Szent István Egyetem Állatorvos-tudományi Kar
Biomatematikai és Számítástechnikai Tanszék

10. előadás

Valószínűség-számítás II.

Fodor János

Biomatematika 2001

Copyright © 2001 jfodor@univet.hu

Last Revision Date: November 2, 2001 Version 1.10

Table of Contents

1 Diszkrét valószínűségi változók (folyt.) 4

1.1 Diszkrét valószínűségi változó várható értéke 4

1.2 A medián és a módusz 5

1.3 A variancia és a szórás 6

1.4 A várható érték és a variancia tulajdonságai 13

2 Néhány fontos diszkrét eloszlás 16

2.1 A binomiális eloszlás 17

2.2 A hipergeometrikus eloszlás 22
2.3 A Poisson-eloszlás 26

3 Folytonos valószínűségi változók 31
3.1 Várható érték, variancia, medián és módusz 35

4 A normális eloszlás 37
4.1 A standard normális eloszlás 43

1. Diszkrét valószínűségi változók (folyt.)

1.1. Diszkrét valószínűségi változó várható értéke

Legyen X diszkrét valószínűségi változó, lehetséges értékei $x_1, x_2, \ldots, x_k, \ldots$, az ezekhez tartozó valószínűségek pedig $p_k = P(X = x_k)$.

Az X **várható értékét** $M(X)$ vagy μ_X jelöli, ahol

$$\mu_X := M(X) := \sum_{\text{minden } x_k\text{-ra}} x_k P(X = x_k)$$

1.2. A medián és a módusz

Egy eloszlás közepének jellemzésére van két másik mennyiség is: a módusz és a medián.

Az X diszkrét valószínűségi változó **módusza** egy olyan x_k lehetséges értéke X-nek, amelyhez tartozó p_k valószínűség a legmagasabb.

A módusz nem feltétlenül egyértelmű. Ha X-nek csak egy módusza van, akkor eloszlását **unimodálisnak** nevezzük.

A **medián** olyan x_{med} szám, amelyre

$$P(X < x_{\text{med}}) \leq 1/2, \quad \text{és } P(X > x_{\text{med}}) \leq 1/2.$$

Belátható, hogy ilyen érték mindig létezik, bár nem mindig egyértelmű.

1.3. A variancia és a szórás

Adott várható értékű valószínűségi változók eloszlása sokféle lehet. Például minden origóra szimmetrikus eloszlású valószínűségi változó várható értéke nulla, annak ellenére, hogy lehetséges értékei illetve azok

valószínűségei a nulla ,,körül" más-más módon ,,tömörülhetnek".

Egy évfolyamról véletlenszerűen választott hallgató érdemjegyének várható értéke akkor is közepes, ha mindenki közepes, és akkor is, ha a társaság fele elégtelen, fele jeles osztályzatú.

A várható értéktől való átlagos eltérés (a változékonyság) mértékének jellemzésére vezetjük be a variancia (szórásnégyzet) illetve a szórás fogalmát.

Egy X diszkrét valószínűségi változó **varianciája** a $V(X)$-szel vagy σ_X^2-szel jelölt szám, amely a következő:

$$\sigma_X^2 := V(X) := M[(X - M(X))^2],$$

feltéve, hogy ez a várható érték létezik.

Most az alábbi egyszerűsített jelölést használjuk:

$$\sigma_X^2 = V(X) = \sum [(X - \mu_X)^2 \cdot P(X)].$$

Nyilvánvaló, hogy $V(X) \geq 0$. Ezért értelmes a következő fogalom.

Az X **szórása** a variancia négyzetgyöke, azaz
$$\sigma_X := \sqrt{V(X)}.$$

Példa.

Egy autókölcsönző cég az elmúlt 20 hét adatait összegyűjtve az alábbi táblázatban foglalta össze a heti kikölcsönzött autók számát:

Kölcsönzött autók száma	Hetek száma (gyakoriság)
10	5
11	6
12	7
13	2
Összesen:	20

Konvertáljuk a táblázat adatait úgy, hogy valószínűségeket kapjunk. Ezt mutatja a következő táblázat.

Kölcsönzött autók száma	Valószínűség
10	0.25
11	0.30
12	0.35
13	0.10
Összesen:	1

Számítsuk ki a hetente kikölcsönzött átlagos autó-számot:

$$\begin{aligned}\mu_X &= \sum X \cdot P(X) \\ &= (10)(0.25) + (11)(.30) + (12)(0.35) + (13)(0.10) \\ &= 11.3.\end{aligned}$$

Számítsuk ki a varianciát:

$$\begin{aligned}\sigma_X^2 &= \sum [(X - \mu_X)^2 \cdot P(X)] \\ &= 0.4225 + 0.0270 + 0.1715 + 0.2890 \\ &= 0.91.\end{aligned}$$

1.4. A várható érték és a variancia tulajdonságai

Legyen X és Y valószínűségi változó, és a, b valós számok. Ekkor

1. $M(X+Y) = M(X) + M(Y)$**.**

2. Ha X és Y **függetlenek**, akkor

$$V(X+Y) = V(X) + V(Y).$$

3. Ha $Y = a \cdot X + b$ akkor

- $\mu_Y = a\mu_X + b$;
- $\sigma_Y^2 = a^2\sigma_X^2$;
- $\sigma_Y = |a|\sigma_X$.

4. Ha $X_1, X_2, \ldots, X_n$ független valószínűségi változók, melyek varianciája azonos ($V(X_1) = V(X_2) = \ldots = V(X_n) = \sigma^2$), akkor

$$V(X_1 + X_2 + \ldots + X_n) = n \cdot \sigma^2$$

és

$$D(X_1 + X_2 + \ldots + X_n) = \sqrt{n} \cdot \sigma.$$

2. Néhány fontos diszkrét eloszlás

Diszkrét valószínűségi változó eloszlása:

- a változó lehetséges értékei, és
- a hozzájuk tartozó valószínűségek.

Vannak olyan eloszlások, amelyek igen sok gyakorlati problémával kapcsolatban előkerültek, ezért az idők során nevet kaptak, és ma már név szerint ismertek.

Amikor ilyen vagy olyan eloszlást említünk, mindig eloszlás **típusról** van szó, amelynek sok különböző

tagja van. Egy típuson belül pedig az egyes konkrét eloszlásokat a **paraméterek** jellemzik.

2.1. A binomiális eloszlás

Ez az eloszlás a valószínűségszámításban nagyon fontos, mert a leggyakrabban használt, illetve feltételezett mintavétel a **visszatevéses mintavétel**.

Ha egy olyan (véges vagy végtelen) populációból, amelyben egy bizonyos tulajdonsággal rendelkező egyedek aránya p, visszatevéssel kiválasztunk n ele-

met („egy n elemű mintát veszünk”), a mintában lévő, a tulajdonsággal rendelkező elemek X száma olyan valószínűségi változó, amelynek lehetséges értékei szintén a 0 és n közötti egész számok, egy k érték ($k = 0, 1, 2, ..., n$) valószínűsége pedig

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Ekkor X-et n, p paraméterű **binomiális eloszlású** valószínűségi változónak nevezzük.

Egy másik gyakori modell ugyancsak a binomiális eloszláshoz vezet. Nyilvánvaló ugyanis, hogyha n-szer (egymástól függetlenül) megismétlünk egy kísérletet, amelyben egy bennünket érdeklő E esemény bekövetkezésének valószínűsége p, és megszámoljuk, hogy az n megfigyelés során E hányszor következett be, akkor egy binomiális eloszlású valószínűségi változóhoz jutunk.

Némi számolás után megkapható, hogy

$$\mu_X = np, \qquad \sigma_X^2 = np(1-p).$$

Példa.

Egy urnában 10 golyó van, közülük **3 piros** és **7 kék**. Legyen S az az esemény, hogy véletlenszerűen húzva egy golyót, az éppen piros.

Ha visszatevéssel húzunk, akkor minden egyes alkalommal $P(S) = 0.3$. Ha mondjuk $n = 20$-szor húzunk, és X jelöli a piros golyók számát, akkor

például

$$\begin{aligned} P(X=5) &= \binom{20}{5} 0.3^5 (1-0.3)^{20-5} \\ &= 15504(0.3^5)(0.7^{15}) \\ &= 0.1789. \end{aligned}$$

A várható érték és a variancia ekkor

$$\mu_X = 20(0.3) = 6, \quad \sigma_X^2 = 20(0.3)(0.7) = 4.2.$$

VISSZATEVÉSES MINTAVÉTEL = BINOMIÁLIS ELOSZLÁS

2.2. A hipergeometrikus eloszlás

Egy N egyedből álló populációból, amelyben egy bizonyos tulajdonsággal K egyed rendelkezik, egy n **különböző** elemből álló mintát veszünk. Ekkor a mintában lévő, az adott tulajdonsággal rendelkező elemek X száma valószínűségi változó, amelynek lehetséges értékei a 0 és n közötti egész számok, egy k értékhez $(k = 0, 1, 2, ..., n)$ tartozó valószínűség pedig

$$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}.$$

Ezt az eloszlást **hipergeometrikus eloszlásnak** nevezzük.

A hipergeometrikus eloszlásra

$$M(X)=n\cdot\frac{K}{N},\ V(X)=\frac{N-n}{N-1}\cdot n\cdot\frac{K}{N}\cdot\left(1-\frac{K}{N}\right).$$

Felismerhető bizonyos hasonlóság a binomiális elosz-

lással (ha $p := \frac{K}{N}$), de a hipergeometrikus eloszlás esetén annak valószínűsége, hogy az adott tulajdonsággal rendelkező elem jön ki, kísérletről kísérletre változik.

Példa.

Tekintsünk egy csomag francia kártyát. Ez 52 lapból áll, amelyek közül 16 olyan van, amely nem számot, hanem valamilyen figurát tartalmaz. Egy embernek 10 lapot osztunk. Mennyi annak valószínűsége, hogy ezek között pontosan 4 figura lesz?

Tudjuk, hogy összesen $\binom{52}{10}$ lehetséges módon osztthatunk 10 lapot, és ezek között $\binom{16}{4}\binom{36}{6}$ olyan leosztás van, amely pontosan 4 figurát tartalmaz. Tehát

$$P(4 \text{ figura}) = \frac{\binom{16}{4}\binom{36}{6}}{\binom{52}{10}}.$$

VISSZATEVÉS NÉLKÜLI MINTAVÉTEL = HIPERGEOMETRIKUS ELOSZLÁS

2.3. A Poisson-eloszlás

A Poisson-eloszlás szoros kapcsolatban van a binomiális eloszlással.

Egy X valószínűségi változót **Poisson-eloszlásúnak** nevezünk, ha lehetséges értékei a nem negatív egészek, és egy k értékhez $(k = 0, 1, 2, 3, ...)$ tartozó valószínűség

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

Itt λ pozitív szám, az eloszlás paramétere.

A Poisson-eloszlást leggyakrabban olyan helyzetben használják, ahol egy adott intervallumban bekövetkező események számát vizsgálják. Például:

- egy telefonkezelő által fogadott hívások száma egy 10 perces időintervallumban,
- egy titkárnő által oldalanként okozott gépelési hibák száma.

A Poisson-eloszlást gyakran olyan binomiális elosz-

lású változók közelítésére használják, amelyek n paramétere igen nagy, p paramétere pedig igen kicsi. Tehát ha egy **nagyon ritka** esemény bekövetkezéseit számoljuk egy kísérlet **nagyon nagyszámú ismétlése** során, akkor ennek a változónak az eloszlása jól közelíthető a Poisson-eloszlással.

Matematikailag a λ paraméterű Poisson-eloszlás binomiális eloszlások olyan sorozatának a határértéke, amelyben az np szorzatok sorozata λ-hoz tart.

Egy λ paraméterű Poisson-eloszlású X valószínűségi változóra

$$M(X) = \lambda, \quad V(X) = \lambda.$$

Példa.

Budapest egyik kerületében a téves tűzriasztások átlagos száma 2.1 naponta. Jelölje X a téves riasztások számát egy adott napon. Mennyi annak valószínűsége, hogy egy adott napon 4 téves riasztás történik?

Nyilván X eloszlása Poisson, paramétere $\lambda = 2.1$. Ezért

$$P(X = 4) = \frac{2.1^4 e^{-2.1}}{4!} = 0.0992.$$

3. Folytonos valószínűségi változók

Vannak olyan véletlen változók is, amelyek értékkészlete a számegyenes egy folytonos (véges vagy végtelen) intervalluma, és így lehetséges értékei nem megszámlálhatóan végtelen sokan vannak. Például:

(a) egy kosárlabda játékos magassága

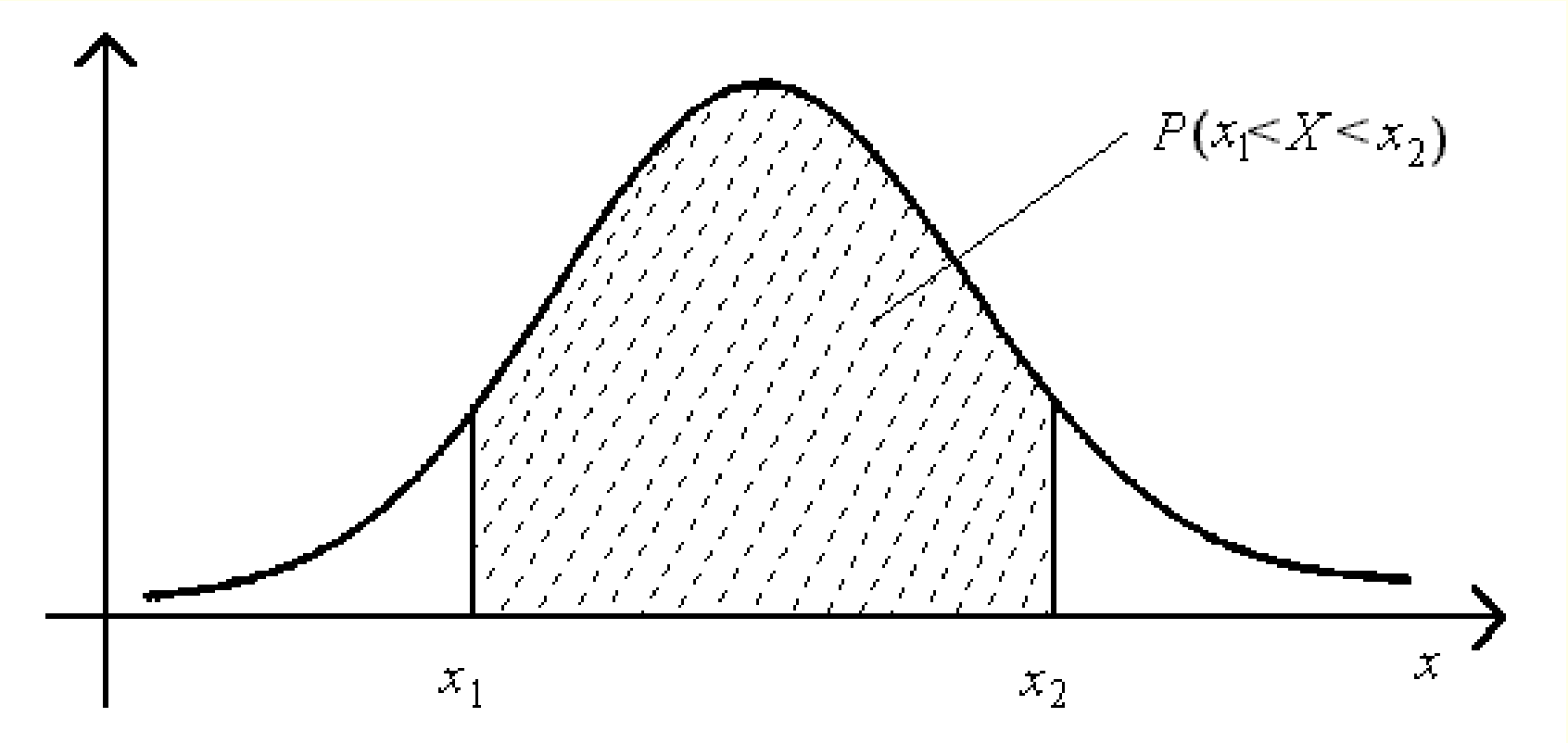
(b) egy ebéd utáni szieszta időtartama.

Egy ilyen változónak valamennyi lehetséges értéke 0 valószínűségű, pozitív valószínűségek csak értéktartományokhoz tartoznak (az egyszerűség kedvéért

gondolhatunk például intervallumokra). Az ilyen változókat **folytonos valószínűségi változóknak** nevezzük.

Folytonos változó eloszlásának megadásához az összes lehetséges tartomány valószínűségét meg kellene adni, ami gyakorlatilag lehetetlen. Ezért folytonos változókra egy olyan függvényt szokás megadni, amelynek segítségével bármely tartományba esés valószínűsége megkapható.

Egy folytonos valószínűségi változó **sűrűségfüggvénye** egy olyan függvény, amelynek függvénygörbe alatti területe (integrálja) bármely tartományon egyenlő a változónak ahhoz a tartományhoz tartozó valószínűségével.



$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx.$$

A valószínűség tulajdonságaiból következik, hogy egy sűrűségfüggvény

- sehol sem negatív,
- a teljes számegyenesen az integrálja 1.

Lássuk most a diszkrét változókra eddig megismert fogalmak értelmezését folytonos változókra!

3.1. Várható érték, variancia, medián és módusz

A folytonos esetben az összegzésnek, és így az átlagolásnak is az **integrálás** a megfelelője. Ezért a

várható értéket és a szórásnégyzetet is integrálként definiáljuk. Legyen f az X változó sűrűségfüggvénye.

$$\mu_X := M(X) := \int_{-\infty}^{\infty} x f(x) dx.$$

$$\sigma_X^2 := M((X - \mu_X)^2).$$

Egy folytonos eloszlású valószínűségi változó **módusza** olyan x érték, ahol a változó sűrűségfüggvényének (lokális) maximuma van.

Folytonos változókra sem mindig egyértelmű, az eloszlás itt is lehet bimodális vagy multimodális.

A **medián** olyan x_{med} érték, amelyre

$$P(X < x_{\text{med}}) = P(X > x_{\text{med}}) = 1/2.$$

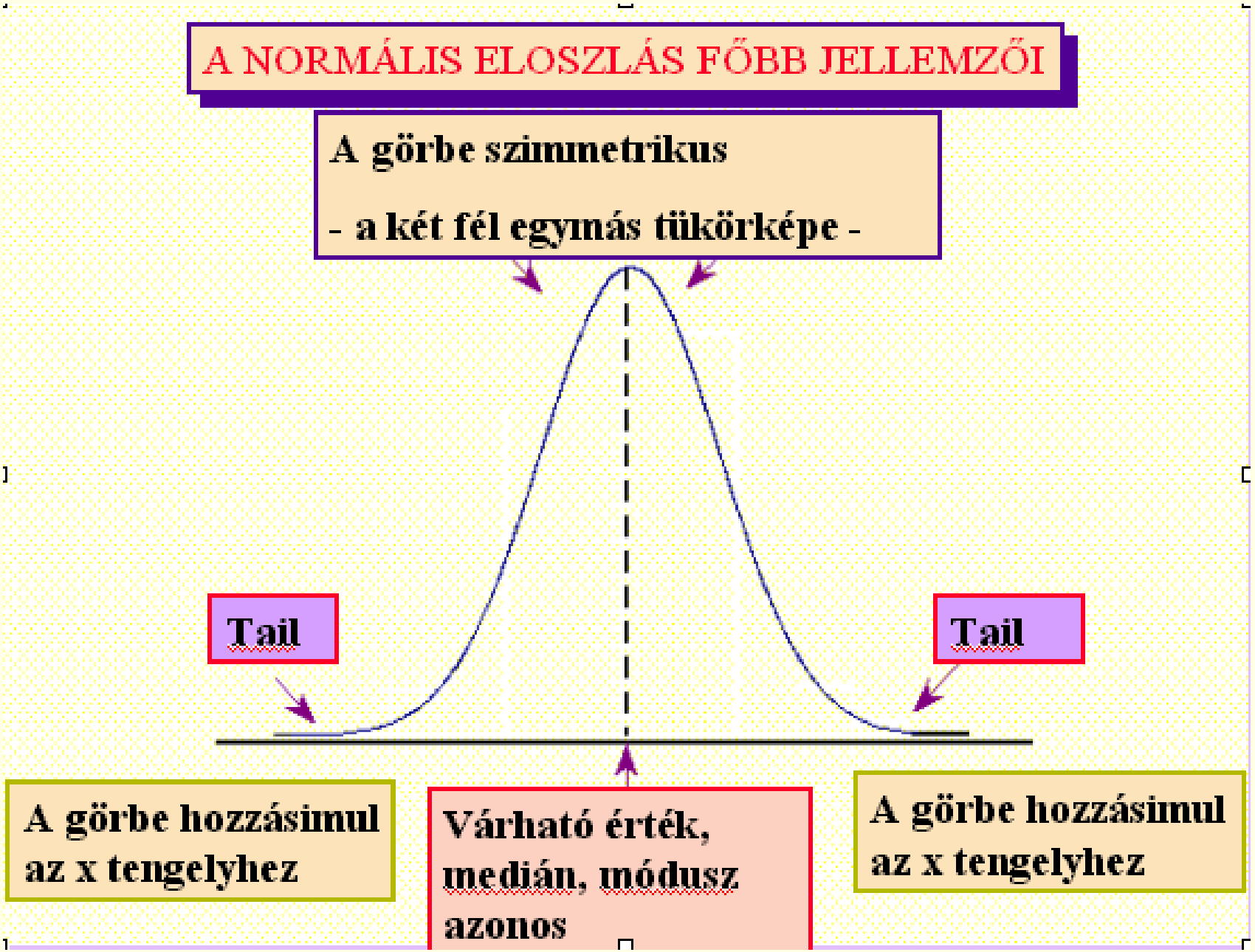
4. A normális eloszlás

A legfontosabb, a gyakorlatban leggyakrabban használt folytonos eloszlás a **normális eloszlás**. Ez is (mint a megismert diszkrét eloszlások) egy **eloszlás család**, tagjai két paraméterrel jellemezhetők. A

sűrűségfüggvényt meglehetősen bonyolult képlet írja le (a paraméterek μ és σ):

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

A sűrűségfüggvény görbéje az úgynevezett **harang-görbe** vagy **Gauss-görbe**.

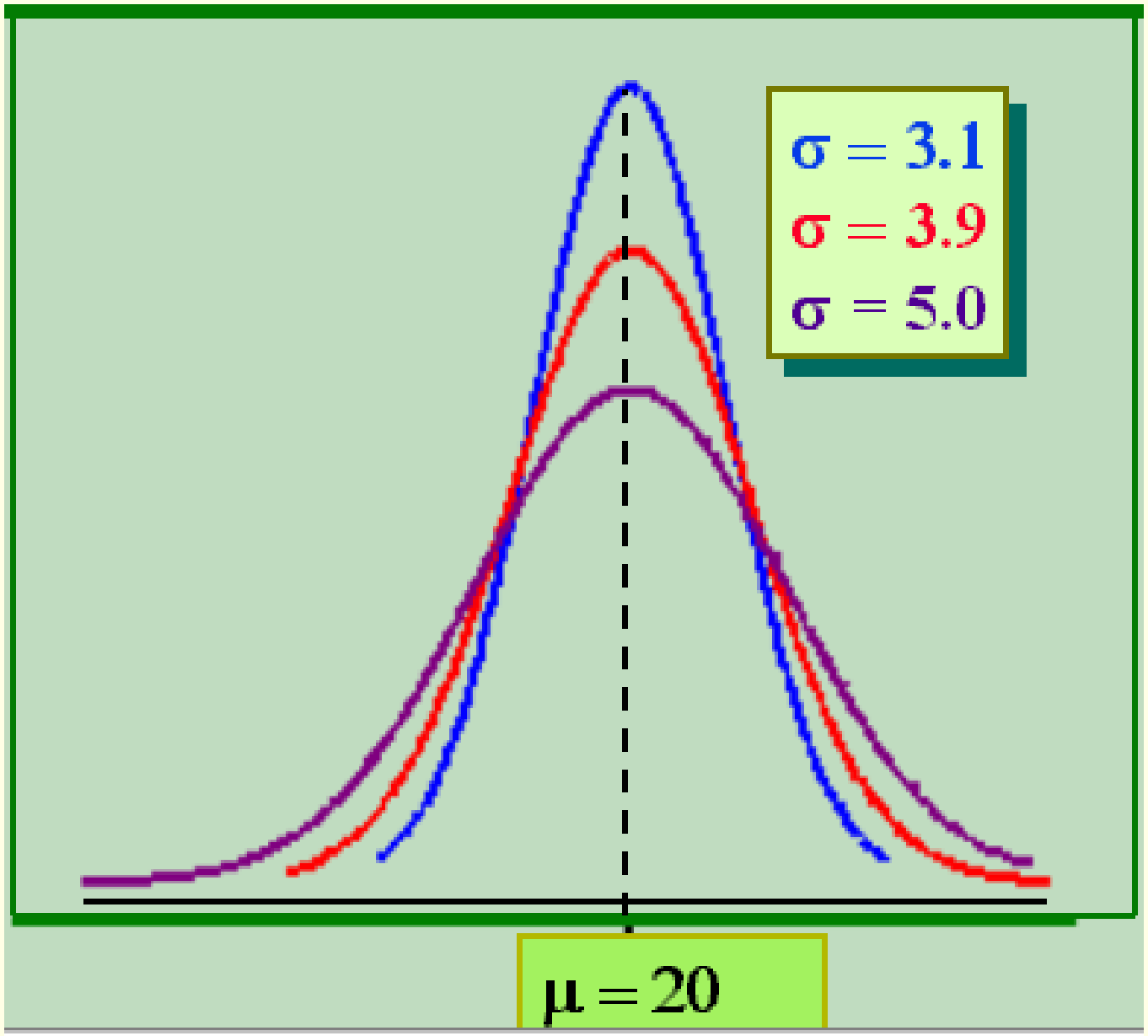

A NORMÁLIS ELOSZLÁS FŐBB JELLEMZŐI
A görbe szimmetrikus
- a két fél egymás tükörképe -
Tail
Tail
A görbe hozzásimul az x tengelyhez
Várható érték, medián, módusz azonos
A görbe hozzásimul az x tengelyhez

A normális eloszlás paramétereinek jelentése:

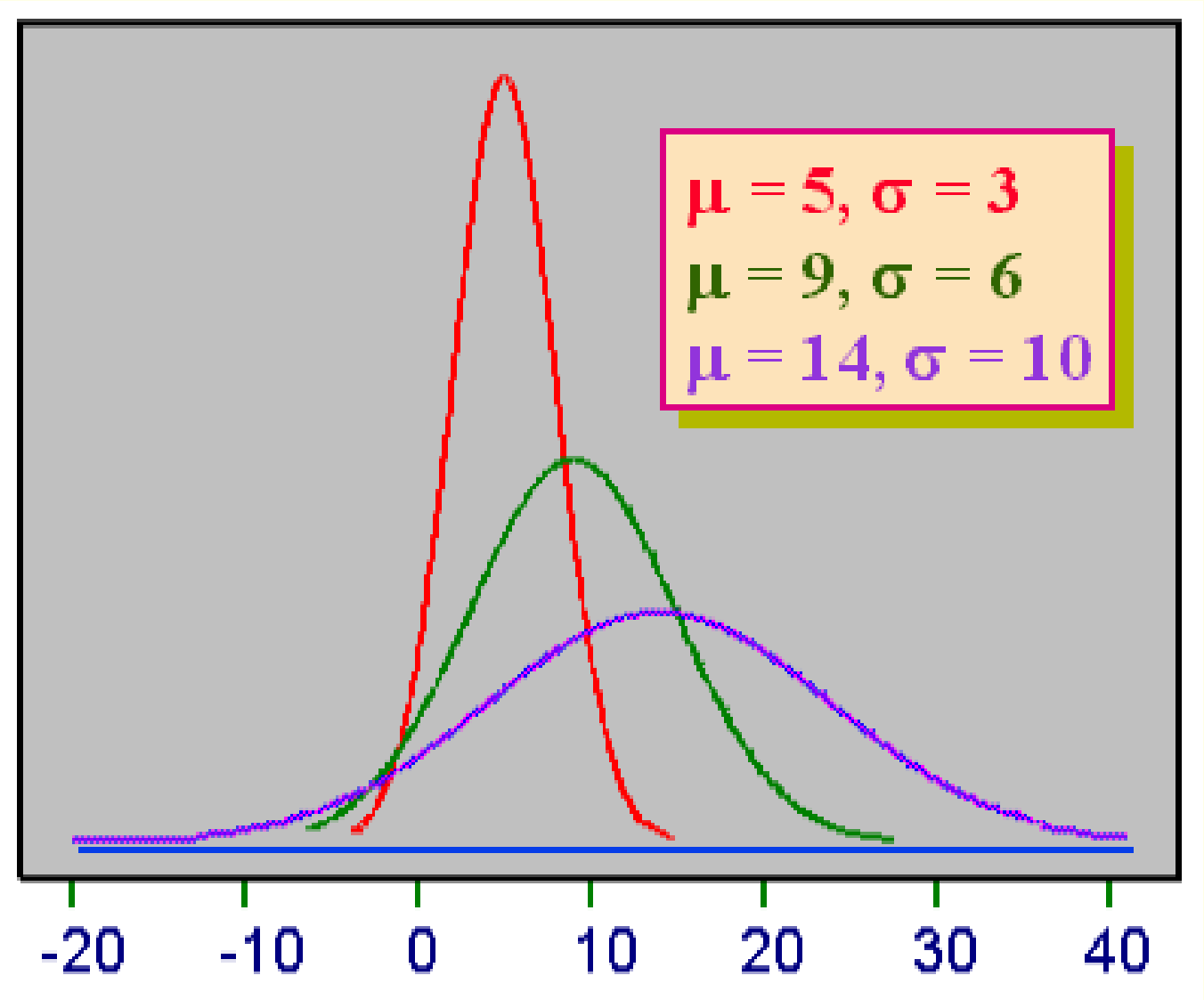
- μ az eloszlás **középértéke**, méghozzá többféle értelemben is, a normális eloszlásnál ugyanis - a görbe szimmetriája miatt - egybeesik a módusz, a medián és a várható érték;
- σ az eloszlás **szórása**.

Jelölés: $X \sim N(\mu, \sigma^2)$ olyan valószínűségi változót jelöl, amely normális eloszlású μ várható értékkel és σ^2 varianciával.

Figyeljük meg a következő ábrán, hogy kis szórás a várható érték körül koncentrálódó eloszlást jelent.


σ = 3.1
σ = 3.9
σ = 5.0
μ = 20

Különböző várható érték és szórás:



4.1. A standard normális eloszlás

A normális eloszlások családjának $\mu = 0$, $\sigma = 1$ paraméterű tagját **standard normális eloszlásnak** nevezik.

Eloszlástáblázatot csak ehhez készítenek, mert a többi mind egyszerűen visszavezethető a standard normálisra. Igaz ugyanis a következő.

Ha $X \sim N(\mu, \sigma^2)$, akkor bármely, belőle lineáris transzformációval származó $Y = aX + b$ változó is normális eloszlású lesz, méghozzá $a\mu + b$ és $|a|\sigma$ paraméterekkel. Jelben: $Y \sim N(a\mu + b, (a\sigma)^2)$

Például az $Y = 2X+3$ változó paraméterei $2\mu+3$ és 2σ, az $Y = 5X$ változóé 5μ és 5σ, az $Y = 0.1X-2$ változóé $0.1\mu - 2$ és 0.1σ, stb.

Ennek a tulajdonságnak a felhasználásával egy $X \sim N(\mu, \sigma^2)$ valószínűségi változót az alábbi lineáris transzformációval transzformálhatunk egy Z standard normális eloszlású változóvá:

$$Z = \frac{X - \mu}{\sigma}.$$

Ezt a transzformációt **standardizálásnak** nevezik.

Gyakran ennek a fordítottjára is szükség van: egy standard normális Z változóból az alábbi lineáris transzformációval kaphatunk $X \sim N(\mu, \sigma^2)$ változót:

$$X = \sigma Z + \mu.$$

Példa.

Egy péknél a 2 kg-os kenyerek tömege normális eloszlású valószínűségi változó $\mu = 2$ kg várható értékkel és $\sigma = 0.03$ kg szórással.

a) Mennyi a valószínűsége, hogy egy találomra kiválasztott kenyér tömege **kevesebb 1.95** kg-nál?

b) Mennyi a valószínűsége, hogy egy találomra kiválasztott kenyér tömege **több 1.98** kg-nál?

c) Milyen – a várható értékre, azaz 2 kg-ra szimmetrikus – határok között van a kenyerek 90%-a?

Jelöljük X-szel a találomra kiválasztott kenyér tö-

megét. Az a) feladatban a $P(X < 1.95)$, a b)-ben a $P(X > 1.98)$ valószínűséget kell meghatároznunk, a c) feladatban pedig olyan h értéket kell találnunk, amelyre $P(2 - h < X < 2 + h) = 0.9$.

a) Standardizáljuk X-et: $Z = (X - 2)/0.03$ standard normális eloszlású. Nyilvánvaló, hogy

$$\begin{aligned} P(X < 1.95) &= P(Z < (1.95 - 2)/0.03) \\ &= P(Z < -1.67) \\ &= P(Z > 1.67) \end{aligned}$$

(ez utóbbi a standard normális eloszlás szimmetriája miatt).

Ez a valószínűség a standard normális eloszlás táblázatából kiolvasható. A jegyzet végén szereplő táblázatban a

$$\boxed{P(Z > z)}$$

alakú valószínűségek találhatók.

A táblázatot a tömörség kedvéért úgy készítették, hogy a második tizedesjegy az oszlopok fejlécében szerepel, vagyis az 1.67 értékhez tartozó valószínűséget az 1.6 érték sorának és a 0.07 oszlopának a metszéspontjában találjuk.

A táblázatban itt 0.047 áll, tehát

$$\boxed{P(X < 1.95) = 0.047.}$$

Ennyi a valószínűsége annak, hogy egy kenyér 1.95 kg-nál kisebb tömegű.

b) Standardizálással kezdünk:

$P(X > 1.98) = P(Z > (1.98 - 2)/0.03)$.

Ezután a kapott eseményt olyan alakra hozzuk, amelynek valószínűsége a táblázatban szerepel:

$$\begin{aligned} P(Z > (1.98 - 2)/0.03) &= P(Z > -0.67) \\ &= P(Z < 0.67) \\ &= 1 - P(Z > 0.67). \end{aligned}$$

A táblázatban a 0.6 sorának és a 0.07 oszlopának metszéspontjában a 0.251 áll, tehát

$$P(X > 1.98) = 1 - 0.251 = 0.749.$$

Ennyi a valószínűsége annak, hogy egy kenyér tömege több, mint 1.98 kg.

c) Ha olyan, a 2 kg-ra szimmetrikus határokat keresünk, amelyek között van a kenyerek 90%-a, akkor olyan h értéket kell találnunk, amelyre

$$P(2 - h < X < 2 + h) = 0.9.$$

Erre a h-ra a szimmetria miatt

$$P(X > 2 + h) = 0.05.$$

A megoldás alapötlete az, hogy keressünk a standard normális eloszlás táblázatából egy olyan v-t, amelyre $P(Z > v) = 0.05$, majd ebből a standardizálás fordított transzformációjával határozzuk meg a keresett h-t.

Megfelelő v-t úgy találhatunk a táblázatban, hogy a táblázat belsejében, ahol a valószínűségek szerepelnek, megkeressük a 0.05-öt (vagy a hozzá legközelebbi valószínűséget), majd a táblázat szélén leolvassuk a v értékét (az első tizedesig a sor elején, a második tizedest pedig az oszlop tetején).

A táblázatban a 0.05-höz legközelebbi két valószínűség a **0.049** és a **0.051**, amelyekhez az **1.65**, illetve az **1.64** értékeket olvashatjuk le.

Válasszuk mondjuk az 1.65-öt. Ezzel tehát $P(Z > 1.65) = 0.049$, ahonnan

$$P(-1.65 < Z < 1.65) = 0.902.$$

A Z-ből az $X = 0.03Z + 2$ transzformációval állíthatunk elő 2 várható értékű, 0.03 szórású normális eloszlású változót. A $(-1.65 < Z < 1.65)$ és a $(2 - 1.65 \cdot 0.03 < X < 2 + 1.65 \cdot 0.03)$ események pontosan ugyanakkor következnek be, mivel az e-

gyenlőtlenségek mindkét oldalát ugyanúgy transzformáltuk. Ezért

$$\begin{aligned} 0.902 &= P(-1.65 < Z < 1.65) \\ &= P(2 - 1.65 \cdot 0.03 < X < 2 + 1.65 \cdot 0.03) \\ &= P(2 - 0.0495 < X < 2 + 0.0495) \\ &= P(1.9505 < X < 2.0495), \end{aligned}$$

tehát a kenyereknek kb. 90%-a 1.9505 és 2.0495 kg közé esik.