A november 18.-i gyakorlat témája

Egy F eloszlásfüggvénnyel rendelkező ξ valószínűségi változó $E\xi$ illetve adva egy $g(x)$ függvény a ξ valószínűségi változó $g(\xi)$ függvényének az $Eg(\xi)$ várható értékét az

$$E\xi = \int_{-\infty}^{\infty} u dF(u), \qquad Eg(\xi) = \int_{-\infty}^{\infty} g(u)\, dF(u) \tag{a}$$

képletek segítségével számíthatjuk ki. A fenti képletben úgynevezett Stieltjes integrál szerepel, ami közvetlenül nehezen számolható. Ezért fontos megérteni, hogy egyszerűsödik a képlet bizonyos fontos speciális esetekben. Ezek egyike az az eset, amikor létezik sűrűségfüggvény. Meg kell értenünk, hogyan tudjuk megállapítani, hogy létezik sűrűségfüggvény, és sűrűségfüggvény létezése esetén hogyan számíthatjuk ki az (a) formulában szereplő képletet. Ezért érdemes a következő tényeket tudni:

Tétel sűrűségfüggvény létezéséről. *Legyen $F(x)$, monoton, minden pontjában folytonos függvény, amelyik véges sok kivételével deriválható. Tegyük fel továbbá azt is, hogy $\lim\limits_{x\to-\infty} F(x) = 0$. Definiáljuk az $f(x) = \dfrac{dF(x)}{dx}$ függvényt azokban a pontokban, ahol $F(x)$ differenciálható, és tetszőleges módon abban a (véges sok) pontban, ahol $F(x)$ nem differenciálható. Ekkor $F(x) = \int_{-\infty}^{\infty} fu)\, du$ minden $-\infty < x < \infty$ pontban, és*

$$\int_{-\infty}^{\infty} u dF(u) = \int_{-\infty}^{\infty} u f(u)\, du, \quad Eg(\xi) = \int_{-\infty}^{\infty} g(u)\, dF(u) = \int_{-\infty}^{\infty} g(u) f(u)\, du.$$

Megjegyzés: A fenti tételben szereplő $F(x)$ függvény akkor eloszlásfüggvény, ha a fent említett tulajdonságok mellett teljesíti az $\lim\limits_{x\to\infty} F(x) = 1$ feltételt. Ezt a feltételt azért nem követeltük meg, mert mint látni fogjuk a fenti kissé általánosabban megfogalmazott tétel jobban használható bizonyos esetekben.

1.) Dobjunk le egy pontot egyenletes eloszlással valamely $[a, b]$ intervallumba, azaz annak valószínűsége, hogy a pont valemely $[u, v] \subset [a, b]$ intervallumba esik legyen $\frac{v-u}{b-a}$. Ekkor a ledobott pont ξ helyének az eloszlásfüggvénye $F(x) = 0$, ha $x \le a$, $F(x) = \frac{x-a}{b-a}$, ha $a \le x \le b$, $F(x) = 1$, ha $x \ge b$. $E\xi = \frac{1}{b-a} \int_a^b u\, du$, és $Eg(\xi) = \frac{1}{b-a} \int_a^b g(u)\, du$ tetszőleges $g(x)$ függvényre.

   Megoldás: Nem nehéz ellenőrizni, hogy a fenti ξ valószínűségi változónak a feladatban definiált $F(x)$ függvény az eloszlásfüggvénye. Ez az $F(x)$ függvény minden pontjában folytonos, és két pont, $x = a$ és $x = b$ pontok, kivételével differenciálható, $f(x) = \frac{dF(x)}{dx} = \frac{1}{b-a}$, ha $a < x < b$, és $f(x) = \frac{dF(x)}{dx} = 0$, ha $x < a$ vagy $x > b$. Ezért az előző tétel alapján $E\xi = \int_{-\infty}^{\infty} u f(u)\, du = \int_a^b \frac{1}{b-a} u\, du$, és $Eg(\xi) = \int_{-\infty}^{\infty} g(u) f(u)\, du = \int_a^b \frac{1}{b-a} g(u)\, du$.

A másik külön tárgyalandó eset az, amikor diszkrét eloszlású valószínűségi változó várható értékét akarjuk kiszámolni. Ezen eset vizsgálata érdekében fogalmazzuk meg a következő eredményt.

Tétel diszkrét eloszlású valószínűségi változó várható értékének a kiszámolásáról. *Legyen ξ olyan valószínűségi változó, amelyik véges sok $a_1 < a_2 < \cdots < a_k$ értéket vesz fel, $P(\xi = a_j) = p_j$, $1 \le j \le k$, $\sum\limits_{j=1}^{k} p_j = 1$. Ekkor a xi valószínűségi változó $F(x)$ eloszlásfüggvénye a következő alakú: $F(x) = 0$, ha $x \le a_1$, $F(x) = \sum\limits_{l=1}^{j} p_j$, ha $a_j < x < a_{j+1}$, $1 \le j \le k-1$, $F(x) = 1$, ha $x > a_k$.*

Legyen kissé általánosabban $F(x)$ olyan függvény, amely megadható valamilyen $a_1 < a_1 < \cdots < a_k$ és $p_1 > 0$ $p_2 > 0, \dots$ $p_k > 0$ sorozatok segítségével úgy, hogy $F(x) = 0$, ha $x \le a_1$, $F(x) = \sum\limits_{l=1}^{j} p_l$, ha $a_j < x < a_{j+1}$, $1 \le j \le k-1$, $F(x) = \sum\limits_{l=1}^{k} p_l$, ha $x > a_k$. Ekkor $\int u\, dF(u) = \sum\limits_{l=1}^{k} a_k$, és $\int g(u)\, dF(u) = \sum\limits_{l=1}^{k} g(a_k)$.

Megjegyzés: A fenti tétel második bekezdésében nem követeltük meg a $\sum\limits_{j=1}^{k} p_j = 1$ feltételt. Ez az enyhe általánosítás lehetővé teszi, hogy az előző tétel és az alább kimondott lemma segítségével ki tudjuk számítani a várható értéket kissé általánosabb, de szintén érdekes esetekben. Erre mutat példát az alábbi 2. feladat.

Lemma. *Legyen F_1 és F_2 két olyan monoton függvény a számegyenesen, amelyekre $\lim\limits_{x\to-\infty} F_j(x) = 0$ és $\lim\limits_{x\to\infty} F_j(x) < \infty$, $j = 1,2$. Ekkor léteznek az $\int_{-\infty}^{\infty} g(u)\, dF_j(u)$, $j = 1,2$ és $\int_{-\infty}^{\infty} g(u)\, d(F_1(u) + F_2(u))$ integrálok minden szép $g(u)$ függvényre. Továbbá,*

$$\int_{-\infty}^{\infty} g(u)\, d(F_1(u) + F_2(u)) = \int_{-\infty}^{\infty} g(u)\, d(F_1(u) + \int_{-\infty}^{\infty} g(u)\, d(F_1(u)$$

A következő példában olyan valószínűségi változóra mutatunk példát, melynek $F(x)$ eloszlásfüggvénye felírható $F(x) = F_1(x) + F_2(x)$ alakban úgy, hogy $F_1(x)$ és $F_2(x)$ teljesíti a fenti két tétel feltételeit. Ezért ki tudjuk számolni a ξ valószínűségi változó függvényeinek a várható értékét.

2.) Dobjunk le egy pontot egyenletes eloszlással a $[0,4]$ intervallumra. Írjuk fel e pont értékét, ha a ledobott pont a $[0,2]$ intervallumba esik, írjuk le a 2 számot, ha a pont a $[2,3]$, a 3 számot, ha a pont a $(3,4]$ intervallumba esik. Mutassuk meg, hogy a ξ valószínűségi változó $F(x)$ eloszlásfüggvénye a következő: $F(x) = 0$, ha $x \le 0$, $F(x) = \frac{x}{4}$, ha $0 \le x \le 2$, $F(x) = \frac{3}{4}$, ha $x < 2 \le 3$, és $F(x) = 1$, ha $x \le 3$. Mutassuk meg továbbá, hogy $F(x)$ felírható $F(x) = F_1(x) + F_2(x)$ alakban a következő módon: $F_1(x) = \int_{-\infty}^{x} f(u)\, du$, $f(u = \frac{1}{2}$, ha $0 \le x \le 2$, $f(u) = 0$, ha $u < 0$ vagy $u > 2$, és $F_2(x) = 0$, ha $x \le 2$, $F(x) = \frac{1}{4}$, ha $2 < x \le 3$, $F(x) = \frac{1}{2}$, ha $x \ge 3$. Ezért $Eg(\xi) = \frac{1}{4} \int_0^2 g(u)\, du + \frac{1}{4} g(2) + \frac{1}{4} g(3)$.

Megoldás: Hasonlóan a múlt órán tárgyalt 2. feladathoz ellenőrizhetjük, hogy a ξ valószínűségi változó eloszlásfüggvénye a feladatban megadott $F(x)$ eloszlásfüggvény. Az $F(x)$ függvényt felírhatjuk $F(x) = F_1(x) + F_2(x)$ alakban, ahol $F_1(x) = 0$, ha $x \le 0$, $F_1(x) = \frac{1}{4}x$, ha $0 \le x \le 2$, $F_1(x) = \frac{1}{2}$, ha $x \ge 2$, és $F_2(x) = 0$, ha $x \le 2$, $F_2(x) = \frac{1}{4}$, ha $2 < x \le 3$, $F_2(x) = \frac{1}{2}$, ha $x > 3$. Tulajdonképpen azt tettük, hogy leválasztottuk a tisztán ugró függvény részt az $F(x)$ eloszlásfüggvényről, így kaptuk az $F_2(x)$ függvényt, a maradék $F_1(x) = F(x) - F_2(x)$ függvény pedig mindenütt folytonos. Az $F_1(x)$ függvény minden pontban folytonos, és 0 és 2 pont kivételével mindenütt deriválható. Ezért a sűrűségfüggvény létezéséről szóló tétel segítségével felírhatjuk, hogy $F_1(x) = \int_{-\infty}^{\infty} f(u)\,du$, és $\int g(u)\,dF_1(u) = \int_{-\infty}^{\infty} g(u)f(u)\,du$ minden $-\infty < x < \infty$ számra, ahol $f(x) = \frac{1}{4}$, ha $0 \le x \le 2$, és $f(x) = 0$, ha $x < 0$ vagy $x > 2$. Másrészt a diszkrét eloszlású valószínűségi változó várható értékének a kiszámolásáról szóló tétel alapján $F_2(x)$ felírható az ott megadott módon $a_1 = 2$, $a_2 = 3$, $p_1 = p_2 = \frac{1}{4}$ paraméterekkel. Ezért $\int_{-\infty}^{\infty} g(u)\,dF_2(u) = \frac{1}{4}g(2) + \frac{1}{4}g(3)$. Ezután a Lemma segítségével kapjuk a feladat végső állítását.

A fenti feladat megoldási módszere azt jelzi, hogy az eloszlásfüggvény felbontásának segítségével egy tisztán ugró és folytonos részre a kívánt várható értéket ki lehet számolni a fenti két tétel és lemma segítségével. Felmerülhet a kérdés, hogy van-e olyan eloszlásfüggvény, amelynek esetében a fenti ismeretek nem elegendőek a várható érték kiszámolásához. A válasz az, hogy lehet rosszindulatú és mesterkélt módon ilyen példákat konstruálni, de a gyakorlatban előforduló összes feladat megoldásához elegendő a fenti eredmények használata (és a megfelelő integrálok és összegek kiszámítása).

Házi feladat:

Feldobunk egy szabályos dobókockát. Ha a dobás értéke 1, 2, 3, 4 vagy 5 akkor felírjuk a dobás értékét egy papírlapra. Ha dobás értéke 6, akkor ledobunk egy pontot véletlenül egyenletes eloszlással a $[0, 2]$ intervallumra, és a ledobott pont helyének az értékét írjuk fel a papírlapra. Számoljuk ki a felírt pont várható értékét és szórásnégyzetét.

3.) Legyen ξ valószínűségi változó $f(u)$ sűrűségfüggvénnyel, a és b valós számok. Határozzuk meg az $a\xi + b$, $a > 0$, és ξ^2 valószínűségi változók sűrűségfüggvényét.

Megoldás: Legyen $F(x) = \int_{-\infty}^{\infty} f(u)\,du$ ξ eloszlásfüggvénye. Ekkor $\eta = a\xi + b$ $G(x)$ eloszlásfüggvénye $G(x) = P(a\xi + b < x) = P\left(\xi < \frac{x-b}{a}\right) = F\left(\frac{x-b}{a}\right)$ sűrűségfüggvénye pedig ennek deriváltja, $\frac{d}{dx}G(x) = \frac{d}{dx}F\left(\frac{x-b}{a}\right) = \frac{1}{a}f\left(\frac{x-a}{b}\right)$. A ξ^2 valószínűségi változó $H(x)$ eloszlásfüggvénye $H(x) = P(\xi^2 < x) = P(-\sqrt{x} < \xi < \sqrt{x}) = P(\xi < \sqrt{x}) - P(\xi \le -\sqrt{x})$, ha $x \ge 0$, és $H(x) = P(\xi^2 < x) = 0$, ha $x < 0$. A ξ^2 valószínűségi változó $h(x)$ sűrűségfüggvénye pedig a $H(x)$ függvény deriváltja, $h(x) = 0$, ha $x < 0$, $h(x) = \frac{1}{2\sqrt{x}}\left(f(\sqrt{x}) + f(-\sqrt{x})\right)$.

Megtárgyaljuk a többváltozós eloszlásfüggvény fogalmát, és az ezzel kapcsolatos legfontosabb fogalmakat és eredményeket. Ezek közül külön felhívom a figyelmet a valószínűségi változók függetlenségének fogalmára az általános (nem feltétlenül diszkrét eloszlású) valószínűségi változók esetében.

Többdimenziós eloszlásfüggvény definiciója. *Legyen adva k valós értékű $\xi_1, \dots, \xi_k$ valószínűségi változó egy $(\Omega, \mathcal{A}, P)$ valószínűségi mezőn. Ezek (együttes) eloszlásfüggvé-*

nye az

$$F(x_1, \dots, x_k) = P(\xi_1 < x_1, \dots, \xi_k < x_k),$$

k változós függvény, ahol $-\infty < x_j < \infty$ *minden* $1 \le j \le k$ *indexre.*

Tétel. *Egy* $F(u_1, \dots, u_k)$ *függvény akkor és csak akkor eloszlásfüggvénye alkalmas* $\xi_1, \dots, \xi_k$ *valószínűségi változóknak valamely* $(\Omega, \mathcal{A}, P)$ *valószínűségi mezőn, ha ez az* F *függvény teljesíti a következő (i)—(iv) tulajdonságokat.*

(i) $F(u_1, \dots, u_k)$ *minden változójának balról folytonos függvénye.*

(ii) $\displaystyle\lim_{\substack{u_j \to \infty \\ \text{minden } j=1,\dots,k \text{ számra}}} F(u_1, \dots, u_k) = 1.$

(iii) $\displaystyle\lim_{\substack{u_j \to -\infty \\ \text{valamely } 1 \le j \le k \text{ számrra}}} F(u_1, \dots, u_k) = 0.$ *(Ez úgy értendő, hogy az összes* u_s*,* $1 \le s \le k$*,* $s \ne j$ *koordinátát rögzítjük, és* $u_j \to -\infty$*.)*

Végül definiáljuk egy az R^k *téren definiált* F *függvényre és egy* $\mathbf{K} = [a_1, b_1) \times \cdots \times [a_k, b_k)$ *téglatestre a*

$$\mu(\mathbf{K}) = \mu_F(\mathbf{K}) = \sum_{\substack{u_j = \, a_j \text{ vagy } b_j \\ j=1,\dots,k}} (-1)^{\chi(u_1, \dots, u_k)} F(u_1, \dots, u_k)$$

mennyiséget, ahol $\chi(u_1, \dots, u_k)$ *jelöli az* a_j*-k számát az* $u_1, \dots, u_k$ *sorozatban. Ekkor*

(iv) $\mu_F(\mathbf{K}) \ge 0$ *minden* $\mathbf{K}$ *téglatestre.*

Több-dimenziós eloszlásfüggvény sűrűségfüggvényének definiciója. *Azt mondjuk, hogy egy* $F(x_1, \dots, x_k)$ *többváltozós eloszlásfüggvénynek létezik* $f(u_1, \dots, u_k)$ *sűrűségfüggvénye, ha teljesül az*

$$F(x_1, \dots, x_k) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} f(u_1, \dots, u_k) \, du_1 \dots du_k$$

azonosság minden $-\infty < x_j < \infty$*,* $1 \le j \le k$ *számra.*

Tétel. *Ha* $(\xi_1, \dots, \xi_k)$ *valószínűségi változók* $F(x_1, \dots, x_k)$ *eloszlásfüggvényének létezik* $f(u_1, \dots, u_k)$ *sűrűségfüggvénye, akkor ez teljesíti minden* $g(\cdot)$ *k-változós (mérhető) függvényre a következő azonosságot:*

$$\begin{aligned} Eg(\xi_1, \dots, \xi_k) &= \int g(u_1, \dots, u_k) \mu_F(\, du_1, \dots, \, du_k) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(u_1, \dots, u_k) f(u_1, \dots, u_k) \, du_1 \dots \, du_k, \end{aligned}$$

ahol μ_F *jelöli az* F *eloszlásfüggvény által meghatározott Stieltjes mértéket. Ez az azonosság úgy értendő, hogy az annak két oldalán szereplő kifejezés egyszerre létezik vagy* *nem létezik, ha mind a két kifejezést mint Lebesgue integrált értelmezzük. Ha a jobboldalon szereplő integrál létezik mint (közönséges) Riemann integrál, akkor ez az integrál tekinthető úgy is mint (egy a Riemann integrál értékével megegyező) Lebesgue integrál. Tehát az azonosság ebben a fontos speciális esetben is érvényes.*

Az egydimenziós esethez hasonlóan a többdimenziós sűrűségfüggvényeket is lehet jellemezni. Igaz a következő tétel.

Tétel. *Egy k-változós $f(u_1,\dots,u_k)$ függvény akkor és csak akkor sűrűségfüggvénye egy alkalmas k-dimenziós eloszlásfüggvénynek, ha $f(u_1,\dots,u_k)\ge 0$ majdnem minden $(u_1,\dots,u_k)$ pontban, és*

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f(u_1,\dots,u_k)\,du_1\dots\,du_k=1.$$

Valószínűségi változók függetlenségének a definiciója. *Legyenek $\xi_1,\,\dots,\xi_n$ valószínűségi változók egy $(\Omega,\mathcal{A},P)$ valószínűségi mezőn. Azt mondjuk, hogy ezek a valószínűségi változók függetlenek, ha minden $x_1,\dots,x_n$ valós számra*

$$P(\xi_1<x_1,\dots,\xi_n<x_n)=P(\xi_1<x_1)\cdots P(\xi_n<x_n).$$

Tétel. *Legyenek $\xi_1,\dots,\xi_k$ független valószínűségi változók, amelyek mindegyikének létezik várható értéke, azaz $E|\xi_j|<\infty$. Ekkor a $\xi_1\cdots\xi_k$ szorzatnak is létezik várható értéke, és*

$$E\xi_1\cdots\xi_k=E\xi_1\cdots E\xi_k.$$

Tétel. *Legyenek $\xi_1,\dots,\xi_k$ független valószínűségi változók, $B_1,\dots,B_k$ a számegyenes Borel mérhető részhalmazai. Ekkor*

$$P(\xi_1\in B_1,\dots,\xi_k\in B_k)=P(\xi_1\in B_1)\cdots P(\xi_k\in B_k).$$

Valószínűségi változók kovarianciájának a definiciója. *Legyen ξ és η két valószínűségi változó ugyanazon a valószínűségi mezőn, (amelyekre teljesül az $E\xi^2<\infty$, $E\eta^2<\infty$ feltétel.) A ξ és η valószínűségi változók kovarianciafüggvénye*

$$\mathrm{Cov}\,(\xi,\eta)=E[(\xi-E\xi)(\eta-E\eta)].$$

Lemma.

$$\mathrm{Cov}\,(\xi,\eta)=E\xi\eta-E\xi E\eta.$$

Ha ξ és η független valószínűségi változók, akkor $\mathrm{Cov}\,(\xi,\eta)=0$.