Exakt konfidensintervall för binomialfördelning och Poissonfördelning

Gunnar Englund
Matematisk statistik
KTH

Ht 2001

Man kan göra exakta konfidensintervall för p då vi har en observation x från en $\mathrm{Bin}(n, p)$-fördelning. Innan vi genomför kalkylen kan vi studera ett bekant fall nämligen normalfördelade observationer med känd standardavvikelse σ_0 och se att vi kan betrakta konfidensintervallet som mängden av alla värden för vilka vi inte skulle ha förkastat en lämpligt vald nollhypotes. Detta är på sätt och vis samma sak som vi utnyttjar då vi utför hypotesprövningar med hjälp av konfidensmetoden.

Om vi har n observationer $x_1, x_2, \cdots, x_n$ som är utfall av oberoende $\mathrm{N}(m, \sigma_0)$ där σ_0 är känt får vi det enkelsidiga konfidensintervallet

$$(-\infty, \bar{x} + \lambda_\alpha \frac{\sigma_0}{\sqrt{n}})$$

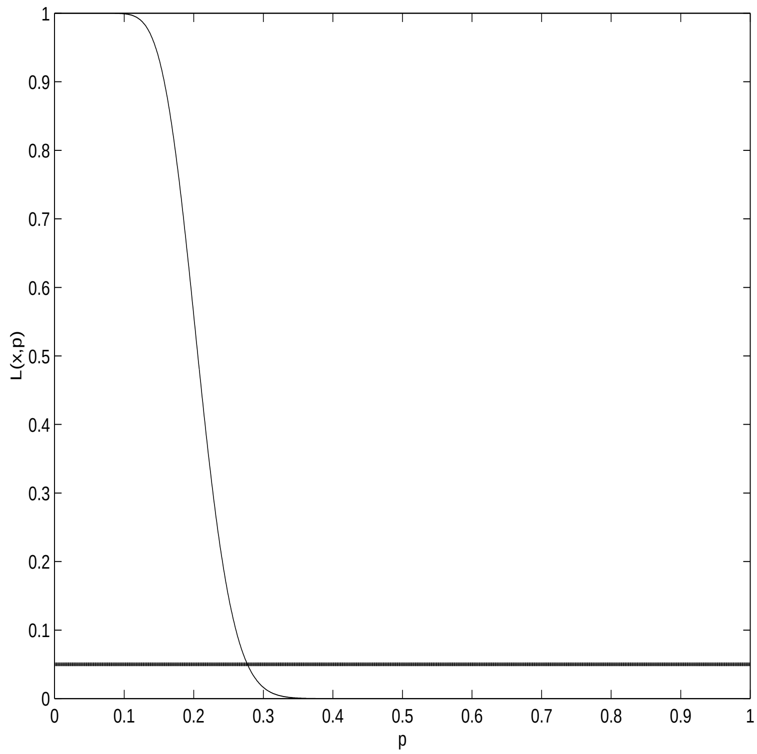
med konfidensgraden $1 - \alpha$.

Vi förkastar $H_0 : m \geq m_0$ till förmån för $H_1 : m < m_0$ på nivån α om m_0 ligger till höger om konfidensintervallet. Detta betyder att vi kan se konfidensintervallet med konfidensgrad $1 - \alpha$ som mängden av de m_0-värden för vilka vi inte skulle ha förkastat H_0 på nivån α. Detta gäller helt generellt och är principen bakom konfidensmetoden.

Vi har nu observerat utfallet x av X som är $\mathrm{Bin}(n, p)$ och där vi vill beräkna ett uppåt begränsat konfidensintervall av typen $(0, \widetilde{p})$ för p med konfidensgrad $1 - \alpha$. Vi bildar nu funktionen

$$L(x, p) = P(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} p^i (1 - p)^{n-i}.$$

För $n = 100$ och $x = 20$ är funktionen uppritad i figuren där vi dessutom ritat in den horisontella linjen $\alpha = 0.05$.
Vi kan alltid lösa ekvationen $L(x, \widetilde{p}) = \alpha$ som i exemplet enligt figuren ger $\widetilde{p} \approx 0.279$. Om $p_0 \geq \widetilde{p}$ så skulle vi förkasta $H_0 : p \geq p_0$ till förmån för $H_1 : p < p_0$ och alltså utgör $(0, \widetilde{p})$ ett enkelsidigt konfidensintervall för p med konfidensgrad



Figur 1: $L(x, p)$ för $\text{Bin}(n, p)$ där $n = 100$ och vi observerat $x = 20$

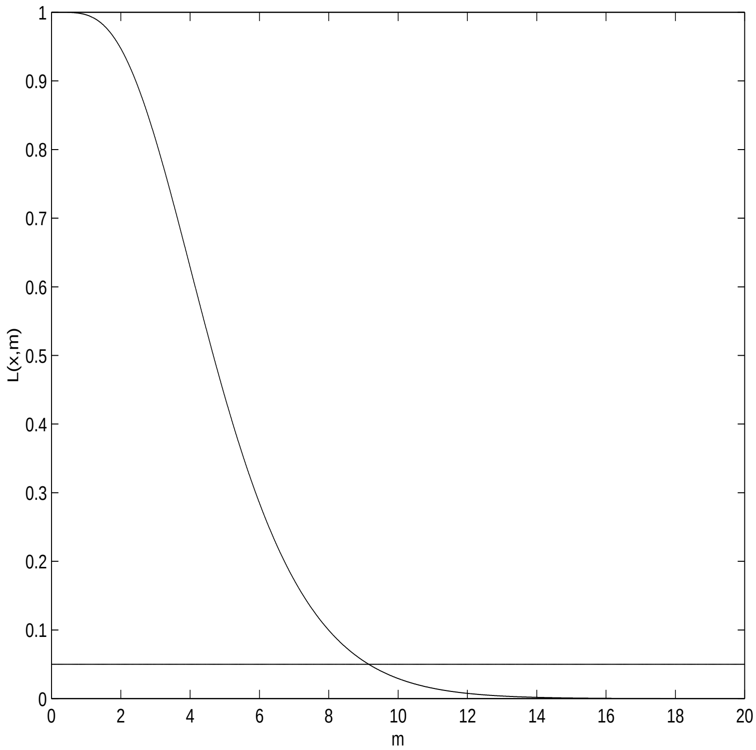
$1-\alpha$. Detta helt i analogi med förfarandet i normalfördelningssituationen som beskrevs ovan.

Vi kan naturligtvis göra nedåt begränsade konfidensintervall på motsvarande sätt. Vi studerar då $L(x, p) = P(X \geq x)$ och löser ekvationen $L(x, \widehat{p}) = \alpha$ och får konfidensintervallet $(\widehat{p}, 1)$ eller om vi gör båda sakerna (med α bytt mot $\alpha/2$) kan vi få dubbelsidiga intervall.

Samma typ av metodik kan användas om vi har en observation x från en $\text{Po}(m)$-fördelad variabel X och studerar

$$L(x, m) = P(X \leq x) = \sum_{i=0}^{x} \frac{m^i}{i!} e^{-m}$$

som också är en funktion som avtar monotont från 1 till 0 då $0 \leq m < \infty$. Se figur.
Vi får med $x = 4$ ett 95%-igt konfidensintervall för m till $(0, 9.17)$.



Figur 2: $L(x, m)$ för Po(m) och vi observerat $x = 4$