Om konfidensintervall för medianer

Gunnar Englund
Matematisk statistik
KTH

Ht 2001

Man kan faktiskt göra ett konfidensintervall för medianen med konfidensgrad minst lika med $1-\alpha$ helt utan några som helst antaganden om den bakomliggande fördelningen (utom möjligen att den har täthet och att medianen är entydigt definierad). Låt $x_1, x_2, \cdots, x_n$ vara utfall av oberoende likafördelade (kontinuerliga) stokastiska variabler $X_1, X_2, \cdots, X_n$ med medianen $\widetilde{m}$, dvs som uppfyller $P(X_i \leq \widetilde{m})$=$P(X_i \geq \widetilde{m})$=1/2. Vi vill konstruera ett numeriskt intervall $(-\infty, a)$ som är ett konfidensintervall med konfidensgrad (åtminstone) $1-\alpha$.

Låt $x_{(1)}, x_{(2)}, \cdots, x_{(n)}$ vara observationerna ordnade i storleksordning och motsvarande stokastiska variabler betecknar vi $X_{(1)}, X_{(2)}, \cdots, X_{(n)}$.

Vi kommer att låta konfidensintervallet vara $(0, x_{(k)})$ där vi väljer k listigt så att konfidensgraden blir åtminstone $1-\alpha$. Notera att $P(X_i \leq \widetilde{m}) = 1/2$. Låt Y =antalet X_i:n $\leq \widetilde{m}$. Vi ser att Y är Bin$(n, 1/2)$ och får

$$P(\widetilde{m} \leq X_{(k)}) = P(Y < k) = P(Y \leq k-1) = \sum_{i=0}^{k-1} \binom{n}{i} (1/2)^n$$

Om vi väljer k så att $P(Y \leq k-1) \geq 1-\alpha$ så betyder det att $(0, x_{(k)})$ är ett uppåt begränsat konfidensintervall för medianen $\widetilde{m}$ med konfidensgrad åtminstone $1-\alpha$.

<u>Exempel:</u> Om vi har de 13 observationerna 1,2,7,11,13,18,20,22,25,35,36,40,47 (sorterade i storleksordning) kan vi skatta medianen med det 7:te i storleksordning, dvs 20. Vi vill nu göra ett enkelsidigt uppåt begränsat konfidensintervall för $\widetilde{m}$ med konfidensgrad åtminstone 95%. Vi kan se att vi får $P(Y \leq 8) = 0.8666$ och $P(Y \leq 9) = 0.9539$ och alltså kan vi ta $k-1 = 9$ eller annorlunda uttryckt $k = 10$, dvs att $(-\infty, x_{(10)}) = (0, 35)$ är ett 95.39%-igt konfidensintervall, dvs intervallet har konfidensgrad åtminstone 95%.

På precis samma sätt kan man göra ett nedåt begränsat konfidensintervall för medianen (man tar $(x_{(n-k+1)}, \infty)$ eller ett dubbelsidigt intervall av typen $(x_{(k)}, x_{(n-k+1)})$. Den observante inser lätt att man med samma typ av metodik kan göra konfidensintervall för godtyckliga percentiler, t ex 10%-punkten i fördelningen.