eftersom $np_j = 96 \cdot 1/6 = 16 \geq 5$, för alla j , så är approximationen som testet bygger på god. Testvariabeln blir

$$Q_{\text{obs}} = \frac{(15 - 16)^2}{16} + \frac{(7 - 16)^2}{16} + \frac{(9 - 16)^2}{16} + \frac{(20 - 16)^2}{16} + \frac{(26 - 16)^2}{16} + \frac{(19 - 16)^2}{16} = 16.0,$$

och antalet frihetsgrader är $r - 1 = 5$. Eftersom $\chi_{0.05}^2(5) = 11.1$, $\chi_{0.01}^2(5) = 15.1$ och $\chi_{0.001}^2(5) = 20.5$, 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 $\theta = (\theta_1, \dots, \theta_s)$, exvis $\text{Po}(\mu)$. Testet som vi nu ska beskrivas kallas för **test av fördelning med skattade parametrar**. Nollhypotesen blir i denna situation

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

För att testa H_0 skattar vi först $\theta_{\text{obs}}^* = \theta^*(x_1, \dots, x_r)$ från stickprovet under antagandet att H_0 är sann. Sedan sätter vi in θ_{obs}^* istället för θ i nollhypotesen och beräknar testvariabeln

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

Om H_0 är sann, så är Q_{obs} ett utfall av en approximativt $\chi^2(r - s - 1)$ -fördelad s.v. 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 $\text{Po}(\mu)$ -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

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

Datat kan sammanfattas i en **kontingenstabell**.

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

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

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

Notera att p_j^* är den bästa möjliga skattningen av p_j givet att försöksserierna är homogena (dvs om H_0 är sann). Om H_0 är sann, så är Q_{obs} ett utfall av en approximativt $\chi^2((r-1)(s-1))$ -fördelad s.v. för stora n .

Exempel. 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. Tentamensresultaten för de båda grupperna redovisas i följande kontingenstabell.

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

Vi vill testa nollhypotesen att undervisningsmetoderna är likvärdiga och får testvariabeln

$$Q_{\text{obs}} = \frac{(22 - 100 \cdot 54/202)^2}{100 \cdot 54/202} + \frac{(46 - 100 \cdot 98/202)^2}{100 \cdot 98/202} + \frac{(32 - 100 \cdot 50/202)^2}{100 \cdot 50/202} + \frac{(32 - 102 \cdot 54/202)^2}{102 \cdot 54/202} + \frac{(52 - 102 \cdot 98/202)^2}{102 \cdot 98/202} + \frac{(18 - 102 \cdot 50/202)^2}{102 \cdot 50/202} = 5.83.$$

Vi förkastar H_0 på 5%-nivån om $Q_{\text{obs}} > \chi_{0.05}^2((3-1)(2-1)) = \chi_{0.05}^2(2) = 5.99$. Enligt testet kan vi inte förkasta nollhypotesen. På 5%-nivån finns därmed ingen signifikant skillnad mellan metoderna. Notera att testet är applicerbart eftersom $n_i p_j^* \geq 100 \cdot 50/202 \approx 24.8 \geq 5$ för alla i, j .

Ett annat typ av χ^2 -test är **oberoendetest**, som används 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 kan indelas i kategorierna $A_1, \dots, A_r$ och egenskapen B i kategorierna $B_1, \dots, B_s$, så totalt finns det rs olika kombinationer av utfall $B_i \cap A_j$ med okända sannolikheter p_{ij} . Marginalsannolikheterna för varje utfall A_j och B_i ges av

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

Vi vill testa nollhypotesen att egenskaperna A och B är oberoende, dvs att

$$H_0 : p_{ij} = \mathbb{P}(B_i \cap A_j) = \mathbb{P}(B_i) \mathbb{P}(A_j) = p_{i.} p_{.j}, \text{ för alla } i, j.$$

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

$$p_{\cdot j}^* = \frac{1}{n} \sum_{i=1}^s x_{ij} \quad \text{respektive} \quad p_{i \cdot}^* = \frac{1}{n} \sum_{j=1}^r x_{ij}.$$

Vi bildar testvariabeln

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

som, om H_0 är sann, är ett utfall av en approximativt $\chi^2((r-1)(s-1))$ -fördelad s.v. för stora n .

Notera att datat i ett oberoendetest kan skrivas i en kontingenstabell på samma sätt som vid homogenitetstest. Skillnaden är att i oberoendetestet, så är antalet observationer på varje rad slumpmässigt, medan det i ett homogenitetstest är givet på förhand. Räkningarna är identiska för de båda testen, men slutsatserna som kan dras från testen skiljer sig åt.