

Пример. Нека је X посматрано обележје, чија расподела вероватноћа припада фамилији $\mathcal{P} = \{\mathcal{B}(1, p) : p \in (0, 1)\}$ допустивих расподела. На основу простог случајног узорка $(X_1, X_2, \dots, X_n)$ из обележја X , за задати ниво поверења β , одредити егзактан интервал поверења за непознату вероватноћу p .

Решење.

Тачкаста оцена за непознату вероватноћу 'успеха' p је узорачка средина $\bar{X}_n$ [предавања - питање 23], која представља узорачку релативну учестаност успеха. Нека је са S_n означен број успеха у n независних експеримената у Бернулијевој шеми. Тада важи: $S_n = n\bar{X}_n \in \mathcal{B}(n, p)$. Функција расподеле G случајне величине $\bar{X}_n$ дата је са

$$\begin{aligned} G(x, p) &= P\{\bar{X}_n \leq x\} = P\{n\bar{X}_n \leq nx\} = P\{S_n \leq nx\} \\ &= \sum_{j=0}^{[nx]} \binom{n}{j} p^j (1-p)^{n-j} = 1 - \sum_{j=[nx]+1}^n \binom{n}{j} p^j (1-p)^{n-j} \\ &= 1 - \binom{n}{[nx]+1} \cdot ([nx]+1) \int_0^p t^{[nx]} (1-t)^{n-([nx]+1)} dt \end{aligned}$$

где се последња једнакост, у којој се појављује некомплетна бета функција, најлакше може објаснити из везе биномне расподеле и статистика поретка из $\mathcal{U}[0, 1]$ расподеле (видети: ***).

Применом Лајбницевог правила за диференцирање под знаком интеграла добија се

$$\frac{\partial}{\partial p} G(x, p) = - \binom{n}{[nx]+1} \cdot ([nx]+1) p^{[nx]} (1-p)^{n-([nx]+1)} < 0,$$

одакле се закључује да је $G(x, p)$ строго опадајућа функција од p , за свако x .

Нека су α_1 и α_2 задате позитивне константе такве да важи $\alpha_1 + \alpha_2 < 0.5$ и нека су Y_1 и Y_2 решења једначина

$$\begin{aligned} G(\bar{X}_n, Y_2) &= \sum_{j=0}^{S_n} \binom{n}{j} Y_2^j (1-Y_2)^{n-j} = \alpha_1 \\ G(\bar{X}_{n-}, Y_1) &= \sum_{j=0}^{S_n-1} \binom{n}{j} Y_1^j (1-Y_1)^{n-j} = 1 - \sum_{j=S_n}^n \binom{n}{j} Y_1^j (1-Y_1)^{n-j} = 1 - \sum_{j=0}^{n-S_n} \binom{n}{j} (1-Y_1)^j Y_1^{n-j} \\ &= 1 - \alpha_2, \end{aligned}$$

где је $\bar{X}_{n-}$ статистика таква да је $\bar{X}_n = \frac{k}{n}$ ако и само ако је $\bar{X}_{n-} = \frac{k-1}{n}$. Тада је $[Y_1, Y_2]$ двострани интервал поверења за p са нивоом поверења бар β , где је $\beta = 1 - (\alpha_1 + \alpha_2)$. Приметити да се вредношћу α_1 контролише вероватноћа у вези са бројем успеха, а вредношћу α_2 вероватноћа у вези са бројем неуспеха.

Једначине којима се дефинишу Y_1 и Y_2 решавају се нумеричким методама, односно применом итеративних алгоритама.

Нумеричка илустрација.

Нека је $n = 30$ и $\bar{x}_{30} = 0.6$ (тј. $s_n = 18$), при чему се задаје $\alpha_1 = \alpha_2 = 0.05$. Задатак се своди на то да се одреде бројеви y_1 и y_2 тако да важи

$$\sum_{j=0}^{17} \binom{30}{j} y_1^j (1-y_1)^{30-j} = 0.95 \quad \text{и} \quad \sum_{j=0}^{18} \binom{30}{j} y_2^j (1-y_2)^{30-j} = 0.05.$$

У наставку је дата функција у статистичком софтверу R, која одређује приближно решење, са задатом тачношћу, једначина горњег облика за унапред одређене (грубо одабране) горњу p_1 и доњу p_2 границу решења, при чему је `value` - вредност са десне стране једначине.

```
binomintpov = function(sn,n,p1,p2,value,maxstp=100,eps=.00001){
  y1 = pbinom(sn,n,p1) #racuna vrednost f-je raspodele B(n,p1) raspodele u tacki sn
  y2 = pbinom(sn,n,p2)
  ic1 = 0
  ic2 = 0
  if(y1 >= value){ic1=1}
  if(y2 <= value){ic2=1}
  if((ic1*ic2) > 0){
    istep = 0
    while(istep < maxstp){
      istep = istep + 1
      p3 = (p1 + p2)/2
      y3 = pbinom(sn,n,p3)
      if(y3 > value){
        p1 = p3
        y1 = y3
      } else {
        p2 = p3
        y2 = y3
      }
      if(abs(p1-p2) < eps){istep = maxstp}
    }
    list(solution=p3,valatsol = y3)
  } else {
    list(error="Bad_start!")
  }
}
```

У нашем примеру:

за решење прве једначине погодно је одабрати $p_1 = 0.4$ јер је $P_{p_1}\{S_{30} \leq 17\} = 0.9787601$ и $p_2 = 0.45$ јер је $P_{p_2}\{S_{30} \leq 17\} = 0.9286106$, `value` = 0.95;

за решење прве једначине погодно је одабрати $p_1 = 0.7$ јер је $P_{p_1}\{S_{30} \leq 18\} = 0.1593218$ и $p_2 = 0.8$ јер је $P_{p_2}\{S_{30} \leq 18\} = 0.009493122$, `value` = 0.05.

Применом функције добија се интервал поверења за p (0.43394,0.75047) са нивоом поверења бар 90%. Апроксимативни 90%-интервал поверења [предавања - питање 23] је (0.45101,0.73254), односно (0.45331,0.74669). Као што се да приметити разлике између ових интервала поверења су занемарљиве у овом случају. Оне долазе до изражаја само када се процењује **мала вероватноћа на малом обиму узорка** и тада се може користити **само** описани **егзактан метод**. △