Curs 17

Cristian Niculescu

1 Testarea semnificaţiei ipotezei 0 III

1.1 Scopurile învăţării

1. Fiind date ipotezele şi datele, să poată identifica un test de semnificaţie potrivit dintr-o listă de teste uzuale.
2. Fiind date ipotezele, datele şi un test de semnificaţie sugerat, să ştie cum să caute detaliile şi să aplice testul.

1.2 Introducere

Toate testele de semnificaţie au acelaşi model de bază în proiectarea şi implementarea lor.
Proiectarea unui test al semficaţiei ipotezei 0 (NHST):
Specificăm ipoteza 0 şi ipotezele alternative.
Alegem o statistică a testului ale cărei repartiţie 0 şi repartiţie alternativă (repartiţii alternative) sunt cunoscute.
Specificăm o regiune de respingere. Cel mai adesea acest lucru este făcut implicit specificând un nivel de semnificaţie α şi o metodă pentru calculul p-valorilor bazată pe cozile repartiţiei 0.
Calculăm puterea folosind repartiţia alternativă (repartiţiile alternative).
Folosirea unui NHST:
Colectăm datele şi calculăm statistica testului.
Verificăm dacă statistica testului este în regiunea de respingere. Cel mai adesea acest lucru este făcut implicit verificând dacă $p < \alpha$. Dacă este aşa, "respingem ipoteza 0 în favoarea ipotezei alternative". Altfel concluzionăm "datele nu sprijină respingerea ipotezei 0".
Observaţi formularea prudentă: când eşuăm în a respinge H_0, nu concluzionăm că H_0 este adevărată. Eşecul respingerii poate avea alte cauze. De exemplu, poate nu avem destule date pentru a distinge clar H_0 şi H_A, în timp ce mai multe date ar indica respingerea lui H_0.

1.3 Parametrii populaţiei şi statistici de selecţie

Exemplul 1. Dacă alegem aleator 10 bărbaţi dintr-o populaţie şi le măsurăm înălţimile, spunem că am selectat înălţimile din populaţie. În acest caz media de selecţie, să spunem $\overline{x}$, este media înălţimilor selectate. Este o statistică şi îi ştim explicit valoarea. Pe de altă parte, adevărata înălţime medie a populaţiei, să spunem μ, este necunoscută şi putem doar să estimăm valoarea ei. Numim μ parametru al populaţiei.
Principalul scop al testării semnificaţiei este folosirea statisticilor de selecţie pentru a trage concluzii despre parametrii de selecţie. De exemplu, putem testa dacă înălţimea medie a bărbaţilor dintr-o populaţie dată este mai mare ca 1.78 m.

1.4 O galerie a testelor de semnificaţie uzuale legate de repartiţia normală

Toate aceste teste presupun date normale.
Repartiţiile 0 pentru aceste teste sunt toate legate de repartiţia normală prin formule explicite. Argumentele sunt accesibile oricui ştie analiză matematică şi este interesat de înţelegerea lor. Dat fiind numele oricărei repartiţii, putem căuta online detaliile construcţiei şi proprietăţilor ei. De asemenea putem folosi `R` pentru a explora repartiţia numeric şi grafic.
Când analizăm datele cu oricare din aceste teste, un lucru de importanţă cheie este să verificăm că presupunerile sunt adevărate sau cel puţin aproximativ adevărate. De exemplu, nu ar trebui să folosim un test care presupune că datele sunt normale până nu am verificat că datele sunt aproximativ normale.

1.4.1 Testul z

Utilizare: Testăm dacă media populaţiei este egală cu o medie ipotetică.
Date: $x_1, x_2, ..., x_n$.
Presupuneri: Datele sunt selecţii normale independente:

$$x_i \sim N(\mu, \sigma^2) \text{ unde } \mu \text{ este necunoscută, dar } \sigma \text{ este cunoscută.}$$

H_0 : Pentru un μ_0 specificat, $\mu = \mu_0$.
H_A:

$$\begin{aligned} \text{Bilaterală: } & \mu \neq \mu_0; \\ \text{unilaterală mai mare: } & \mu > \mu_0; \\ \text{unilaterală mai mică: } & \mu < \mu_0. \end{aligned}$$

Statistica testului: $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$.
Repartiţia 0: $f(z|H_0)$ este pdf a lui $Z \sim N(0,1)$.
valoarea p:

Bilaterală:	$p = P(\|Z\| > \|z\|)$	$=$	$2 * (1 - \texttt{pnorm}(\texttt{abs}(\texttt{z}), 0, 1))$;
unilaterală mai mare:	$p = P(Z > z)$	$=$	$1 - \texttt{pnorm}(\texttt{z}, 0, 1)$;
unilaterală mai mică:	$p = P(Z < z)$	$=$	$\texttt{pnorm}(\texttt{z}, 0, 1)$.

Codul `R`: Nu pare a fi o singură funcţie `R` pentru a folosi un tezt z. Desigur, se poate folosi `R` pentru a calcula z şi valoarea p.
Exemplul 2. Vezi exemplul 13 din cursul 16.

1.4.2 Testul t de o selecţie pentru medie

Utilizare: Testăm dacă media populaţiei este egală cu o medie ipotetică.
Date: $x_1, x_2, ..., x_n$.
Presupuneri: Datele sunt selecţii normale independente:

$$x_i \sim N(\mu, \sigma^2) \text{ unde atât } \mu \text{ cât şi } \sigma \text{ sunt necunoscute.}$$

H_0 : Pentru un μ_0 specificat, $\mu = \mu_0$.
H_A:

$$\begin{aligned} \text{Bilaterală: } & \mu \neq \mu_0; \\ \text{unilaterală mai mare: } & \mu > \mu_0; \\ \text{unilaterală mai mică: } & \mu < \mu_0. \end{aligned}$$

Statistica testului: $t = \frac{\overline{x}-\mu_0}{s/\sqrt{n}}$,
unde s^2 este dispersia de selecţie: $s^2 = \frac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2$.
Repartiţia 0: $f(t|H_0)$ este pdf a lui $T \sim t(n-1)$ (repartiţia t a lui Student cu $n-1$ grade de libertate).
Valoarea p:

Bilaterală:	$p = P(\|T\| > \|t\|)$	$=$	$2 * (1 - \texttt{pt}(\texttt{abs}(\texttt{t}), \texttt{n} - 1))$;
unilaterală mai mare:	$p = P(T > t)$	$=$	$1 - \texttt{pt}(\texttt{t}, \texttt{n} - 1)$;
unilaterală mai mică:	$p = P(T < t)$	$=$	$\texttt{pt}(\texttt{t}, \texttt{n} - 1)$.

Exemplu de cod `R`: Pentru datele $x = 1, 3, 5, 7, 2$ putem face un test t de o selecţie cu H_0: $\mu = 2.5$ folosind comenzile în `R`:
```
x=c(1,3,5,7,2)
t.test(x, mu=2.5)
```

Acestea vor da câteva informaţii, incluzând media datelor, valoarea t şi valoarea p bilaterală. Vezi ajutorul pentru această funcţie pentru alte setări ale argumentelor.
Exemplul 3. Vezi exemplul 2 din partea a 2-a a cursului 16.

1.4.3 Testul t cu 2 selecţii pentru compararea mediilor

Cazul dispersiilor egale
Începem descriind testul presupunând dispersii egale.
Utilizare: Testăm dacă mediile populaţiei din 2 populaţii diferă printr-o cantitate ipotetică.
Date: $x_1, x_2, ..., x_n$ şi $y_1, y_2, ..., y_m$.
Presupuneri: Ambele grupe de date sunt selecţii normale independente:

$$x_i \sim N(\mu_x, \sigma^2),$$
$$y_j \sim N(\mu_y, \sigma^2),$$

unde atât μ_x cât şi μ_y sunt necunoscute şi posibil diferite. Dispersia σ^2 este necunoscută, dar aceeaşi pentru ambele grupe.
H_0: Pentru un μ_0 specificat: $\mu_x - \mu_y = \mu_0$.
H_A:

Bilaterală:	$\mu_x - \mu_y \neq \mu_0$;
unilaterală mai mare:	$\mu_x - \mu_y > \mu_0$;
unilaterală mai mică:	$\mu_x - \mu_y < \mu_0$.

Statistica testului: $t = \frac{\overline{x}-\overline{y}-\mu_0}{s_p}$, unde s_x^2 şi s_y^2 sunt dispersiile de selecţie ale datelor x, respectiv y şi s_p^2 este (uneori numită) dispersia de selecţie combinată:

$$s_p^2 = \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right) \text{ şi } df = n+m-2.$$

Repartiţia 0: $f(t|H_0)$ este pdf a lui $T \sim t(df)$, repartiţia t cu $df = n+m-2$ grade de libertate.
Valoarea p:

Bilaterală:	$p = P(\lvert T\rvert > \lvert t\rvert)$	$=$	$2 * (1 - \texttt{pt}(\texttt{abs}(\texttt{t}), \texttt{df}))$;
unilaterală mai mare:	$p = P(T > t)$	$=$	$1 - \texttt{pt}(\texttt{t}, \texttt{df})$;
unilaterală mai mică:	$p = P(T < t)$	$=$	$\texttt{pt}(\texttt{t}, \texttt{df})$.

Codul `R`: Funcţia din `R` `t.test` va rula un test t cu 2 selecţii cu dispersii egale setând argumentul `var.equal=TRUE`.
Exemplul 4. Vezi exemplul 3 din a 2-a parte a cursului 16.

Observaţii. 1. Cel mai adesea testul este făcut cu $\mu_0 = 0$. Adică, ipoteza 0 este că mediile sunt egale, i.e. $\mu_x - \mu_y = 0$.
2. Dacă datele x şi y au aceeaşi lungime, n, atunci formula pentru s_p^2 devine mai simplă:

$$s_p^2 = \frac{s_x^2 + s_y^2}{n}.$$

Cazul dispersiilor inegale
Există o formă a testului t pentru cazul când dispersiile nu sunt presupuse egale. Uneori este numit testul t al lui Welch.
Acesta arată exact la fel cu excepţia unor mici schimbări în presupuneri şi în formula dispersiei combinate:
Utilizare: Testăm dacă mediile populaţiei din 2 populaţii diferă printr-o cantitate ipotetică.
Date: $x_1, x_2, ..., x_n$ şi $y_1, y_2, ..., y_m$.
Presupuneri: Ambele grupe de date sunt selecţii normale independente:

$$x_i \sim N(\mu_x, \sigma_x^2),$$
$$y_j \sim N(\mu_y, \sigma_y^2),$$

unde atât μ_x cât şi μ_y sunt necunoscute şi posibil diferite. Dispersiile σ_x^2 şi σ_y^2 sunt necunoscute şi nu sunt presupuse a fi egale.
H_0, H_A: Exact aceleaşi ca în cazul dispersiilor egale.
Statistica testului: $t = \frac{\overline{x}-\overline{y}-\mu_0}{s_p}$, unde s_x^2 şi s_y^2 sunt dispersiile de selecţie ale datelor x, respectiv y şi s_p^2 este (uneori numită) dispersia de selecţie combinată:

$$s_p^2 = \frac{s_x^2}{n} + \frac{s_y^2}{m} \text{ şi } df = \frac{(s_x^2/n + s_y^2/m)^2}{(s_x^2/n)^2/(n-1) + (s_y^2/m)^2/(m-1)}.$$

Repartiţia 0: $f(t|H_0)$ este pdf a lui $T \sim t(df)$, repartiţia t cu df grade de libertate.
Valoarea p: Exact aceeaşi ca în cazul dispersiilor egale.
Codul `R`: Funcţia din `R` `t.test` va rula un test t în acest caz setând argumentul `var.equal=FALSE`.
Testul t cu 2 selecţii în perechi
Când datele vin firesc în perechi (x_i, y_i), putem folosi testul t cu 2 selecţii în perechi (după de verificăm că presupunerile sunt valide!)
Exempul 5. Pentru a măsura eficienţa unei medicaţii de scăderea colesterolului, putem testa fiecare subiect înainte şi după tratamentul cu medicamentul. Deci pentru fiecare subiect avem o pereche de măsurători: x_i = nivelul de colesterol înainte de tratament şi y_i = nivelul de colesterol după

tratament.

Exemplul 6. Pentru a măsura eficienţa unui tratament pentru cancer putem pune în pereche fiecare subiect care a primit treatamentul cu unul care nu l-a primit. În acest caz am vrea să punem în perechi subiecţii care sunt similari în termeni de stadiu al bolii, vârstă, sex, etc.

Utilizare: Testăm dacă diferenţa medie dintre valorile perechilor dintr-o populaţie este egală cu o valoare ipotetică.

Date: $x_1, x_2, ..., x_n$ şi $y_1, y_2, ..., y_n$ trebuie să aibă aceeaşi lungime.

Presupuneri: Diferenţele $w_i = x_i - y_i$ sunt independente dintr-o repartiţie normală $N(\mu, \sigma^2)$, unde μ şi σ sunt necunoscute.

OBSERVAŢIE: Acesta este chiar un test t de o selecţie folosind w_i.

H_0 : Pentru un μ_0 specificat, $\mu = \mu_0$.

H_A:

$$\begin{aligned} \text{Bilaterală: } & \mu \neq \mu_0; \\ \text{unilaterală mai mare: } & \mu > \mu_0; \\ \text{unilaterală mai mică: } & \mu < \mu_0. \end{aligned}$$

Statistica testului: $t = \frac{\overline{w} - \mu_0}{s/\sqrt{n}}$,

unde s^2 este dispersia de selecţie: $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (w_i - \overline{w})^2$.

Repartiţia 0: $f(t|H_0)$ este pdf a lui $T \sim t(n-1)$ (repartiţia t a lui Student cu $n-1$ grade de libertate).

Valoarea p:

$$\begin{array}{lll} \text{Bilaterală:} & p = P(|T| > |t|) & = \mathtt{2 * (1 - pt(abs(t), n - 1));} \\ \text{unilaterală mai mare:} & p = P(T > t) & = \mathtt{1 - pt(t, n - 1);} \\ \text{unilaterală mai mică:} & p = P(T < t) & = \mathtt{pt(t, n - 1).} \end{array}$$

Cod `R`: Funcţia `t.test` din `R` va face un test cu 2 selecţii în perechi dacă punem argumentul `paired=TRUE`. Putem de asemenea să facem un test t de o selecţie pentru $x - y$.

Exemplul 7. Acest exemplu este din John Rice, *Mathematical Statistics and Data Analysis*, 2nd edition, p. 412.

Pentru a studia efectul fumatului asupra agregării trombocitelor, P.H. Levine (An acute effect of cigarette smoking on platelet function, *Circulation*, 48, 619-623, 1973) a luat sânge de la 11 subiecţi înainte şi după ce au fumat o ţigară şi a măsurat gradul în care trombocitele s-au agregat. Iată datele:

Before	25	25	27	44	30	67	53	53	52	60	28
After	27	29	37	56	46	82	57	80	61	59	43
Difference	2	4	10	12	16	15	4	27	9	-1	15

Ipoteza 0 este că fumatul nu a avut efect asupra agregării trombocitelor, i.e. diferenţa ar trebui să aibă media $\mu_0 = 0$. Iată codul `R` pentru un test t cu 2 selecţii pentru a testa această ipoteză:

```
before.cig=c(25,25,27,44,30,67,53,53,52,60,28)
after.cig=c(27,29,37,56,46,82,57,80,61,59,43)
t.test(after.cig,before.cig,mu=0,paired=TRUE)
```

Iată rezultatul:

```
Paired t-test

   data:  after.cig and before.cig
t = 4.2716, df = 10, p-value = 0.001633
alternative hypothesis:  true difference in means is not equal to
0
95 percent confidence interval:
4.91431 15.63114
sample estimates:
mean of the differences
10.27273
```

Obţineam aceleaşi rezultate cu testul t de o selecţie:

```
t.test(after.cig-before.cig,mu=0)
```

1.4.4 ANOVA unidirecţională (Testul F pentru medii egale)

Utilizare: Testăm dacă mediile populaţiilor din n grupe coincid.
Date (n grupe, m date din fiecare grup)

$$\begin{array}{cccc} x_{11}, & x_{12}, & \ldots, & x_{1m} \\ x_{21}, & x_{22}, & \ldots, & x_{2m} \\ & & \ldots & \\ x_{n1}, & x_{n2}, & \ldots, & x_{nm} \end{array}$$

Presupuneri: Datele pentru fiecare grup sunt independente normale din repartiţii cu medii (posibil) diferite, dar aceeaşi dispersie:

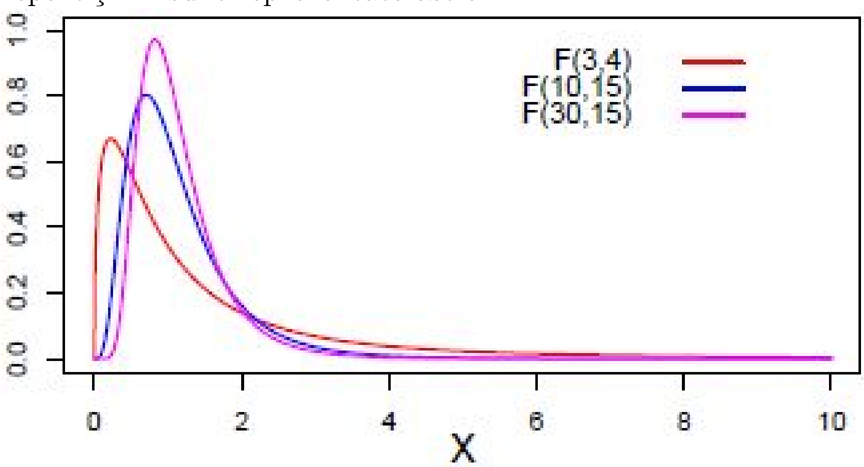
$$\begin{array}{c} x_{1j} \sim N(\mu_1, \sigma^2) \\ x_{2j} \sim N(\mu_2, \sigma^2) \\ \ldots \\ x_{nj} \sim N(\mu_n, \sigma^2) \end{array}$$

Mediile grupurilor μ_i sunt necunoscute şi posibil diferite. Dispersia σ este cunoscută, dar aceeaşi pentru toate grupurile.
H_0: Toate mediile sunt identice $\mu_1 = \mu_2 = \ldots = \mu_n$.

H_A: Nu toate mediile sunt egale.
Statistica testului: $w = \frac{MS_B}{MS_W}$, unde

$$\begin{aligned}
\overline{x}_i &= \text{ media grupului } i \\
&= \frac{x_{i1} + x_{i2} + \ldots + x_{im}}{m}. \\
\overline{x} &= \text{ media tuturor datelor.} \\
s_i^2 &= \text{ dispersia de selecţie a grupului } i \\
&= \frac{1}{m-1}\sum_{j=1}^{m}(x_{ij} - \overline{x}_i)^2. \\
MS_B &= \text{ dispersia între grupuri} \\
&= m \cdot \text{ dispersia de selecţie a mediilor grupurilor} \\
&= \frac{m}{n-1}\sum_{i=1}^{n}(\overline{x}_i - \overline{x})^2. \\
MS_W &= \text{ media dispersiilor grupurilor} \\
&= \text{ media de selecţie a } s_1^2, \ldots, s_n^2 \\
&= \frac{s_1^2 + s_2^2 + \ldots + s_n^2}{n}.
\end{aligned}$$

Ideea: Dacă μ_i sunt toate egale, acest raport ar trebui să fie aproape de 1. Dacă ele nu sunt egale, atunci MS_B ar trebui să fie mai mare, în timp ce MS_W ar trebui să rămână cam aceeaşi, deci w ar trebui să fie mai mare.
Repartiţia 0: $f(w|H_0)$ este pdf a lui $W \sim F(n-1, n(m-1))$.
Aceasta este repartiţia F cu $n-1$ şi $n(m-1)$ grade de libertate. Câteva repartiţii F sunt reprezentate astfel:



Valoarea p: $p = P(W > w) = 1 - \mathtt{pf}(\mathtt{w}, \mathtt{n}-1, \mathtt{n}*(\mathtt{m}-1))$.
Observaţii. 1. ANOVA testează dacă toate mediile sunt egale. Nu testează

dacă unele din medii sunt egale.
2. Există un test unde dispersiile nu sunt presupuse egale.
3. Există un test unde grupurile nu au toate acelaşi număr de date.
4. `R` are o funcţie `aov()` pentru teste ANOVA. Vezi:
`https://personality-project.org/r/r.guide/r.anova.html#oneway`
`http://en.wikipedia.org/wiki/F-test`.
Exemplul 8. Tabelul arată nivelul de durere perceput de pacienţi (pe o scară de la 1 la 6) după 3 proceduri medicale diferite.

T_1	T_2	T_3
2	3	2
4	4	1
1	6	3
5	1	3
3	4	5

(1) Faceţi un test F care să compare mediile acestor tratamente.
(2) Bazându-vă pe test, ce puteţi concluziona despre tratamente?
Răspuns. Folosind codul de mai jos, statistica F este 0.325 şi valoarea p este 0.729. La orice nivel rezonabil de semnificaţie nu respingem ipoteza 0 că nivelul mediu de durere este acelaşi pentru toate 3 tratamentele.
Nu este rezonabil să concluzionăm că ipoteza 0 este adevărată. Cu doar 5 date pe procedură ne poate lipsi puterea de a distinge mediile diferite.
Codul `R` pentru test

```
# DATE ----
T1=c(2,4,1,5,3)
T2=c(3,4,6,1,4)
T3=c(2,1,3,3,5)
procedure=c(rep('T1',length(T1)),rep('T2',length(T2)),rep('T3',length(T3)))
pain=c(T1,T2,T3)
data.pain=data.frame(procedure,pain)
aov.data=aov(pain~procedure,data=data.pain) # face analiza dispersiei
print(summary(aov.data))# arată tabelul rezumatului
```

Rezumatul arată o valoare p (arătată ca `Pr(>F)`) de 0.729. De aceea nu respingem ipoteza 0 că toate 3 mediile sunt egale.

1.4.5 Testul χ^2 pentru bunătatea potrivirii

Acesta este un test despre cât de bine o repartiţie de probabilitate ipotetică se potriveşte cu datele. Statistica testului este numită statistică χ^2 şi repartiţia

0 asociată statisticii χ^2 este repartiţia χ^2. Ea este notată cu $\chi^2(df)$, unde parametrul df este numărul de grade de libertate.
Presupunem că avem o funcţie masă de probabilitate necunoscută dată de următorul tabel.

Outcomes	ω_1	ω_2	...	ω_n
Probabilities	p_1	p_2	...	p_n

În testul χ^2 pentru bunătatea potrivirii ipoteza este o mulţime de valori pentru probabilităţi. Tipic, ipoteza este că probabilităţile au o repartiţie cunoscută cu anumiţi parametri, de exemplu binomială, Poisson, multinomială. Apoi testul încearcă să determine dacă această mulţime de probabilităţi ar fi putut genera rezonabil datele colectate.
Utilizare: Testăm dacă date discrete se potrivesc cu o anumită funcţie masă de probabilitate finită.
Date: Câte un număr observat O_i pentru fiecare rezultat posibil ω_i.
Presupuneri: Niciuna.
H_0: Datele provin dintr-o anumită repartiţie discretă.
H_A: Datele provin dintr-o repartiţie diferită.
Statistica testului: Datele constau din numerele observate O_i pentru fiecare ω_i. Din tabelul de probabilitate al ipotezei 0 obţinem o mulţime de numere aşteptate E_i. Sunt 2 statistici pe care le putem folosi:

$$\text{Statistica raportului de verosimilitate } G = 2\sum O_i \ln\left(\frac{O_i}{E_i}\right)$$
$$\text{Statistica } \chi^2 \text{ a lui Pearson } X^2 = \sum \frac{(O_i - E_i)^2}{E_i}.$$

Există o teoremă că sub ipoteza 0 $X^2 \approx G$ şi ambele sunt aproximativ χ^2. Înainte de calculatoare, X^2 era folosită deoarece era mai uşor de calculat. Acum, este mai bine să folosim G deşi încă veţi vedea X^2 folosită destul de des.
Grade de libertate df: Pentru testele χ^2 numărul gradelor de libertate poate fi un pic complicat. În acest caz $df = n - 1$. Este calculat ca numărul de celule care pot fi completate liber în concordanţă cu statisticile folosite pentru a calcula numerele aşteptate din celule presupunând H_0.
Repartiţia 0: Presupunând H_0, ambele statistici au (aproximativ) o repartiţie χ^2 cu df grade de libertate. Adică atât $f(G|H_0)$ cât şi $f(X^2|H_0)$ au aceeaşi pdf ca $Y \sim \chi^2(df)$.
Valoarea p:

$$\begin{aligned} p &= P(Y > G) &= 1 - \mathtt{pchisq}(\mathtt{G}, \mathtt{df}) \\ p &= P(Y > X^2) &= 1 - \mathtt{pchisq}(\mathtt{X}^2, \mathtt{df}) \end{aligned}$$

Codul `R`: Funcţia `R` `chisq.test` poate fi folosită pentru a face calcule pentru un test χ^2 folosind X^2. Pentru G, trebuie să facem calculele "cu mâna" (ajutaţi de `R`) sau să găsim un pachet care are o funcţie. (Probabil va fi numită `likelihood.test` sau `G.test`.)

Observaţie. Când statistica raportului de verosimilitate G este folosită, testul este de asemenea numit test G sau testul raportului de verosimilitate.

Exemplul 9. Primul exemplu χ^2. Presupunem că avem un experiment care produce date numerice. Pentru acest experiment rezultatele posibile sunt 0, 1, 2, 3, 4, 5 sau mai mult. Facem 51 de încercări şi numărăm frecvenţa fiecărui rezultat, obţinând următoarele date:

Outcomes	0	1	2	3	4	≥ 5
Observed counts	3	10	15	13	7	3

Presupunem că ipoteza noastră H_0 este că datele provin din 51 de încercări ale repartiţiei binomiale(8,0.5) şi ipoteza noastră alternativă H_A este că datele provin dintr-o altă repartiţie.

1. Faceţi un tabel al numerelor observate şi aşteptate.
2. Calculaţi statistica raportului de verosimilitate G şi statistica χ^2 a lui Pearson X^2.
3. Calculaţi numărul de grade de libertate ale repartiţiei 0.
4. Calculaţi valorile p corespunzătoare lui G şi X^2.

Răspuns. 1. Presupunând H_0, datele provin dintr-o repartiţie binomială(8,0.5). Avem 51 de observaţii, deci numărul aşteptat pentru fiecare rezultat este $51\cdot$ probabilitatea lui. Folosind `R` se obţine:

Rezultate	0	1	2	3	4	≥ 5
Numere observate	3	10	15	13	7	3
Probabilităţi H_0	0.00390625	0.03125	0.109375	0.21875	0.2734375	0.3632812
Numere aşteptate	0.1992188	1.59375	5.578125	11.15625	13.9453125	18.52734

2. Folosind formulele de mai sus calculăm $X^2 = 116.4056$ şi $G = 66.08122$.
3. Singura statistică folosită în calculul numerelor aşteptate a fost numărul total de observaţii 51. Deci, numărul de grade de libertate este 5, i.e. putem pune 5 dintre numerele din celule liber şi ultimul este determinat de cerinţa ca totalul să fie 51.
4. Valorile p sunt $p_G = 1 - \mathtt{pchisq}(\mathtt{G}, 5)$ şi $p_{X^2} = 1 - \mathtt{pschisq}(\mathtt{X}^2, 5)$. Ambele valori p sunt efectiv 0. Pentru aproape orice nivel de semnificaţie respingem H_0 în favoarea H_A.

Exemplul 10. (Grade de libertate.) Presupunem că avem aceleaşi date ca în exemplul precedent, dar ipoteza 0 a noastră este că datele vin din încercări independente ale repartiţiei binomiale$(8, \theta)$, unde θ poate fi oricât între 0 şi 1. H_A este că datele vin dintr-o altă repartiţie. În acest caz trebuie să estimăm

θ din date, de exemplu folosind MLE. În total am calculat 2 valori din date: numărul total de date şi estimarea lui θ. Deci, numărul de grade de libertate este $6 - 2 = 4$.

Exemplul 11. Experimentele genetice ale lui Mendel. (Adaptat din Rice, *Mathematical Statistics and Data Analysis, 2nd ed.*, exemplul C, p. 314).

În unul din experimentele lui pe mazăre, Mendel a încrucişat 556 masculi netezi, galbeni cu femele verzi zbârcite. Presupunând că genele netede şi zbârcite apar cu frecvenţă egală ne-am aştepta ca 1/4 din populaţia de mazăre să aibă 2 gene netede (SS), 1/4 să aibă 2 gene zbârcite (ss) şi 1/2 rămasă să fie heterozigoţi Ss. De asemenea aşteptăm aceste fracţii pentru gene galbene (Y) şi verzi (y). Dacă genele de culoare şi netezime sunt moştenite independent şi cele netede şi galbene sunt ambele dominante ne-am aştepta la următorul tabel de frecvenţe pentru fenotipuri.

	Yellow	Green	
Smooth	9/16	3/16	3/4
Wrinkled	3/16	1/16	1/4
	3/4	1/4	1

Tabelul de probabilitate pentru ipoteza 0

Deci, din 556 de încrucişări, numărul aşteptat de păstăi netede galbene este $556 \cdot 9/16 = 312.75$. Analog pentru celelalte posibilităţi. Iată un tabel care dă numerele observate şi aşteptate din experimentele lui Mendel.

	Observed count	Expected count
Smooth yellow	315	312.75
Smooth green	108	104.25
Wrinkled yellow	102	104.25
Wrinkled green	31	34.75

Ipoteza 0 este că numerele observate sunt selecţii aleatoare repartizate conform tabelului de frecvenţe de mai sus. Folosim numerele să calculăm statisticile noastre.

Statistica raportului de verosimilitate este

$$
\begin{aligned}
G &= 2\sum O_i \ln\left(\frac{O_i}{E_i}\right)\\
&= 2\left(315\ln\left(\frac{315}{312.75}\right) + 108\ln\left(\frac{108}{104.25}\right) + 102\ln\left(\frac{102}{104.25}\right) + 31\ln\left(\frac{31}{34.75}\right)\right)\\
&= 0.6184391.
\end{aligned}
$$

Statistica χ^2 a lui Pearson este

$$
\begin{aligned}
X^2 &= \sum \frac{(O_i - E_i)^2}{E_i}\\
&= \frac{(315-312.75)^2}{312.75} + \frac{(108-104.25)^2}{104.25} + \frac{(102-104.25)^2}{104.25} + \frac{(31-34.75)^2}{34.75}\\
&= 0.6043165.
\end{aligned}
$$

Cele 2 statistici sunt foarte apropiate. Acesta este cazul de obicei. În general statistica raportului de verosimilitate este mai robustă şi ar trebui preferată. Numărul de grade de libertate este 3, deoarece sunt 4 cantităţi observate şi o relaţie între ele: suma lor este 556. Deci, sub ipoteza 0, G are o repartiţie $\chi^2(3)$. Folosind `R` pentru a calcula valoarea p obţinem

$$p = 1 - \texttt{pchisq}(\texttt{0.6184391}, \texttt{3}) = \texttt{0.8921985}.$$

Nu vom respinge ipoteza 0 la orice nivel rezonabil de semnificaţie.
Valoarea p folosind statistica lui Pearson este 0.8954435 - aproape identică.

1.4.6 Testul χ^2 pentru omogenitate

Acesta este un test pentru a vedea dacă unele seturi independente de date provin din aceeaşi repartiţie. (Înţelesul omogenităţii în acest caz este că toate repartiţiile sunt aceleaşi.)
Utilizare: Testăm dacă m mulţimi independente de date discrete provin din aceeaşi repartiţie.
Rezultate: $\omega_1, \omega_2, ..., \omega_n$ sunt rezultate posibile. Acestea sunt aceleaşi pentru fiecare set de date.
Date: Presupunem m seturi independente de date dând numere pentru fiecare rezultat posibil. Adică, pentru setul de date i avem un număr observat O_{ij} pentru fiecare rezultat posibil ω_j.
Presupuneri: Niciuna.
H_0: Fiecare set de date provine din aceeaşi repartiţie. (Nu specificăm care este această repartiţie.)

H_A: Seturile de date nu provin din aceeaşi repartiţie.
Statistica testului: Vezi exemplul de mai jos. Sunt mn celule conţinând numerele pentru fiecare rezultat pentru fiecare set de date. Folosind Repartiţia 0 putem estima numerele aşteptate pentru fiecare din seturile de date. Statisticile X^2 şi G sunt calculate exact ca mai sus.
Numărul de grade de libertate $df = (m-1)(n-1)$. (Vezi exemplul de mai jos.)
Repartiţia 0: $\chi^2(df)$. Valorile p sunt calculate ca în testul χ^2 pentru bunătatea potrivirii.
Codul `R`: Funcţia `R` `chisq.test` poate fi folosită pentru a face calcule pentru un test χ^2 folosind X^2. Pentru G, trebuie să facem calculele "cu mâna" (ajutaţi de `R`) sau să găsim un pachet care are o funcţie. (Probabil va fi numită `likelihood.test` sau `G.test`.)
Exemplul 12. Cineva pretinde că a găsit o piesă de teatru pierdută de mult a lui William Shakespeare. Vă cere să testaţi dacă piesa a fost sau nu scrisă de Shakespeare.
Mergeţi la `http://www.opensourceshakespeare.org` şi alegeţi aleator 12 pagini din *Regele Lear* şi număraţi folosirea cuvintelor uzuale. Faceţi acelaşi lucru pentru "piesa pierdută de mult". Obţineţi următorul tabel.

Word	a	an	this	that
King Lear	150	30	30	90
Long lost work	90	20	10	80

Folosind aceste date, faceţi un test de semnificaţie pentru pretenţia că piesa pierdută de mult este de William Shakespeare. Folosiţi un nivel de semnificaţie de 0.1.
Răspuns. H_0: Pentru cele 4 cuvinte numărate piesa pierdută de mult are aceleaşi frecvenţe relative ca numerele luate din *Regele Lear*.
Totalul cuvintelor numărate în ambele piese este 500, deci estimarea de verosimilitate maximă a frecvenţelor relative presupunând H_0 este numărul total pentru fiecare cuvânt împărţit la 500.

Word	a	an	this	that	Total count
King Lear	150	30	30	90	300
Long lost work	90	20	10	80	200
totals	240	50	40	170	500
rel. frequencies under H_0	240/500	50/500	40/500	170/500	500/500

Acum, numerele aşteptate pentru fiecare piesă sub H_0 sunt numerele totale pentru acea piesă înmulţite cu frecvenţele relative din tabelul de mai sus. Următorul tabel dă numerele: (observate, aşteptate) pentru fiecare piesă. Pe ultima linie, în loc de ”249” se va citi ”240”.

Word	a	an	this	that	Totals
King Lear	(150, 144)	(30, 30)	(30, 24)	(90, 102)	(300, 300)
Long lost work	(90, 96)	(20, 20)	(10, 16)	(80, 68)	(200, 200)
Totals	(249, 240)	(50, 50)	(40, 40)	(170, 170)	(500, 500)

Statistica χ^2 este

$$\begin{aligned} X^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \frac{6^2}{144} + \frac{0^2}{30} + \frac{6^2}{24} + \frac{12^2}{102} + \frac{6^2}{96} + \frac{0^2}{20} + \frac{6^2}{16} + \frac{12^2}{68} \\ &\approx 7.904412. \end{aligned}$$

Sunt 8 celule şi toate numerele marginale sunt fixate deoarece a fost nevoie de ele pentru a determina numerele aşteptate. Am putea să punem liber valorile din 3 celule din tabel, de exemplu cele 3 celule albastre, apoi restul celulelor sunt determinate pentru a face corecte totalele marginale. Astfel, $df = 3$. (Sau ne puteam reaminti că $df = (m-1)(n-1) = 3 \cdot 1 = 3$, unde m este numărul de coloane şi n este numărul de linii.)
Folosind `R` găsim `1-pchisq(7.904412,3)=0.04802909`. Deoarece aceasta este mai mică decât nivelul nostru de semnificaţie de 0.1, respingem ipoteza 0 că frecvenţele relative ale cuvintelor sunt aceleaşi în ambele piese.
Dacă facem în plus presupunerea că toate piesele lui Shakespeare au frecvenţe similare ale cuvintelor (ceea ce este ceva ce putem verifica) concluzionăm că piesa pierdută probabil nu este a lui Shakespeare.

2 Comparaţie între deducţiile frecvenţionistă şi Bayesiană

2.1 Scopul învăţării

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

2.2 Introducere

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

2.3 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.

2.4 Critici şi apărări

2.4.1 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.

2.4.2 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.

2.4.3 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.

2.4.4 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?

2.5 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.

2.6 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

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

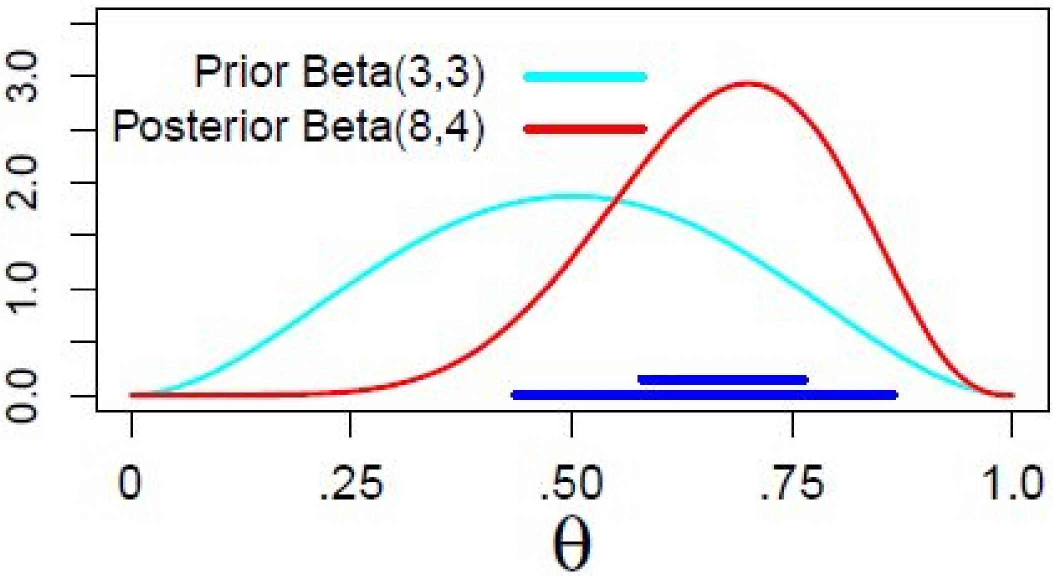
Experimentul 2: Repartiţia 0 este geometrică(0.5) deci, valoarea p unilaterală este probabilitatea a 5 sau mai multe aversuri înaintea primului revers. Folosind `R` obţinem

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

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) formeză 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), probabiltiatea a posteriori ca moneda să fie deplasată spre avers este 0.8867188.

2.7 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 absolvească 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.