Matematikai alapok és valószínűségszámítás

Valószínűségi eloszlások

Binomiális eloszlás

Bevezetés

A tudományos életben megfigyeléseket teszünk, kísérleteket végzünk. Ezek többféle különböző eredményre vezethetnek, és a konkrét eredmény a véletlenen is múlik.

Például, ha az iskolai végzettséget vizsgáljuk, akkor egy személy kiválasztása az kísérlet, amelynek eredménye a változó valamely értéke.

A megfigyelési egység kiválasztása jellemzően véletlenszerűen történik, így a változó bekövetkező értéke is véletlenszerű, ezért a statisztikai változókat ***véletlen változó***knak is nevezzük.

Bevezetés

Egy statisztikai (véletlen) változó akkor jól definiált, ha:

- Ismert az értékkészlete (mik a változó lehetséges értékei)
- Minden megfigyelési egységhez hozzárendelhető a változó egy, és csakis egy értéke.

A valószínűség eloszlása

Egy megfigyelési egység véletlenszerű kiválasztásakor a vizsgálandó változónak bekövetkezik valamilyen értéke, ami egy eseménynek tekinthető. A változó értékei tehát események, egymást kölcsönösen kizáró (diszjunkt) események. Ezen diszjunkt események uniója a biztos eseményt adja.

Ha a populációból véletlenszerűen kiválasztunk egy megfigyelési egységet, akkor a változó különböző értékei bizonyos valószínűséggel következnek be. A változó értékei valószínűségének összege 1.

A valószínűség eloszlása

Egy statisztikai változót pontos megismeréséhez, tudnunk kell az adott változó eloszlását, azaz azt, hogy milyen módon oszlik meg az egységnyi valószínűség a változó különböző értékei között.

A változók eloszlása elméletileg végtelen sokféle lehet, azonban a gyakorlatban kezelhető számú speciális, jól definiált eloszlás valamelyike jellemző a változók túlnyomó többségére.

Diszkrét változók valószínűségi eloszlása

Egy statisztikai változót akkor tekinthetünk diszkrétnek, ha csak véges (kis) számú, egymástól jól elkülönülő értéket vehet fel.

Diszkrét változók esetén a változó eloszlását ismerni annyit jelent, mint ismerni az adott változó értékeit és az értékekhez tartozó valószínűségeket.

Például, ha ismert egy populációban a nemek aránya (relatív gyakorisága), vagyis valószínűsége, akkor azt mondhatjuk, hogy ismerjük a biológiai nem változót.

A Binomiális eloszlás

Tekintsük az alábbi, gyakori kísérleti elrendezést:

- ***n*** számú kísérletet, vagy próbát végzünk
- Minden kísérlet eredménye sikerként vagy kudarcként fogható fel
- A siker valószínűsége, ***p***, próbáról próbára állandó
- Az egyes próbák egymástól függetlenek

Az ilyen kísérleti elrendezés esetén a változó Binomialis eloszlást követ ***n*** és ***p*** paraméterekkel.

Jelölése: B(n,p).

A különböző paraméterekkel jellemezhető binomiális eloszlások egymástól különbözőek lesznek.

A Binomiális eloszlás

Példa: pénzérmét feldobunk egymás után négyszer

Kérdés: hányszor lesz fej a dobás eredménye a négy dobásból?

Statisztikai változót: ‘fejek száma’.

Lehetséges kimenetek: 0, 1, 2, 3, 4

? Mennyi a valószínűsége ezeknek a kimeneteknek?

A Binomiális eloszlás

Példa: pénzérmét feldobunk egymás után négyszer

Kérdés: hányszor lesz fej a dobás eredménye a négy dobásból?

Statisztikai változót: ‘fejek száma’.

Lehetséges kimenetek: 0, 1, 2, 3, 4

? Mennyi a valószínűsége ezeknek a kimeneteknek?

4 fej p(4)=1/16

3 fej p(3)=4/16

2 fej p(2)=6/16

A további kimenetekre is kiszámolható, hogy milyen valószínűséggel fordulhatnak elő:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

A Binomiális eloszlás

A binomiális eloszlás két fontos paramétere:

próbák száma, ***n***

siker (azaz a bennünket érdeklő kimenet) valószínűsége, ***p***.

Ahogy *n* nő, úgy nő a lehetséges kimenetek száma, úgy oszlik meg egyre több érték között a valószínűség.

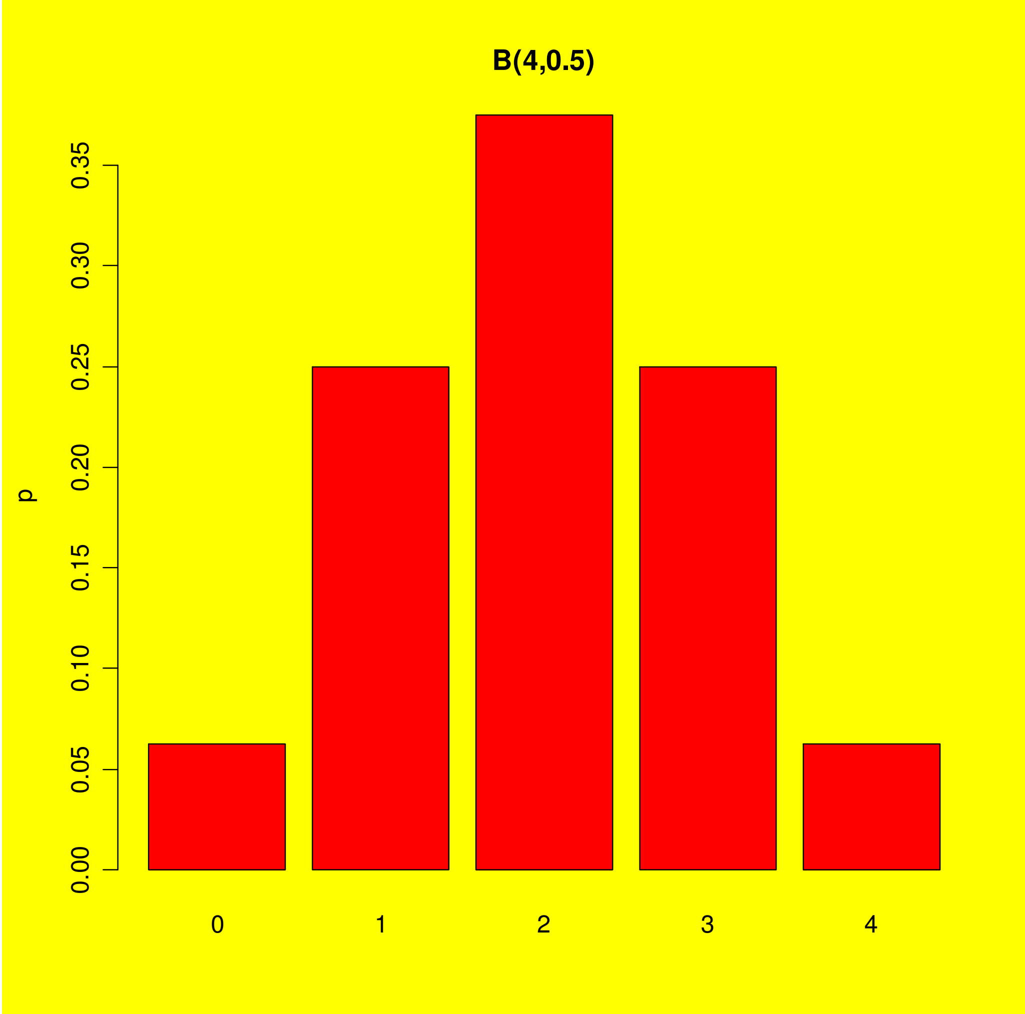
Amennyiben p = 0.5 a binomiális eloszlás szimmetrikus lesz, egyéb esetekben pedig aszimmetrikus, annál aszimmetrikusabb, minél inkább eltér a valószínűség 0.5-től.

Fontos, hogy az egyes kimenetek diszjunkt események, uniójuk pedig a teljes eseményt adja, azaz uniójuk valószínűsége 1.

Példánkban a binomiális eloszlás a következőképpen adható meg:

fejek száma (X; x = 0, 1, 2, 3, 4)	valószínűség $p(X = x)$
0	0.0625
1	0.25
2	0.375
3	0.25
4	0.0625

A B(4, 0.5) eloszlás grafikus reprezentációja



Kumulatív valószínűség

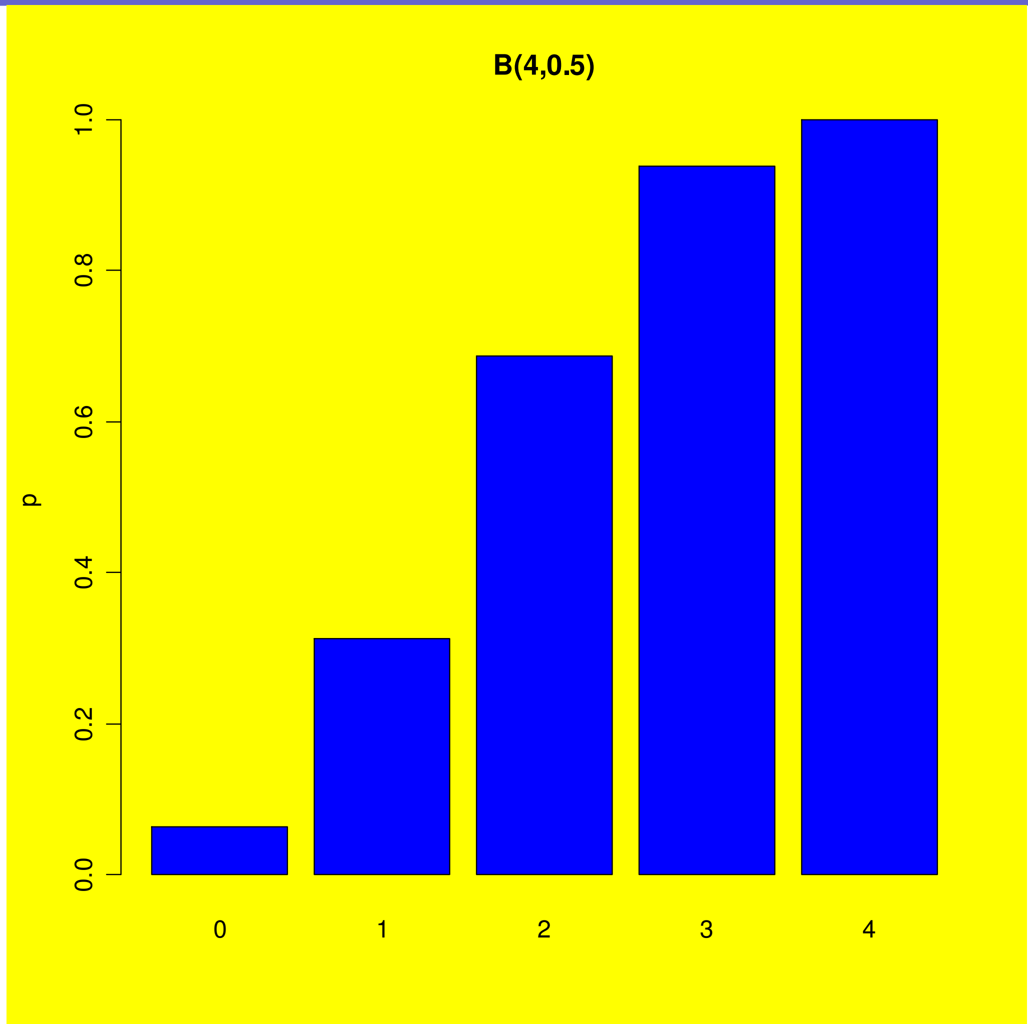
Az egyes értékek valószínûsége mellett a kumulatív valószínûség is meghatározható, ami egy adott érték vagy annál kisebb érték bekövetkezésének valószínûségét adja meg.

A kumulatív valószínûség fontos szerepet játszik a statisztikai hipotézisvizsgálatok során.

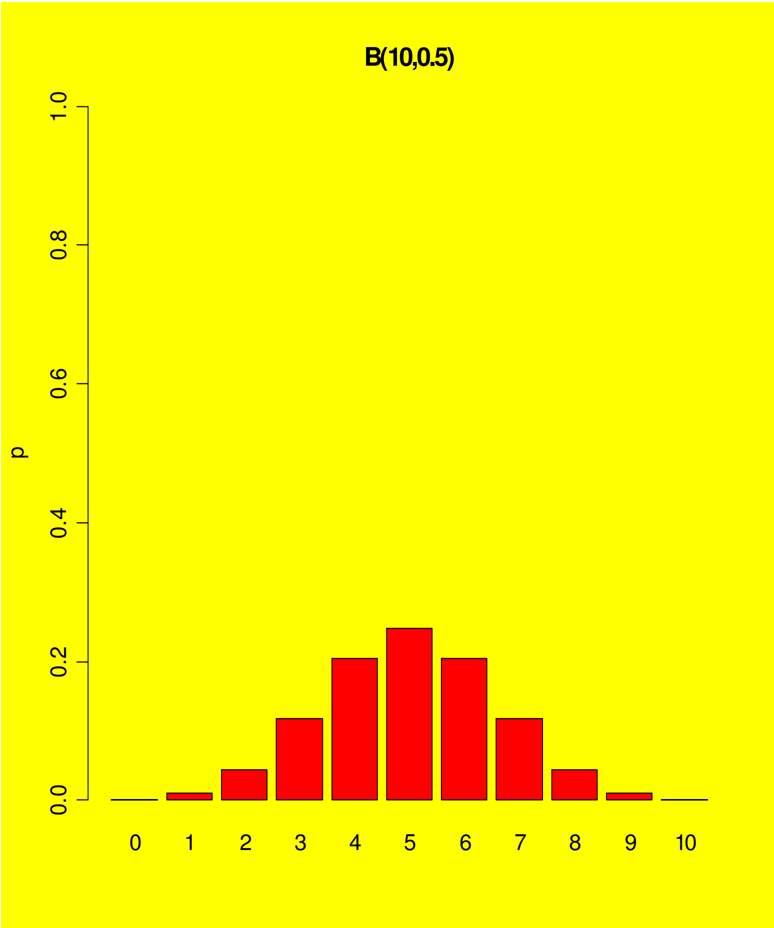
A B(4, 0.5) eloszlás kumulatív valószínűségei

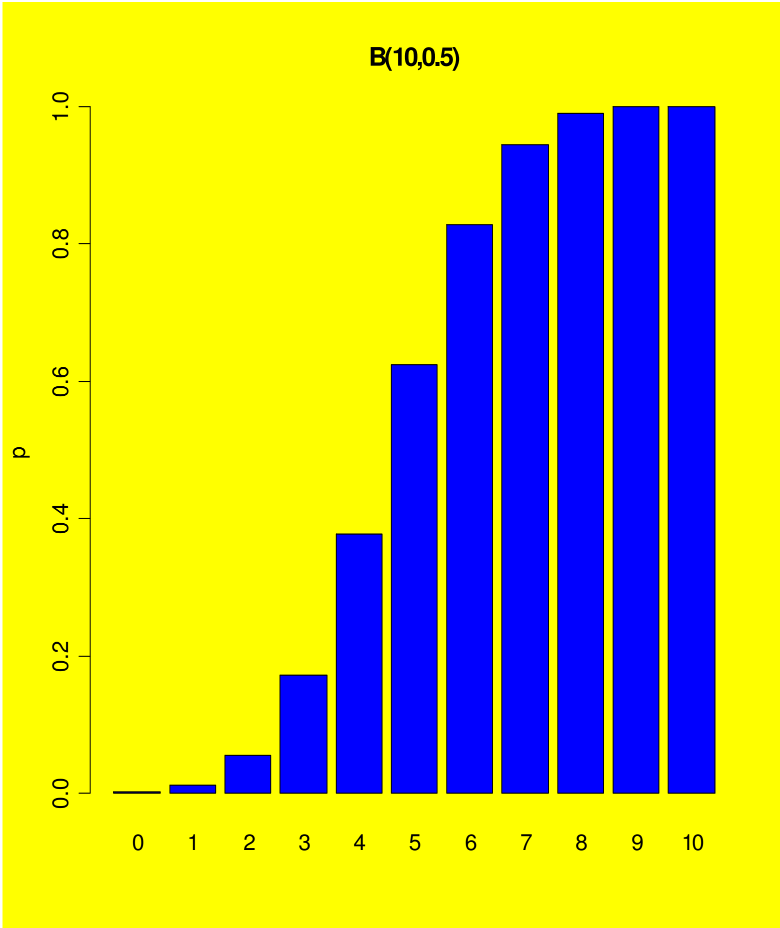
fejek száma (X; x = 0, 1, 2, 3, 4)	valószínűség $p(X = x)$	kumulatív valószínűség $p(X \leq x)$
0	0,0625	0.0625
1	0,25	0,3125
2	0,375	0,6875
3	0,25	0,9375
4	0,0625	1

A B(4, 0.5) eloszlás kumulatív valószínűségei grafikusan

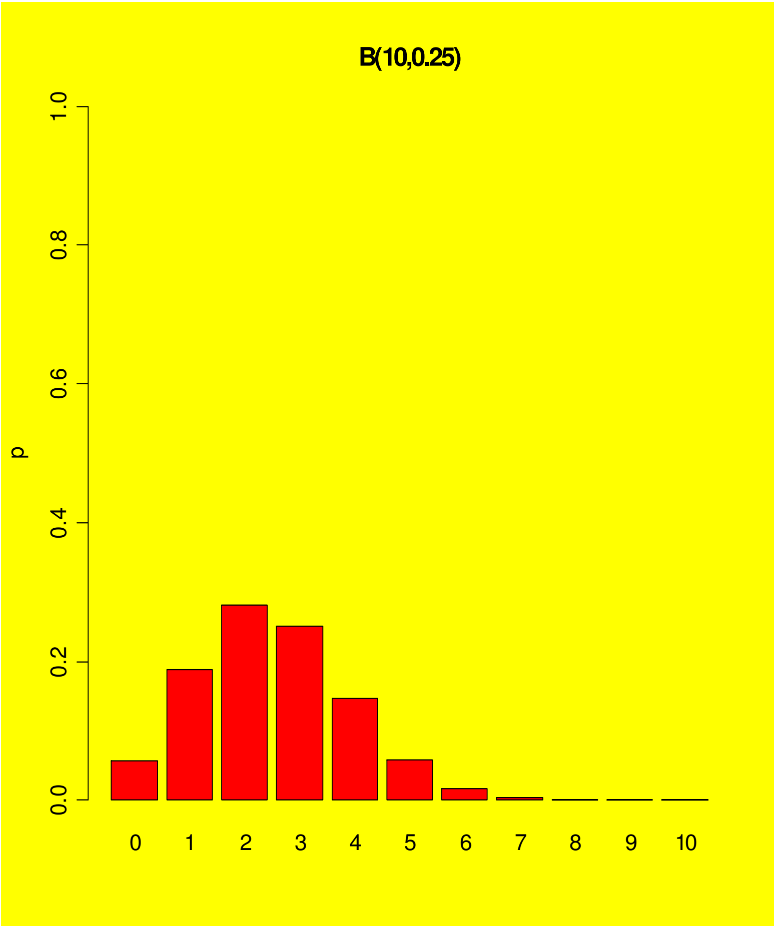


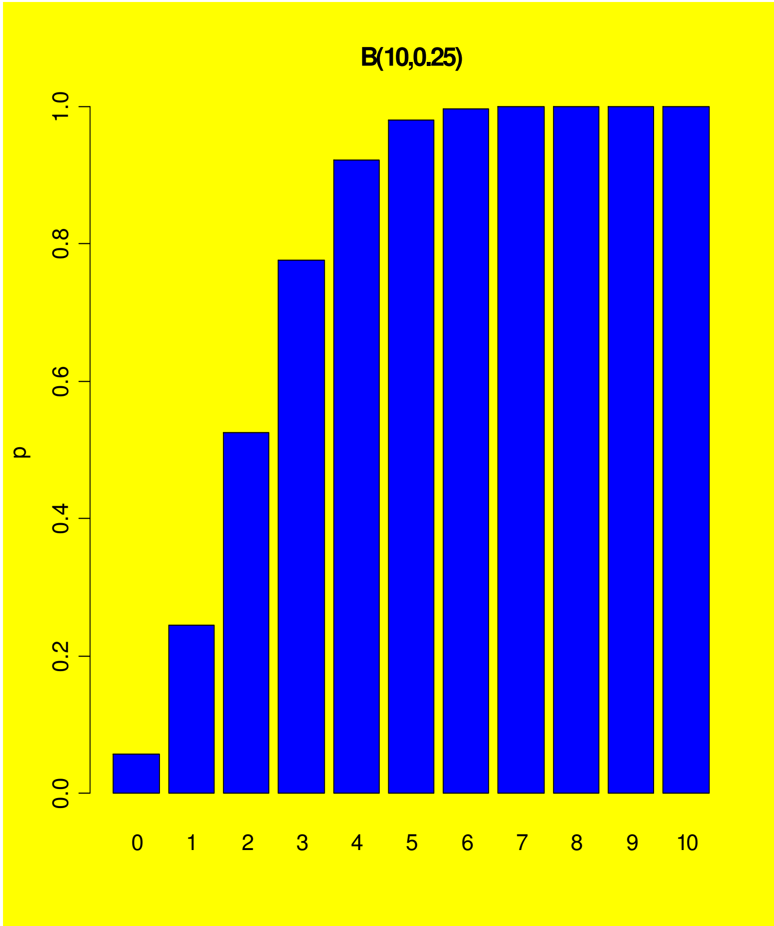
Néhány Binomiális eloszlás grafikus reprezentációja: B(10, 0.5)



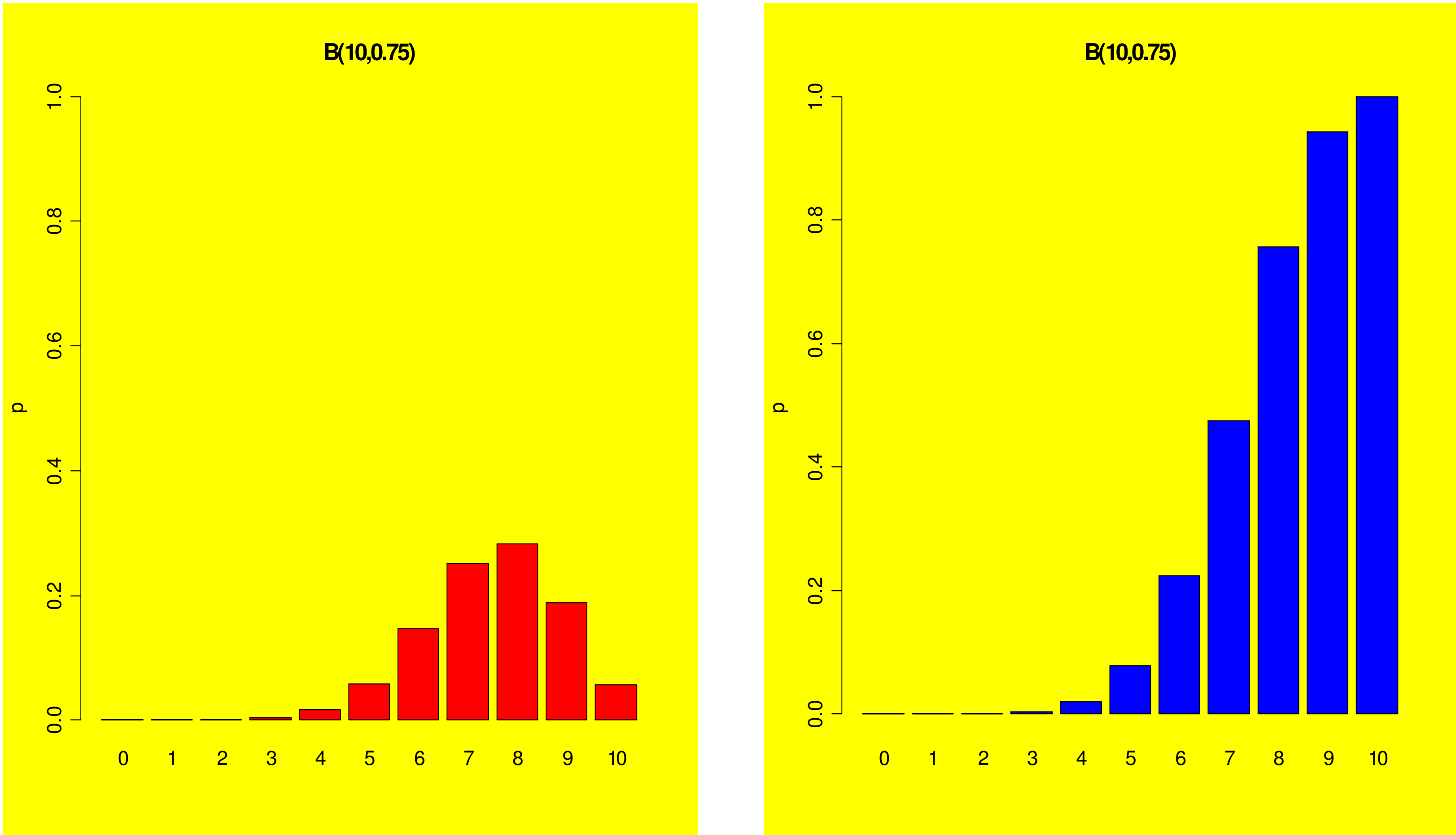


B(10, 0.25)


B(10,0.25)
p
0.0
0.2
0.4
0.6
0.8
1.0
0 1 2 3 4 5 6 7 8 9 10


B(10,0.25)
p
0.0
0.2
0.4
0.6
0.8
1.0
0 1 2 3 4 5 6 7 8 9 10

B(10, 0.75)


B(10,0.75)
p
0.0
0.2
0.4
0.6
0.8
1.0
0 1 2 3 4 5 6 7 8 9 10
B(10,0.75)
p
0.0
0.2
0.4
0.6
0.8
1.0
0 1 2 3 4 5 6 7 8 9 10

Kumulatív valószínűségi táblázat és használata

A különböző binomiális eloszlások értékeihez tartozó kumulatív valószínűségek megtalálhatók táblázatba foglalva, illetve kiszámíthatóak például az R statisztikai szoftver használatával.

Táblázatok használatakor meg kell keresni azt a táblázatot, amelyik a keresett paraméterekkel rendelkező binomiális eloszláshoz tartozik, és ebben meg kell keresni a kérdéses értékhez tartozó kumulatív valószínűségi értéket.

Az R szoftver esetén a pbinom(érték,n,p) parancs alkalmazásával kaphatjuk meg az ‘érték’-hez tartozó kumulatív valószínűséget, az n és p paraméterekkel leírható Binomiális eloszlás esetén.