Zeinuen proba

Josemari Sarasola

Gizapedia

Zeinuen proba

Definizioa

Zeinuen proba proba ez-parametrikoa* da, proba binomial* baten kasu berezia, non $p = 1/2$, aplikazio hauek dituena:

- lagin bakar bateko datuetarako, populazioak mediana zehatza duen probatzeko; adibidez, $3.1, 2.4, 5.6, 6.7, 4.2, 3.8$ lagina $Me = 3.5$ duen populazio batetik hartua den erabakitzeko.
- lagin binakatuetan, medianen diferentzia 0 probatzeko, eta hedaduraz lagin batek bestea baino balio handiagoak ematen dituen erabakitzeko; adibidez, ongarri berezia eman ondoren, landare baten errendimendua hobea den erabakitzeko;
- medianari buruzko konfiantza-tarteak ezartzeko.

- Proba ez-parametrikoak populazioari buruzko banaketa jakinik ezartzen ez duten probak dira, eta beraz malguagoak dira aplikatzean.
- Proba binomialak banaketa binomial batean p probabilitatea balio jakina hartzen duen erabakitzeko probak dira.

Medianari buruzko proba (I)

- Mediana azpitik datuen %50ak, probabilitatearen %50a, uzten duen balioa da.
- Beraz, populazio batean datu bat medianaren azpitik izateko probabilitatea 0.5 da.
- Proba garatzeko aldagai estatistikoa jarraitua izan behar da, hau da, balio ezberdin asko hartu behar ditu.
- Hipotesi nuluan, medianaren balio jakin bat zehazten da: $H_0 : Me = m$. Aukeran, hori baztertuz, $H_a : Me \neq m$.

Medianari buruzko proba (II)

- $Me = m$ azpitik dagoen datuari, $-$ zeinua esleitzen zaio; $Me = m$ gainetik dagoenari, $+$ zeinua.
- $-$ eta $+$ zeinuen kopuruak zenbatu: r^- eta r^+.
- Proba alde bikoa da: $H_0 : Me = m$ hipotesi nulua, r^- eta r^+ zeinuen kopuruak oso handiak zein oso txikiak direnean baztertzen da. Beraz, alde biko proba da.
- Baina, modu finko batean egiteagatik r^- eta r^+ kopuruetatik txikiena hartzen da: r deituko diogu kopuru honi.
- Txikiena denez, hipotesi nulua soilik r oso txikia denean baztertuko da.

Medianari buruzko proba (III)

- Lagin tamaina txikietarako r^* balio kritikoa bilatzen da tauletan, α adierazgarritasun maila ezberdinetarako.
- Hipotesi nulua, medianaren balioa alegia, $r \leq r^*$ betetzen denean baztertzen da.
- Aldagai jarraitua denez, hipotesi nuluko medianarekin balioarekin bat datorren daturik ezin da izan, baina egongo balitz datu hori ezabatuko genuke.

Medianari buruzko proba (IV): p-balioaren kalkulua

- Datu bakoitza medianaz behera zein gora izateko probabilitatea 0.5 da, medianaren definiziotik.
- n datuetan, medianaz gorako eta beherako datu kopurua $B(n, 0.5)$ banatzen da.
- n datuetan, zeinu kopurua minimoa x izango da, medianatik gorako datuak x direnean edo medianatik beherako datuak x direnean.
- Beraz, horren probabilitatea hau izango da:

$$
\begin{aligned}
P[min = x] &= P[Me - tik\ gora = x] + P[Me - tik\ behera = x] \\
&= 0.5^x 0.5^{n-x} \frac{n!}{x!(n-x)!} + 0.5^x 0.5^{n-x} \frac{n!}{x!(n-x)!} \\
&= 2 \times 0.5^n \frac{n!}{x!(n-x)!}
\end{aligned}
$$

- p-balioak ebidentziaren edo zerbait arraroagoren probabilitatea ematen du. Minimoa hartzen dugunez, arraroa behetik dago. Beraz:

$$p = P[X \leq x] = 2 \sum_{i=1}^{x} 0.5^n \frac{n!}{i!(n-i)!} = 2 \times 0.5^n \sum_{i=1}^{x} \frac{n!}{i!(n-i)!}$$

- p-balioa α-rekin alderatzen da, horren kalkuluan zeinu kopuru minimoa medianatik gorakoena zein beherakoena izan daitekeela kontuan hartu dugulako.

Medianari buruzko proba (V): adibidea

- Datuak: 6.14-6.45-8.33-11.05-5.67-5.22-7.52-10.08-12.34
- Mediana populazioan 7 den probatu nahi da: $H_0 : Me = 7$. $\alpha = \%10$.
- Datuei zeinuak esleitu: $-,-,+,+,-,-,+,+,+$.
- $r^- = 4; r^+ = 5$. Txikiena hartzen da: $r = 4$.
- Tauletan balio kritikoa bilatzen da: $r^* = 1$.
- r balioa gainetik dagoenez, hipotesi nulua onartu, eta beraz datuen populaziorako mediana 7 dela baiezta daiteke.
- p-balioa *medianatik behera* (eta ez gora) 4 datu edo gutxiago izateko probabilitatea da:

$$p = P[X \leq 4] = P[X=0] + P[X=1] + P[X=2] + P[X=3] + P[X=4]$$
$$= 0.5^0 \cdot 0.5^9 \cdot \frac{9!}{0!9!} + 0.5^1 \cdot 0.5^8 \cdot \frac{9!}{1!8!} + 0.5^2 \cdot 0.5^7 \cdot \frac{9!}{2!7!}$$
$$+ 0.5^3 \cdot 0.5^6 \cdot \frac{9!}{3!6!} + 0.5^4 \cdot 0.5^5 \cdot \frac{9!}{4!3!} = \sum_{i=1}^{4} 0.5^9 \frac{9!}{i!(9-i)!} = 0.5$$

- p-balioa hemen $\alpha/2$ balioarekin alderatzen dugu, minimoa medianatik gora ere gerta zitekeela ez dugulako kontuan hartu. $p > \alpha/2$ dugunez, $H_0 : Me = 7$ onartu egiten da.

Medianari buruzko proba (VI): alde bakarreko proba

- Datuak: 1.34-2.45-0.33-4.73-0.68-1.02-3.13-3.45
- Mediana 0.5 baino handiagoa den probatu nahi da: $H_0 : Me \geq 0.5$. $\alpha = \%10$
- Medianaren 0.5 balioa harturik, datuei zeinuak esleitu:
  $+,+,-,+,+,+,+,+ \rightarrow r^- = 1; r^+ = 7$
- $H_0 : Me > 0.5$ baztertuko da r^- aski handia denean (r^+ txikia denean).
- Txikitik begiratzen dugu, taulak erabili ahal izateko. Beraz, r^+ kopuruari erreparatu beharko zaio.
- Taulatik balio kritikoa hartzen da (α bider 2 egin behar da)*:
  $n = 8, \alpha = 0.20 \rightarrow r^* = 1$
- $r^+ = 7$ estatistikoa $r^* = 1$ balio kritikoa baino handiagoa denez, hipotesi nulua onartu eta mediana 0.5 baino handiagoa dela adieraz daiteke.

* α adierazgarritasun-mailako alde bakarreko probak eta 2α adierazgarritasun-mailako alde biko probak maila kritiko berdina dute. Taula alde biko probarako egina dagoenez, α bider 2 egin behar da. Ikus, hobeto ulertzeko, eranskina.

Medianari buruzko proba (VII): alde bakarreko proba

Laburpen moduan,

- $H_0 : Me < m$ hipotesian, r^- balioari erreparatu.
- $H_0 : Me > m$ hipotesian, r^+ balioari erreparatu.
- Ondoren, tauletan r^* balio kritikoa bilatu, ezarritako α bider 2 eginez, eta erabakia hartu.

Zeinuen proba

Lagin binakatuak (I)

- Bi lagin binakatuak (ingelesez, *paired*; gaztelaniaz, *apareadas*) direla esaten da, elementu berdinen gainean jaso direnean, gehienetan tratamendu baten aurretik eta ondoren.
- Adibidez, lagin binakatuak talde bateko gaixoek odolean duten substantzia-kopurua tratamendu baten aurretik eta ondoren; eta gela bateko ikasleen matematika-trebetasuna ikasturtearen hasieran eta bukaeran.
- Hipotesi nuluak izan daitezke:
  $H_0 : Me(bukaera) - M(hasiera) = 0$
  (mediana biak berdinak dira, beraz hori baztertzen badugu, adieraziko dugu biak desberdinak direla, norabide jakinik eman gabe);
  $H_0 : Me(bukaera) - M(hasiera) \geq 0$ (mediana bukaeran handiagoa da)
  edo
  $H_0 : Me(bukaera) - M(hasiera) \leq 0$ (mediana bukaeran txikiagoa da).
- Jasotako aldagaia jarraitua izan behar da, berdinketak saiheste aldera.

Lagin binakatuak (II): prozedura

- Bi lagin binakatuetan diferentziak kalkulatu.
- Diferentzia positiboa bada, $+$ zeinuaz adierazi; negatiboa bada, $-$ zeinuaz.
- Garatu alde bakarreko probetan bezala.

Lagin binakatuak (III): adibidea

- Gela batean ikasleen trebetasun matematikoa jaso da ikasturtearen hasieran nahiz bukaeran.
- Hasieran: 4.2-6.5-7.2-5.4-8.4-5.6-7.4-7.2
- Bukaeran: 5.4-6.6-7.4-7.0-8.2-6.8-7.3-7.8
- Galdera: trebeagoak al dira ikasleak bukaeran? $\alpha = 0.10$
- Diferentziak kalkulaturik zeinuak eman:++++-+-+
- 6 zeinu positibo eta 2 negatibo dira. Beraz, bada aukera baieztatzeko bukaeran hobe izan zela. Alderantziz (6-,2+) izan balitz ordea, ez dago horretarako aukerarik, eta problema bukatuko genuke, berdintasuna dagoela adieraziz.
- Txikiena hartzen dugu: $r = 2$.
- Balio kritikoa bilatzen dugu taulan, alde bakarreko proba dela jakinik:

$$n = 8, \alpha = 0.20(2 - sided) \rightarrow r^* = 1$$

- $r > r^*$ dugunez, ezin da baieztatu bukaeran trebeagoak direnik (horretarako 1 negatibo behar genuke gehienez).

Zeinuen probaren aplikazioa: medianari buruzko konfiantza-tartea

Pausoak adibide bati jarraiki

- Honako datu hauetarako %95eko konfiantza tartea eratu medianari buruz.
- Datu ordenatuak: 6.14-6.45-8.33-11.05-5.67-5.22-7.52-10.08-12.34
- Lagin tamaina: 9.
- $B(n = 9, p = 0.5)$ banaketan probabilitate metatuen taula eratu.
- $(x = 0, p = 0.0009)$;$(x = 1, p = 0.01)$;$(x = 2, p = 0.054)$;...
- Azpitik *gehienez* %2.5eko probabilitatea (konfiantza eza, %5a alegia, bi muturretan banatzen delako) uzten duen balioa bilatu. Gehienez, konfiantza maila gutxienez ulertzen delako.
- Balio hori $x = 1$ da. $x + 1 = 2$ hartzen dugu.
- $x + 1 = 2$gn datua hartzen da hasieratik eta bukaeratik.
- Beraz, %98ko konfiantzaz $(1 - 0.01 \times 2)$ mediana 6.45 eta 10.08 balioen artean dagoela esan daiteke.

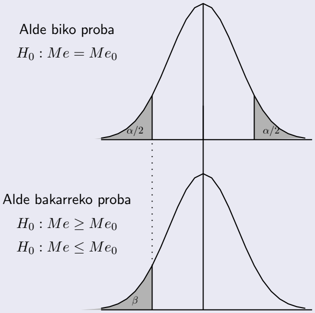
Zeinuen proba: lagin tamaina handiak

Lagin tamaina handiak

Lagin tamaina handia denean, n balioak tauletatik at daudenean zehazkiago, p-balioak eta balio kritikoak kalkulatzeko, De Moivre-Laplace teorema (banaketa normalaren eta konbergentzia estokastikoaren gaian ikasiko da) erabiltzen da, banaketa binomialaren ordez, banaketa normala baliatuz hurbilketa gisa.

Oharra: Gai hori ikasten denean ebatziko da lagin tamaina handietako problemaren bat.

Eranskina: Zergatik hartzen taulan 2α alde bakarreko probetan



$$\beta = \alpha/2 \rightarrow \alpha = 2\beta$$