Hipotézisvizsgálat.

Ez a jegyzet csak rövid áttekintést ad a Bolla–Krámli könyv Hipotézsvizsgálat című könyv 4. fejezetéről. A legfontosabb fogalmakat és eredményeket ismertetem, és egyes helyeken, ahol ezt hasznosnak tartom, teszek néhány megjegyzést az eredmények bizonyításáról is.

Az alapfeladat a következő. Megfigyelünk egy mintát, azaz független, egyforma eloszlású valószínűségi változók valamilyen $X_1, \dots, X_n$ sorozatát. Feltesszük, hogy az X_j, $j = 1, \dots, n$, valószínűségi változók az R^k Euklideszi térben veszik fel az értéküket. A könyv 4. fejezete csak a $k = 1$ esettel foglalkozik. Azt akarjuk ellenőrizni a minta alapján, hogy a megfigyelt valószínűségi változók eloszlása rendelkezik-e bizonyos tulajdonságokkal. Pontosabban megfogalmazva, van eloszlásoknak egy valamilyen módon definiált családja, és azt akarjuk ellenőrizni, hogy az általunk megfigyelt valószínűségi változók eloszlása ebbe a családba esik-e. Azt a feltételezést, hogy ez így van, nevezzük nullhipotézisnek. Van az eloszlásoknak egy másik, a feladat jellegétől függő ettől diszjunkt családja, és tudjuk, hogy a megfigyelt valószínűségi változók eloszlása e két eloszláscsalád valamelyikébe esik. Azt a feltételezést, hogy a minta elemeinek az eloszlása ebbe a másik családba esik nevezzük ellenhipotézisnek. A megfigyelt minta alapján döntést hozunk arról, hogy szerintünk teljesül-e a nullhipotézis. Ha egy minta elemeinek az eloszlása teljesíti a nullhipotézist, és mi olyan döntést, hozunk, hogy nem teljesíti azt, akkor olyan hibát követünk el, amit az irodalomban elsőfajú hibának neveznek. Azt a hibát, amit akkor követünk el, ha egy olyan minta elemeinek az eloszlásáról, amelyik nem teljesíti a nullhipoézist azt mondjuk, hogy teljesíti a nullhipotézist másodfajú hibának nevezik. Célunk természetesen az, hogy olyan döntést hozzunk, amelyiknek mind az elsőfajú mind a másodfajú hibája kicsi. De az elsőfajú hiba csökkentésének az az ára, hogy a másodfajú hiba nő. Ezért ezt a feladatot pontosabban meg kell fogalmazni.

Először néhány egyéb terminológiát vezetek be. Ha mind a null mind az ellenhipotézisben szereplő eloszlások egy Euklideszi tér részhalmazának a pontjaival vannak indexelve, akkor paraméteres próbáról beszélünk. Ha ezen eloszlások családja sokkal gazdagabb, és ezért elemeit nem lehet ilyen módon indexelni, akkor nem paraméteres próbáról beszélünk. E jegyzetben a paraméteres próbákról szóló eredményeket ismertetem, bár a könyv tartalmaz néhány eredményt nem paraméteres próbákról is. Egy teszt abból áll, hogy a minta lehetséges értékeinek a halmazát két diszjunkt részre egy

$\mathcal{X}_e$ elfogadási és egy $\mathcal{X}_k$ kritikus halmazra osztjuk. Ha a minta az $\mathcal{X}_e$ halmazba esik, akkor elfogadjuk, ha az $\mathcal{X}_k$ halmazba esik, akkor elutasítjuk a nullhipotézist. Bizonyos esetekben ennél kissé általánosabb eljárást alkalmazunk. Nevezetesen, a minta bizonyos értékei esetén, egy a mintától független kisérletet teszünk, és ezen kisérlet eredménye alapján p valószínűséggel elfogadjuk, és $1-p$ valószínűséggel elutasítjuk a nullhipotézist. Az, hogy milyen valószínűséggel utasítjuk el a nullhipotézist az $\mathcal{X}_e$ és $\mathcal{X}_k$ halmazok megválasztásával, függ attól, hogy a nullhipotézis melyik eleme következett be. Akkor mondjuk, hogy a próba ereje ε, ha a legkellemetlenebb esetben is az elsőfajú hiba értéke nem nagyobb, mint ε. Formálisan, jelölje $\{P_\vartheta,\ \vartheta\in\Theta\}$, a nullhipotézis által kijelölt eloszlások családját, ahol Θ valamilyen paraméteter tartomány. Akkor mondjuk, hogy az $\mathcal{X}_e$ és $\mathcal{X}_k$ halmazok által meghatározott próba ereje ε, ha $\sup_{\vartheta\in\Theta} P_\vartheta((X_1,\dots,X_n)\in\mathcal{X}_k)=\varepsilon$. A könyv az ilyen tulajdonságú próbákat ε terjedelmű próbáknak nevezi. A tipikus statisztikai feladatok azt jelentik, hogy rögzítünk egy ε hibahatárt, és olyan próbát keresünk, amelynek az ereje kisebb vagy egyenlő ε, és a másodfajú hibája (ami természetesen függ attól, hogy az ellenhipotézis melyik eloszlása jelenik meg) viszonylag kicsi.

A könyv 4. fejezetének az első fontos eredménye a Neyman–Pearson lemma. Ez azzal a speciális esettel foglalkozik, amikor mind a nullhipotézis mind az ellenhipotézis egyetlen elemből, egy P_{ϑ_0} illetve P_{ϑ_1} valószínűségi mértékből áll. Rögzítünk egy $\varepsilon>0$ számot. A Neyman–Pearson lemma megadja azt, hogy hogyan néz ki az a próba, amelynek elsőfajú hibája kisebb vagy egyenlő ε, és eme megkötés mellett a másodfajú hibája a lehető legkisebb. Ezen eredmény megfogalmazása érdekében bevezetünk néhány jelölést.

Jelölje $\mathbf{X}=(X_1,\dots,X_n)$ a megfigyelt mintát. Legyen adva egy olyan μ mérték, amelyre nézve mind a P_{ϑ_0} mind a P_{ϑ_1} mérték abszolút folytonos, és legyen a sűrűségfüggvényük ezen μ mérték szerint $p_{\vartheta_0}(x)$ és $p_{\vartheta_1}(x)$. Jelölje $L_{\vartheta_0}(\mathbf{X})=\prod_{j=1}^n p_{\vartheta_0}(X_j)$ a P_{ϑ_0} és $L_{\vartheta_1}(\mathbf{X})=\prod_{j=1}^n p_{\vartheta_1}(X_j)$ a P_{ϑ_1} mérték szerinti likelihoodfüggvényét az $\mathbf{X}=(X_1,\dots,X_n)$ mintának.

A $\mathbf{X}=(X_1,\dots,X_n)$ minta egy csak 0 vagy 1 értéket felvevő $\psi(\mathbf{X})$ függvényét próbafüggvénynek nevezzük, amely a következő próbát határozza meg. A nullhipotézist fogaduk el abban az esetben, ha $\psi(\mathbf{X})=0$, és az ellenhipotézist fogadjuk el abban az esetben, ha $\psi(\mathbf{X})=1$. Érdemes a próbafüggvényt kissé általánosabban definiálni, és olyan definiciót bevezetni, amelyben bizonyos minták megjelenése esetén $1-p$ valószínűséggel a nullhipotézist, és p valószínűséggel az ellenhipotézist fogadjuk el. Ennek alapján, legyen $0\le\psi(\mathbf{X})\le 1$, és $\psi(\mathbf{X})=p$ esetén $1-p$ valószínűséggel a nullhipotézist és p valószínűséggel az ellenhipotézist fogadjuk el (egy a mintától független kisérlet alapján). Egy próba definiciója a $\psi(\mathbf{X})$ próbafüggvény megadását jelenti, amelynek segítségével a fent megadott módon hozzuk meg a döntésünket. Egy $\psi(\mathbf{X})$ próbafügvény által meghatározott próba elsőfajú hibája $E_{\vartheta_0}(1-\psi(\mathbf{X}))$, másodfajú hibája $E_{\vartheta_1}\psi(\mathbf{X})$. A most bevezetett fogalmak és jelölések segítségével megfogalmazzuk a Neyman–Pearson lemmát.

Neyman–Pearson lemma. *Legyen adva egy $\mathbf{X}=(X_0,\dots,X_n)$ minta a nullhipotézis szerint P_{ϑ_0} az ellenhipotézis szerint P_{ϑ_1} eloszlással, $P_{\vartheta_0}\neq P_{\vartheta_1}$. Minden $0<\varepsilon<1$ számhoz létezik olyan $c=c(\varepsilon)>0$ és $p=p(\varepsilon)$, $0\le p<1$, szám, hogy a segítségükkel definiált $\psi(\mathbf{X})=\psi_{c,p}(\mathbf{X})$ próbafüggvény, amelyet a*

$$\psi(\mathbf{X})=\begin{cases} 0 & ha\ \frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}<c, \\ p & ha\ \frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}=c, \\ 1 & ha\ \frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}>c. \end{cases}$$

képlet definiál teljesíti a $E_{\vartheta_0}(1-\phi(\mathbf{X}))=\varepsilon$ azonosságot, azaz a $\psi(\mathbf{X})$ próbafüggvény által megadott próba elsőfajú hibája ε. Másrészt, ha egy $\psi'(\mathbf{X})$ próbafüggvény által meghatározott próbának az elsőfajú hibája kisebb vagy egyenlő ε, akkor e próba másodfajú hibája nagyobb vagy egyenlő, mint a $\psi(\mathbf{X})$ próbafüggvény által meghatározott próba másodfajú hibája. Képletben kifejezve: Ha $E_{\vartheta_0}(1-\psi'(\mathbf{X}))\le\varepsilon$, akkor $E_{\vartheta_1}\psi'(\mathbf{X})\ge E_{\vartheta_1}\psi(\mathbf{X})$.

Eltekintek e lemma bizonyításától, csak egy rövid heurisztikus magyarázatot adok arra, hogy miért igaz ez az eredmény. Ha adva van két $\mathbf{X}=(X_1,\dots,X_n)$ és $\mathbf{X}'=(X_1',\dots,X_n')$ minta, amelyekre $\frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}\le\frac{L_{\theta_1}(\mathbf{X}')}{L_{\theta_0}(\mathbf{X}')}$, akkor az $\mathbf{X}$ minta megjelenése esetén inkább döntenénk a nullhipotézis mellett, mint a $\mathbf{X}'$ minta megjelenése esetén. Ez azt sugallja, hogy olyan próbát válasszunk, amelyben egy alkalmas $c>0$ konstanst választunk, és $\frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}<c$ esetén a nullhipotézis, $\frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}>c$ esetén az ellenhipotézis érvényessége mellett döntünk. A c konstanst úgy akarjuk választani, hogy az elsőfajú hiba pontosan ε legyen. Lehetséges, hogy a választandó c szám olyan lesz, hogy a $P_{\vartheta_0}\left(\frac{L_{\theta_1}(\mathbf{X})}{L_{\theta_0}(\mathbf{X})}=c\right)$ valószínűség pozitív. Ezen esemény bekövetkezte esetén véletleníteni kell, bizonyos valószínűséggel a nullhipotézist, és bizonyos valószínűséggel az ellenhipotézist kell választani annak érdekében, hogy az elsőfajú hiba pontosan ε legyen.

Tekintünk statisztikai modelleket a következő tulajdonságokkal. A H_0

nullhipotézis olyan P_ϑ valószínűségi mértékekből áll amelyek ϑ indexére $\vartheta \in \Theta_0$ valamely Θ_0 indexhalmazzal, a H_1 ellenhipotézis olyan P_ϑ valószínűségi mértékekből áll, amelyekre $\vartheta \in \Theta_1$ egy másik Θ_1 indexhalmazzal, ezek az indexhalmazok diszjunktak, azaz $\Theta_0 \cap \Theta_1 = \emptyset$. Továbbá ezeknek a mértékeknek van közös domináló mértékük, azaz bevezetve a $\Theta = \Theta_0 \cup \Theta_1$ jelölést, van olyan μ mérték, hogy minden P_ϑ, $\vartheta \in \Theta$, mértéknek van $p_\vartheta(\cdot)$ sűrűségfüggvénye e szerint a μ mérték szerint. A könyv bevezet ilyen modellekben egy próbát, amit likelihood hányados próbának nevez. Ez egy plauzibilis eljárás, és bár a könyv nem ismertet olyan eredményeket, amelyek e módszer előnyös tulajdonságairól szólnak, mégis érdemes ezt ismertetni. A könyv több olyan statisztikai feladatot tárgyal, ahol a jó statisztikai próbát úgy találjuk meg, mint az adott modellben megjelenő likelihood hányados próbát.

Vezessük be az $L_\vartheta(\mathbf{X}) = \prod_{j=1}^n p_\vartheta(X_j)$ likelihood függvényt minden $\mathbf{X} = (X_1, \dots, X_n)$ (n-elemű) mintára és P_ϑ, $\vartheta \in \Theta$, valószínűségi mértékre. Definiáljuk a

$$\lambda_n(\mathbf{X}) = \frac{\sup_{\vartheta \in \Theta_0} L_\vartheta(\mathbf{X})}{\sup_{\vartheta \in \Theta} L_\vartheta(\mathbf{X})}$$

kifejezést. Ekkor $0 \le \lambda_n(\mathbf{X}) \le 1$, és azt várjuk, hogy természetes modellekben, akkor lesz ez a kifejezés nagy, azaz közel 1 nagy valószínűséggel, ha $\vartheta \in \Theta_0$. Ugyanis $L_{\vartheta'}(\mathbf{X}) = \exp\left\{\sum_{j=1}^n \log p_{\vartheta'}(X_j)\right\} \sim \exp\{nE_\vartheta \log p_{\vartheta'}(X_1)\}$ minden $\vartheta' \in \Theta$ paraméterre a nagy számok törvénye alapján, és azt várjuk, hogy természetes modellekben, abban az esetben, ha az $\mathbf{X} = (X_1, \dots, X_n)$ minta elemei P_ϑ eloszlásúak, és összehasonlítunk két $\vartheta \in \Theta$ és $\vartheta' \in \Theta$, $\vartheta \ne \vartheta'$, paraméterhez tartozó P_ϑ és $P_{\vartheta'}$ mértéket, akkor $E_\vartheta \log p_\vartheta(X_1) > E_\vartheta \log p_{\vartheta'}(X_1)$. Ez azt sugallja, hogy akkor fogadjuk el a nullhipotézist, ha $\lambda_n(\mathbf{X}) \ge c$, ahol a $c = c(\varepsilon)$ konstanst úgy választjuk, hogy az elsőfajú hiba ε legyen. A most leírt módszer alapján kapott próbákat hívják likelihood hányados próbának.

A Neyman–Pearson lemmában a következő problémával foglalkoztunk. Megfigyelünk egy mintát, azaz független egyforma eloszlású valószínűségi változók egy sorozatát, amelyiknek elemeiről tudjuk, hogy az eloszlásuk két ismert eloszlás valamelyike, de nem tudjuk melyik. A minta ismerete alapján erről akarunk dönteni. A Neyman–Pearson lemmában megadtuk, hogy melyik az a döntési eljárás, amelyben megköveteljük, hogy az elsőfajú hiba (az a hiba, hogy abban az esetben, amikor a minta az első eloszlásból jött, mi mégis úgy itéljük meg, hogy az a második eloszlásból jött) valószínűsége egy előírt számnál kisebb legyen, és a másodfajú hiba (az a hiba, hogy abban az esetben, amikor a minta a második eloszlásból jött, mi mégis úgy itéljük meg, hogy az az első eloszlásból jött) valószínűsége eme feltétel teljesülése esetén a lehető legkisebb. Most ennek a problémának egy változatát tárgyaljuk.

Ebben a problémában is két ismert eloszlás van, és el akarjuk dönteni, hogy egy minta elemei e két eloszlás melyikéből származnak. Most is meg tudjuk figyelni a minta elemeit, de azok elemszáma nem egy fix szám, hanem elkezdjük a minta elemeit figyelni, és mi döntjük el, hogy mikor hagyjuk abba a megfigyelést. Természetesen minél több elemet figyelünk meg, annál jobb döntést tudunk hozni, jó döntési eljárás esetén annál kisebb elsőfajú és másodfajú hibát fogunk elkövetni. Most olyan módszert kívánunk kidolgozni, amelyikben mind az elsőfajú mind a másodfajú hiba kisebb, mint egy előírt szám. Ezt azáltal tudjuk elérni, hogy sokáig figyeljük a minta elemeit. Viszont a célunk az, hogy a megfigyelt minta (véletlen) elemszáma legyen minél kisebb. Ezt az elvárást pontosítani kell. Mikor tekintünk egy véletlen elemszámot kicsinek? Azt várjuk el, hogy a megfigyelt elemek számának a várható értéke legyen kicsi. De még ezzel az elvárással sem fogalmaztuk meg a feladatot pontosan. Milyen eloszlás szerinti várható értéket kívánunk kicsivé tenni? Két természetes jelölt van. Az a két eloszlás, amelyek valamelyike megegyezik a megfigyelt minta elemeinek az eloszlásával. Az alább ismertetett eredményből ki fog derülni, hogy egy olyan eljárást tudunk kidolgozni, amely szerint mind a két mérték szerint minimális a várható értéke a javasolt döntési módszerben megfigyelt minta elemszámának azon döntési módszerekben megfigyelt minta elemszáma között, amelyekben mind az elsőfajú mind a másodfajú hiba kisebb, mint ebben a módszerben.

Ebben a feladatban úgynevezett szekvenciális eljárásokkal foglalkozunk. Minden egyes mintaelem megjelenése után (a nulladik mintaelemtől kezdve) döntést hozunk, és az háromféle lehet.

1.) Úgy döntünk, hogy elfogadjuk a nullhipotézist, és befejezzük a minta megfigyelését.

2.) Úgy döntünk, hogy elutasítjuk a nullhipotézist, és befejezzük a minta megfigyelését.

3.) Úgy döntünk, hogy megfigyeljük a következő mintaelem értékét is.

Az, hogy az n-ik lépésben az 1.), 2.) és 3.) lehetőségek közül melyiket választjuk az $\mathbf{X}_n = (X_1, \dots, X_n)$ véletlen vektor értékétől függ. Részletesebben megfogalmazva, jelölje N azt a véletlen időpontot, amikor befejezzük a mintaelemek megfigyelését. Tekintsük az $\{\omega\colon N(\omega) \geq n-1\}$ eseményt, amely az $\mathbf{X}_n = (X_1, \dots, X_n)$, (valójában csak az $\mathbf{X}_{n-1} = (X_1, \dots, X_{n-1})$) véletlen vektortól függ, és vegyük ennek a halmaznak egy olyan három A, B és C halmazból álló particióját, amelyikre ezen három halmaz mindegyike mérhető a $\sigma(X_1, \dots, X_n)$ σ-algebra szerint. A szekvenciális eljárás n-ik lépése abból áll, hogy definálunk három olyan A, B és C halmazt, amelyek teljesítik a fenti feltételeket, és az A halmaz bekövetkezése esetén az 1.), a B halmaz bekövetkezése esetén a 2.), a C halmaz bekövetkezése esetén a 3.) lehetőséget választjuk.

Célunk olyan szekvenciális eljárás kidolgozása, amelyikben mind az elsőfajú mind a másodfajú hiba kisebbb, mint egy előírt szám, és a megfigyelt mintaelemek számának a várható értéke (a minta elemszámainak mindkét lehetséges eloszlása szerint) viszonylag kicsi.

Bevezetem a Wald-féle valószínűséghányados próbákat, amelyek szekvenciális próbák, és tartalmaznak bizonyos paramétereket. Ezek alkalmas megválasztásával megkapjuk a kívánt tulajdonságú kis hibájú, nem túl nagy számú megfigyelést tartalmazó tesztelési eljárásokat. Annak érdekében, hogy ezeket ismertessem, először be kell vezetni néhány jelölést.

Legyen adva egy $(\Omega, \mathcal{A}.P)$ valószínűségi mezőn független, egyforma eloszlású valószínűségi változók egy $X_1, X_2, \dots$, sorozata, amelyeknek a sűrűségfüggvénye egy μ domináló mértékre nézve a H_0 nullhipotézis szerint egy (ismert) $f_{\vartheta_0}(\cdot)$ függvény, a H_1 ellenhipotézis szerint pedig egy (ismert) $f_{\vartheta_1}(\cdot)$ függvény. Feltesszük, hogy ez a két sűrűségfüggvény különböző. Definiáljuk minden $n = 1, 2, \dots$ indexre és $\mathbf{X}_n = (X_1, \dots, X_n)$ mintára az $L_{\theta_0}(\mathbf{X}_n) = \prod_{j=1}^n f_{\vartheta_0}(X_j)$ és $L_{\theta_1}(\mathbf{X}_n) = \prod_{j=1}^n f_{\vartheta_1}(X_j)$ likelihood függvényeket, valamint a

$$V_n = V_n(\mathbf{X}_n) = \frac{L_{\vartheta_1}(\mathbf{X}_n)}{L_{\vartheta_0}(\mathbf{X}_n)}, \quad n = 1, 2, \dots$$

valószínűségi változókat. Rögzítsünk két $0 < A < 1 < B < \infty$ számot, és definiáljuk segítségükkel a következő szekvenciális próbát. Ezt nevezzük Wald-féle valószínűséghányados próbának. Az $n = 1, 2, \dots,$ indexre egymás után a következő eljárást követjük egészen addig a (véletlen) indexig, amikor úgy döntünk, hogy befejezzük a minta megfigyelését.

1.) Ha $V_n(\mathbf{X}_n) \leq A$, akkor úgy döntünk, hogy elfogadjuk a nullhipotézist, és befejezzük a minta megfigyelését.

2.) Ha $V_n(\mathbf{X}_n) \ge B$, akkor úgy döntünk, hogy elutasítjuk a nullhipotézist, és befejezzük a minta megfigyelését.

3.) Ha $A < V_n(\mathbf{X}_n) < B$, akkor úgy döntünk, hogy megfigyeljük a következő mintaelem értékét is, és folytatjuk az eljárást.

A Wald-féle valószínűséghányados próba javasolása hipotézisek vizsgálatára hasonló elven működik mint a Neyman–Pearson lemma indoklása. Ha a $V_n = V_n(\mathbf{X}_n) = \frac{L_{\vartheta_1}(\mathbf{X}_n)}{L_{\vartheta_0}(\mathbf{X}_n)}$ hányados kicsi, akkor valószínű, hogy a nullhipotézis teljesül, tehát érdemes elfogadni azt, ha ez a hányados nagy, akkor valószínűleg az ellenhipotézis teljesül, tehát érdemes a nullhipotézist elutasítani. Ha ez a hányados se nem túl nagy se nem túl kicsi, akkor érdemes új mintaelemet kérni, hogy nagyobb biztonsággal tudjunk jó döntést hozni. Ha elég sok mintaelemet figyelünk meg, akkor olyan állapotba jutunk, hogy tudunk döntést hozni. Ennek indoklására elég megmutatni, hogy

$$\lim_{n\to\infty} \log V_n(\mathbf{X}_n) = -\infty \quad 1 \text{ valószínűséggel},$$

ha a nullhipotézis teljesül, és

$$\lim_{n\to\infty} \log V_n(\mathbf{X}_n) = \infty \quad 1 \text{ valószínűséggel},$$

ha az ellenhipotézis teljesül.

Ezen álllítás következik a nagy számok törvényéből, ha azt a $\log V_n(\mathbf{X}_n) = \sum_{j=1}^n \log \frac{f_{\vartheta_1}(X_j)}{f_{\vartheta_0}(X_j)}$ összegre alkalmazzuk, és megmutatjuk, hogy

$$E_{\vartheta_1} \log \frac{f_{\vartheta_1}(X_j)}{f_{\vartheta_0}(X_j)} > 0, \quad \text{és} \quad E_{\vartheta_0} \log \frac{f_{\vartheta_1}(X_j)}{f_{\vartheta_0}(X_j)} < 0.$$

A 4.1. tétel egy olyan eredményt fogalmaz meg, amelyből következik, hogy a Wald-féle valószínűséghányados próbákban a megfigyelt elemek számának a várható értéke minden $0 < A < 1 < B < \infty$ paraméterválasztás esetén véges mind a nullhipotézis mind az ellenhipotézis mértéke szerint, sőt ugyanez elmondható a megfigyelt elemek számának tetszőleges momentumáról. A 4.2. tétel, az ún. Wald-azonosság egy olyan azonosságot fogalmaz meg, amely lehetőséget ad ezen várható érték aszimptotikájának a meghatározására. (Lásd a könyv 4. fejezetének (4.3) képletét.) E jegyzetben csak a 4.2. tételt fogalmazom meg. Ezután beszélek e tétel bizonyításának a hátteréről.

4.2. Tétel. Wald azonosság. *Legyenek $Z_1, Z_2, \dots$ független egyforma eloszlású valószínűségi változók, amelyekre $E|Z_1| < \infty$, N megállási szabály, amelyre $EN < \infty$. Ekkor*

$$E\left(\sum_{j=1}^{N} Z_j\right) = ENEZ_1.$$

Ez az azonosság tipikus martingálokról szóló eredmény. Az

$$U_n = \sum_{j=1}^{n}(Z_n - EZ_n), \quad n = 0, 1, 2, \dots,$$

sorozat martingál. Ezért szép N megállási szabályokra $EU_N = 0$. Ez igaz például akkor, ha olyan megállási szabályt tekintünk, amelynek értéke egy valószínűséggel kisebb egy adott számnál, például az $N_k = \min(N, k)$ megállási szabályt tetszőleges $k = 1, 2, \dots$ számmal. A tétel feltételelei lehetővé teszik, hogy $k \to \infty$ határátmenettel megkapjuk az $EU_N = 0$ azonosságot. Innen következik a Wald azonosság.

Érdemes a 4.2. tételt a $Z_j = \log \frac{f_{\vartheta_1}(X_j)}{f_{\vartheta_0}((X_j)}$, $j = 1, 2, \dots$, választással alkalmazni a P_{ϑ_0} vagy P_{ϑ_1} mértékkel, továbbá azt az N megállási szabályt venni, amelyiket a Wald-féle valószínűséghányados próbában alkalmaztunk. Ez a most bevezetett Z_j valószínűségi változókkal úgy is megadható, mint a legkisebb olyan n szám, amelyre vagy $\sum_{j=1}^{n} Z_j \le \log A$ vagy $\sum_{j=1}^{n} Z_j \ge \log B$. Ilyen választással kapjuk a (4.3) relációt $a = \log A$ és $b = \log B$ jelöléssel. (Ha abban az $N = n$ időpontban, amikor a $\sum_{j=1}^{n} Z_n$ összeg kilép az (a, b) intervallumból ennek az összegnek az értéke mindig pontosan a vagy b volna, akkor pontos azonosságot írhatnánk. Az általános esetben csak közelítő azonosság érvényes. Megjegyzem, hogy a gyakorlatban érdekes esetekben ε nagyon kicsi, ezért az $|a|$ és b számokat nagyra kell választani. Ezért van a (4.3) formulában közelítő azonosság.

A 4.3. tételben megadnak egy jó becslést arra, hogy hogyan kell az A és B konstansokat választani a Wald-féle valószínűséghányados próbában, ha azt akarjuk, hogy az elsőfajú hiba ε_1, a másodfajú hiba ε_2 legyen. Pontos értéket nem tudunk megadni, de az ismertetett eredmények gyakorlati szempontból elegendőek. Ismertetem ezeket az eredményeket, és megadom a vázlatos bizonyításukat.

4.3. Tétel. *A Wald-féle valószínűséghányados próbában a $0 < \varepsilon_1 < 1$ elsőfajú és $0 < \varepsilon_2 < 1$ másodfajú hibája valamint a próba definiciójában szereplő A és B konstansok között a következő reláció érvényes:*

$$A \geq \frac{\varepsilon_2}{1-\varepsilon_1} \quad \text{és} \quad B \leq \frac{1-\varepsilon_2}{\varepsilon_1}.$$

A fenti tételnek van a következő hasznos következménye.

A 4.3. Tétel következménye. *Adva az ε_1 és ε_2 jelöltek az elsőfajú és másodfajú hibára definiáljuk segítségükkel a következő A' és B' konstansokat, amelyek választása a Wald-féle valószínűséghányados próbában az ún. $(\varepsilon_1, \varepsilon_2)$ stratégiát adja.*

$$A' = \frac{\varepsilon_2}{1-\varepsilon_1} \quad \text{és} \quad B' = \frac{1-\varepsilon_2}{\varepsilon_1}.$$

Az A' és B' konstanssal konstruált Wald-féle valószínűséghányados próba ε_1' elsőfajú és ε_2' másodfajú hibája teljesíti az $\varepsilon_1' + \varepsilon_2' \leq \varepsilon_1 + \varepsilon_2$ valamint az

$$\varepsilon_2' \leq \frac{\varepsilon_2}{1-\varepsilon_1} \quad \text{és} \quad \varepsilon_1' \leq \frac{\varepsilon_1}{1-\varepsilon_2}$$

egyenlőtlenségeket.

A 4.3. tétel bizonyítása. Jelölje N azt a véletlen időpontot, amikor befejezzük a minta megfigyelését. Ekkor $P_{\vartheta_0}(N) < \infty) = 1$, $P_{\vartheta_1}(N) < \infty) = 1$, és

$$E_N = \left\{ \omega\colon \frac{L_{\vartheta_1}(\mathbf{X}_N(\omega))}{L_{\vartheta_0}(\mathbf{X}_N(\omega))} \leq A \right\}$$

az a halmaz, ahol a nullhipotézist elfogadjuk, és

$$K_N = \left\{ \omega\colon \frac{L_{\vartheta_1}(\mathbf{X}_N(\omega))}{L_{\vartheta_0}(\mathbf{X}_N(\omega))} \geq B \right\}$$

az a halmaz, ahol a nullhipotézist elutasítjuk.

Ezért $P_{\vartheta_0}(E_N) = 1-\varepsilon_1$, $P_{\vartheta_0}(K_N) = \varepsilon_1$, $P_{\vartheta_1}(E_N) = \varepsilon_2$ és $P_{\vartheta_1}(K_N) = 1-\varepsilon_2$. Továbbá, $L_{\vartheta_1}(\mathbf{X}_N(\omega)) \geq BL_{\vartheta_0}(\mathbf{X}_N(\omega))$ a K_N halmazon, és integrálva ezen egyenlőtlenség mindkét oldalát a domináló μ mérték szerint azt kapjuk, hogy $P_{\vartheta_1}(K_N) \geq BP_{\vartheta_0}(K_N)$. Hasonlóan, $P_{\vartheta_1}(E_N) \leq AP_{\vartheta_0}(E_N)$. Ez azt jelenti, hogy $B\varepsilon_1 \leq 1-\varepsilon_2$, és $\varepsilon_2 \leq A(1-\varepsilon_1)$, és ezt kellett bizonyítani.

A következő Wald–Wolfowitz tétel bizonyítása érdekes gondolatokat tartalmaz. Ezért, bár a részletes bizonyítást nem dolgozom ki, elmagyarázom annak legfontosabb lépéseit. A tétel azt mondja ki, hogy a Wald-féle valószínűséghányados próbák optimálisak a következő értelemben. Tekintsünk egy Wald-féle valószínűséghányados próbát valamilyen A és B paraméterrel. Ha egy másik szekvenciális próbának mind az elsőfajú mind a másodfajú hibája nem nagyobb, mint ennek a próbának, akkor a próbában felhasznált mintaelemek számának a várható értéke nagyobb vagy egyenlő, mint a tekintett Wald-féle valószínűséghányados próbában. Ez igaz mind a P_{ϑ_0} mind a P_{ϑ_1} mérték szerint vett várható értékre.

4.4. Tétel. Wald és Wolfowitz tétele. *Rögzítsünk egy $0 < A < 1 < B < \infty$ számpárt, és tekintsük az általa meghatározott Wald-féle valószínűséghányados próbát. Legyen e próba elsőfajú hibája ε_1, és másodfajú hibája ε_2. Jelölje a próba során megfigyelt elemek számát N. Tekintsünk egy másik szekvenciális próbát, amelyiknek az elsőfajú hibája kisebb vagy egyenlő ε_1, és másodfajú hibája kisebb vagy egyenlő ε_2. Jelölje az e próba során megfigyelt elemek számát N'. Legyen $E_{\vartheta_0}(N') < \infty$ és $E_{\vartheta_1}(N') < \infty$. Ekkor $E_{\vartheta_0}(N) \le E_{\vartheta_0}(N')$, és $E_{\vartheta_1}(N) \le E_{\vartheta_1}(N')$.*

E feladat megoldása érdekében tekintünk egy alkalmasan megfogalmazott optimális döntési eljárásról szóló feladatot, és vizsgáljuk annak megoldását. Ez nagy segítséget nyújt Wald és Wolfowitz tételének a bizonyításában. Ezt a feladatot a valószínűségszámítás klasszikus nyelvén fogalmazom meg, amely jobban megvilágítja a problémát, mint a könyv statisztika nyelvén megfogalmazott tárgyalása.

Az optimális döntési eljárásról szóló feladatban két független, egyforma eloszlású $Y_1, Y_2, \dots$ és $Z_1, Z_2 \dots$ valószínűségi változokból álló sorozatot tekintünk egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Az $Y_1, Y_2, \dots$ sorozat elemeinek az eloszlása P_{ϑ_0}, a Z_1, Z_2,... sorozat elemeinek az eloszlása P_{ϑ_1}. Ezt a két P_{ϑ_0} és P_{ϑ_1} mértéket a továbbiakban rögzítjük. Adjunk meg ezenkívül egy $0 < \pi < 1$ számot, és legyen adva egy olyan U valószínűségi változó, amelyre $P(U = 0) = 1 - \pi$, $P(U = 1) = \pi$, és U független az $Y_1, Y_2, \dots,$ és $Z_1, Z_2 \dots$ sorozatoktól. Legyenek ezenkívül rögzítve bizonyos $c > 0$, $w_1 > 0$ és $w_2 > 0$ számok, és definiáljuk e paraméterek segítségével a következő feladatot.

Van egy rendszer, amelyik a 0 állapotba kerül, ha $U = 0$, és az 1 állapotba kerül, ha $U = 1$. Ha a rendszer a 0 állapotban van, akkor kibocsájtja egymás után az $Y_1, Y_2, \dots$ valószínűségi változókat, ha a rendszer az 1 állapotban van, akkor kibocsájtja egymás után a $Z_1, Z_2, \dots$ valószínűségi változókat.

Ismerjük a π számot, és a P_{ϑ_0} és P_{ϑ_1} eloszlásokat, de nem ismerjük az U, $Y_1, Y_2, \ldots$ és $Z_1, Z_2, \ldots$ valószínűségi változók értékeit. Lehetőségünk van megfigyelni a rendszer által egymás után kibocsájtott valószínűségi változók értékét, azaz az $Y_1, Y_2, \ldots$ sorozat egymás utáni elemeit, ha a rendszer a 0 állapotban van, és a $Z_1.Z_2, \ldots$ sorozat egymás utáni elemeit, ha a rendszer az 1 állapotban van, de minden egyes megfigyelésért c forintot kell fizetni. Jogunk van eldönteni, hogy mennyi ideig folytatjuk a megfigyelést. Ezután meg kell tippelnünk, hogy a rendszer a 0 vagy az 1 állapotban van-e, de a helytelen tippért fizetni kell. Annak a hibának, hogy a 0 állapotra tippeltünk az 1 állapot helyett w_1 forint az ára, annak a hibának, hogy az 1 állapotra tippeltünk az 0 állapot helyett w_2 forint az ára. A feladat az optimális döntési eljárásnak, azaz annak az eljárásnak a megadása, amelyre az előbb leírt procedúra költségének a várható értéke minimális.

Érdemes megérteni e feladat kapcsolatát a szekvenciális próbák vizsgálatával. Vezessük be az X_n, $n = 1, 2, \ldots$, valószínűségi változókat az $X_n = Y_n$, ha $U = 0$, és $X_n = Z_n$, ha $U = 1$, $n = 1, 2, \ldots$, képletekkel. Ezzel a jelöléssel az $X_1, X_2, \ldots$ valószínűségi változókat figyeljük meg. Megadjuk e sorozat feltételes eloszlását feltéve az U valószínűségi változó értékét. Az $U = 0$ esetben ez a feltételes eloszlás megegyezik független P_{ϑ_0} eloszlású valószínűségi változók sorozatának az eloszlásával. Az $U = 1$ esetben ez a feltételes eloszlás megegyezik független P_{ϑ_1} eloszlású valószínűségi változók sorozatának az eloszlásával. Egy döntési eljárás megadása azt jelenti, hogy először megadunk egy az $X_1, X_2, \ldots$ sorozat elemeinek értékétől függő N megállási szabályt azaz az egy olyan pozítiv egész értéket felvevő N valószínűségi változót, amelyre $\{\omega\colon\ N(\omega) = k\} \in \sigma(X_1, \ldots, X_k) = \sigma(\mathbf{X}_k)$ minden $k = 1, 2, \ldots$ indexre, majd definiálunk egy olyan $T = T(X_1, \ldots, X_{N(\omega)}) = T(\mathbf{X}_{N(\omega)})$ valószínűségi változót, amelyre $T = 0$ vagy $T = 1$. A T valószínűségi változót döntési függvénynek nevezzük. A $T = 0$ esemény bekövetkezésekor arra tippelünk, hogy a rendszer a 0 állapotban van, míg a $T = 1$ esemény bekövetkezésekor arra tippelünk, hogy a rendszer az 1 állapotban van. Teszünk még egy technikai kiegészítést a definicióban. Megengedjük azt is, hogy a nulladik időpontban, azaz az X_n mintaelemek megfigyelése nélkül hozzunk döntést, ami vagy a determinisztikus $T = 0$ vagy a determinisztikus $T = 1$ függvény. Formálisan, legyen $\mathbf{X}_0$ az üres sorozat, az általa generált $\sigma(\mathbf{X}_0)$ a triviális $\{\emptyset, \Omega\}$, σ-algebra. Az $N(\omega)$ megállási szabály a nulla értéket is felveheti, de $\{\omega\colon\ N(\omega) = 0\} \in \sigma(\mathbf{X}_0)$, és a T döntési függvény is konstans az $\{\omega\colon\ N(\omega) = 0\}$ halmazon.

Egy N megállási szabállyal és T döntési függvénnyel megadott döntési eljárás költségének a várható értéke

$$K(N,T) = K(N,T,\pi) = \pi(cE_{\vartheta_0}(N) + w_1 P_{\vartheta_0}(T=1)) \\ +(1-\pi)(c(E_{\vartheta_1}(N) + w_2 P_{\vartheta_1}(T=0)), \tag{1}$$

és a feladat annak az (N,T) párral megadott döntési szabálynak a megtalálása, amelyre $K(N,T)$ minimális.

Egy szekvenciális próbát hasonlóan egy az $X_1, X_2, \dots$ sorozat szerinti N megállási szabállyal és egy $T = T(X_1,\dots,X_N)$ 0 vagy 1 értéket felvevő valószínűségi változóval adunk meg. A különbség az, hogy ebben az esetben minket a $P_{\vartheta_0}(T=1)$ elsőfajú és $P_{\vartheta_1}(T=0)$ másodfajú hiba nagysága érdekel, illetve az $E_{\vartheta_0}(N)$ és $E_{\vartheta_1}(N)$ várható értékek, amelyek a megfigyelt mintaelemek számának a várható értékével egyenlőek a H_0 nullhipotézis illetve a H_1 ellenhipotézis teljesülése esetén. Azt a feladatot vizsgáljuk, hogy ha csak olyan szekvenciális próbákat választhatunk, amelyekben a $P_{\vartheta_0}(T=1)$ elsőfajú és $P_{\vartheta_1}(T=0)$ másodfajú hiba kisebb, mint két előírt szám, akkor milyen kicsi lehet $E_{\vartheta_0}(N)$ és $E_{\vartheta_1}(N)$ alkalmas próba választása estén. Az optimális döntési szabály kereséséről szóló feladat valójában azt jelenti, hogy olyan szekvenciális próbát keresünk, amelyre a $P_{\vartheta_0}(T=1)$, $P_{\vartheta_1}(T=0)$, $E_{\vartheta_0}(N)$ és $E_{\vartheta_1}(N)$ mennyiségek valamely lineáris kombinációja minimális értéket vesz fel.

Informálisan azt mondhatjuk, hogy a fent megfogalmazott optimális döntési eljárásról szóló feladatban ugyanazokat a stratégiákat vizsgáljuk, mint a szekvenciális próbákban, csak ezeket a stratégiákat másképpen értékeljük ki. Olyan eljárásokat tekintünk, ahol először eldöntjük meddig folytatjuk a minta megfigyelését, aztán a megfigyelések alapján megtippeljük, hogy melyik állapotban vagyunk. Ezért az optimális döntési eljárásról szóló eredmények hasznosak lehetnek a szekvenciális próbák vizsgálatában is. Az első fontos eredmény az optimális döntésről szóló feladatban, az a 4.5. lemmában megfogalmazott állítás, amelyik azt mondja ki, hogy minden $0 < \pi < 1$, $c > 0$ és $w_1 > 0$, $w_2 > 0$ paraméterválasztás esetében az optimális stratégia a Wald-féle valószínűséghányados próba olyan alkalmas A és B konstansokkal, amelyek a fenti paraméterektől függnek. Egy másik fontos eredmény szerint minden $0 < \pi < 1$ paraméter és $0 < A < 1 < B < \infty$ konstansok esetén a $c > 0$ és $w_1 > 0$, $w_2 > 0$ paramétereket megválaszthatjuk úgy, hogy az optimális döntés a vizsgált feladatban megegyezik az ezen A és B konstansok által meghatározott Wald-féle valószínűséghányados próbával. Feltehetjük azt is, hogy a w_1 és w_2 konstansokat úgy választjuk, hogy $w_1 + w_1 = 1$.

Bár a könyv csak implicit módon tartalmazza ezt az eredményt, én megfogalmazom explicit módon az alábbi 4.4′. tételben. A 4.4. tétel ezen állítás egyszerű következménye.

4.4′. Tétel. *Legyenek adva bizonyos* $0 < A < 1 < B < \infty$ *és* $0 < \pi < 1$ *számok. Meg tudunk választani olyan* $c > 0$*, és* $w_1 > 0$*,* $w_2 > 0$ *számokat, amelyekre* $w_1 + w_2 = 1$*, és ha az előbb definiált optimális döntési eljárásról szóló feladatot ezekkel a* π*,* c*,* w_1 *és* w_2 *paraméterekkel tekintjük, akkor az optimális döntési eljárás e feladatban megegyezik a Wald-féle valószínűséghányados próbával* $0 < A < 1 < B < \infty$ *paraméterekkel.*

A 4.4. tétel bizonyítása a 4.4′. tétel segítségével. Rögzítsünk egy $0 < \pi < 1$ számot, és válasszuk a $c > 0$, $w_1 > 0$, $w_2 > 0$, $w_1 + w_2 = 1$ paramétereket oly módon, hogy ezzel a választással a korábban definiált optimális döntési eljárásról szóló feladat optimális döntési eljárása megegyezik a Wald-féle valószínűséghányados próbával a 4.4 tétel megfogalmazásában szereplő $0 < A < 1 < B < \infty$ paraméterekkel. Tekintsük azt az N megállási szabályból és T nulla vagy egy értéket felvevő döntési függvényből álló (N, T) párt, amely meghatározza ezt a próbát. Ekkor $P_{\vartheta_0}(T = 1) = \varepsilon_1$, és $P_{\vartheta_1}(T = 0) = \varepsilon_2$. Tekintsünk egy másik valamilyen (N', T') megállási szabály és döntési függvénypárral definiált szekvenciális próbát. Legyen $P_{\vartheta_0}(T' = 1) = \varepsilon'_1$. és $P_{\vartheta_1}(T' = 0) = \varepsilon'_2$. Ekkor mind az (N, T) mind az (N', T') pár meghatároz egy döntési eljárást, és az (N, T) pár által megadott döntési eljárásnak a 4.4′. tételben megfogalmazott optimalitási tulajdonságából következik az alábbi egyenlőtlenség.

$$\begin{aligned}&\pi(cE_{\vartheta_0}(N) + w_1\varepsilon_1) + (1 - \pi)(c(E_{\vartheta_1}(N) + w_2\varepsilon_2) \\ &\qquad \le \pi(cE_{\vartheta_0}(N') + w_1\varepsilon'_1) + (1 - \pi)(c(E_{\vartheta_1}(N') + w_2\varepsilon'_2).\end{aligned}$$

A 4.4. tételben olyan (N', T') párok által meghatározott szekvenciális próbát tekintettünk, amelyre $\varepsilon'_1 \le \varepsilon_1$ és $\varepsilon'_2 \le \varepsilon_2$. Ilyen esetben a fenti egyenlőtlenségből következik az alábbi egyenlőtlenség is.

$$\pi E_{\vartheta_0}(N) + (1 - \pi)E_{\vartheta_1}(N) \le \pi E_{\vartheta_0}(N') + (1 - \pi)E_{\vartheta_1}(N').$$

Mivel ez az egyenlőtlenség minden $0 < \pi < 1$ paraméterre igaz, innen következnek az $E_{\vartheta_0}(N) \le E_{\vartheta_0}(N')$ és $E_{\vartheta_1}(N') \le E_{\vartheta_1}(N)$ egyenlőtlenségek is. A 4.4. tételt beláttuk.

A 4.4. tétel bizonyításának legfontosabb lépése az optimális stratégia megtalálása az előbb megfogalmazott optimális döntési eljárásról szóló feladatban. A könyv csak vázlatosan ismerteti ezen probléma megoldását.

Ezért én is megelégszem azzal, hogy megadjam annak legfontosabb gondolatait, illetve megfogalmazzam a legfontosabb megoldandó matematikai problémákat. Először megadom a keresett optimális stratégiát.Ennek érdekében bevezetem a következő mennyiségeket.

Jelöljön (a könyv jelöléseit követve δ egy lehetséges (N,T) döntési eljárást, és jelölje $\mathcal{D}$ az összes olyan lehetséges $\delta=(N,T)$ döntés halmazát, amelyre $E_{\vartheta_0}(N)<\infty$, és $E_{\vartheta_1}(N)<\infty$. Legyen minden $\delta=(N,T)\in\mathcal{D}$ döntés költsége $r(\pi,\delta)=K(N,T,\pi)$ az (1) formulában definiált $K(N,T,\pi)$ függvénnyel. Jelölje $\mathcal{D}_1$ azon $\delta=(N,T)$ döntési eljárások halmazát, amelyekre $P(N\ge n)=1$, és legyen

$$\rho(\pi)=\inf_{\delta\in\mathcal{D}_1} r(\pi,\delta). \tag{2}$$

Mint a könyv megmutatja, a $\rho(\pi)$ függvény alulról korlátos, konkáv, ezért a $(0,1)$ intervallumon folytonos függvény. (Vegyük észre, hogy az $r(\pi,\delta)$ függvény folytonos minden $\delta\in\mathcal{D}$ döntési eljárásra. Ezt felhasználva meg lehet mutatni, hogy $\rho(\pi)$ mérhető függvény.) Nem nehéz belátni, hogy

$$\lim_{\pi\to 0}\rho(\pi)=\lim_{\pi\to 1}\rho(\pi)=c.$$

A $\rho(\cdot)$ függvény segítségével meg tudjuk mondani, milyen π paramétereknél állunk le az optimális döntési eljárásban az első megfigyelés előtt, azaz a nulladik lépésben, és ebben az esetben milyen döntest hozunk.

Legyen $\delta_0=(N_0,T_0)$, ahol $N_0\equiv 0$, $T_0\equiv 0$, és $\delta_1=(N_0,T_1)$, ahol $N_0\equiv 0$, $T_1\equiv 1$. Ekkor $r(\pi,\delta_0)=\pi w_1$ és $r(\pi,\delta_1)=(1-\pi)w_0$. Minden $0<\pi<1$ paraméterre az optimális stratégia a δ_0, δ_1 és az optimális $\delta\in\mathcal{D}_1$ stratégia valamelyike, (ez utóbbi költsége $\rho(\pi)$), és ezek közül azt kell választani, amelyiknek minimális a költsége. Ez azt jelenti, hogy ha $\rho\left(\frac{w_2}{w_1+w_2}\right)\ge\frac{w_1w_2}{w_1+w_2}$, akkor a $\pi\le\frac{w_1}{w_1+w_2}$ esetben a δ_0, és a $\pi>\frac{w_1}{w_1+w_2}$ esetben a δ_1 stratégiát alkalmazzuk. Ha $\rho\left(\frac{w_2}{w_1+w_2}\right)<\frac{w_1w_2}{w_1+w_2}$, akkor van egy olyan $0<\pi'<\frac{w_2}{w_1+w_2}$ szám, amelyre $\pi'w_1=\rho(\pi')$, és egy olyan $\frac{w_2}{w_1+w_2}<\pi''<1$ szám, amelyre $(1-\pi'')w_2=\rho(\pi'')$. (Lásd a könyvben a 4.1 ábrát.) Ebben az esetben az optimális stratégia a δ_0 stratégia, ha $\pi\le\pi'$, a δ_1 stratégia, ha $\pi\ge\pi''$, míg a $\pi'<\pi<\pi''$ esetben $N\ge 1$, és meg kell keresni a legjobb $\delta\in\mathcal{D}_1$ stratégiát.

Ilyen módon megadtuk az optimális stratégia nulladik lépését. A további lépések megtalálása hasonló elveken alapul, bár kidolgozásuk technikailag bonyolultabb. Ahhoz, hogy ezeket megadjuk be kell vezetni a következő feltételes valószínűségeket, amelyeket a könyv aposteriori valószínűségeknek nevez.

Definiáljuk minden $\mathbf{X}_n = (X_1, \dots, X_n)$ sorozatra a

$$\pi_n = \pi_n(\mathbf{X}_n) = P(U = 1 | X_1, \dots, X_n), \quad n = 1, 2, \dots,$$

feltételes valószínűségeket, és legyen $\pi_0 = \pi$. Be lehet látni, hogy

$$\pi_n = \pi_n(\mathbf{X}_n) = \frac{\pi L_{\vartheta_1}(\mathbf{X}_n)}{\pi L_{\vartheta_1}(\mathbf{X}_n) + (1 - \pi) L_{\vartheta_0}(\mathbf{X}_n)}, \quad n = 1, 2, \dots, \tag{3}$$

ahol $L_{\vartheta_0}(\mathbf{X}_n) = \prod_{j=1}^n f_{\vartheta_0}(Y_j)$, és $L_{\vartheta_1}(\mathbf{X}_n) = \prod_{j=1}^n f_{\vartheta_1}(Z_j)$. Innen következik az alábbi rekurziv formula.

$$\pi_n(\mathbf{X}_n) = \frac{\pi_{n-1}(\mathbf{X}_{n-1}) f_{\vartheta_1}(X_n)}{\pi_{n-1}(\mathbf{X}_{n-1}) f_{\vartheta_1}(X_n) + (1 - \pi_{n-1}(\mathbf{X}_{n-1})) f_{\vartheta_0}(X_n)}.$$

Ki tudjuk számolni az U és $X_n, X_{n+1}, \dots$ valószínűségi változók együttes feltételes eloszlását is feltéve az $\mathbf{X}_n = (X_1, \dots, X_n)$ véletlen vektor által generált σ-algebrát. Azt kapjuk, hogy

$$\begin{aligned} &P(U = 0, X_{n+1} \in B_1, X_{n+2} \in B_2, \dots | X_1, \dots, X_n) \\ &\quad = (1 - \pi_n(\mathbf{X}_n)) P_{\vartheta_0}(X_1 \in B_1) P_{\vartheta_0}(X_2 \in B_2) \cdots \ , \\ &P(U = 1, X_{n+1} \in B_1, X_{n+2} \in B_2, \dots | X_1, \dots, X_n) \\ &\quad = \pi_n(\mathbf{X}_n) P_{\vartheta_1}(X_1 \in B_1) P_{\vartheta_1}(X_2 \in B_2) \cdots \end{aligned}$$

$B_1, B_2, \dots$ mérhető halmazok minden sorozatára.

Ez azt jelenti, hogy minden $n = 1, 2, \dots$ időpontban a jövőbeli X_{n+1}, $X_{n+2}, \dots$, valószínűségi változók együttes feltételes eloszlása feltéve a múltbeli $X_1, \dots, X_n$ valószínűségi változók értékeit megegyezik az eredeti $X_1, X_2, \dots$ valószínűségi változók együttes eloszlásával, ha annak definiciójában az $U = 1$ esemény π valószínűségét a $\pi_n(\mathbf{X}_n)$ feltételes valószínűséggel helyettesítjük a feltételes eloszlás kiszámolásában. Ez azt sugallja, hogy az optimális stratégia n-ik lépésében hasonló eljárást kell folytatni, mint a kiinduló lépésben, csak a π valószínűséget a $\pi_n(\mathbf{X}_n)$ feltételes valószínűséggel kell helyettesíteni. Az alábbi 4.5. lemma fogalmazza meg a fenti heurisztikus érvelés pontos jelentését és helyességét.

4.5. Lemma. *Tekintsük a korábban definiált optimális döntési eljárásról szóló feladatot valamilyen* $0 < \pi < 1$, $c > 0$, $w_1 > 0$, $w_2 > 0$ *paraméterekkel* *és P_{ϑ_0} és P_{ϑ_1} eloszlásokkal. Tekintsük a (2) formulában a segitségükkel definiált $\rho(\pi)$, $0 < \pi < 1$, függvényt, és legyen π' és π'' a $\pi w' = \rho(\pi')$ és $(1-\pi'')w_2 = \rho(\pi'')$ egyenlet megoldása, feltéve, hogy van ilyen megoldás. Ha ezen egyenleteknek van megoldása, akkor a keresett optimális döntési eljárás a következő. Tekintintsük a $\pi_n(\mathbf{X}_n)$, $n = 0, 1, 2, \ldots$, feltételes valószínűségek sorozatát, és legyen N az a legkisebb index, amelyre vagy $\pi_N(\mathbf{X}_N) \le \pi'$ vagy $\pi_N(\mathbf{X}_N) \ge \pi''$, és legyen $T = 0$, ha $\pi_N(\mathbf{X}_N) \le \pi'$, $T = 1$, ha $\pi_N(\mathbf{X}_N) \ge \pi''$. Ha a fenti egyenleteknek nincs megoldása, akkor $N \equiv 0$, és $T = 0$, ha $\pi \le \frac{w_1}{w_1+w_2}$, és $T = 1$, ha $\pi > \frac{w_1}{w_1+w_2}$.*

A valószínűségszámításban kidolgozták az optimális megállások elméletét. A következő problémát vizsgálták. Tekintünk egy (diszkrét idejű) sztochasztikus folyamatot, és amennyiben egy adott időpontban a folyamatot megállítjuk, akkor fizetni kell egy a megállítás időpontjától és a sztochasztikus folyamat ezen időpontbeli és múltbeli állapotaitól függő költséget. A feladat annak az optimális megállási időnek a megtalálása, amelynek választása esetén költségünk várható értéke minimális. A most tárgyalt feladat tekinthető ilyen tipusú feladatnak, mert a fő probléma az N megállási időpont megtalálása. Az N időpontbeli megálláshoz tartozó jó T döntési függvény megtalálása és a felmerülő költség meghatározása egyszerű.

A most említett elmélet eredményeinek és az itt tárgyalt formulák, észrevételek segítségével be lehet bizonyítani a 4.5. lemmát. Erről teszek majd néhány észrevételt, de a részletes bizonyítást nem tárgyalom. Viszont érdemes a 4.5. lemmában definiált N megállási szabályt olyan ekvivalens formában megadni, amelyik mutatja e megoldás kapcsolatát a Wald-féle valószínűséghányados próbával. Az N megállási szabály másik megadását a feltételes valószínűség kiszámításáról szóló (3) formula segítségével tesszük meg.

A (3) formulában definiált $\pi_n(\mathbf{X}_n)$ felírható $\pi_n(\mathbf{X}) = \frac{\pi V_n(\mathbf{X}_n)}{\pi V_n(\mathbf{X}_n)+(1-\pi)}$ alakban is a $V_n(\mathbf{X}_n) = \frac{L_{\vartheta_1}(\mathbf{X}_n)}{L_{\vartheta_0}(\mathbf{X}_n)}$ függvénnyel. E formula segítségével a $\pi_n(\mathbf{X}_n) \le \pi'$ egyenlőtlenség úgy írható, hogy $V_n(\mathbf{X}_n) \le \frac{1-\pi}{\pi}\frac{\pi'}{1-\pi'}$, és a $\pi_n(\mathbf{X}_n) \ge \pi''$ egyenlőtlenség úgy írható, hogy $V_n(\mathbf{X}_n) \ge \frac{1-\pi}{\pi}\frac{\pi''}{1-\pi''}$. Ez azt jelenti, hogy a 4.5. lemmában a π' és π'' számok segítségével megadott optimális döntési eljárás megegyezik a Wald-féle valószínűséghányados próbával az $A = \frac{1-\pi}{\pi}\frac{\pi'}{1-\pi'}$ és $B = \frac{1-\pi}{\pi}\frac{\pi''}{1-\pi''}$ paraméterválasztással.

Szükségünk van még egy eredményre, amelyik azt biztosítja, hogy az optimális döntési eljárásról szóló feladatban a paramétereket meg tudjuk választani a számunkra megfelelő módon. Ez a könyv 4.6.lemmája. E lemma bizonyításával a könyv nem foglalkozik. Ehelyett Lehmann E. L. Testing statistical hypotheses, John Wiley and Sons, New York–London 1959., könyv 3.12. paragrafusára hívatkozik.

4.6. Lemma. *Legyen adva egy* $0 < \pi < 1$ *szám. Minden olyan* π', π'' *számpárhoz, amelyre* $0 < \pi' < \pi < \pi'' < 1$, *léteznek olyan* $c > 0$, $w_1 > 0$, $w_2 > 0$, $w_1 + w_2 = 1$ *számok, hogy a (2) formulában a segitségükkel definiált* $\rho(\pi)$, $0 < \pi < 1$, *függvényre és ezekre a* π' *és* π'' *számokra teljesülnek a* $\pi w' = \rho(\pi')$ *és* $(1 - \pi'')w_2 = \rho(\pi'')$ *egyenletek.*

A 4.4′. tételt nem nehéz belátni a 4.5. és 4.6. lemmák segítségével. Legyenek adva valamilyen $0 < A < 1 < B < \infty$ és $0 < \pi < 1$ számok. Tudunk választani olyan π' és π'' számokat, $0 < \pi' < \pi < \pi'' < 1$, amelyekre $A = \frac{1-\pi}{\pi}\frac{\pi'}{1-\pi'}$ és $B = \frac{1-\pi}{\pi}\frac{\pi''}{1-\pi''}$. Válasszunk olyan $c > 0$, $w_1 > 0$, $w_2 > 0$, $w_1 + w_2 = 1$ számokat, amelyekre ezek a π' és π'' számok teljesítik a 4.6. lemmában megfogalmazott egyenleteket. Ezután alkalmazzuk a 4.5. lemmát ezekkel a $0 < \pi < 1$, $c > 0$, $w_1 > 0$ és $w_2 > 0$ paraméterekkel. Az így kapott optimális dóntési eljárás teljesíti a 4.4′. tételt.

Végül teszek néhány megjegyzést a 4.5. lemma bizonyításáról. A lemma felhasználja a következő gondolatot. Ha eljutottunk egy állapotba valahányadik lépésben, akkor el kell dönteni, figyelembe véve az eddig összegyüjtött információkat, hogy mit érdemes csinálni. Érdemes-e azonnal leállni és definiálni az optimális T döntésfüggvényt, vagy előnyösebb-e egy új mintaelemet kérni, és azután annak figyelembe vételével keresni az optimális stratégiát a következő lépésben. Az optimális stratégia ezen két lehetőség közül a jobbik választása, és annak feltételes várható értéke e két stratégia feltételes várható értékének a minimuma. Meg kívánjuk fogalmazni ezt az állítást pontosabban is.

Ezen állítás pontos megfogalmazása lehetséges, de az új fogalmak bevezetését és azokkal kapcsolatos eredmények bizonyítását igényli. A nulladik lépésben ezt az elvet alkalmaztuk a (2) formulában bevezetett $\rho(\pi)$ függvény segítségével. Ez a függvény várható értékek infimuma volt, és az infimumban több mint megszámlálható elem szerepelt. A későbbi lépésekben hasonló definiciót kellene bevezetni, de ott megszámlálhatónál több feltételes várható érték infimumát kellene venni. Ez problematikus. E problémát a lényeges szuprémum és lényeges infimum bevezetésével és egy velük kapcsolatos tétel bizonyításával lehet megoldani. Ebben a definicióban azt használjuk ki, hogy két valószínűségi változót, amelyek egy valószínűséggel megegyeznek azonosnak tekintünk, és akkor mondjuk, hogy egy valószínűségi változó nagyobb, mint egy másik, ha egy valószinűséggel nagyobb nála.

Lényeges szuprémum és lényeges infimum definiciója. *Legyen adva ξ_t, $t \in T$, valószínűségi változók halmaza egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, ahol T tetszőleges indexhalmaz. Akkor mondjuk, hogy az η valószínűségi változó a ξ_t, $t \in T$, valószínűségi változók lényeges szuprémuma, (lényeges infimuma), ha $P(\eta \geq \xi_t) = 1$, ($P(\eta \leq \xi_t) = 1$), minden $t \in T$ indexre, és ha $P(\eta' \geq \xi_t) = 1$, ($P(\eta' \leq \xi_t) = 1$), minden $t \in T$ indexre, valamely η' valószínűségi változóra, akkor $P(\eta' \geq \eta) = 1$, ($P(\eta' \leq \eta) = 1$).*

Tétel a lényeges szuprémum és lényeges infimum tulajdonságairól. *Legyen adva ξ_t, $t \in T$, valószínűségi változók halmaza egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, amelyek értékeiket a $\pm\infty$ pontokkal kibővített számegyenesen veszik fel. Létezik a ξ_t, $t \in T$, valószínűségi változóknak η lényeges szuprémuma, (lényeges infimuma), és ez az η lényeges szuprémum, (lényeges infimum), előállítható a következő módon. Létezik olyan $T_0 = \{t_1, t_2, \dots\} \subset T$ megszámlálható halmaz, amelyre $\eta = \sup_{t \in T_0} \xi_t$ ($\eta = \inf_{t \in T_0} \xi_t$).*

Megadom azt az eredményt az optimális megállási feladatok megoldásáról, amely segít a 4.5. lemma bizonyításában. Ebben infimum helyett szuprémumot keresünk, de -1-gyel való szorzással a feladat visszavezethető ilyen problémára.

A feladat a következő: Legyen adva egy $(\xi_n, \mathcal{F}_n)$, $n = 0, 1, 2, \dots$, rendszer egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn, ahol $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{A}$ növekvő σ-algebrák sorozata, ξ_n $\mathcal{F}_n$ mérhető valószínűségi változó, jelentése a nyeremény értéke az n-ik lépésben. Olyan τ megállási szabályt keresünk erre a növekvő σ-algebra rendszerre, amelyre az $E\xi_\tau$ nyeremény a lehető legnagyobb. Feltesszük, hogy $E\min(\xi_n, 0) > -\infty$ minden $n = 0, 1, 2, \dots$ indexre. Olyan rekurziós azonosságot keresünk, amely kapcsolatot teremt azon optimális megállási szabályok nyereménye között, amelyekben csak az n-ik és azok között, amelyekben csak az $n+1$-ik lépés után állhatunk meg.

Megadunk egy ilyen rekurziós formulát, amelyik nagyon hasznos a vizsgálatokban. Annak érdekében, hogy ezt megfogalmazzuk bevezetjük a következő jelöléseket.

$$C_n = \{\tau\colon \tau \text{ megállási szabály, } P(\tau \geq n) = 1,\ E\xi_\tau < \infty\}, \quad n = 0, 1, 2, \dots,$$

$$\gamma_n = \text{ess sup}_{\tau \in C_n} E(\xi_\tau | \mathcal{F}_n),$$

ahol ess sup lényeges szuprémumot jelöl, és

$$v_n = \sup_{\tau \in C_n} E\xi_\tau.$$

A következő tétel érvényes.

Az optimális megállás rekurziós tulajdonságairól szóló tétel.

$$\begin{aligned}\gamma_n &= \max(\xi_n, E(\gamma_{n+1}|\mathcal{F}_n)), \quad n = 0, 1, 2, \ldots, \\ v_n &= E\gamma_n, \quad n = 0, 1, 2, \ldots.\end{aligned}$$

A 4.5. lemmában ezt az eredményt a következő szereposztással alkalmazzuk. $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$, és

$$\xi_0 = \min(w_1\pi, w_2(1-\pi)),$$

$$\xi_n = -cn - \min_{A,B} \left(w_1(P_{\vartheta_0}((X_1, \ldots, X_n) \in A) + w_2(P_{\vartheta_1}((X_1, \ldots, X_n) \in B) \right),$$

$n = 1, 2, \ldots$, ahol a minimum az olyan (A, B) halmazpárokra vétetik fel, amelyek az $\mathbf{X}_n = (X_1, \ldots, X_n)$ véletlen vektor értékkészletének $\mathcal{F}_n$ mérhető particiói. Erre a $(\xi_n, \mathcal{F}_n)$, $n = 0, 1, 2, \ldots$ rendszerre keresünk maximális nyereményt biztosító $\tau \in C_0$ megállási szabályt.

A ξ_n, $n = 1, 2, \ldots$ valószínűségi változók definicióját egyszerűbb alakra lehet hozni. Be lehet látni, hogy

$$\xi_n = -cn - \min[w_1\pi_n(\mathbf{X}_n), w_2(1 - \pi_n(\mathbf{X}_n))], \quad n = 1, 2, \ldots.$$

A továbbiakban $X_n = X_n(\pi)$, $\mathbf{X}_n = \mathbf{X}_n(\pi)$, $\gamma_n = \gamma_n(\pi)$, $v_n = v_n(\pi)$ jelölést fogunk használni akkor, ha azt a jegyzetben konstruált modellt tekintjük, amelyikben $P(U = 0) = 1 - \pi$, $P(U = 1) = \pi$. Tekinteni fogjuk ezt a modellt minden $0 < \pi < 1$ paraméterre. Legyen továbbá $\mathcal{F}_n = \sigma(\mathbf{X}_n(\pi))$.

Ezzel a jelöléssel $\rho(\pi) = -v_1(\pi)$ a (2) formulában definiált $\rho(\pi)$ függvénnyel. Be lehet látni, felhasználva a $\pi_n(\mathbf{X}_n(\pi)), X_{n+1}(\pi), X_{n+2}(\pi), \ldots)$ vektornak a $\sigma(\mathbf{X}_n)$ σ-algebra szerinti feltételes eloszlásáról megfogalmazott állítást, hogy

$$E(\gamma_{n+1}(\mathbf{X}_{n+1}(\pi))|\mathcal{F}_n) = -\rho(\pi_n(\mathbf{X}_n(\pi)) - c(n-1). \tag{4}$$

Speciálisan, $E(\gamma_{n+1}(\mathbf{X}_{n+1}(\pi))|\mathcal{F}_n)$ a $\pi_n(\mathbf{X}_n(\pi))$ valószínűségi változó függvénye.

Legyen π' és π'' a $\pi w' = \rho(\pi')$ és $(1-\pi'')w_2 = \rho(\pi'')$ egyenlet megoldása. Foglalkozzunk egyelőre csak azzal az esettel, amikor ezen egyenleteknek van megoldása. Vezessük be a következő τ_0 megállási szabályt.

$$\tau_0(\omega) = \min(n\colon\ n \geq 0,\ \pi_n(\pi)(\omega) \leq \pi' \text{ vagy } \pi_n(\pi)(\omega) \geq \pi'').$$

Tudjuk, hogy $\tau_0 \in C_0$. Be akarjuk látni, hogy τ_0 az optimális megállási szabály, azaz $v_0 = E\tau_0$.

Vezessük be a $D_n = \{\omega\colon\ \tau_0(\omega) = n\}$ és $F_n = \{\omega\colon\ \tau_0(\omega) \geq n\}$, halmazokat, $0 \leq n < \infty$. Azt állítom, hogy

$$\gamma_m(\omega)I_{F_m}(\omega) = E(\xi_{\tau_0}(\omega)(\pi)I_{F_m}(\omega)|\mathcal{F}_m)(\omega), \quad 0 \leq m < \infty.$$

Ezen azonosság bizonyításában felhasználjuk a (4) formulát, az optimális megállás rekurziós tulajdonságairól szóló tételt és a $\rho(\cdot)$ függvény (2) definícióját, illetve annak tulajdonságait. Ezek segítségével be lehet látni, hogy $\gamma_n(\omega) = \xi_n(\omega)$ a D_n halmazon, és $\gamma_n(\omega) = E(\gamma_{n+1}|\mathcal{F}_n)(\omega)$ az F_{n+1} halmazon. (Az F_{n+1} halmaz $\mathcal{F}_n$ mérhető.) Továbbá

$$\gamma_m(\omega)I_{D_n}(\omega) = E(\xi_n|\mathcal{F}_m)(\omega)I_{D_n}(\omega), \quad \text{ha } n \geq m.$$

Ezeket az egyenleteket összeadva $n = m, m+1, \dots$-re kapjuk, hogy

$$\gamma_m(\omega)I_{F_m}(\omega) = E(\xi_{\tau_0}(\omega)|\mathcal{F}_m)(\omega)I_{F_m}(\omega) = E(\xi_{\tau_0}(\omega)I_{F_m}(\omega)|\mathcal{F}_m)(\omega),$$

ahogy állítottam. A kapott azonosságot alkalmazva $m = 0$-ra, és várható értéket véve kapjuk a 4.5 lemma állítását abban az esetben, ha a $\pi w' = \rho(\pi')$ és $(1-\pi'')w_2 = \rho(\pi'')$ egyenleteknek van megoldása. Ugyanis azt kapjuk, hogy $v_0 = E\gamma_0 = E\xi_{\tau_0}$. (Megjegyzem,hogy $F_0 = \Omega$.)

Az optimális T döntési függvény választása, (amelynek választásával a nyereményfüggvényünk ξ_n, $n = 0, 1, 2, \dots$) viszonylag egyszerű. $T(\omega) = T_n(\omega) = 0$, ha $\tau_0(\omega) = n$, és $\pi_n(\omega) \leq \pi'$. Másrészt $T(\omega) = T_n(\omega) = 1$, ha $\tau_0(\omega) = n$, és $\pi_n(\omega) \geq \pi''$.

Ha a fenti egyenletnek nincs megoldása akkor $N \equiv 0$ az optimális megállási szabály, és szintén a 4.5 lemmában megfogalmazott stratégiát kell alkalmazni. Ennek az esetnek a vizsgálata egyszerűbb, (és valójában erre az eredményre nincs szükségünk).