Curs 12

Cristian Niculescu

1 Comparaţie între deducţiile frecvenţionistă şi Bayesiană

1.1 Scopul învăţării

Să poată să explice diferenţa dintre valoare p şi probabilitate a posteriori unui doctor.

1.2 Introducere

Am aflat despre cele 2 şcoli de deducţie statistică: Bayesiană şi frecvenţionistă. Ambele abordări permit evaluarea dovezilor despre ipoteze concurente.

1.3 Formula lui Bayes ca piatră de încercare

Formula lui Bayes este o propoziţie abstractă despre probabilităţi condiţionate ale evenimentelor:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Am început deducţia Bayesiană reinterpretând evenimentele din formula lui Bayes:

$$P(\mathcal{H}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{H})P(\mathcal{H})}{P(\mathcal{D})}.$$

Acum $\mathcal{H}$ este o ipoteză şi $\mathcal{D}$ sunt datele care pot da dovezi pentru sau împotriva lui $\mathcal{H}$. Fiecare termen din formula lui Bayes are un nume şi un rol.
A priori $P(\mathcal{H})$ este probabilitatea ca $\mathcal{H}$ să fie adevărată înainte de considerarea datelor.
A posteriori $P(\mathcal{H}|\mathcal{D})$ este probabilitatea ca $\mathcal{H}$ să fie adevărată după considerarea datelor.
Verosimilitatea $P(\mathcal{D}|\mathcal{H})$ este dovada despre $\mathcal{H}$ furnizată de datele $\mathcal{D}$.
$P(\mathcal{D})$ este probabilitatea totală a datelor luând în considerare toate ipotezele

posibile.
Dacă a priori şi verosimilitatea sunt cunoscute pentru toate ipotezele, atunci formula lui Bayes calculează a posteriori exact. Acesta a fost cazul când am aruncat un zar selectat aleator dintr-o urnă al cărei conţinut îl ştiam. Numim aceasta logica deductivă a teoriei probabilităţilor şi ea dă un mod direct de a compara ipotezele, a trage concluzii şi a lua decizii.
În cele mai multe experimente, probabilităţile a priori ale ipotezelor nu sunt cunoscute. În acest caz, recurgem la arta deducţiei statistice: ori inventăm o a priori (Bayesienii), ori facem tot ce putem folosind doar verosimilitatea (frecvenţioniştii).
Şcoala Bayesiană modelează incertitudinea printr-o repartiţie de probabilitate peste ipoteze. Abilitatea cuiva de a face deducţii depinde de gradul de încredere în a priori aleasă şi robusteţea constatărilor pentru a alterna repartiţiile a priori poate fi relevantă şi importantă.
Şcoala frecvenţionistă foloseşte doar repartiţii condiţionate ale datelor cunoscând ipoteze specifice. Presupunerea este că o ipoteză (parametru specificând repartiţia condiţionată a datelor) este adevărată şi că datele observate provin din acea repartiţie. În particular, abordarea frecvenţionistă nu depinde de o a priori subiectivă care poate varia de la un investigator la altul.
Aceste 2 şcoli pot fi comparate mai departe după cum urmeză:

Deducţia Bayesiană

foloseşte probabilităţi atât pentru ipoteze cât şi pentru date;
depinde de a priori şi verosimilitatea datelor observate;
cere să se ştie sau să se construiască o ”a priori subiectivă”;
a dominat practica statistică înainte de secolul 20;
poate fi intens calculată din cauza integrării peste mulţi parametri.

Deducţia frecvenţionistă (NHST)

nu foloseşte niciodată probabilitatea unei ipoteze (nu a priori sau a posteriori);
depinde de verosimilitatea $P(\mathcal{D}|\mathcal{H})$ atât pentru datele observate cât şi pentru cele neobservate;
nu cere o a priori;
practica statistică dominantă de-a lungul secolului 20;
tinde să fie mai puţin intensă din punctul de vedere al calculelor.

Măsuri frecvenţioniste ca valori p sau intervale de încredere continuă să domine cercetarea, în special în ştiinţele vieţii. Totuşi, în era actuală a computerelor puternice şi datelor mari, metodele Bayesiene au avut o renaştere enormă în domenii ca învăţarea automată şi genetica. Acum sunt un număr de studii clinice mari, în curs de desfăşurare folosind protocoale Bayesiene, ceea ce ar fi fost greu de imaginat cu o generaţie în urmă.
În timp ce împărţirile profesionale rămân, consensul care se formează printre

statisticienii de top este că cea mai eficace abordare a problemelor complexe adesea foloseşte cele mai bune perspective din ambele şcoli lucrând împreună.

1.4 Critici şi apărări

1.4.1 Critica deducţiei Bayesiene

1. Principala critică a deducţiei Bayesiene este că o a priori este subiectivă. Nu există o singură metodă de a alege a priori, deci persoane diferite pot produce a priori diferite şi de aceea pot ajunge la a posteriori şi concluzii diferite.
2. Mai departe, sunt obiecţii filozofice la atribuirea de probabilităţi ipotezelor, deoarece ipotezele nu sunt rezultate ale experimentelor repetabile în care se poate măsura frecvenţa pe termen lung. Mai degrabă, o ipoteză este ori adevărată ori falsă, indiferent dacă se ştie care este cazul. O monedă este ori corectă, ori incorectă; tratamentul 1 este ori mai bun, ori mai rău ca tratamentul 2; soarele va răsări sau nu va răsări mâine.

1.4.2 Apărarea deducţiei Bayesiene

1. Probabilitatea ipotezelor este exact ce ne trebuie pentru a lua decizii. Când doctorul îmi spune că un test a ieşit pozitiv vreau să ştiu care este probabilitatea că aceasta înseamnă că sunt bolnav. Adică, vreau să ştiu probabilitatea ipotezei "Sunt bolnav".
2. Folosirea teoremei lui Bayes este riguroasă logic. Odată ce avem a priori toate calculele noastre au certitudinea logicii deductive.
3. Încercând diferite a priori, putem vedea cât de sensibile sunt rezultatele noastre de alegerea a priori.
4. Este uşor de comunicat un rezultat încadrat în termeni de probabilităţile ipotezelor.
5. Chiar dacă a priori pot fi subiective, se pot specifica presupunerile folosite pentru a ajunge la ele, care permit altor oameni să le folosească sau să încerce alte a priori.
6. Dovezile derivate din date sunt independente de noţiuni despre "datele mai extreme", care depind de schema experimentală exactă.
7. Datele pot fi folosite pe măsură ce vin. Nu există o cerinţă ca fiecare contingent să fie planificat înainte.

1.4.3 Critica deducţiei frecvenţioniste

1. Este ad-hoc şi nu are forţa logicii deductive. Noţiuni ca "date mai extreme" nu sunt bine definite. Valoarea p depinde de schema experimentală exactă.

2. Experimentele trebuie complet specificate înainte. Aceasta poate conduce la rezultate aparent paradoxale. Vezi "istoria voltmetrului" în:
`http://en.wikipedia.org/wiki/Likelihood_principle`.
3. Valoarea p şi nivelul de semnificaţie sunt notoriu predispuse la interpretare greşită. Un nivel de semnificaţie de 0.05 înseamnă că probabilitatea unei erori de tipul I este 5%. Adică, dacă ipoteza 0 este adevărată, atunci în 5% din timp va fi respinsă din cauza caracterului aleatoriu. Mulţi oameni gândesc eronat că o valoare p de 0.05 înseamnă că probabilitatea ipotezei 0 este 5%. Aţi putea argumenta că aceasta nu este o critică a deducţiei frecvenţioniste ci, mai degrabă, o critică a ignoranţei populare. Totuşi, subtilitatea ideilor contribuie sigur la problemă.

1.4.4 Apărarea deducţiei frecvenţioniste

1. Este obiectivă: toţi statisticienii vor fi de acord cu valoarea p. Orice individ poate apoi decide dacă valoarea p garantează respingerea ipotezei 0.
2. Testarea ipotezei folosind testarea semnificaţiei frecvenţionistă este aplicată în analiza statistică a investigaţiilor ştiinţifice, evaluând puterea dovezilor împotriva unei ipoteze 0 cu date. Interpretarea rezultatelor este lăsată utilizatorului testelor. Diferiţi utilizatori pot aplica diferite nivele de semnificaţie pentru a determina semnificaţia statistică. Statistica frecvenţionistă nu pretinde că dă un mod de a alege un nivel de semnificaţie; mai degrabă descrie explicit compromisul între erorile de tipul I şi de tipul II.
3. Proiectarea experimentului frecvenţionistă cere o descriere atentă a experimentului şi metodelor de analiză înaintea startului. Aceasta ajută controlul pentru părtinirea experimentatorului.
4. Abordarea frecvenţionistă a fost folosită de peste 100 de ani şi am văzut progres ştiinţific extraordinar. Cu toate că frecvenţionistul însuşi n-ar pune o probabilitate pe încrederea că metodele frecvenţioniste sunt valoroase, n-ar trebui ca această istorie să dea Bayesianului o încredere a priori puternică în utilitatea metodelor frecvenţioniste?

1.5 Fiţi atenţi la p-urile voastre

Facem un test t cu 2 selecţii pentru medii egale, cu $\alpha = 0.05$ şi obţinem o valoare p de 0.04. Care sunt şansele ca cele 2 selecţii să provină din repartiţii cu aceeaşi medie?
(a) 19/1 (b) 1/19 (c) 1/20 (d) 1/24 (e) necunoscute.
Răspuns. (e) necunoscute. Metodele frecvenţioniste dau doar probabilităţi ale statisticilor condiţionate de ipoteze. Ele nu dau probabilităţi ale ipotezelor.

1.6 Reguli de oprire

Când derulăm o serie de încercări, avem nevoie de o regulă despre când să ne oprim. 2 reguli uzuale sunt:
1. Facem exact n încercări şi ne oprim.
2. Facem încercări până vedem un rezultat şi apoi ne oprim.
În acest exemplu vom considera 2 experimente de aruncarea monedei.
Experimentul 1: Aruncăm moneda exact de 6 ori şi raportăm numărul aversurilor.
Experimentul 2: Aruncăm moneda până apare primul revers şi raportăm numărul aversurilor.
Jon este îngrijorat că moneda lui este deplasată spre aversuri, deci, înainte s-o utilizeze la cursuri, el îi testează corectitudinea. El face un experiment şi raportează lui Jerry că secvenţa lui de aruncări a fost $HHHHHT$. Dar Jerry nu este atent şi uită ce experiment a făcut Jon să producă datele.
Abordarea frecvenţionistă
Deoarece a uitat ce experiment a făcut Jon, Jerry frecvenţionistul decide să calculeze valorile p pentru ambele experimente cunoscând datele lui Jon.
Fie θ probabilitatea aversului. Avem ipotezele 0 şi alternativă unilaterală

$$H_0 : \theta = 0.5, \ \ H_A : \theta > 0.5.$$

Experimentul 1: Repartiţia 0 este binomială(6, 0.5) deci valoarea p unilaterală este probabilitatea a 5 sau 6 aversuri în 6 aruncări. Folosind `R` obţinem

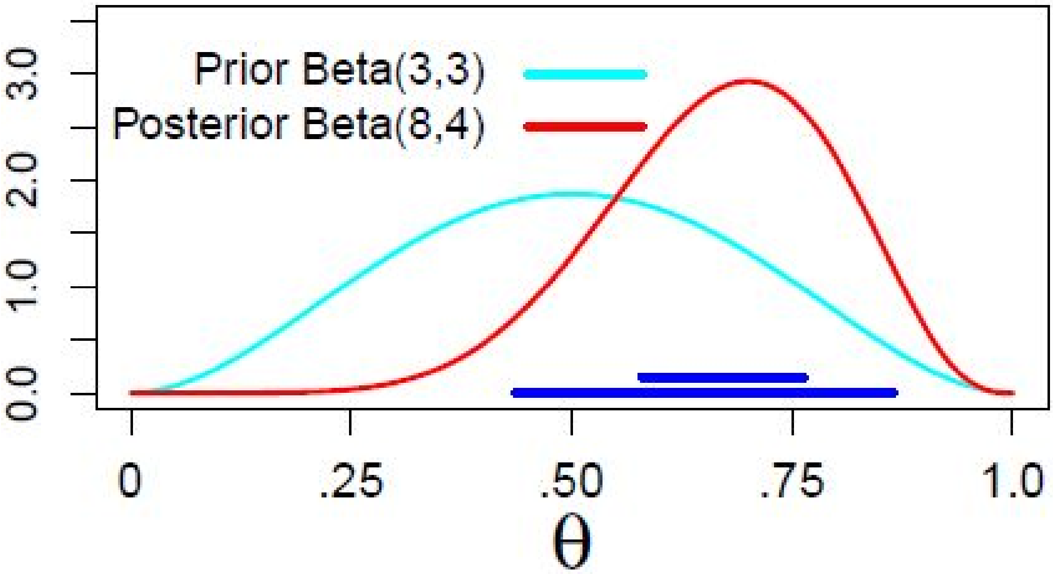
$$p = 1 - \texttt{pbinom}(4, 6, 0.5) = 0.109375.$$

Experimentul 2: Repartiţia 0 este geometrică(0.5) deci, valoarea p unilaterală este probabilitatea a 5 sau mai multe aversuri înaintea primului revers. Folosind `R` obţinem

$$p = 1 - \texttt{pgeom}(4, 0.5) = 0.03125.$$

Folosind nivelul de semnificaţie tipic de 0.05, aceleaşi date duc la concluzii diferite! Am respinge H_0 în experimentul 2, dar nu în experimentul 1.
Frecvenţionistul nu are nicio problemă cu aceasta. Mulţimea rezultatelor este diferită pentru experimentele diferite, deci noţiunea de date extreme şi de aceea, de valoare p, este diferită. De exemplu, în experimentul 1 am considera $THHHHH$ a fi la fel de extrem ca $HHHHHT$. În experimentul 2 n-am vedea niciodată $THHHHH$ deoarece experimentul s-ar termina după primul revers.
Abordarea Bayesiană.
Jerry Bayesianul ştie că nu contează care dintre cele 2 experimente a fost

făcut de Jon, deoarece funcţiile (coloanele) de verosimilitate binomială şi geometrică pentru datele $HHHHHT$ sunt proporţionale. În orice caz, el trebuie să inventeze o a priori şi alege Beta(3, 3). Aceasta este o a priori relativ plată concentrată peste intervalul $0.25 \leq \theta \leq 0.75$.
Vezi `http://mathlets.org/mathlets/beta-distribution`.
Deoarece repartiţiile beta şi binomială (sau geometrică) formează o pereche conjugată, actualizarea Bayesiană este simplă. Datele de 5 aversuri şi 1 revers dau o repartiţie a posteriori Beta(8,4). Iată graficul a priori şi a posteriori. Liniile albastre de jos sunt intervale de probabilitate 50% şi 90% pentru a posteriori.



Repartiţiile a priori şi a posteriori cu intervale de probabilitate 0.5 şi 0.9

Iată calculele relevante din `R`:
Interval de probabilitate 50% a posteriori:
`qbeta(c(0.25,0.75),8,4)` = [0.579529, 0.7635974].
Interval de probabilitate 90% a posteriori:
`qbeta(c(0.05,0.95),8,4)` = [0.4356258, 0.8649245].
$P(\theta > 0.5|\text{date}) = 1 - \texttt{pbeta}(0.5, 8, 4) = 0.8867188$.
Plecând de la a priori Beta(3,3), probabilitatea a posteriori ca moneda să fie deplasată spre avers este 0.8867188.

1.7 Luarea deciziilor

Destul de des scopul deducţiei statistice este de a ajuta la luarea deciziilor, de exemplu dacă sau nu să fie supus unei intervenţii chirurgicale, cât de mult să investească într-o acţiune, dacă sau nu să mergă să termine şcoala, etc.
În teoria deciziilor statistice, consecinţele acţiunilor sunt măsurate printr-o funcţie de utilitate. Funcţia de utilitate atribuie o pondere fiecărui rezultat posibil; în limbajul probabilităţilor este pur şi simplu o variabilă aleatoare.
De exemplu, în investiţiile mele aş putea atribui o utilitate de d rezultatului unui câştig de d lei pe acţiune (dacă $d < 0$, utilitatea mea este negativă). Pe de altă parte, dacă toleranţa mea pentru risc este mică, voi atribui mai multă utilitate negativă pentru pierderi decât pentru câştiguri (să zicem, $-d^2 - 1$ dacă $d < 0$ şi d dacă $d \geq 0$).
O regulă de decizie combină utilitatea medie cu dovezile pentru fiecare ipoteză cunoscând datele (de exemplu, valori p sau repartiţii a posteriori) într-un cadru statistic formal pentru luarea deciziilor.
În acest cadru, frecvenţionistul va considera media utilităţii dată fiind o ipoteză

$$E(U|\mathcal{H}),$$

unde U este variabila aleatoare reprezentând utilitatea. Există metode frecvenţioniste pentru combinarea utilităţii medii cu valori p ale ipotezelor pentru a ghida deciziile.
Bayesianul poate combina $E(U|\mathcal{H})$ cu a posteriori (sau a priori dacă este înaintea colectării datelor) pentru a crea o regulă de decizie Bayesiană.
În orice cadru, 2 persoane considerând aceeaşi investiţie pot avea funcţii de utilitate diferite şi lua decizii diferite. De exemplu, o acţiune riscantă (cu un potenţial mai mare de a fluctua) va fi mai atrăgătoare în raport cu prima funcţie de utilitate de mai sus decât în raport cu a 2-a.
Un rezultat teoretic semnificativ este că pentru orice regulă de decizie există o regulă de decizie Bayesiană care este, într-un sens precis o regulă cel puţin la fel de bună.

2 Intervale de încredere bazate pe date normale

2.1 Scopurile învăţării

1. Să poată să determine dacă o expresie defineşte o statistică de interval validă.

2. Să poată calcula intervale de încredere z şi t pentru media datelor normale.
3. Să poată calcula intervalul de încredere χ^2 pentru dispersia datelor normale.
4. Să poată defini nivelul de încredere al unui interval de încredere.
5. Să poată explica relaţia dintre intervalul de încredere z (şi nivelul de încredere) şi regiunea de nerespingere z (şi nivelul de încredere) în NHST.

2.2 Introducere

Intervalele de încredere sunt unelte ale statisticii frecvenţioniste. Presupunem că avem un model (repartiţie de probabilitate) pentru date observate cu un parametru necunoscut. Am văzut cum NHST foloseşte datele pentru a testa ipoteza că parametrul necunoscut are o valoare particulară.
Am văzut de asemenea cum estimări punctuale ca MLE folosesc datele pentru a da o estimare a parametrului necunoscut. Pe cont propriu, o estimare punctuală ca $\overline{x} = 2.2$ nu poartă informaţie despre acurateţea ei; este doar un singur număr, indiferent dacă este bazat pe 10 date sau 1000000 de date.
Din acest motiv, statisticienii măresc estimările punctuale la intervale de încredere. De exemplu, pentru a estima o medie necunoscută μ am putea spune că cea mai bună estimare a mediei este $\overline{x} = 2.2$ cu un interval de 95% încredere $[1.2, 3.2]$. Alt mod de a descrie intervalul este $\overline{x} \pm 1$.
Luate împreună, lungimea intervalului şi nivelul de încredere dau o măsură a puterii dovezilor care sprijină ipoteza că μ este aproape de estimarea noastră $\overline{x}$. Nivelul de încredere al unui interval este analog nivelului de semnificaţie al unui NHST. Nu este un accident că adesea vedem nivelul de semnificaţie $\alpha = 0.05$ şi nivelul de încredere $0.95 = 1 - \alpha$.
Intervalele de încredere z şi t pentru medie şi χ^2 pentru dispersie sunt uşor de calculat. Provocarea cu intervalele de încredere nu este calculul lor, ci mai degrabă interpretarea lor corectă şi a şti cum să le folosim în practică.

2.3 Statistici interval

Reamintim definiţia noastră de lucru: o statistică este orice se poate calcula din date. În particular, formula pentru o statistică nu poate include cantităţi necunoscute.
Exemplul 1. Presupunem că $x_1, ..., x_n$ provin din $N(\mu, \sigma^2)$, unde μ şi σ sunt necunoscute.
(i) $\overline{x}$ şi $\overline{x} - 5$ sunt statistici.
(ii) $\overline{x} - \mu$ nu este o statistică deoarece μ este necunoscută.
(iii) Dacă μ_0 este o valoare cunoscută, atunci $\overline{x} - \mu_0$ este o statistică. Acest

caz apare când considerăm ipoteza 0 $\mu = \mu_0$. De exemplu, dacă ipoteza 0 este $\mu = 5$, atunci statistica $\overline{x} - \mu_0$ este chiar $\overline{x} - 5$ de la (i).
Putem juca acelaşi joc cu intervale pentru a defini statistici interval.
Exemplul 2. Presupunem că $x_1, ..., x_n$ provin din $N(\mu, \sigma^2)$ unde μ este necunoscută.
(i) Intervalul $[\overline{x} - 2.2, \overline{x} + 2.2] = \overline{x} \pm 2.2$ este o statistică interval.
(ii) Dacă σ este cunoscută, atunci $\left[\overline{x} - \frac{2\sigma}{\sqrt{n}}, \overline{x} + \frac{2\sigma}{\sqrt{n}}\right]$ este o statistică interval.
(iii) Pe de altă parte, dacă σ este necunoscută, atunci $\left[\overline{x} - \frac{2\sigma}{\sqrt{n}}, \overline{x} + \frac{2\sigma}{\sqrt{n}}\right]$ **nu** este o statistică interval.
(iv) Dacă s^2 este dispersia de selecţie, atunci $\left[\overline{x} - \frac{2s}{\sqrt{n}}, \overline{x} + \frac{2s}{\sqrt{n}}\right]$ este o statistică interval deoarece s^2 este calculată din date.
Tehnic, o statistică interval nu este mai mult decât o pereche de statistici punctuale care dau marginile inferioară şi superioară ale intervalului.
Observaţii.
1. Intervalul este aleator - noi date aleatoare vor produce un nou interval.
2. Intervalul nu depinde de valoarea unui parametru necunoscut.
3. Pdf pentru unele repartiţii este funcţie pară, pentru altele nu.
Exemplul 3. Pdf-urile pentru $N(0, 1)$ şi pentru t sunt funcţii pare, i.e. graficul este simetric faţă de axa Oy, dar pentru χ^2 nu.

2.4 Intervale de încredere z pentru medie

Presupunem că avem date normal repartizate:

$$x_1, x_2, ..., x_n \sim N(\mu, \sigma^2)$$

2.4.1 Definiţia intervalelor de încredere z pentru medie

Definiţie. Presupunem că datele $x_1, x_2, ..., x_n \sim N(\mu, \sigma^2)$, cu medie necunoscută μ şi dispersie cunoscută σ^2. Intervalul de încredere $(1 - \alpha)$ pentru μ este

$$\left[\overline{x} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \overline{x} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right], \tag{1}$$

unde $z_{\alpha/2}$ este valoarea critică din dreapta, pentru care $P(Z > z_{\alpha/2}) = \alpha/2$. De exemplu, dacă $\alpha = 0.05$, atunci $z_{\alpha/2} = \texttt{qnorm}(1 - 0.05/2) \approx 1.96$, deci intervalul de încredere 0.95 (sau 95%) este

$$\left[\overline{x} - \frac{1.96\sigma}{\sqrt{n}}, \overline{x} + \frac{1.96\sigma}{\sqrt{n}}\right].$$

Următoarea aplicaţie generează date normale şi arată intervalele de încredere z sau t pentru medie.
`http://mathlets.org/mathlets/confidence-intervals/`.
Exemplul 4. Presupunem că colectăm 100 de date dintr-o repartiţie $N(\mu, 3^2)$ şi media de selecţie este $\overline{x} = 12$. Daţi intervalul de încredere 95% pentru μ.
Răspuns. Folosind formula, intervalul de încredere 95% pentru μ este

$$\left[\overline{x} - \frac{1.96\sigma}{\sqrt{n}}, \overline{x} + \frac{1.96\sigma}{\sqrt{n}}\right] = \left[12 - \frac{1.96 \cdot 3}{10}, 12 + \frac{1.96 \cdot 3}{10}\right] = [11.412, 12.588].$$

2.4.2 Explicarea definiţiei partea 1: regiuni de respingere

Explicăm definiţia (1) plecând de la cunoştinţele noastre despre regiuni de respingere/nerespingere. Exprimarea "regiune de nerespingere" nu este frumoasă, dar o vom folosi în locul exprimării incorecte "regiune de acceptare".
Exemplul 5. Presupunem că $n = 12$ date provin din $N(\mu, 5^2)$, unde μ este necunoscută. Faceţi un test de semnificaţie bilateral pentru H_0: $\mu = 2.71$ folosind statistica $\overline{x}$ la nivelul de semnificaţie $\alpha = 0.05$. Descrieţi regiunile de respingere şi nerespingere.
Răspuns. Sub ipoteza 0 avem $\mu = 2.71$, $x_i \sim N(2.71, 5^2)$ şi deci

$$\overline{x} \sim N(2.71, 5^2/12),$$

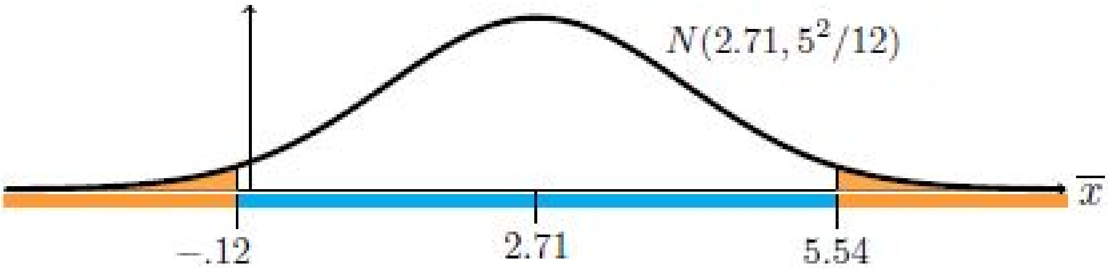
unde $5^2/12$ este dispersia $\sigma_{\overline{x}}^2$ a lui $\overline{x}$. Ştim că semnificaţia $\alpha = 0.05$ corespunde la o regiune de respingere în afara a 1.96 deviaţii standard ale mediei ipotetice. Adică, regiunile de nerespingere şi de respingere sunt separate de valorile critice $2.71 \pm 1.96\sigma_{\overline{x}}$.
Regiunea de nerespingere:

$$\left[2.71 - \frac{1.96 \cdot 5}{\sqrt{12}}, 2.71 + \frac{1.96 \cdot 5}{\sqrt{12}}\right] \approx [-0.12, 5.54].$$

Regiunea de respingere:

$$\left(-\infty, 2.71 - \frac{1.96 \cdot 5}{\sqrt{12}}\right) \cup \left(2.71 + \frac{1.96 \cdot 5}{\sqrt{12}}, \infty\right) \approx (-\infty, -0.12) \cup (5.54, \infty).$$

Următoarea figură arată regiunile de respingere şi nerespingere pentru $\overline{x}$. Regiunile reprezintă domenii pentru $\overline{x}$, deci sunt reprezentate de bare colorate pe axa $\overline{x}$. Aria regiunii colorate de sub grafic este nivelul de semnificaţie.



Regiunile de respingere (portocalie) şi de nerespingere (albastră) pentru $\overline{x}$.

Exemplul 6. Presupunem că n date provin din $N(\mu, \sigma^2)$, unde μ este necunoscută şi σ este cunoscută. Faceţi un test de semnificaţie bilateral al lui H_0: $\mu = \mu_0$ folosind statistica $\overline{x}$ la nivelul de semnificaţie $\alpha = 0.05$. Descrieţi regiunile de respingere şi nerespingere.

Răspuns. Sub ipoteza 0 $\mu = \mu_0$ avem $x_i \sim N(\mu_0, \sigma^2)$ şi deci

$$\overline{x} \sim N(\mu_0, \sigma^2/n),$$

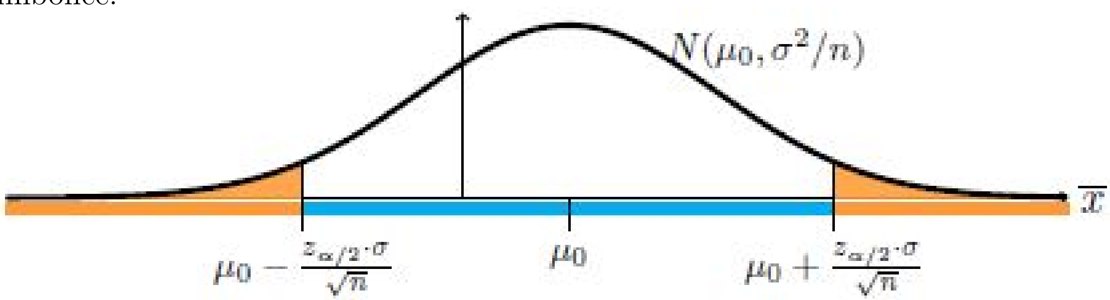
unde σ^2/n este dispersia $\sigma^2_{\overline{x}}$ a lui $\overline{x}$ şi μ_0, σ şi n sunt toate valori cunoscute. Fie $z_{\alpha/2}$ valoarea critică: $P(Z > z_{\alpha/2}) = \alpha/2$. Regiunile de nerespingere şi de respingere sunt separate de valorile lui $\overline{x}$ care sunt la distanţă de $z_{\alpha/2} \cdot \sigma_{\overline{x}}$ de media ipotetică. Deoarece $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$, avem:
Regiunea de nerespingere:

$$\left[\mu_0 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \mu_0 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right]. \tag{2}$$

Regiunea de respingere:

$$\left(-\infty, \mu_0 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right) \cup \left(\mu_0 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \infty\right).$$

Obţinem aceeaşi figură ca mai sus, cu numerele explicite înlocuite de valori simbolice.



Regiunile de respingere (portocalie) şi de nerespingere (albastră) pentru $\overline{x}$.

2.4.3 Manipularea intervalelor: pivotarea

Pivotarea este ideea că $\overline{x} \in [\mu_0 - a, \mu_0 + a] \iff \mu_0 \in [\overline{x} - a, \overline{x} + a], \forall a > 0$.
Exemplul 7. Presupunem că avem media de selecţie $\overline{x}$ şi media ipotetică $\mu_0 = 2.71$. Presupunem de asemenea că repartiţia 0 este $N(\mu_0, 3^2)$. Atunci, cu un nivel de semnificaţie de 0.05, avem:
$\mu_0 + 1.96\sigma = 2.71 + 1.96 \cdot 3 = 2.71 + 5.88$ este valoarea critică 0.025;
$\mu_0 - 1.96\sigma = 2.71 - 1.96 \cdot 3 = 2.71 - 5.88$ este valoarea critică 0.975.
Regiunea de nerespingere este centrată în $\mu_0 = 2.71$. Adică, nu respingem H_0 dacă $\overline{x}$ este în intervalul

$$[\mu_0 - 1.96\sigma, \mu_0 + 1.96\sigma] = [2.71 - 5.88, 2.71 + 5.88].$$

Intervalul de încredere este centrat în $\overline{x}$. Intervalul de încredere 0.95 are aceeaşi lungime cu regiunea de nerespingere. Este intervalul

$$[\overline{x} - 1.96\sigma, \overline{x} + 1.96\sigma] = [\overline{x} - 5.88, \overline{x} + 5.88].$$

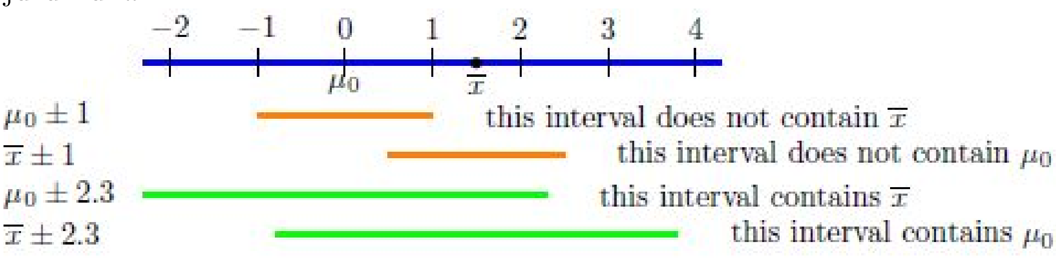
Aici este o simetrie:
$\overline{x} \in [2.71 - 1.96\sigma, 2.71 + 1.96\sigma] \iff 2.71 \in [\overline{x} - 1.96\sigma, \overline{x} + 1.96\sigma]$.
Această simetrie este numită pivotare. Iată câteva exemple numerice de pivotare:
Exemplul 8. **(i)** $1.5 \in [0 - 2.3, 0 + 2.3] \implies 0 \in [1.5 - 2.3, 1.5 + 2.3]$.
(ii) Analog $1.5 \notin [0 - 1, 0 + 1] \implies 0 \notin [1.5 - 1, 1.5 + 1]$.
Simetria poate fi mai clară dacă vorbim în termeni de distanţe: afirmaţia

$$"1.5 \in [0 - 2.3, 0 + 2.3]"$$

spune că distanţa de la 1.5 la 0 este cel mult 2.3. Analog, afirmaţia

$$"0 \in [1.5 - 2.3, 1.5 + 2.3]"$$

spune exact acelaşi lucru, i.e. distanţa de la 0 la 1.5 este cel mult 2.3.
Iată o vizualizare a *pivotării* din intervalele din jurul lui μ_0 în intervalele din jurul lui $\overline{x}$.



Distanţa dintre $\overline{x}$ şi μ_0 este 1.5. Deoarece $1 < 1.5$, $[\mu_0 - 1, \mu_0 + 1]$ nu se întinde destul de departe pentru a conţine $\overline{x}$. Analog, intervalul $[\overline{x} - 1, \overline{x} + 1]$ nu se întinde destul de mult pentru a conţine μ_0. Deoarece $2.3 > 1.5$, $\overline{x} \in [\mu_0 - 2.3, \mu_0 + 2.3]$ şi $\mu_0 \in [\overline{x} - 2.3, \overline{x} + 2.3]$.

2.4.4 Explicarea definiţiei, partea a 2-a: translatarea regiunii de nerespingere într-un interval de încredere

În exemplele precedente avem o ipoteză 0. Dar dacă nu avem o ipoteză 0? În acest caz, avem estimarea punctuală $\overline{x}$, dar încă vrem să folosim datele pentru a estima un interval pentru media necunoscută. Adică, vrem o statistică interval. Aceasta este dată de un interval de încredere.
Folosind numerele din exemplul 5, spunem că la nivelul de semnificaţie de 0.05, nu respingem H_0 dacă

$$\overline{x} \in \left[2.71 - \frac{1.96 \cdot 5}{\sqrt{12}}, 2.71 + \frac{1.96 \cdot 5}{\sqrt{12}}\right]. \tag{3}$$

Rolurile lui $\overline{x}$ şi 2.71 sunt simetrice. Nu respingem dacă

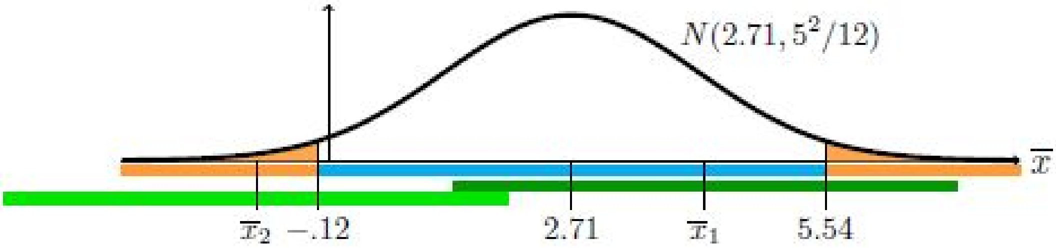
$$2.71 \text{ este în intervalul } \overline{x} \pm \frac{1.96 \cdot 5}{\sqrt{12}}. \tag{4}$$

Am ajuns la scopul nostru al unei statistici interval care estimează media necunoscută. Putem rescrie (4) astfel: la nivelul de semnificaţie de 0.05 nu respingem dacă

$$2.71 \in \left[\overline{x} - \frac{1.96 \cdot 5}{\sqrt{12}}, \overline{x} + \frac{1.96 \cdot 5}{\sqrt{12}}\right]. \tag{5}$$

Astfel, diferite valori ale lui $\overline{x}$ generează diferite intervale.
Intervalul din (5) este exact intervalul de încredere definit în (1).
Facem observaţii despre intervalul de încredere:
1. Depinde doar de $\overline{x}$, deci este o statistică.
2. Nivelul de semnificaţie $\alpha = 0.05$ înseamnă că, presupunând că ipoteza 0 că $\mu = 2.71$ este adevărată, datele aleatoare ne vor duce la respingerea ipotezei 0 în 5% din timp (o eroare de tipul I).
3. Din nou, **presupunând** că $\mu = 2.71$, atunci 5% din timp intervalul de încredere nu va conţine 2.71 şi reciproc, 95% din timp va conţine 2.71.
Următoarea figură ilustrează cum nu respingem H_0 dacă intervalul de încredere centrat în $\overline{x}$ conţine μ_0 şi respingem H_0 dacă intervalul de încredere nu conţine μ_0.
1. Am plecat cu figura de la exemplul 5, care arată repartiţia 0 pentru $\mu_0 = 2.71$ şi regiunile de respingere şi nerespingere.
2. Am adăugat 2 valori posibile ale statisticii $\overline{x}$, $\overline{x}_1$ şi $\overline{x}_2$ şi intervalele lor de încredere. Observaţi că lungimea fiecărui interval este exact aceeaşi ca lungimea regiunii de nerespingere deoarece ambele folosesc $\pm\frac{1.96 \cdot 5}{\sqrt{12}}$.
Prima valoare, $\overline{x}_1$ este în regiunea de nerespingere şi intervalul ei de încredere conţine $\mu_0 = 2.71$. Aceasta ilustrează că nerespingerea lui H_0 corespunde faptului că intervalul de încredere conţine pe μ_0.

A 2-a valoare, $\overline{x}_2$, este în regiunea de respingere şi intervalul ei de încredere nu conţine pe μ_0. Aceasta ilustrează că respingerea lui H_0 corespunde faptului că intervalul de încredere nu conţine pe μ_0.



Regiunea de nerespingere (albastru) şi cele 2 intervale de încredere (verde).

Alegerea noastră pentru ipoteza 0 $\mu = 2.71$ a fost complet arbitrară. Dacă înlocuim $\mu = 2.71$ cu altă ipoteză $\mu = \mu_0$, atunci intervalul (5) va fi acelaşi. Numim intervalul (5) un interval de încredere 95% deoarece, presupunând $\mu = \mu_0$, în medie, va conţine μ_0 în 95% din experimentele aleatoare.

2.4.5 Explicarea definiţiei partea a 3-a: translatarea unei regiuni de nerespingere generale într-un interval de încredere

Valorile specifice ale lui σ şi n din exemplul precedent pot fi înlocuite prin litere. În acest fel putem trece exemplul 6 prin aceiaşi paşi ca exemplul 5.
În cuvinte, relaţia (2) şi figura corespunzătoare spune că nu respingem dacă

$$\overline{x} \in \left[\mu_0 - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \mu_0 + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right].$$

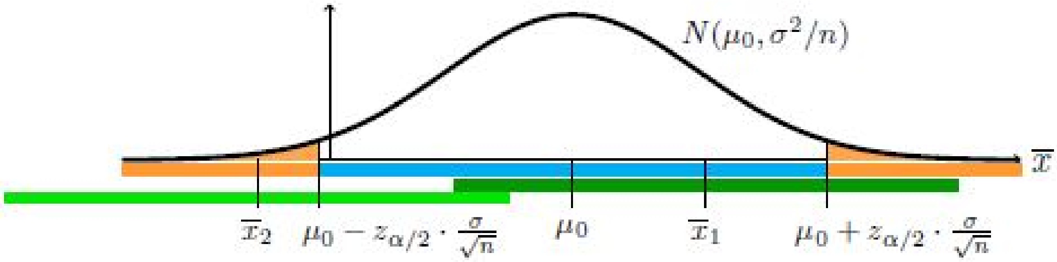
Aceasta este echivalent cu a spune că nu respingem dacă

$$\mu_0 \text{ este în intervalul } \overline{x} \pm \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}. \tag{6}$$

Putem rescrie relaţia (6) astfel: la nivelul de semnificaţie α nu respingem dacă

$$\mu_0 \in \left[\overline{x} - \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}, \overline{x} + \frac{z_{\alpha/2} \cdot \sigma}{\sqrt{n}}\right]. \tag{7}$$

Numim intervalul din (7) un interval de încredere $(1-\alpha)$ deoarece, presupunând $\mu = \mu_0$, în medie va conţine μ_0 în fracţia de $(1-\alpha)$ din experimentele aleatoare.
Figura următoare ilustrează că μ_0 este în intervalul de încredere $(1-\alpha)$ centrat în $\overline{x} \iff \overline{x}$ este în regiunea de nerespingere (la nivelul de semnificaţie α) pentru H_0: $\mu = \mu_0$.



$\overline{x}_1$ este în regiunea de nerespingere pentru H_0: $\mu = \mu_0 \iff$ intervalul de încredere centrat în $\overline{x}_1$ conţine pe μ_0.

2.4.6 Exemplu de calcul

Exemplul 9. Presupunem că datele 2.5, 5.5, 8.5, 11.5 provin dintr-o repartiţie $N(\mu, 10^2)$ cu medie necunoscută μ.
a) Calculaţi estimarea punctuală $\overline{x}$ pentru μ şi intervalele de încredere 95%, 80% şi 50% corespunzătoare.
b) Considerăm ipoteza 0 $\mu = 1$. Aţi respinge H_0 la $\alpha = 0.05$? $\alpha = 0.2$? $\alpha = 0.5$? Faceţi aceasta în 2 moduri: întâi verificând dacă valoarea ipotetică a lui μ este în intervalul de încredere relevant şi apoi construind o regiune de respingere.
Răspuns. a) $\overline{x} = \frac{2.5+5.5+8.5+11.5}{4} = 7$. Punctele critice sunt

$$z_{0.025} = \texttt{qnorm}(0.975) \approx 1.96,\ z_{0.1} = \texttt{qnorm}(0.9) \approx 1.28,\ z_{0.25} = \texttt{qnorm}(0.75) \approx 0.67.$$

Deoarece $n = 4$, avem $\overline{x} \sim N(\mu, 10^2/4)$, i.e. $\sigma_{\overline{x}} = 5$. Deci avem:

$$\text{interval de încredere 95\%} = [\overline{x} - z_{0.025}\sigma_{\overline{x}}, \overline{x} + z_{0.025}\sigma_{\overline{x}}] \approx [7 - 1.96 \cdot 5, 7 + 1.96 \cdot 5] = [-2.8, 16.8].$$
$$\text{interval de încredere 80\%} = [\overline{x} - z_{0.1}\sigma_{\overline{x}}, \overline{x} + z_{0.1}\sigma_{\overline{x}}] \approx [7 - 1.28 \cdot 5, 7 + 1.28 \cdot 5] = [0.6, 13.4].$$
$$\text{interval de încredere 50\%} = [\overline{x} - z_{0.25}\sigma_{\overline{x}}, \overline{x} + z_{0.25}\sigma_{\overline{x}}] \approx [7 - 0.67 \cdot 5, 7 + 0.67 \cdot 5] = [3.65, 10.35].$$

Fiecare dintre aceste intervale este o estimare interval pentru μ. Dacă nivelul de încredere este mai mare, atunci intervalul corespunzător este mai lung.
b) Deoarece $\mu = 1$ este în intervalele de încredere 95% şi 80%, nu respingem ipoteza 0 la nivelele $\alpha = 0.05$ sau $\alpha = 0.2$. Deoarece $\mu = 1$ nu este în intervalul de încredere 50%, respingem H_0 la nivelul $\alpha = 0.5$.
Construim regiunile de respingere folosind aceleaşi valori critice ca la a). Diferenţa este că regiunile de respingere sunt complementare de intervale centrate în valoarea ipotetică pentru μ: $\mu_0 = 1$ şi intervalele de încredere sunt centrate în $\overline{x}$. Iată regiunile de respingere:

$$\alpha = 0.05 \implies (-\infty, \mu_0 - z_{0.025}\sigma_{\overline{x}}) \cup (\mu_0 + z_{0.025}\sigma_{\overline{x}}, \infty) \approx (-\infty, -8.8) \cup (10.8, \infty).$$
$$\alpha = 0.2 \implies (-\infty, \mu_0 - z_{0.1}\sigma_{\overline{x}}) \cup (\mu_0 + z_{0.1}\sigma_{\overline{x}}, \infty) \approx (-\infty, -5.4) \cup (7.4, \infty).$$
$$\alpha = 0.5 \implies (-\infty, \mu_0 - z_{0.25}\sigma_{\overline{x}}) \cup (\mu_0 + z_{0.25}\sigma_{\overline{x}}, \infty) \approx (-\infty, -2.35) \cup (4.35, \infty).$$

Pentru a face NHST trebuie să verificăm dacă $\overline{x} = 7$ se află sau nu în regiunea de respingere.

$$\alpha = 0.05 : 7 \notin (-\infty, -8.8) \cup (10.8, \infty).$$

Nu respingem ipoteza că $\mu = 1$ la nivelul de semnificaţie de 0.05.

$$\alpha = 0.2 : 7 \notin (-\infty, -5.4) \cup (7.4, \infty).$$

Nu respingem ipoteza că $\mu = 1$ la nivelul de semnificaţie de 0.2.

$$\alpha = 0.5 : 7 \in (-\infty, -2.35) \cup (4.35, \infty).$$

Respingem ipoteza că $\mu = 1$ la nivelul de semnificaţie de 0.5.
Obţinem aceleaşi răspunsuri cu orice metodă.

2.5 Intervale de încredere t pentru medie

În acest cadru σ este necunoscută, deci trebuie să facem următoarele înlocuiri:
1. Folosim $s_{\overline{x}} = \frac{s}{\sqrt{n}}$ în locul lui $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$. Aici s^2 este dispersia de selecţie pe care am folosit-o mai înainte la teste t.
2. Folosim valori critice t în locul valorilor critice z.

2.5.1 Definiţia intervalelor de încredere t pentru medie

Definiţie. Presupunem că $x_1, ..., x_n \sim N(\mu, \sigma^2)$, unde valorile mediei μ şi deviaţiei standard σ sunt ambele necunoscute. Intervalul de încredere $(1-\alpha)$ pentru μ este

$$\left[\overline{x} - \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}, \overline{x} + \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}\right], \tag{8}$$

unde $t_{\alpha/2}$ este valoarea critică din dreapta a. î. $P(T > t_{\alpha/2}) = \alpha/2$ pentru $T \sim t(n-1)$ şi s^2 este dispersia de selecţie a datelor.

2.5.2 Construirea intervalelor de încredere t

Presupunem că n date provin din $N(\mu, \sigma^2)$, unde μ şi σ^2 sunt necunoscute. Sub ipoteza 0 $\mu = \mu_0$, avem $x_i \sim N(\mu_0, \sigma^2)$. Deci media Studentizată are o repartiţie Student t cu $n-1$ grade de libertate:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \sim t(n-1).$$

Fie $t_{\alpha/2}$ valoarea critică a. î. $P(T > t_{\alpha/2}) = \alpha/2$ pentru $T \sim t(n-1)$. Ştim din testul t de o selecţie că regiunea de nerespingere este dată de

$$|t| \leq t_{\alpha/2}.$$

Folosind definiţia statisticii t pentru a scrie regiunea de respingere în termeni de $\overline{x}$, obţinem: la nivelul de semnificaţie α nu respingem dacă

$$\frac{|\overline{x} - \mu_0|}{s/\sqrt{n}} \leq t_{\alpha/2} \iff |\overline{x} - \mu_0| \leq t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

Geometric, partea dreaptă spune că nu respingem dacă

$$\mu_0 \text{ este la cel mult } t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \text{ de } \overline{x}.$$

Aceasta este echivalent cu a spune că nu respingem dacă

$$\mu_0 \in \left[\overline{x} - \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}, \overline{x} + \frac{t_{\alpha/2} \cdot s}{\sqrt{n}}\right].$$

Acest interval este intervalul de încredere definit în (8).
Exemplul 10. Presupunem că datele 2.5, 5.5, 8.5, 11.5 provin dintr-o repartiţie $N(\mu, \sigma^2)$ cu μ şi σ^2 ambele necunoscute.
Daţi estimările interval pentru μ aflând intervalele de încredere 95%, 80% şi 50%.
Răspuns. În `R`:
`x=c(2.5,5.5,8.5,11.5)`
$\overline{x} = \texttt{mean(x)} = 7$
$s^2 = \texttt{var(x)} = 15.$
Punctele critice sunt
$t_{0.025} = \texttt{qt}(0.975, 3) = 3.182446,$
$t_{0.1} = \texttt{qt}(0.9, 3) = 1.637744,$
$t_{0.25} = \texttt{qt}(0.75, 3) = 0.7648923.$

interval de încredere 95% $= [\overline{x} - t_{0.025} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{0.025} \cdot \frac{s}{\sqrt{n}}] = [0.8372198, 13.16278].$

interval de încredere 80% $= [\overline{x} - t_{0.1} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{0.1} \cdot \frac{s}{\sqrt{n}}] = [3.828522, 10.17148].$

interval de încredere 50% $= [\overline{x} - t_{0.25} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{0.25} \cdot \frac{s}{\sqrt{n}}] = [5.518792, 8.481208].$

Toate aceste intervale de încredere dau estimări interval pentru valoarea μ. Din nou, dacă nivelul de încredere este mai mare, atunci intervalul corespunzător este mai lung.

2.6 Intervale de încredere χ^2 pentru dispersie

Definiţie. Presupunem că datele $x_1, ..., x_n$ provin din $N(\mu, \sigma^2)$ cu media μ şi deviaţia standard σ ambele necunoscute. Intervalul de încredere $(1-\alpha)$ pentru dispersia σ^2 este

$$\left[\frac{(n-1)s^2}{c_{\alpha/2}}, \frac{(n-1)s^2}{c_{1-\alpha/2}}\right]. \tag{9}$$

Aici $c_{\alpha/2}$ este valoarea critică din dreapta a. î. $P(X^2 > c_{\alpha/2}) = \alpha/2$ pentru $X^2 \sim \chi^2(n-1)$ şi s^2 este dispersia de selecţie a datelor.
Deducerea acestui interval este aproape identică cu deducerile anterioare, acum plecând de la testul χ^2 pentru dispersie. Faptul de bază de care avem nevoie este că, pentru date provenite din $N(\mu, \sigma^2)$ cu σ cunoscută, statistica

$$\frac{(n-1)s^2}{\sigma^2}$$

are o repartiţie χ^2 cu $n-1$ grade de libertate. Deci, dată fiind ipoteza 0 H_0: $\sigma = \sigma_0$, statistica testului este $(n-1)s^2/\sigma_0^2$ şi regiunea de nerespingere la nivelul de semnificaţie α este

$$c_{1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma_0^2} \leq c_{\alpha/2}.$$

Aceasta este echivalentă cu

$$\frac{(n-1)s^2}{c_{1-\alpha/2}} \geq \sigma_0^2 \geq \frac{(n-1)s^2}{c_{\alpha/2}}.$$

Aceasta spune că nu respingem dacă

$$\sigma_0^2 \in \left[\frac{(n-1)s^2}{c_{\alpha/2}}, \frac{(n-1)s^2}{c_{1-\alpha/2}}\right].$$

Acesta este intervalul nostru de încredere $(1-\alpha)$.

3 Intervale de încredere: 3 puncte de vedere

3.1 Scopurile învăţării

1. Să poată produce intervale de încredere z, t şi χ^2 pe baza statisticilor standardizate corespunzătoare.
2. Să poată să utilizeze un test de ipoteză pentru a construi intervalul de încredere pentru un parametru necunoscut.
3. Să refuze să răspundă la întrebări de genul ”dat fiind un interval de încredere, care este probabilitatea sau şansele ca el să conţină adevărata valoare a parametrului necunoscut?”

3.2 Introducere

Abordarea noastră a intervalelor de încredere a fost o combinaţie între statisticile standardizate şi testarea ipotezelor. Azi vom considera fiecare dintre aceste perspective separat şi vom introduce şi un al 3-lea punct de vedere formal. Fiecare dă propria sa perspectivă.
1. **Statistica standardizată.** Cele mai multe intervale de încredere sunt bazate pe statistici standardizate cu repartiţii cunoscute ca z, t sau χ^2. Aceasta dă un mod direct de a construi şi interpreta intervalele de încredere ca o estimare punctuală plus sau minus o eroare.
2. **Testarea ipotezei.** Intervale de încredere pot fi de asemenea construite din testele de ipoteză. În cazurile unde nu avem o statistică standardizată această metodă încă funcţionează. Este în acord cu abordarea statisticii standardizate în cazurile unde se aplică ambele.
Această perspectivă leagă noţiunile de nivel de semnificaţie α pentru testarea ipotezei şi nivel de încredere $(1 - \alpha)$ pentru intervale de încredere; în ambele cazuri α este probabilitatea de a face o eroare de "tipul 1". Aceasta dă o înţelegere pentru folosirea cuvântului "încredere". Această perspectivă ajută de asemenea la evidenţierea naturii frecvenţioniste a intervalelor de încredere.
3. **Formal.** Definiţia formală a intervalelor de încredere este precisă şi generală. Într-un sens matematic dă o perspectivă în funcţionarea interioară a intervalelor de încredere. Totuşi, deoarece este atât de generală, uneori duce la intervale fără proprietăţi utile.

3.3 Intervale de încredere via statistici standardizate

Strategia de aici este în esenţă aceeaşi ca aceea anterioară. Presupunând date normale, avem ceea ce numim statistici standardizate, ca media standardizată, media Studentizată şi dispersia standardizată. Aceste statistici au repartiţii bine cunoscute care depind de valorile ipotetice ale lui μ şi σ. Apoi folosim calcul algebric pentru a produce intervale de încredere pentru μ şi σ.
Ideea subiacentă intervalelor de încredere este în esenţă simplă: plecăm cu o statistică standardizată (de exemplu z, t sau χ^2) şi folosim ceva calcul algebric pentru a obţine un interval care depinde doar de date şi parametrii cunoscuţi.

3.3.1 Intervale de încredere z pentru μ: date normale cu σ cunoscută

Intervalele de încredere z pentru media datelor normale sunt bazate pe media standardizată, adică pe statistica z. Plecăm cu n date independente normale

$$x_1, x_2, ..., x_n \sim N(\mu, \sigma^2).$$

Presupunem că μ este parametrul necunoscut de interes şi σ este cunoscută. Ştim că media standardizată este normală standard:

$$z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Pentru valoarea critică normală standard $z_{\alpha/2}$ avem:
$P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) = 1 - \alpha$.
De aceea,

$$P\left(-z_{\alpha/2} \leq \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}|\mu\right) = 1 - \alpha.$$

Un pic de algebră pune aceasta în forma unui interval în jurul lui μ:

$$P\left(\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}|\mu\right) = 1 - \alpha.$$

Putem sublinia că intervalul depinde doar de statistica $\overline{x}, n$ şi valoarea cunoscută σ scriind aceasta ca

$$P\left(\mu \in \left[\overline{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right] |\mu\right) = 1 - \alpha.$$

Acesta este intervalul de încredere z $(1 - \alpha)$ pentru μ. Adesea îl scriem folosind prescurtarea

$$\overline{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}.$$

Gândiţi-l ca $\overline{x}\pm$ eroarea.

3.3.2 Intervale de încredere t pentru μ: date normale cu μ şi σ necunoscute

Intervalele de încredere t pentru media datelor normale sunt bazate pe media Studentizată, adică pe statistica t. Plecăm cu n date independente normale

$$x_1, x_2, ..., x_n \sim N(\mu, \sigma^2).$$

Presupunem că μ este parametrul necunoscut de interes şi σ este necunoscută. Ştim că media Studentizată are o repartiţie Student t cu $n-1$ grade de libertate:

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} \sim t(n-1),$$

unde s^2 este dispersia de selecţie.
Acum, tot ce avem de făcut este să înlocuim media standardizată cu media Studentizată şi aceeaşi logică pe care am utilizat-o pentru z ne dă intervalul de încredere t: plecăm cu

$$P\left(-t_{\alpha/2} \leq \frac{\overline{x} - \mu}{s/\sqrt{n}} \leq t_{\alpha/2}|\mu\right) = 1 - \alpha.$$

Un pic de algebră pune aceasta în forma unui interval în jurul lui μ:

$$P\left(\overline{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}|\mu\right) = 1 - \alpha.$$

Putem sublinia că intervalul depinde doar de statisticile $\overline{x}$ şi s şi de n scriind aceasta ca

$$P\left(\mu \in \left[\overline{x} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}\right] |\mu\right) = 1 - \alpha.$$

Acesta este intervalul de încredere t $(1-\alpha)$ pentru μ. Adesea îl scriem folosind prescurtarea

$$\overline{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}.$$

Gândiţi-l ca $\overline{x}\pm$ eroarea.

3.3.3 Intervale de încredere χ^2 pentru σ^2: date normale cu μ şi σ necunoscute

Intervalele de încredere χ^2 pentru dispersia datelor normale sunt bazate pe dispersia standardizată, adică pe statistica χ^2.
Urmăm aceeaşi logică ca mai sus pentru a obţine un interval de încredere χ^2 pentru σ^2.
Presupunem că avem n date normale independente: $x_1, x_2, ..., x_n \sim N(\mu, \sigma^2)$.
Presupunem că μ şi σ sunt ambele necunoscute. Dispersia standardizată este

$$X^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1).$$

Ştim că statistica X^2 are o repartiţie χ^2 cu $n-1$ grade de libertate.
Pentru Z şi t am folosit simetria repartiţiilor pentru a înlocui $z_{1-\alpha/2}$ cu $-z_{\alpha/2}$

şi $t_{1-\alpha/2}$ cu $-t_{\alpha/2}$. Deoarece repartiţia χ^2 nu este simetrică trebuie să fim expliciţi despre valorile critice atât la stânga cât şi la dreapta. Adică,

$$P(c_{1-\alpha/2} \leq X^2 \leq c_{\alpha/2}) = 1 - \alpha,$$

unde $c_{\alpha/2}$ şi $c_{1-\alpha/2}$ sunt valorile critice din coada dreaptă. Deci

$$P\left(c_{1-\alpha/2} \leq \frac{(n-1)s^2}{\sigma^2} \leq c_{\alpha/2}|\sigma\right) = 1 - \alpha.$$

Un pic de algebră pune aceasta în forma unui interval în jurul lui σ^2:

$$P\left(\frac{(n-1)s^2}{c_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{c_{1-\alpha/2}}|\sigma\right) = 1 - \alpha.$$

Putem sublinia că intervalul depinde doar de statistica s^2 şi de n scriind aceasta ca

$$P\left(\sigma^2 \in \left[\frac{(n-1)s^2}{c_{\alpha/2}}, \frac{(n-1)s^2}{c_{1-\alpha/2}}\right]|\sigma\right) = 1 - \alpha.$$

Acesta este intervalul de încredere χ^2 $(1-\alpha)$ pentru σ^2.

3.4 Intervale de încredere via testarea ipotezei

Presupunem că avem date provenite dintr-o repartiţie cu un parametru θ a cărui valoare este necunoscută. Un test de semnificaţie pentru valoarea θ are următoarea descriere scurtă:
1. Fixăm ipoteza 0 H_0: $\theta = \theta_0$ pentru o valoare specială θ_0, de exemplu adesea avem H_0: $\theta = 0$.
2. Folosim datele pentru a calcula valoarea unei statistici a testului, s-o numim x.
3. Dacă x este destul de departe în coada repartiţiei 0 (repartiţia presupunând ipoteza 0), atunci respingem H_0.

În cazul unde nu este o valoare specială pentru a fi testată, putem dori încă să estimăm θ. Acesta este reversul testului de semnificaţie: în loc să vedem dacă ar trebui să respingem o anumită valoare a lui θ deoarece nu se potriveşte cu datele, mai degrabă vrem să găsim un domeniu de valori ale lui θ care, într-un anumit sens, se potrivesc cu datele.
Aceasta ne dă următoarele definiţii:
Definiţie. Fiind dată o valoare x a statisticii testului, intervalul de încredere $(1-\alpha)$ conţine toate valorile θ_0 care nu sunt respinse (la nivelul de semnificaţie α) când aceste valori sunt ipoteza 0.
Definiţie. O eroare de tipul 1 CI apare când intervalul de încredere nu

conţine adevărata valoare a lui θ.
Pentru un interval de încredere $(1-\alpha)$ rata erorii de tipul 1 CI este α.
Exemplul 1. Iată un exemplu care leagă intervalele de încredere şi testele ipotezei. Presupunem că datele x provin dintr-o repartiţie binomială$(12, \theta)$ cu θ necunoscută. Fie $\alpha = 0.1$ şi creăm intervalul de încredere $(1-\alpha) = 90\%$ pentru fiecare valoare posibilă a lui x.
Strategia noastră este să considerăm pe rând câte o valoare posibilă a lui θ şi să alegem regiunile de respingere pentru testul de semnificaţie cu $\alpha = 0.1$. Odată ce acest lucru este făcut, vom şti, pentru orice valoare a lui x, ce valori ale lui θ nu sunt respinse, i.e. intervalul de încredere asociat cu x.
Pentru a începe, facem un tabel de verosimilitate pentru binomiala$(12, \theta)$ (tabelul 1). Fiecare linie arată probabilităţile $p(x|\theta)$ pentru o valoare a lui θ. Pentru a păstra mărimea rezonabilă arătăm θ doar în creşteri de 0.1.

$\theta\backslash x$	0	1	2	3	4	5	6	7	8	9	10	11	12
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00
0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.09	0.23	0.38	0.28
0.8	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.13	0.24	0.28	0.21	0.07
0.7	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.16	0.23	0.24	0.17	0.07	0.01
0.6	0.00	0.00	0.00	0.01	0.04	0.10	0.18	0.23	0.21	0.14	0.06	0.02	0.00
0.5	0.00	0.00	0.02	0.05	0.12	0.19	0.23	0.19	0.12	0.05	0.02	0.00	0.00
0.4	0.00	0.02	0.06	0.14	0.21	0.23	0.18	0.10	0.04	0.01	0.00	0.00	0.00
0.3	0.01	0.07	0.17	0.24	0.23	0.16	0.08	0.03	0.01	0.00	0.00	0.00	0.00
0.2	0.07	0.21	0.28	0.24	0.13	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00
0.1	0.28	0.38	0.23	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Tabelul 1. Tabelul de verosimilitate pentru binomiala$(12, \theta)$

Tabelele 2-4 de mai jos arată regiunea de respingere (în portocaliu) şi regiunea de nerespingere (în albastru) pentru diferite valori ale lui θ. Pentru a sublinia natura linie cu linie a procesului, tabelul 2 arată aceste regiuni pentru $\theta = 1$, apoi tabelul 3 adaugă regiunile pentru $\theta = 0.9$ şi tabelul 4 le arată pentru toate valorile lui θ.
Imediat după tabele explicăm detaliat cum au fost alese regiunile de respingere/nerespingere. "significance"="semnificaţie".

$\theta\backslash x$	0	1	2	3	4	5	6	7	8	9	10	11	12	significance
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.000
0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.09	0.23	0.38	0.28	
0.8	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.13	0.24	0.28	0.21	0.07	
0.7	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.16	0.23	0.24	0.17	0.07	0.01	
0.6	0.00	0.00	0.00	0.01	0.04	0.10	0.18	0.23	0.21	0.14	0.06	0.02	0.00	
0.5	0.00	0.00	0.02	0.05	0.12	0.19	0.23	0.19	0.12	0.05	0.02	0.00	0.00	
0.4	0.00	0.02	0.06	0.14	0.21	0.23	0.18	0.10	0.04	0.01	0.00	0.00	0.00	
0.3	0.01	0.07	0.17	0.24	0.23	0.16	0.08	0.03	0.01	0.00	0.00	0.00	0.00	
0.2	0.07	0.21	0.28	0.24	0.13	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	
0.1	0.28	0.38	0.23	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
0.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

Tabelul 2. Tabelul de verosimilitate pentru binomiala$(12, \theta)$ cu regiunile de respingere/nerespingere pentru $\theta = 1$

$\theta \backslash x$	0	1	2	3	4	5	6	7	8	9	10	11	12	significance
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.000
0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.09	0.23	0.38	0.28	0.026
0.8	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	0.13	0.24	0.28	0.21	0.07	
0.7	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.16	0.23	0.24	0.17	0.07	0.01	
0.6	0.00	0.00	0.00	0.01	0.04	0.10	0.18	0.23	0.21	0.14	0.06	0.02	0.00	
0.5	0.00	0.00	0.02	0.05	0.12	0.19	0.23	0.19	0.12	0.05	0.02	0.00	0.00	
0.4	0.00	0.02	0.06	0.14	0.21	0.23	0.18	0.10	0.04	0.01	0.00	0.00	0.00	
0.3	0.01	0.07	0.17	0.24	0.23	0.16	0.08	0.03	0.01	0.00	0.00	0.00	0.00	
0.2	0.07	0.21	0.28	0.24	0.13	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	
0.1	0.28	0.38	0.23	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
0.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

Tabelul 3. Tabelul de verosimilitate cu regiunile de respingere/nerespingere arătate pentru $\theta = 1$ şi 0.9

$\theta \backslash x$	0	1	2	3	4	5	6	7	8	9	10	11	12	significance
1.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.000
0.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.09	0.23	0.38	0.28	0.026
0.8	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.05	**0.13**	0.24	0.28	0.21	0.07	0.073
0.7	0.00	0.00	0.00	0.00	0.01	0.03	0.08	0.16	**0.23**	0.24	0.17	0.07	0.01	0.052
0.6	0.00	0.00	0.00	0.01	0.04	0.10	0.18	0.23	**0.21**	0.14	0.06	0.02	0.00	0.077
0.5	0.00	0.00	0.02	0.05	0.12	0.19	0.23	0.19	**0.12**	0.05	0.02	0.00	0.00	0.092
0.4	0.00	0.02	0.06	0.14	0.21	0.23	0.18	0.10	0.04	0.01	0.00	0.00	0.00	0.077
0.3	0.01	0.07	0.17	0.24	0.23	0.16	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.052
0.2	0.07	0.21	0.28	0.24	0.13	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.073
0.1	0.28	0.38	0.23	0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.026
0.0	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000

Tabelul 4. Tabelul de verosimilitate cu regiunile de respingere/nerespingere

Alegerea regiunilor de respingere şi nerespingere în tabele Prima problemă cu care ne confruntăm este cum să alegem exact regiunea de respingere. Am folosit 2 reguli:

1. Probabilitatea totală a regiunii de respingere, i.e. semnificaţia, trebuie să fie mai mică sau egală cu 0.1. (Deoarece avem o repartiţie discretă este imposibil să facem semnificaţia exact 0.1.)

2. Construim regiunea de respingere alegând pe rând valorile lui x, totdeauna luând valoarea nefolosită cu cea mai mică probabilitate. Ne oprim când valoarea următoare ar face semnificaţia mai mare ca 0.1.

Sunt şi alte moduri de a alege regiunea de respingere din care ar putea rezulta mici diferenţe.

Tabelul 2 arată regiunile de respingere (portocalie) şi nerespingere (albastră) pentru $\theta = 1$. Acesta este un caz special deoarece cele mai multe probabilităţi de pe această linie sunt 0. Ne mutăm la tabelul următor şi parcurgem procesul pentru acesta.

În tabelul 3, parcurgem paşii pentru a afla regiunile pentru $\theta = 0.9$.

Cea mai mică probabilitate este când $x = 0$ şi aceasta este mai mică decât 0.1, deci $x = 0$ este în regiunea de respingere.
Următoarea cea mai mică probabilitate este când $x = 1$ şi suma celor 2 probabilităţi (pentru $x = 0$ şi $x = 1$) este mai mică decât 0.1, deci $x = 1$ este în regiunea de respingere.
Continuăm cu $x = 2, 3, ..., 8$. În acest punct probabilitatea totală din regiunea de respingere este 0.026.
Următoarea cea mai mică probabilitate este când $x = 9$. Adunarea acestei probabilităţi (0.09) la 0.026 ar face probabilitatea totală peste 0.1. Deci lăsăm $x = 9$ în afara regiunii de respingere şi oprim procesul.
Observăm 3 lucruri pentru linia $\theta = 0.9$:
1. Niciuna din probabilităţile de pe această linie nu este cu adevărat 0, totuşi unele dintre ele sunt suficient de mici astfel încât aproximarea lor la 2 zecimale este 0.
2. Arătăm semnificaţia pentru această valoare a lui θ în marginea din dreapta. Mai precis, arătăm nivelul de semnificaţie al NHST cu ipoteza 0 $\theta = 0.9$ şi regiunea de respingere dată.
3. Regiunea de respingere constă din valori ale lui x. Când spunem că regiunea de respingere este arătată în portocaliu, în realitate ne referim la faptul că regiunea de respingere constă din valorile lui x corespunzătoare probabilităţilor evidenţiate cu portocaliu.
Gândiţi: Întoarceţi-vă la linia $\theta = 1$ şi asiguraţi-vă că înţelegeţi de ce regiunea de respingere este $x \in \{0, 1, ..., 11\}$ şi semnificaţia este 0.
Exemplul 2. Folosind tabelul 4, determinaţi intervalul de încredere 0.9 când $x = 8$.
Răspuns. Intervalul de încredere 90% constă din toate acele θ care nu ar fi respinse de un test al ipotezei cu $\alpha = 0.1$ când $x = 8$. Privind la tabel, intrările albastre (nerespinse) din coloana $x = 8$ corespund la $0.5 \leq \theta \leq 0.8$; intervalul de încredere este $[0.5, 0.8]$.
Observaţie. Scopul acestui exemplu este de a arăta cum sunt legate intervalele de încredere şi testele ipotezei. Deoarece tabelul 4 are doar un număr finit de valori pentru θ, răspunsul nostru este aproape, dar nu exact. Pentru această problemă am putea utiliza `R` pentru a afla că, corect la 2 zecimale, intervalul de încredere este $[0.42, 0.84]$.
Exemplul 3. Explicaţi de ce rata erorii CI de tipul 1 va fi cel mult 0.092, cu condiţia ca adevărata valoare a lui θ să fie în tabel.
Răspuns. Răspunsul scurt este că acesta este maximul semnificaţiilor pentru orice θ din tabelul 4. Mai pe larg: facem o eroare CI de tipul 1 dacă intervalul de încredere nu conţine adevărata valoare a lui θ, s-o numim $\theta_{\text{adevărată}}$. Aceasta se întâmplă exact când data x este în regiunea de respingere pentru $\theta_{\text{adevărată}}$. Probabilitatea ca aceasta să se întâmple este semnificaţia pentru

$\theta_{\text{adevărată}}$ şi aceasta este cel mult 0.092.
Observaţie. Scopul acestui exemplu este să arate cum nivelul de încredere, rata erorii CI de tipul 1 şi semnificaţia pentru fiecare ipoteză sunt legate. Ca în exemplul precedent, putem folosi `R` pentru a calcula semnificaţia pentru mult mai multe valori ale lui θ. Când facem asta aflăm că semnificaţia maximă pentru orice θ este 0.1 apărând când $\theta = 0.04524072$.

Observaţii de sinteză:

1. Începem cu o statistică de test x. Intervalul de încredere este aleator deoarece depinde de x.
2. Pentru fiecare valoare ipotetică a lui θ facem un test de semnificaţie cu nivelul de semnificaţie α alegând regiunile de respingere.
3. Pentru o valoare specifică a lui x intervalul de încredere asociat pentru θ constă din toţi θ care nu au fost respinşi pentru acea valoare, i.e. toţi θ care au x în regiunile lor de nerespingere.
4. Deoarece repartiţia este discretă nu putem atinge totdeauna nivelul exact de semnificaţie, deci intervalul de încredere al nostru este în realitate un "interval de încredere cel puţin 90%".

Exemplul 4. Deschideţi aplicaţia
`http://mathlets.org/mathlets/confidence-intervals/`. Jucaţi-vă cu aplicaţia pentru a înţelege natura aleatoare a intervalelor de încredere şi sensul încrederii ca (1 - rata erorii CI de tipul 1).

a) Citiţi ajutorul. Este scurt şi vă va ajuta să vă orientaţi în aplicaţie. Jucaţi-vă cu diferite setări ale parametrilor pentru a vedea cum afectează ele mărimea intervalelor de încredere.

b) Stabiliţi numărul experimentelor la $N = 1$. Apăsaţi butonul "Run N trials" repetat şi vedeţi că de fiecare dată când datele sunt generate intervalele de încredere sar împrejur.

c) Acum puneţi nivelul de încredere la $c = .5$. Când apăsaţi butonul "Run N trials" ar trebui să vedeţi că aproximativ 50% din intervalele de încredere includ adevărata valoare a lui μ. Valorile "Z correct" şi "t correct" ar trebui să se schimbe în consecinţă.

d) Acum puneţi numărul încercărilor la $N = 100$, cu $c = .8$. Butonul "Run N trials" va face acum 100 de încercări o dată. Doar ultimul interval de încredere va fi arătat în grafic, dar încercările sunt făcute toate şi statistica "percent correct" va fi actualizată pe baza tuturor celor 100 de încercări.
Apăsaţi butonul "Run N trials" repetat. Urmăriţi cum ratele corecte încep să conveargă la nivelul de încredere. Pentru a converge şi mai rapid, puneţi $N = 1000$.

3.5 Viziune formală a intervalelor de încredere

Reamintim: O statistică interval este un interval I_x calculat din datele x. Un interval este determinat de marginile lui inferioară şi superioară şi acestea sunt aleatoare deoarece x sunt aleatoare.
Presupunem că x provin dintr-o repartiţie cu pdf $f(x|\theta)$, unde parametrul θ este necunoscut.
Definiţie: Un interval de încredere $(1-\alpha)$ pentru θ este o statistică interval I_x astfel încât

$$P(\theta_0 \in I_x|\theta = \theta_0) = 1 - \alpha$$

pentru toate valorile posibile ale lui θ_0.
Această definiţie este un mod de a cântări dovezile produse de datele x.
Nivelul de încredere a unei statistici interval este o probabilitate privind un interval aleator şi o valoare ipotetică θ_0 pentru parametrul necunoscut. Exact, este probabilitatea ca intervalul aleator (calculat din datele aleatoare) să conţină valoarea θ_0, dat fiind că parametrul modelului este cu adevărat θ_0. Deoarece adevărata valoare a lui θ este necunoscută, statisticianul frecvenţionist defineşte intervalele de încredere 95% astfel încât probabilitatea 0.95 este validă indiferent care valoare ipotetică a parametrului este de fapt adevărată.

3.6 Comparaţie cu intervalele de probabilitate Bayesiene

Intervalele de încredere sunt o noţiune frecvenţionistă şi frecvenţioniştii nu atribuie probabilităţi ipotezelor, de exemplu valorii unui parametru necunoscut. Mai degrabă ei calculează verosimilităţi; adică, probabilităţi despre date sau statistici asociate dată fiind o ipoteză (observaţi condiţia $\theta = \theta_0$ din viziunea formală a intervalelor de încredere). Construcţia intervalelor de încredere provine în întregime din tabelul complet de verosimilitate.
În contrast, intervalele de probabilitate a posteriori Bayesiene dau într-adevăr probabilitatea ca valoarea parametrului necunoscut să fie în domeniul raportat. Adăugăm avertismentul uzual că aceasta depinde de alegerea specifică a unei a priori Bayesiene (posibil subiectivă).
Această distincţie între cele 2 este subtilă deoarece intervalele de probabilitate a posteriori Bayesiene şi intervalele de încredere frecvenţioniste împart următoarele proprietăţi:
1. Ele pleacă de la un model $f(x|\theta)$ pentru datele observate x cu parametrul necunoscut θ.
2. Date fiind datele x, ele dau un interval $I(x)$ specificând un domeniu de valori pentru θ.
3. Ele vin cu un număr (să zicem 0.95) care este probabilitatea a ceva.

În practică, mulţi oameni interpretează greşit intervalele de încredere ca intervale de probabilitate Bayesiene, uitând că frecvenţioniştii nu pun niciodată probabilităţi peste ipoteze (aceasta este analoaga greşelii că valoarea p în NHST este probabilitatea că H_0 este falsă). Dauna acestei interpretări greşite este cumva atenuată de faptul că, fiind date destule date şi o a priori rezonabilă, intervalele Bayesian şi frecvenţionist adesea se dovedesc a fi destul de similare.
Pentru un exemplu amuzant ilustrând cât de diferite pot fi, vezi primul răspuns de aici (care implică fursecuri cu ciocolată!):
`http://stats.stackexchange.com/questions/2272/whats-the-difference-between-a-.`
Acest exemplu foloseşte definiţii formale şi este în realitate despre mulţimi de încredere în locul intervalelor de încredere.