11. Feltételes várható érték

Az előző fejezetben vizsgált lineáris regresszió egyik hátulütője, hogy csak a változók közötti lineáris összefüggést fogja meg, mélyebb relációkat nem. Példán érzékeltetve, hiába lehet X értékéből tökéletesen meghatározni X^2 értékét, az X^2 lineáris regressziója X-re nem feltétlenül lesz jó közelítése X^2-nek. Természetes tehát a kérdés, nem lehetne ezt jobban csinálni valahogy? Persze a válasz igenlő, amiben a feltételes várható érték fogalma, illetve az ebből adódó regressziós függvény lesz segítségünkre.

11.1. Feltételes várható érték, diszkrét regresszió

Idézzük fel a feltételes valószínűség definícióját: ha A olyan esemény, amire $\mathbb{P}(A) > 0$, akkor

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$

a B-nek az A-ra vett feltételes valószínűsége. Szemléletesen, ez a B esemény valószínűsége arra az esetre fókuszálva, amikor A bekövetkezik. A második fejezetben beláttuk, hogy $B \mapsto \mathbb{P}(B \mid A)$ is valószínűségi mérték. Vagyis bármi, amit valószínűségi mértékekre bizonyítottunk, rá is alkalmazható.

⚠ **11.1.1. Definíció.** Legyen Y valószínűségi változó és A olyan esemény, amire $\mathbb{P}(A) > 0$. Ekkor az Y-nak az A-ra vett **feltételes várható értéke** az Y változó $\mathbb{P}(. \mid A)$ valószínűségi mérték szerinti várható értéke. Jelölés: $\mathbb{E}(Y \mid A)$.

A feltételes valószínűséghez hasonlóan, $\mathbb{E}(Y \mid A)$ jelentése az Y átlagos értéke a teljes eseménytér helyett az A eseményre szorítkozva.

11.1.2. Lemma. *Legyen Y egyszerű valószínűségi változó, és A esemény, amire $\mathbb{P}(A) > 0$. Ekkor*

$$\mathbb{E}(Y \mid A) = \sum_{k \in \operatorname{Ran}(Y)} k \cdot \mathbb{P}(Y = k \mid A). \tag{10}$$

Vagyis elég kicserélnünk a várható érték definíciójában a $\mathbb{P}$-t $\mathbb{P}(. \mid A)$-ra. Általánosabban, az alábbi módon tudjuk a feltételes várható értéket visszavezetni várható érték számolására.

11.1.3. Állítás. *Legyen Y valószínűségi változó, és $\mathbb{P}(A) > 0$. Ekkor $\mathbb{E}(Y \mid A) = \frac{1}{\mathbb{P}(A)} \mathbb{E}(Y \mathbf{1}_A)$.*[48]

Mi ennek a fogalomnak az értelme? Több esetben a feltételes várható érték nem kiszámolandó cél, hanem a feladat megfogalmazásának eszköze, ahogy ezt a feltételes valószínűség esetén is láttuk. Tegyük fel például, hogy egy eszközből kétféle márka is elérhető, az egyik átlagosan 3 évig, míg a másik 4 évig nem romlik el. Válasszunk a kettő közül véletlenszerűen egyet, és jelölje Y az élettartamát. Ekkor a feladatban lévő információink $\mathbb{E}(Y \mid \{\text{elsőt választjuk}\}) = 3$ és $\mathbb{E}(Y \mid \{\text{másodikat választjuk}\}) = 4$.

Feltétel eseményként használhatjuk egy másik valószínűségi változó $\{X < x\}$ nívóhalmazát is, azaz nézhetjük $\mathbb{E}(Y \mid X < x)$-et valamilyen X valószínűségi változóra.

11.1.4. Példa. Legyen Y és Z egy-egy szabályos kockadobás eredménye, $X = Y + Z$ és $x = 7$. Ekkor

$$\mathbb{P}(Y = k \mid X < 7) = \frac{\mathbb{P}(Y = k, X < 7)}{\mathbb{P}(X < 7)} = \frac{6-k}{36} \Big/ \frac{15}{36} = \frac{6-k}{15},$$

$$\mathbb{E}(Y \mid X < 7) = \sum_{k=1}^{6} k \cdot \mathbb{P}(Y = k \mid X < 7) = \sum_{k=1}^{6} k \cdot \frac{6-k}{15} = \frac{6 \cdot 21 - 91}{15} = \frac{7}{3}.$$

Továbbá, ha $\mathbb{P}(X = x) > 0$, akkor $\mathbb{E}(Y \mid X = x)$ is értelmes. Az előző példánál maradva

$$\mathbb{E}(Y \mid X = 5) = \sum_{k=1}^{6} k \cdot \mathbb{P}(Y = k \mid X = 5) = \sum_{k=1}^{4} k \cdot \frac{1}{4} = 2{,}5.$$

Szemléletesen, $\mathbb{E}(Y \mid X = 5)$ az Y átlagos értéke abban az esetben, ha tudjuk, hogy X értéke 5.

[48] Az állítás következménye, hogy ha $\mathbb{E}(Y)$ létezik és véges, akkor emiatt $\mathbb{E}(Y \mid A)$ is létezik és véges.

Az $\mathbb{E}(Y \mid X = x)$ mennyiségre úgy is tekinthetünk, mint egy függvényre az x változóban. Ez egy determinisztikus függvény, azaz nem valószínűségi változó, hiszen tetszőleges x valós szám esetén $\mathbb{E}(Y \mid X = x)$ valós szám (feltéve, hogy értelmes és véges).

11.1.5. Definíció. Legyenek X és Y valószínűségi változók, ahol X diszkrét. Jelölje S_X az X lényeges értékeinek a halmazát, azaz

$$S_X \stackrel{\text{def}}{=} \{x \in \mathbb{R} \mid \mathbb{P}(X = x) > 0\}.$$

Tekintsük a

$$g : S_X \to \mathbb{R}, \qquad g(x) = \mathbb{E}(Y \mid X = x)$$

valós függvényt. Ekkor az Y-nak az X-re vett (diszkrét) **regresszió**ja a $g(X)$ valószínűségi változó. Jelölése: $\mathbb{E}(Y \mid X)$.

11.1.6. Példa. Dobunk egy szabályos kockával, majd az eredménynek megfelelő számú szabályos érmével. Legyen X a kockadobás értéke, Y pedig a fejek száma az érmedobások között. Mi Y regressziója X-re? Ha ismerjük X értékét, akkor Y binomiális eloszlású, paraméterei: X értéke és $\frac{1}{2}$. Binomiális eloszlású változónak ismerjük a várható értékét (a paraméterek szorzata), ezért $\mathbb{E}(Y \mid X = x) = x\frac{1}{2}$, ahol $x \in \{1, 2, 3, 4, 5, 6\}$.

Az elnevezés motivációja a lineáris regresszió elnevezéséből kiindulva érthető meg: Y lineáris regressziója X-re egy olyan $\beta X + \alpha$ alakú lineáris függvénye az X valószínűségi változónak, ami a „legjobb lineáris közelítést” adja Y-ra. Ezt általánosítva, az Y változó X-re vett regressziója nem szorítkozik lineáris függvényekre, hanem azt a $g(X)$ függvényét adja vissza az a X valószínűségi változónak, amely a „legjobb közelítést” adja Y-ra. Ez a „legjobb” g függvény történetesen a fenti $x \mapsto \mathbb{E}(Y \mid X = x)$, hiszen mi lehetne jobb közelítés, mint az Y átlagos értéke abban az esetben, amikor X értékéről tudjuk, hogy x.

Hogy mit értünk „legjobb közelítés” alatt? Intuitívan fogalmazva az $\mathbb{E}(Y \mid X)$ valószínűségi változó mindent tud az Y-ról, amit X ismeretében tudni lehet, és ezzel a „lehető legtöbb” információval ad közelítést Y-ra.[49] Ezt precízen úgy tudjuk átfogalmazni, hogy ha további, X-re vonatkozó feltétel esetén nézzük Y feltételes várható értékét, akkor az már kiszámolható $\mathbb{E}(Y \mid X)$ regresszióból is, nem kell hozzá az eredeti Y-t ismernünk. Még precízebben:

11.1.7. Állítás. *Legyen X diszkrét valószínűségi változó. Tegyük fel, hogy $\mathbb{P}(X < x) > 0$ és $\mathbb{E}(Y)$ véges. Ekkor*

$$\mathbb{E}\big(g(X) \mid X < x\big) = \mathbb{E}(Y \mid X < x),$$

ahol $g(X) = \mathbb{E}(Y \mid X)$.

Bizonyítás. A fenti (10) egyenletből, és a feltételes valószínűség definíciójából adódóan

$$\mathbb{E}\big(g(X) \mid X < x\big) \stackrel{(10)}{=} \sum_{k \in \mathrm{Ran}(X)} g(k) \cdot \mathbb{P}(X = k \mid X < x) =$$

$$= \sum_{k<x} g(k) \frac{\mathbb{P}(X = k)}{\mathbb{P}(X < x)} = \sum_{k<x} \mathbb{E}(Y \mid X = k) \frac{\mathbb{P}(X = k)}{\mathbb{P}(X < x)}.$$

Itt felhasználhatjuk a fenti állítást és a várható érték additivitását:

$$= \sum_{k<x} \frac{1}{\mathbb{P}(X = k)} \mathbb{E}(Y \mathbf{1}_{\{X=k\}}) \frac{\mathbb{P}(X = k)}{\mathbb{P}(X < x)} = \frac{1}{\mathbb{P}(X < x)} \mathbb{E}\Big(Y \sum_{k<x} \mathbf{1}_{\{X=k\}}\Big)$$

$$= \frac{1}{\mathbb{P}(X < x)} \mathbb{E}\big(Y \mathbf{1}_{\{X<x\}}\big) = \mathbb{E}(Y \mid X < x).$$

Ez éppen a belátandó állítás. □

[49] A „legjobb közelítés” másik megfogalmazását lásd a harmadik alfejezet megfelelő állításában.

11.2. Folytonos regresszió

A regresszió fogalmát abban az esetben is szeretnénk bevezetni, ha X folytonos valószínűségi változó. Ezzel az a lényeges probléma, hogy $\mathbb{P}(X = x) = 0$ minden x érték esetén, ezért $\mathbb{E}(Y \mid X = x)$ a fenti definícióval értelmetlen. De ne adjuk fel rögtön, ugyanis az „Y legjobb közelítése X függvényében” fogalom viszont nem tűnik értelmetlennek.

A regressziót általánosan az előző alfejezet utolsó állításával definiálhatjuk.

11.2.1. Definíció. Legyenek X és Y valószínűségi változók. Az Y-nak az X-re vett **regresszió**ja az a $g(X)$ alakú[50] valószínűségi változó, amire

$$\mathbb{E}\big(g(X) \mid X < x\big) = \mathbb{E}(Y \mid X < x) \tag{11}$$

minden olyan $x \in \mathbb{R}$ esetén, amire $\mathbb{P}(X < x) > 0$ (ami miatt $\mathbb{E}(. \mid X < x)$ definiált az előző alfejezet feltélteles várható érték definíciójával). A $g(X)$ valószínűségi változó jelölése: $\mathbb{E}(Y \mid X)$.

Mivel a diszkrét regresszió is regresszió (lásd előző alfejezet utolsó állítása), így a diszkrét jelzőt elhagyjuk, és a továbbiakban minden regresszió alatt ezt a definíciót értjük. Emellett itt jegyezzük meg, hogy az $\mathbb{E}(Y \mid X)$ regresszió, mint valószínűségi változó, nem egyértelmű (de azért majdnem). Ugyanúgy, ahogy a sűrűségfüggvény sem egyértelmű, ugyanis g értékét néhány olyan ponton büntetlenül megváltoztathatjuk, amiket 0 valószínűséggel vesz fel X. Egy ilyen változtatás $g(X)$ értékét is megváltoztatja, de a definícióban szereplő egyenlet érvényben marad.

A g függvény[51] jelölése: $x \mapsto \mathbb{E}(Y \mid X = x)$, elnevezése **regressziós függvény**. A jelölés több értelemben sem pontos. Egyrészt a fejezet legelején felírt feltételes várható érték értelemben az $\mathbb{E}(Y \mid X = x)$ nem feltétlenül van értelmezve, hiszen $\mathbb{P}(X = x)$ lehet nulla is. Ettől függetlenül a jelölés szemléletes, hiszen informálisan $g(x)$ az Y átlagos legjobb közelítése az $\{X = x\}$ feltétel esetén.

Másrészt a jelölés abban az értelemben sem precíz, hogy g egyáltalán nem egyértelmű, így rögzített x-re a függvénynek nincs jóldefiniált értéke. Például ha X nemnegatív, akkor g a negatív félegyenesen akárhogy megválasztható. Mivel g kiszámolása tipikusan egy köztes lépés az $\mathbb{E}(Y \mid X)$ valószínűségi változó kiszámolásához, amely valószínűségi változó már lényegében egyértelmű (lásd előző bekezdés), így g nem egyértelmű voltával nem fogunk a továbbiakban foglalkozni.

Megjegyzés. Ha $\mathbb{E}(|Y|) < \infty$, akkor $\mathbb{E}(Y \mid X)$ is létezik. Ezt nem bizonyítjuk.

Rendben, most már definiálva van a regresszió. De hogyan lehet kiszámolni akár a g regressziós függvényt, akár a $g(X) = \mathbb{E}(Y \mid X)$ regressziót? Ha X egyszerű, akkor ez az előző alfejezet (10) egyenletéből világos:

$$\mathbb{E}(Y \mid X = x) = \sum_{y \in \mathrm{Ran}(Y)} y \cdot \mathbb{P}(Y = y \mid X = x)$$

minden $x \in S_X$ esetén, hiszen ez a $\mathbb{P}(. \mid X = x)$ valószínűségi mérték szerinti várható érték. Folytonos esetben ehelyett a következőt tudjuk mondani.

11.2.2. Definíció. Legyen (X, Y) folytonos valószínűségi vektorváltozó, és jelölje együttes sűrűségfüggvényét $f_{X,Y}$. Ekkor Y-nak az X-re vett **feltételes sűrűségfüggvénye**:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int_{-\infty}^{\infty} f_{X,Y}(x,u)\mathrm{d}u},$$

olyan $x, y \in \mathbb{R}$ számokra értelmezve, amire $f_X(x) \neq 0$, és $f_{Y|X}(y \mid x) = 0$ ha $f_X(x) = 0$.

[50] A g nem akármilyen függvény, hiszen $g(X)$ valószínűségi változó kell legyen. Ehhez általában azt követelik meg, hogy g úgynevezett Borel-mérhető függvény legyen; ekvivalensen, folytonos függvények pontonkénti határértéke.

[51] A g értelmezési tartományát nem specifikáltuk, nem is igazán tudjuk, hiszen g nem egyértelmű. Ami lényeges feltétel a g értelmezési tartományára, hogy tartalmazza azon $x \in \mathbb{R}$ pontokat, aminek bármilyen kis környezetébe pozitív eséllyel esik X, legfeljebb egy 0 mértékű halmaz kivételével.

A definíció megjegyzésében segíthet, ha észrevesszük a hasonlóságát a $\mathbb{P}(B \mid A) = \frac{\mathbb{P}(B\cap A)}{\mathbb{P}(A)}$ egyenlőséggel.

11.2.3. Állítás. *Legyen (X, Y) folytonos valószínűségi vektorváltozó. Ekkor*

$$\mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\mathrm{d}y$$

az Y-nak az X-re vett regressziós függvénye.

Bizonyítás. Jelölje g az $z \mapsto \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid z)\mathrm{d}y$ függvényt ($z \in \mathbb{R}$). Azt kell leellenőriznünk, hogy g-re teljesül a (11) egyenlet.

Legyen $x \in \mathbb{R}$ olyan, amire $\mathbb{P}(X < x) > 0$. A fejezet első állítása miatt

$$\mathbb{E}\big(g(X) \mid X < x\big) = \frac{1}{\mathbb{P}(X < x)}\mathbb{E}\big(g(X)\mathbf{1}_{\{X<x\}}\big),$$

ahol

$$\mathbb{E}\big(g(X)\mathbf{1}_{\{X<x\}}\big) = \int_{-\infty}^{\infty} g(z)\mathbf{1}_{\{z<x\}} f_X(z)\mathrm{d}z = \int_{-\infty}^{x} g(z) f_X(z)\mathrm{d}z.$$

Ha $f_X(z) > 0$, akkor g definíciója szerint

$$g(z)f_X(z) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid z)\mathrm{d}y \cdot f_X(z) = \int_{-\infty}^{\infty} y \cdot \frac{f_{X,Y}(z, y)}{f_X(z)} f_X(z)\mathrm{d}y = \int_{-\infty}^{\infty} y \cdot f_{X,Y}(z, y)\mathrm{d}y.$$

Ha $f_X(z) = 0$, akkor $g(z)f_X(z) = 0$. Leellenőrizhető, hogy a jobb oldali integrál szintén 0, ha $f_X(z) = 0$, felhasználva, hogy $\int_{-\infty}^{\infty} f_{X,Y}(z, y)\mathrm{d}y = f_X(z)$. Összességében, az előző három egyenletből az adódik,

$$\mathbb{E}\big(g(X) \mid X < x\big) = \frac{1}{\mathbb{P}(X < x)} \int_{-\infty}^{x}\int_{-\infty}^{\infty} y \cdot f_{X,Y}(z, y)\mathrm{d}y\mathrm{d}z.$$

Itt felhasználhatjuk a valószínűségi vektorváltozó transzformáltjának várható értékére vonatkozó állítást, és a fejezet első állítását, így a fentit folytatva

$$= \frac{1}{\mathbb{P}(X < x)}\mathbb{E}(Y\mathbf{1}_{\{X<x\}}) = \mathbb{E}(Y \mid X < x),$$

ami éppen a belátandó állítás. □

11.2.4. Példa. Legyen az (X, Y) valószínűségi vektorváltozó együttes sűrűségfüggvénye $15x^2y$, ha $0 < x < y < 1$, és 0 egyébként. Határozzuk meg az $\mathbb{E}(Y \mid X)$ regressziót.
Először számoljuk ki X sűrűségfüggvényét a $0 < x < 1$ esetekben:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\mathrm{d}y = \int_{x}^{1} 15x^2y\mathrm{d}y = \frac{15}{2}(x^2 - x^4).$$

Tehát a feltételes sűrűségfüggvény értéke $0 < x < y < 1$ esetén

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{15x^2y}{\frac{15}{2}(x^2 - x^4)} = \frac{2y}{1 - x^2},$$

és $f_{Y|X}(y \mid x) = 0$ egyébként. Tehát $0 < x < 1$ esetén

$$\mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\mathrm{d}y = \int_{x}^{1} y \cdot \frac{2y}{1 - x^2}\mathrm{d}y = \frac{2}{3} \cdot \frac{1 - x^3}{1 - x^2}.$$

Behelyettesítve, az $\mathbb{E}(Y \mid X)$ regresszió a $\frac{2}{3} \cdot \frac{1-X^3}{1-X^2}$ valószínűségi változó.[52]

[52] Érdekes utánaszámolni, hogy a regresszió egyáltalán nem szimmetrikus X-ben és Y-ban, azaz $\mathbb{E}(X \mid Y)$ egyáltalán nem biztos, hogy hasonlít $\mathbb{E}(Y \mid X)$-re.

11.3. Regresszió tulajdonságai, teljes várható érték tétele

A regresszió könnyebb meghatározásához érdemes megvizsgálnunk a tulajdonságait, ahogy azt a korábbi fogalmak esetében is tettük.

11.3.1. Állítás. *Legyenek X, Y, Z valószínűségi változók. Ekkor teljesülnek a következők:*

(1) Tetszőleges $a, b \in \mathbb{R}$ esetén $\mathbb{E}(aY + bZ \mid X) = a\mathbb{E}(Y \mid X) + b\mathbb{E}(Z \mid X)$.
(2) Tetszőleges h folytonos[53] *függvény esetén $\mathbb{E}\big(h(X)Y \mid X\big) = h(X)\mathbb{E}(Y \mid X)$.*
(3) Ha X és Y függetlenek, akkor $\mathbb{E}(Y \mid X) = \mathbb{E}(Y)$.

A linearitás nem meglepő. A második tulajdonság szemléletesen azt állítja, hogy mivel $h(X)$ értéke meghatározható X-ből, ezért ha a $h(X)Y$-ra keressük a legjobb becslést X függvényében, akkor elég az Y becslését megoldanunk, a $h(X)$ szorzóként kiemelhető a várható értékből. A harmadik tulajdonság jelentése, hogy ha X és Y függetlenek, akkor X-ből nem tudunk jobb becslést faragni Y-ra, mint a legjobb konstans becslés, ami a várható értéke.

Az első alfejezetben párhuzamot vontunk a regresszió és a lineáris regresszió közt: mindkét módszer jó közelítést keres X függvényében Y-ra. A lineáris regresszió esetében pontosan meg is fogalmaztuk, hogy mi az az optimalizálási probléma, amit a lineáris regresszió megold (sőt, ez volt a definíció). Hasonlóan, a regresszió is felírható optimalizálási probléma megoldásaként.[54]

11.3.2. Állítás. *Tegyük fel, hogy $\mathbb{E}(Y^2)$ véges. Ekkor a*

$$\mathbb{E}\Big(\big(Y - g(X)\big)^2\Big)$$

várható érték pontosan akkor minimális a g függvényben, ha $g(X)$ és $\mathbb{E}(Y \mid X)$ 1-valószínűséggel megegyeznek.

Röviden, itt nem X lineáris függvényével próbáljuk minimalizálni az Y-tól való átlagos négyzetes eltérést, hanem ennél általánosabb függvények segítségével. Ilyen értelemben a regresszió a lineáris regresszió javítása.

Azt gondolhatnánk, hogy a regresszió szinte sosem egyezik meg a lineáris regresszió esetével, csak teljesen elfajuló esetekben. A következő állítás mutatja, hogy ez nem igaz.

11.3.3. Állítás. *Legyenek $Z_1, \dots, Z_n$ független, normális eloszlású valószínűségi változók, $X = \sum_{i=1}^n a_i Z_i$ és $Y = \sum_{i=1}^n b_i Z_i$ valamilyen $a_1, \dots, a_n, b_1, \dots, b_n$ valós számokra. Ekkor $\mathbb{E}(Y \mid X)$ megegyezik az Y változó X-re vett lineáris regressziójával.*

Hogyan tudjuk hasznosítani a regressziót olyan problémában, ahol a probléma kérdésfelvetésében nem szerepel feltételes várható érték? Úgy, hogy a szokásos várható érték is számolható regresszió segítségével ugyanúgy, ahogy a második előadáson a szokásos valószínűséget is számoltunk feltételes valószínűségekkel (lásd pl. teljes valószínűség tétele). Ilyen módszer a teljes várható érték tétele is.

11.3.4. Állítás (Teljes várható érték tétele)**.** *Legyen X és Y valószínűségi változó, amikre $\mathbb{E}(X)$ és $\mathbb{E}(Y)$ véges. Ekkor $\mathbb{E}\big(\mathbb{E}(Y \mid X)\big) = \mathbb{E}(Y)$.*

Hát ez mi? Ez nem is úgy néz ki, mint ahogy a teljes valószínűség tétele kinézett. Hol a szumma? Minek két várható érték jel egymás után? Nos, a szumma a külső várható értékben van elrejtve. A várható értékre pedig azért van szükség, mert – ahogy feljebb megvizsgáltuk – $\mathbb{E}(Y \mid X)$ egy valószínűségi változó, így ha számot szeretnénk kapni belőle, ehhez vehetjük a várható értékét.

További érv amellett, hogy a fenti állítás elnevezése találó, hogy ha a feltételes várható értéket kiírjuk a diszkrét esetre, akkor pont olyan formulát kapunk, mint a teljes valószínűség tételében.

---

[53] A folytonosság itt elégséges feltétel, de valójában csak arra a gyengébb feltételre van szükségünk, hogy $h(X)$ is valószínűségi változó legyen.

[54] Hogy nem egy minimalizálási problémával definiáltuk a regressziót, annak az az oka, hogy a karakterizáció feltételezi, hogy $\mathbb{E}(Y^2)$ véges, míg a regresszió létezéséhez elég, ha $\mathbb{E}(Y)$ véges.

11.3.5. Állítás (Teljes várható érték tétele, diszkrét eset)**.** *Legyen X diszkrét valószínűségi változó, és $\{x_1, \ldots, x_n, \ldots\}$ az X értékkészletének azon pontjai, amire $\mathbb{P}(X = x_i) > 0$. Ha $\mathbb{E}(Y) < \infty$, akkor*

$$\mathbb{E}(Y) = \sum_{i=1}^{\infty} \mathbb{E}(Y \mid X = x_i)\mathbb{P}(X = x_i).$$

11.3.6. Példa. Számoljuk ki a geometriai eloszlás szórását. (Oké, már kiszámoltuk máshogy, de ez rövidebb.) Dobáljunk fel egy cinkelt pénzérmét, amíg fejet nem kapunk, ahol a fej esélye p. Jelölje a szükséges dobások számát Y, és legyen $X = 1$, ha az első dobás fej, különben 0. Ekkor

$$\mathbb{E}(Y^2) = \mathbb{E}(Y^2 \mid X = 1)\mathbb{P}(X = 1) + \mathbb{E}(Y^2 \mid X = 0)\mathbb{P}(X = 0).$$

Ha az első dobás írás, akkor összesen eggyel többet kell majd várnunk, mintha most kezdenénk a dobálást, a geometriai eloszlás örökifjú tulajdonsága miatt. Egyenlettel $\mathbb{P}(Y = k \mid X = 0) = \mathbb{P}(Y + 1 = k)$ minden k pozitív egészre. Vagyis $\mathbb{E}(Y^2 \mid X = 0) = \mathbb{E}\big((Y + 1)^2\big)$, hiszen Y-nak a $\mathbb{P}(. \mid X = 0)$ valószínűségi mérték szerinti eloszlása megegyezik az Y szokásos értelemben vett eloszlásával. Következésképp,

$$\begin{aligned}\mathbb{E}(Y^2) &= 1 \cdot \mathbb{P}(X = 1) + \mathbb{E}\big((Y + 1)^2\big)\mathbb{P}(X = 0) = \\ &= p + \mathbb{E}(Y^2 + 2Y + 1)(1 - p),\end{aligned}$$

amit átrendezve $\mathbb{E}(Y^2) = \frac{2-p}{p^2}$ adódik, felhasználva hogy $\mathbb{E}(Y) = \frac{1}{p}$. Így $\mathbb{D}^2(Y) = \frac{1-p}{p^2}$.

A tétel teljes eseményrendszerre is kimondható.

11.3.7. Állítás (Teljes várható érték tétele, teljes eseményrendszerrel)**.** *Legyen $A_1, \ldots, A_n$ teljes eseményrendszer Ω-n, amire $\mathbb{P}(A_i) > 0$ minden i-re. Ha $\mathbb{E}(Y) < \infty$, akkor*

$$\mathbb{E}(Y) = \sum_{i=1}^{n} \mathbb{E}(Y \mid A_i)\mathbb{P}(A_i).$$

Folytonos esetben pedig a következőképp írható.

11.3.8. Állítás (Teljes várható érték tétele, folytonos eset)**.** *Legyen X folytonos valószínűségi változó, $\mathbb{E}(Y) < \infty$ és jelölje X sűrűségfüggvényét f_X. Ekkor*

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} \mathbb{E}(Y \mid X = x) f_X(x)\mathrm{d}x,$$

ahol $\mathbb{E}(Y \mid X = x)$ az Y-nak az X-re vett regressziós függvénye.