Lastnosti testov

Nataša Kejžar

Povzetek

Funkcija slučajne spremenljivke (Rice 2.3)

Če lahko zvezno spremenljivko Y izrazimo s spremenljivko X, za katero poznamo porazdelitev (npr. gostoto) kot funkcijo $g(X) = Y$, kjer je g odvedljiva, strogo monotono naraščajoča funkcija na nekem intervalu, potem lahko zapišemo

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

To velja, saj lahko, ko računamo s kumulativno porazdelitveno funkcijo, napišemo

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) \\ &= F_X(g^{-1}(y)). \end{aligned}$$

Ko kumulativno porazdelitveno funkcijo odvajamo, dobimo gostoto.

Lastnosti statističnih testov

Velikost testa

- Pri stopnji značilnosti α je zaželjeno, da je velikost testa α; tega ne dosežemo vedno (npr. diskretne spremenljivke)
- Izračun (za test, ki zavrača za majhne vrednosti testne statistike T):
$$P(T \leq t | \mathbf{H_0}) = \alpha.$$
t označuje mejo zavrnitve oz. kritično vrednost testne statistike.
- Če je velikost testa večja od α, je test liberalen, zavrača prevečkrat, kar je slabo (imamo večjo napako I. vrste kot mislimo). Sprejemljivo je, da je velikost testa manjša od α; test je v tem primeru konservativen.

Moč testa

- Računa se za točno določeno alternativno domnevo.
- Izračun (za test, ki zavrača za majhne vrednosti testne statistike T):
$$P(T \leq t | \mathbf{H_A}) = 1 - \beta.$$
t označuje mejo zavrnitve oz. kritično vrednost testne statistike.
- Ko primerjamo dva statistična testa pri enaki H_0, nas zanima, kateri ima večjo moč. Vsak od testov ima lahko za različno alternativno domnevo različno moč.

Naloge

1. Pokažite, kako (po kakšni porazdelitvi) je porazdeljena spremenljivka Z, če velja $F_X(X) = Z$ in je F_X^{-1} definiran za vse vrednosti X.

   a. To preverite grafično s simulacijami za različne porazdelitve.
   b. Bi lahko porazdelitev vrednosti p pod H_0 tudi pojasnili na ta način? Kako?

2. Gumbelova porazdelitev ekstremnih vrednosti ima kumulativno porazdelitveno funkcijo

$$F(x) = \exp\left[-\exp\left(-\frac{x-a}{b}\right)\right] \qquad ; b > 0.$$

   a. Izpeljite inverzno kumulativno porazdelitveno funkcijo $F^{-1}(x)$.
   b. Pokažite, zakaj se inverzna kumulativna porazdelitvena funkcija (z enakomerno porazdeljeno spremenljivko) lahko uporabi kot generator naključnih vrednosti spremenljivke, ki je porazdeljena po $f(x)$ (v našem primeru po Gumbelovi porazdelitvi ekstremnih vrednosti). *Namig: Pokažite, da lahko uporabite rezultat naloge 1.*
   c. S pomočjo inverzne kumulativne porazdelitvene funkcije generirajte vzorec 100 enot iz Gumbelove porazdelitve za primerna parametra a in b. (Generirati morate vzorec iz enakomerne porazdelitve in na njem uporabiti inverzno porazdelitveno funkcijo.)
   d. Narišite histogram tako dobljene porazdelitvene funkcije za vrednosti $a = 2$ in $b = 0.5$.

3. Gostota eksponentno porazdeljene spremenljivke X ima obliko

$$f(x) = \lambda e^{-\lambda x} \qquad ; x \geq 0, \quad \lambda > 0.$$

   a. Izpeljite inverzno porazdelitveno funkcijo, torej:

$$F_X^{-1}(U) = X,$$

   kjer je $U \sim Unif(0,1)$.
   b. S pomočjo inverzne porazdelitvene funkcije generirajte vzorec $N = 100$ iz porazdelitve $Exp(4)$. Narišite vzorec in mu dodajte teoretično krivuljo z **dexp**.

4. Psihologi preučujejo vedenje kadilcev, ki se odločijo, da bodo prenehali kaditi. Ve se, da po enem letu kadi še približno polovica teh ljudi. V svoji novi raziskavi bi radi ugotovili, ali ima na dejansko prenehanje kajenja po enem letu kakšen vpliv vrsta psihoterapije, ki se uporablja pri zdravljenju hude depresije. Kako bo delovala ta terapija na kadilce, je povsem neznano. Lahko bi terapija pomagala, čeprav je možno tudi, da bi celo povečala hrepenenje po nikotinu. Zato menijo, da bi se moralo za pomemben učinek vsaj 15% ljudi odločiti drugače, kot se brez terapije. Ker vedo, da je ljudi za tako raziskavo zelo težko dobiti, postavijo zgornjo mejo velikosti vzorca na 50 oseb.

   a. Zapišite ničelno in alternativno domnevo za vaš primer. Komentirajte.
   b. Povejte, kaj je vaša testna statistika in kako je porazdeljena.
   c. Izračunajte in prikažite na grafu moč testa (za strokovno pomembno razliko) za velikosti vzorca od 10 do 50 (povečujte velikost vzorca za 1).
   d. Komentirajte graf in **razložite**, zakaj dobite tako nenavadno obliko krivulje. Kakšno velikost vzorca bi morali raziskovalci vzeti, če bi želeli imeti moč testa vsaj 30%?
   e. Kaj bi svetovali raziskovalcem, ki bi k vam prišli glede izračuna velikosti vzorca?

5. Raziskovalci iz prejšnje naloge vidijo, da ima terapija vpliv samo na ženske in ker imajo v vzorcu 40 oseb samo 18 žensk, ne vidijo tako velikega učinka, kot so ga napovedali. Vseeno jih zanima, ali obstaja med moškimi in ženskami razlika v deležu nekadilcev (po 1 letu terapije). To želijo preveriti s statističnim testom za primerjavo dveh deležev (gl. **http://www.dummies.com/how-to/content/how-to-compare-two-population-proportions.html**).

   a. Zapišite ničelno in alternativno domnevo za primer. Komentirajte.
   b. Povejte, kaj je vaša testna statistika in kako je porazdeljena. Razložite, zakaj.

c. Komentirajte izbiro testa.
d. Zakaj se za izračun SE ne uporabi p_1 in p_2?
e. Recimo, da je na podvzorcu moških delež nekadilcev točno 0,5, pri ženskah pa je ravno toliko več nekadilk, da bi bil učinek terapije na vzorcu nad 15%. Izračunajte vrednost testne statistike in statistično in vsebinsko interpretirajte rezultat.
f. Kako veliko moč testa imamo, če bi želeli dokazati razliko za vsaj 5%? Moč testa poskušajte izračunati teoretično in jo tudi **simulirajte**. Komentirajte.
g. Kako velik vzorec bi morali imeti, če bi želeli dokazati razliko za vsaj 5%? Spremenite simulacije iz prejšnje točke in komentirajte rezultat.
h. Kako veliko razliko bi lahko dokazali (moc testa naj bo 80%) z velikostjo vzorca, ki jo imamo?

6. Raziskava javnega mnenja v ZDA je pokazala, da 47% ljudi podpira Busha, 44% ljudi pa Al Gore-a. V vzorec so vzeli 2207 ljudi. Testirajte pri $\alpha = 0.05$, ali več Američanov podpira Busha.

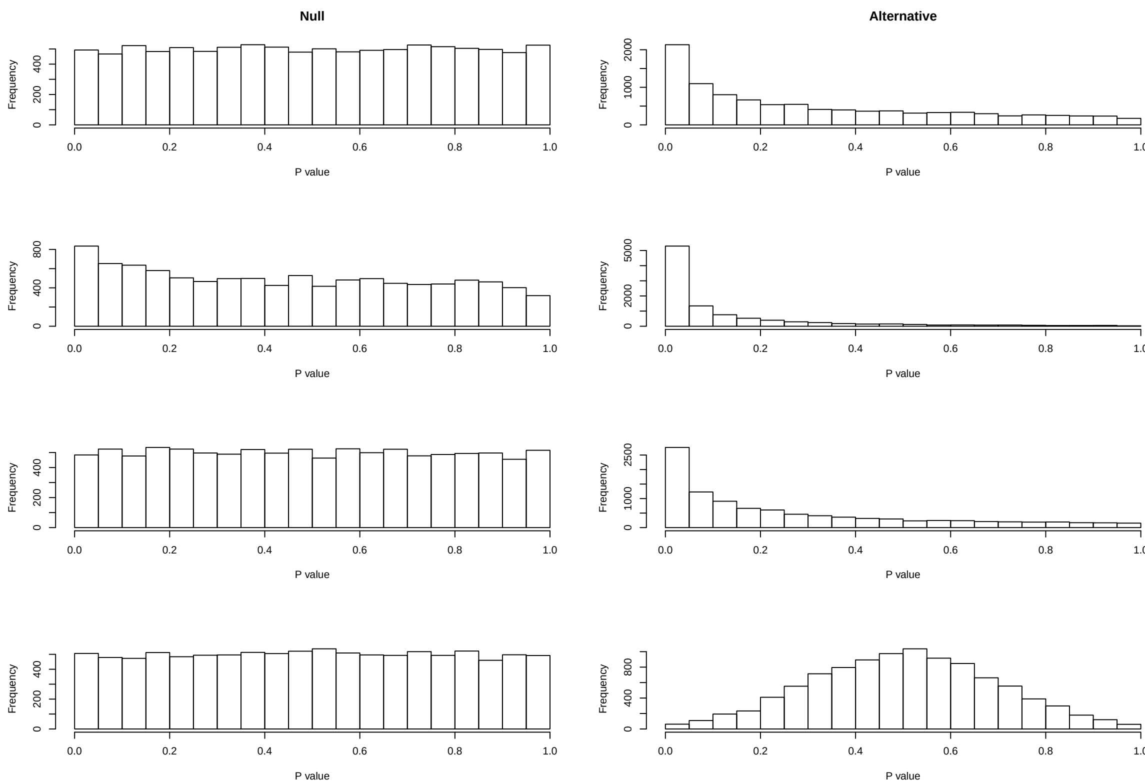
    a. Zapišite ničelno hipotezo. Ali je enostranska ali dvostranska?
    b. Uvedite novo spremenljivko *razlika deležev*, ki bo vsebovala informacijo o raziskavi javnega mnenja. Kako se ta spremenljivka izraža s spremenljivkama X_A (*delež podpore* Al Gore-u) in X_B (*delež podpore* Bushu)?
    c. Izračunajte pričakovano vrednost nove spremenljivke. Kaj pa varianca?
    d. Izračunajte vzorčno varianco nove spremenljivke in ob pomoči CLI zapišite testno statistiko.
    e. Vsebinsko interpretirajte dobljeni rezultat.

7. Klinična dietetičarka želi primerjati dve različni dieti za diabetike. Predpostavlja, da je dieta A uspešnejša kot dieta B (v smislu zniževanja koncentracije krvne glukoze). Načrtuje naključni vzorec diabetičnih bolnikov, ki jih bo naključno razporedila na eno od dveh diet. Na koncu raziskave, ki bo trajala 6 tednov, bo izmerila krvni sladkor (vzet na tešče) vsakemu pacientu. Pričakuje, da bo povprečna razlika koncentracije glukoze 10 mg/dl. Pričakuje, da bosta standardna odklona krvnega sladkorja za obe dieti enaki (15 mg/dl). Dietetičarka bi rada izvedela, kakšno število pacientov naj vzame za posamezno dieto, da bo imela moč testa najmanj 0,8 ($\alpha = 0,05$). Privošči si lahko največ 100 pacientov.
   Zanima nas torej moč testa za različni velikosti skupin. Privzemimo najprej, da imamo natanko 100 pacientov. Preverimo, kaj se dogaja z močjo in velikostjo testa za različni velikosti skupin z dieto A in dieto B.

    a. Kateri statistični test je primeren za raziskovalkin problem?
    b. S simulacijami preverite, da test za različne velikosti skupin (n_1 in $n_2 = 100 - n_1$) vedno zavrača s stopnjo značilnosti α. Velikost n_1 povečujte za 10.
    c. S simulacijami preverite še moč testa za različne velikosti skupin (glej prejšnjo točko). Kaj ugotovite?
    d. Kaj boste svetovali dietetičarki? Kakšno velikost vzorca naj izbere in kako naj vzorec razdeli na skupini?
    e. Pokažite teoretično, da ugotovitve iz vaših simulacij (glede razdelitve vzorca v 2 skupini) veljajo.

```r
# funkcija za izracun p-vrednosti
pttest <- function(n1,n,muA,mu0=0,sd0=15){
  x1 = rnorm(n1,mu0,sd0)
  x2 = rnorm(n-n1,mu0+muA,sd0)
  x = c(x1,x2)
  skupina = as.factor(c(rep(1,n1),rep(2,n-n1)))
  return(t.test(x~skupina,var.equal=TRUE)$p.value)
}
```

8. S simulacijami smo primerjali štiri alternativne testne statistike za preverjanje neke ničelne domneve. Vsako izmed njih smo preverili pod ničelno domnevo in neko dano alternativno domnevo - v vsakem primeru smo generirali 10000 ponovitev in shranili vrednosti p. Spodaj so prikazani dobljeni grafi vrednosti p pod ničelno (levo) in pod alternativno domnevo (desno), vsaka vrstica predstavlja svoj test. Za uporabo katerega testa bi se odločili na podlagi teh grafov? **Utemeljite.**



9. Primerjati želimo povprečji vzorcev enake velikosti n iz neodvisnih normalnih porazdelitev, predpostavimo, da imata enako varianco, ničelna domneva je $H_0 : \mu_1 = \mu_2$. Zanima nas dvostranska alternativna domneva. Naj $\bar{X}_1$ in $\bar{X}_2$ označujeta povprečji vzorcev, σ_p pa koren skupne variance. Mejo zavrnitve pri stopnji značilnosti α zapišemo kot $z_{1-\alpha/2}$, velja torej

$$
\begin{aligned}
\alpha &= P_{\mu_1=\mu_2}\left(\left|\frac{\bar{X}_1-\bar{X}_2}{\sigma_p\sqrt{2/n}}\right| > z_{1-\alpha/2}\right) \\
&= P_{\mu_1=\mu_2}\left(\frac{\bar{X}_1-\bar{X}_2}{\sigma_p\sqrt{2/n}} > z_{1-\alpha/2}\right) + P_{\mu_1=\mu_2}\left(\frac{\bar{X}_1-\bar{X}_2}{\sigma_p\sqrt{2/n}} < z_{\alpha/2}\right)
\end{aligned}
$$

   a. Razložite, kako bi moč testa v nekem konkretnem primeru preverili s simulacijami (napišite približno kodo v R).
   b. Teoretično **izpeljite moč testa**, če velja $\mu_1 - \mu_2 = \Delta$, vse ostale predpostavke pa ostajajo enake. **Narišite skico**. (Za lažje računanje lahko predpostavite, da je vrednost Δ pozitivna in dovolj velika, da bo prispevek k moči na negativni strani zanemarljiv.)
   c. Razložite dele izraza, ki ga dobite, in opišite, kaj vpliva na moč ter kako.

10. Raziskovalec želi vpeljati nov diagnostični test za odkrivanje mutacije nekega gena pri bolnicah z diagnozo raka dojke. Obstoječi test je sicer 100% zanesljiv za odkrivanje mutacije, vendar ne zna ločiti med posameznimi vrstami te mutacije, kar pa je za zdravnike ravno tako pomembno.
    Novi test lahko razločuje med različnimi vrstami mutacij, seveda pa bo primeren le, če je dovolj zanesljiv pri odkrivanju mutacije.
    Raziskovalec bi torej želel zbrati vzorec bolnic, na katerih bi naredil oba testa za mutacijo gena, rad bi dokazal, da je delež pravilno diagnosticiranih z novim testom nad 0.9. Pri tem med deležema lažno pozitivnih in lažno negativnih diagnoz ne želi razlikovati.

    a. Kaj je ničelna domneva?

b. Zanima ga, kako velik vzorec potrebuje, da bi imel 80% moč ($\alpha = 0.05$). Komentirajte podatke. Kako bi se lotili izračuna?
c. Zanima ga, kako velik vzorec potrebuje, da bi imel 80% moč ($\alpha = 0.05$), ob predpostavki, da je pravilno diagnosticiranih v resnici 96% bolnic, ki imajo eno vrsto mutacije in 93% bolnic z drugo vrsto mutacije. Delež bolnic s prvo mutacijo je znan in je enak 22%. Zapišite rezultat, ki vas zanima, s pomočjo formule za popolno verjetnost.
d. Kako bi generirali vzorec iz predpostavke v prejšnji alineji?

11. Iz teoretičnih izpeljav za nek statistični test vemo, da pod H_0 zavrača točno z velikostjo $\alpha = 0.05$. Test smo sprogramirali, sedaj pa bi se radi s simulacijami prepričali, da nismo naredili kakšne napake. Zanima nas, ali sprogramirani test pod H_0 zares zavrača točno v 5%. Ker pri simulacijah nikoli ne moremo zagotovo nekaj trditi, bi radi pognali tolikšno število simulacij, da bomo lahko velikost tega testa ugotovili na 0.5 odstotnih točk natančno. Zanima nas torej, najmanj koliko simulacij moramo pognati za tako natančnost.

    a. Zapišite, katera spremenljivka nas zanima in kako je porazdeljena.
    b. Za oceno števila simulacij uporabite teoretično porazdelitev (ne asimptotske!).
        i. Narišite graf, ki prikazuje širino intervala, kjer se nahaja srednjih 95% simuliranih vrednosti, v odvisnosti od števila simulacij. Komentirajte ugotovitve.
        ii. Kaj se zgodi v primeru, da test pod H_0 ne zavrača z velikostjo točno α, ampak z večjo (npr. 0.07)? Imamo v tem primeru še vedno dovolj natančne simulacije? Pokažite to (npr. s simulacijami).
        iii. Podajte teoretično razlago za prejšnjo točko.
    c. Preverite, do kakšne mere bi bili rezultati zanesljivi, če bi računanje poenostavili z asimptotsko porazdelitvijo.