2. METROLÓGIA ÉS HIBASZÁMíTÁS

2.1 Metrológiai alapfogalmak

A **metrológia** a mérések tudománya, a mérésekkel kapcsolatos ismereteket fogja össze.

Méréssel egy objektum valamilyen tulajdonságáról számszerű értéket kapunk. A **mérési eredményt** egy számmal (a **mértékszámmal**) és a **mértékegységgel** adjuk meg. A mérés történhet mérőeszközzel vagy műszerrel. A **mérőeszközzel** való mérésnél az objektum valamilyen tulajdonságát **közvetlen összehasonlítással** kapjuk meg (pl. ha hosszúságot méterrúddal vagy tömeget mérleggel mérünk). A **műszerrel** való mérés viszont **közvetett**, az objektum és a műszer valamilyen kölcsönhatásából kalibrálással adja meg a mérni kívánt mennyiség értékét. (Pl. az elektromos áram erősségét mérő Deprez-rendszerű ampermérőben a szögelfordulás az áram, a mágneses tér és a rugó kölcsönhatásának az eredménye, de a műszer skálája már áramerősségre van kalibrálva.) Sok esetben azonban nem is tudjuk a jellemezni kívánt mennyiséget közvetlenül mérni, hanem csak más, vele kapcsolatban álló mennyiséget, illetve mennyiségeket, és az ezekre kapott mérési eredményekből számítással kapjuk a kívánt adatot. Ekkor is mérésről **-** közvetett, illetve **összetett mérésről** beszélünk.

A mérést megismételve általában nem kapunk azonos mérési eredményeket. Egyrészt, mert a mérendő mennyiség változhat az idővel. De ha a mérendő mennyiség állandó is, a mérési eredményt a műszer állapota és a megfigyelést végző ember is befolyásolja. Jelöljük a mérni kívánt mennyiség valóságos értékét x-szel, a mérési eredményt x_m-mel. A **mérés hibája** a mért érték és a valóságos érték különbsége:

$$\Delta x = x_m - x \qquad 2.1$$

és a relatív hiba:

$$\delta x = \Delta x / x \qquad 2.2$$

A mérés megbízhatóságát elsősorban a műszer jellemzői határozzák meg.

<u>A műszerek jellemzői</u>

a./ *Érzékenység*: Meghatározza, hogy a mérendő mennyiség egységnyi változásához a műszerről közvetlenül leolvasható érték milyen változása tartozik. Ez a műszerről leolvasható legkisebb egységnek (skálarésznek) megfelelő mérendő mennyiség reciproka. (Pl. ha egy mutatós mA-mérő végkitérése 250 mA, skálája lineáris és 50 beosztású, a műszer érzékenysége 50 skr / 250 mA = 0,2 skr/mA, és a mutató egy skálarésznyi kitérése 5 mA áramnak felel meg.) Ha a skála nem lineáris, akkor az érzékenység függ attól, hogy milyen érték közelében mérünk.

b./ *Pontosság*: A mért és valódi érték maximális lehetséges eltérése, a hiba abszolút értékének maximuma.

c./ *Reprodukálhatóság*: A műszerrel történő ismételt mérésekkel kapható mérési eredmények lehetséges maximális eltérése egymástól.

<u>Hibatípusok</u>

Véletlen hiba: A mérési eredmények a valóságos értéktől mindkét irányban azonos valószínűséggel, véletlenszerűen térnek el. Nagy számú mérés átlagát véve a véletlen hiba tetszőlegesen csökkenthető.

Rendszeres hiba: A mérési eredmények a valóságos értéktől eltérő érték körül ingadoznak. Oka a hibás vagy rosszul beállított műszer, de rendszeres hibát okoz az is, ha elhanyagolunk vagy rosszul veszünk figyelembe valamilyen, a mérést befolyásoló külső tényezőt (pl. hőmérsékletet vagy nyomást).

2.2. A hibaszámítás alapjai

A hibaszámítás a valószínűségszámítás és matematikai statisztika felhasználásával a mérés során fellépő véletlen hibák becslésére ad módot.

Valószínűségszámítási alapfogalmak; a valószínűségi sűrűség- és eloszlásfüggvény

Méréssel nemcsak egyetlen objektum valamilyen állandó értékű sajátosságára kaphatunk számszerű értéket, hanem jellemezhetjük egy objektum olyan sajátosságát is, mely időben vagy helyileg változik. Vagy jellemezhetünk egy olyan egyedekből álló sokaságot, melyekre nézve a mérendő tulajdonság különböző mértékű. Beszélhetünk a terem hőmérsékletéről, akkor is, ha a hőmérséklet a radiátor mellett más mint az ajtó mellett, éjszaka más mint nappal. Vagy megadhatjuk az évfolyamon a hallgató lányok átlagos magasságát. Egy sokaság esetén az egyedi mérések valamilyen átlaga lesz az egész sokaság jellemzője, de tisztáznunk kell, hogy értjük ezt az átlagot.

A sokaságot jellemző számnak az egyedekre is jellemzőnek kell lenni valamilyen módon, a sokaságra vonatkozó adatból bizonyos mértékig meg kell tudnunk jósolni az egyedi mérések eredményét, azaz a sokaság jellemzésénél azt is meg kell adnunk, hogy a megadott átlagérték körül adott valószínűséggel milyen intervallumban lesznek az egyes mérések eredményei.

Hasonló a probléma a terem hőmérsékletének megadásánál is. Ha a terem hőmérséklete időben változik, a termet különböző időpontokban elképzelve tekinthetjük sokaságnak, melynek elemei az egyes időpontokban a terem pillanatnyi állapotai, és ezekben az állapotokban mérhetjük az egyedi hőmérséklet értékeket. De ha egy jól meghatározott, időben állandó mennyiséget ismételten mérünk, a műszer időben különböző állapotai, a műszer és a mért objektum, valamint a környezet időben változó kölcsönhatása miatt az ismételt mérés eredményei változni fognak. Ezek az elképzelt ismételt mérések megint egy sokaság elemeinek tekinthetők.

Tételezzünk fel egy **sokaságot**, és egy, **a sokaságon értelmezett ξ mennyiséget**. ξ különböző értékeket vehet fel, de egyeseket nagyobb, másokat kisebb valószínűséggel. (Egy diáklány magasságát gyakran mérjük 160 és 170 cm közötti értéknek, és a terem hőmérséklete kevéssé valószínű, hogy -10 °C alatt van.) Azt mondjuk, hogy **ξ valószínűségi változó**. Egy valószínűségi változó lehet **folytonos** (mint a magasság vagy hőmérséklet), és lehet **diszkrét** (mint a kockadobálás eredménye). A **sokaság** lehet **véges elemű** (mint a másodikos vegyészlányok sokasága), vagy **végtelen elemű** (mint a terem állapota tetszőleges időpontokban).

Amikor mérünk, kiválasztjuk egy egyedét a sokaságnak, és ezen végezzük el a mérést. Azt is mondhatjuk, hogy **mintát veszünk** a sokaságból és azon egy **kísérletet végzünk**. A kísérlet eredményes vagy sikeres, ha azt kaptuk, amit vártunk, pl. 6-ost a kockadobálásnál. A **sikeres kísérletet** nevezzük **esemény**nek. Az esemény **valószínűsége P**: az összes lehetséges kedvező kimenetelű mintavétel, kísérlet száma (n_k) osztva az összes lehetséges kísérlet számával ($N_Ö$):

$$P = n_k / N_Ö \qquad 2.3$$

A valószínűséget pontos értékét meghatározhatjuk, ha van információnk az egész sokaságról, különben a sokaság több-kevesebb elemén végzett kísérlet alapján becsüljük P értékét.

Legyen például a sokaság 100 skatulya gyufa, és a ξ *diszkrét* valószínűségi változó a gyufaszálak száma egy dobozban. Ha a 100-ból 10 dobozban van 40 szál gyufa, akkor annak az eseménynek a valószínűsége, hogy egy véletlenül kiválasztott skatulyában pont 40 szál gyufát találjunk,

$$P(\xi=40) = P(40) = 10 / 100 = 0{,}1.$$

Ha a valószínűségi változó *folytonos*, akkor csak azt kérdezhetjük, hogy egy bizonyos intervallumba eső értéket milyen valószínűséggel vehet fel. Pl. ha a valószínűségi változó a lányok magassága, és az évfolyam 50 hallgatólányából 10-nek a magassága 160 és 170 cm közé esik, akkor annak a valószínűsége, hogy egy véletlenszerűen kiválasztott lány h magasságát 160 és 170 cm közöttinek találjuk:

$$P(160<h<170) = 10 / 50 = 0{,}2.$$

Ismételjük meg a kísérletet N-szer, és tegyük fel, hogy n esetben kaptunk kedvező eredményt (pl. 40 szál gyufát a gyufaszámlálási kísérletben). Az esemény **relatív gyakorisága**:

$$q = n / N \qquad 2.4$$

Ha N nő, a relatív gyakoriság a valószínűséghez tart:

$$\lim_{N\to\infty} q = P \qquad 2.5$$

Jelöljük a ξ folytonos valószínűségi változó aktuális mért értékét x-szel.

P ($x_i < x < x_i + \Delta x$) annak valószínűsége, hogy x értéke x_i és $x_i + \Delta x_i$ közé essen. A

$$\lim_{\Delta x\to 0} \frac{P(x_i < x < x_i + \Delta x_i)}{\Delta x} = f(x_i) \qquad 2.6$$

határérték a ξ változó valószínűségi sűrűsége x_i -nél, és f(x) ξ eloszlásának **valószínűségi sűrűségfüggvénye** vagy **frekvencia függvénye**. Annak valószínűsége, hogy ξ x_i és $x_i + \Delta x_i$ közötti értéket vegyen fel, közelítőleg

$$P(x_i \le x \le x_i + \Delta x_i) = f(x_i)\,\Delta x, \qquad 2.7$$

ha Δx elég kicsi. Általában annak valószínűsége, hogy ξ x_1 és x_2 közé essen:

$$P(x_1 \le x \le x_2) = \int_{x_1}^{x_2} f(x)\,dx \qquad 2.8$$

A valószínűségi sűrűségfüggvény integrálja a valószínűségi változó teljes értelmezési tartományára

$$\int_{-\infty}^{\infty} f(x')\,dx' = P(-\infty \le x \le \infty) = 1\ .$$

A valószínűségi sűrűségfüggvény integrálfüggvényét **valószínűségi eloszlásfüggvénynek** nevezzük és F(x)-szel jelöljük:

$$F(x) = \int_{-\infty}^{x} f(x')\,dx' \qquad F(\infty) = 1 \qquad 2.9$$

F(x_i) annak a valószínűsége, hogy a valószínűségi változó értéke nem nagyobb, mint egy adott x_i érték:

$$P(x \le x_i) = F(x_i) \qquad 2.10$$

Az eloszlásfüggvénnyel megadhatjuk annak a valószínűségét, hogy a valószínűségi változó értéke x_1 és x_2 közé esik:

$$P(x_1 \le x \le x_2) = \int_{x_1}^{x_2} f(x)\,dx = \int_{-\infty}^{x_2} f(x)\,dx - \int_{-\infty}^{x_1} f(x)\,dx = F(x_2) - F(x_1)$$

<u>Összetett kísérlet eredményének valószínűsége</u>

Ha két mennyiséget, ξ-t és η-t mérünk egymás után, mi a valószínűsége annak, hogy ξ-re x-et, η-ra y-t kapjunk eredményül?

Ha N(y/x) azoknak az eseteknek a száma, amikor η-ra y értéket kapunk, feltéve, hogy ξ-re már x értéket kaptunk, akkor a valószínűség értelmezéséből

$$P(x,y) = \lim_{N\to\infty} \frac{N(y/x)}{N} = \lim_{N\to\infty} \frac{N(y/x)}{N(x)} \frac{N(x)}{N} = P(y/x)\,P(x). \qquad 2.11$$

P(y/x)-t **feltételes valószínűségnek** nevezzük. **Ha az η változóra vonatkozó kísérlet eredménye nem függ a ξ változótól, akkor**

$$P(y/x) = P(y) \quad \text{és} \quad P(x,y) = P(x)\,P(y) \qquad 2.12$$

Ebben az esetben **a két** esemény vagy kísérlet vagy **valószínűségi változó egymástól független**. **Független események valószínűségei szorzódnak, és ugyanez áll a valószínűségi sűrűségfüggvényekre is.**
A mérést általában megismételjük, hogy biztosabb eredményt kapjunk. Ilyenkor feltesszük, hogy az egyes méréseket egymástól függetlenül végeztük, hogy egy későbbi mérés eredményét nem befolyásolja egy előző eredmény. Vigyázzunk, hogy ez tényleg teljesüljön! Nagyon gyakori a megismételt mérés eredményének önkéntelen szubjektív torzítása!

Az eloszlás paraméterei, a minta jellemzői

A valószínűségi sűrűségfüggvényben szereplő konstansok és az azokból leszármaztatott mennyiségek az **eloszlás paraméterei**. Amikor mintát veszünk a sokaságból és ezen a mintán mérést végzünk, a célunk az, hogy az eloszlásról kapjunk információt. Az eloszlásfüggvény alakja ismert, vagy egy meghatározott függvénnyel közelítjük, a függvényben szereplő konstansokat viszont a mérésből akarjuk meghatározni, vagy megbecsülni.
A mérési eredményekből az eloszlásparaméterekre kapott becsléseket nevezzük a **minta jellemzőinek**.
A mérés lehetséges kimenetele, a jövőbeli mérési eredmény is valószínűségi változó és így valamilyen valószínűségi sűrűségfüggvény rendelhető hozzá. A mérésnél ténylegesen kapott mérési eredmények viszont konkrét számok, melyekkel kapcsolatban már nincs értelme valószínűségről beszélni.
A mintavételnél is beszélhetünk a majdani mintáról mint sokaságról, melynek elemei -az elképzelt mérési eredmények- valószínűségi változók. Ennek az elképzelt mintának mint sokaságnak a jellemzői szintén valószínűségi változók lesznek.
A következőkben az eloszlás paramétereit -melyek a sokaságra jellemzők- görög, a minta jellemzőit pedig latin betűkkel fogjuk jelölni.

A legfontosabb eloszlásparaméterek

Legyen a sokaságon egy ξ valószínűségi változó (egy tulajdonsága a sokaság elemeinek) értelmezve, melynek értékét x-szel jelöljük.
a./ **A várható érték**
Tételezzünk fel egy véges, N elemű sokaságot, melyen értelmezett valószínűségi változó diszkrét értékeket vehet fel. Pl. legyen a sokaság 100 doboz gyufa és a valószínűségi változó a gyufaszálak száma egy dobozban. Legyen ξ értéke a sokaság n_i számú elemén x_i. Akkor x átlagértéke

$$\overline{x} = \sum_{i=1}^{N} (n_i x_i)\,/\,N = \sum_{i=1}^{N} x_i\ P(x_i)\ , \qquad 2.13a$$

mivel $P(x_i) = n_i / N$. 2.13a-t általánosítva, tetszőleges eloszlásra definiálhatunk egy, az eloszlásra jellemző paramétert, az eloszlás **várható értékét**:
Legyen az eloszlás **diszkrét**, de a sokaság nem feltétlenül véges; a ξ valószínűségi változó az x_i értéket P_i valószínűséggel vegye fel. Akkor **a ξ valószínűségi változó várható értékén**, E [x]-en a következő összeget értjük:

$$E[x] = \sum_i x_i \, P(x_i) = \mu_x \,, \qquad 2.13$$

ahol az összegzés a sokaság összes elemére történik.
Ha az eloszlás **folytonos**, és az eloszlás valószínűségi sűrűségfüggvénye f(x), akkor a **várható érték**, 2.13 analógiájára

$$E[x] = \int_{-\infty}^{\infty} x\, f(x)\, dx = \mu_x \qquad 2.14$$

A valószínűségi változó **konstansszorosának** várható értéke a várható érték konstansszorosa:

$$E[cx] = \int_{-\infty}^{\infty} cx\, f(x)\, dx = c \int_{-\infty}^{\infty} x\, f(x)\, dx = c\, E[x] \qquad 2.15$$

Két valószínűségi változó, ξ és η összegének és szorzatának várható értékénél az összetett kísérlet valószínűségével kell számolnunk.
Ha a valószínűségi változók diszkrétek, annak valószínűsége, hogy a ξ tulajdonságra x-et, az η tulajdonságra y-t kapjunk mérési eredményül, 2.11 szerint P(x,y). Folytonos eloszlásnál a valószínűségeket a sűrűségfüggvénnyel határozzuk meg. Annak valószínűsége, hogy a ξ tulajdonság mért értéke x és x + dx közé essen, és az η tulajdonságé y és y + dy közé:

$$P(x \le \xi \le x+dx,\ y \le \eta \le y+dy) = h(x,y)\, dx\, dy \qquad 2.16$$

és 2.12-nek megfelelően, ha a két valószínűségi változó független, h(x,y) = f(x) g(y), vagyis két, csak x-től, illetve csak y-tól függő sűrűségfüggvény szorzata.
Két folytonos valószínűségi változó **összegének** várható értéke a két várható érték összege:

$$E[x+y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y)\, h(x,y)\, dx\, dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} h(x,y)\, dy\right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} h(x,y)\, dx\right) dy =$$

$$= \int_{-\infty}^{\infty} x\, f(x)\, dx + \int_{-\infty}^{\infty} y\, g(y)\, dy = E[x] + E[y]\,. \qquad 2.17$$

Két **független** valószínűségi változó **szorzatának** várható értéke az egyes várható értékek szorzata:

$$E[xy] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, h(x,y)\, dx\, dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f(x)\, g(y)\, dx\, dy =$$

$$= \int_{-\infty}^{\infty} x\, f(x)\, dx \cdot \int_{-\infty}^{\infty} y\, f(y)\, dy = E[x] \cdot E[y] \qquad 2.18$$

Ha a valószínűségi sűrűségfüggvény **szimmetrikus** egy x_0 értékre, akkor $y = x - x_0$ várható értéke E [y] = 0, azaz az eredeti várható érték éppen x_0 -lal egyenlő.
(Bizonyítás:

$$E[y] = \int_{-\infty}^{\infty} y\, f(y)\, dy = \int_{-\infty}^{0} y\, f(y)\, dy + \int_{0}^{\infty} y\, f(y)\, dy\,, \qquad \text{és mivel a szimmetria miatt}$$

f(y) = f(-y), így $\displaystyle \int_{-\infty}^{0} y\, f(y)\, dy + \int_{0}^{\infty} y\, f(y)\, dy = -\int_{0}^{\infty} y\, f(-y)\, dy + \int_{0}^{\infty} y\, f(y)\, dy = 0.$)

b./ **A medián és a módusz**

A eloszlás **mediánja** (μ_e) a valószínűségi változónak az az értéke, melynél kisebb és nagyobb érték is ugyanolyan valószínűségű, azaz ahol az eloszlásfüggvény értéke

$$F(\mu_e) = 0,5.$$

Az eloszlás **módusza** a sűrűségfüggvény maximum helye.

Szimmetrikus eloszlás várható értéke, mediánja és módusza azonos.

c./ **A variancia**

A valószínűségi változó értéke kisebb vagy nagyobb mértékben eltérhet a várható értéktől a sokaság elemein. A **variancia** ennek az eltérésnek a mértéke:

$$\mathrm{Var}\,[x] = \int_{-\infty}^{\infty} (x-\mu_x)^2 f(x)\,dx = E\,[(x-\mu_x)^2] \qquad 2.19$$

Változó **konstansszorosának** varianciája a variancia szorozva a konstans négyzetével:

$$\mathrm{Var}\,[cx] = E\,[cx-c\mu_x]^2 = c^2\, E\,[x-\mu_x] = c^2\, \mathrm{Var}\,[x] \qquad 2.20$$

Két **független** valószínűségi változó **összegének** varianciája az egyes varianciák összege:

$$\mathrm{Var}\,[x+y] = E\,[(x+y-\mu_x-\mu_y)^2] =$$

$$= E\,[(x-\mu_x)^2] + E\,[(y-\mu_y)^2] + 2\,E\,[(x-\mu_x)(y-\mu_y)] = \mathrm{Var}\,[x] + \mathrm{Var}\,[y] \qquad 2.21$$

Itt felhasználtuk az összeg várható értékére vonatkozó 2.17 összefüggést, és mivel ξ és η függetlenek, 2.21 harmadik tagjánál fel tudtuk használni 2.18-at, miszerint

$$E\,[(x-\mu_x)(y-\mu_y)] = E\,[x-\mu_x] \cdot E\,[y-\mu_y] = 0.$$

Két **független** valószínűségi változó **szorzatának** varianciája nem egyezik meg viszont a varianciák szorzatával!

A normális (Gauss) eloszlás

Normális eloszlást mutatnak azok a valószínűségi változók, melyek értékét sok kismértékű véletlenszerű hatás befolyásolja. A normális eloszlásnál a valószínűségi sűrűségfüggvény az ún. **Gauss-függvény**:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad 2.22$$

ahol μ az eloszlás **várható értéke**, σ pedig az úgynevezett **szórás**, a variancia négyzetgyöke.

Az ún. **normalizált Gauss-függvényt** 2.22-ből az

$$u = (x-\mu) / \sigma \qquad 2.23$$

transzformációval kapjuk, és a következő alakú:

$$f(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u^2} \qquad 2.24$$

A normalizált Gauss-eloszlás várható értéke 0, varianciája 1.

Bizonyítás: A várható érték 2.14 és a variancia 2.19 definíciójából:

$$E[u] = \int_{-\infty}^{\infty} u\, f(u)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u\, e^{-\frac{1}{2}u^2}\, du = 0,$$

mert az integrandus páratlan függvény.

$$Var[u] = \int_{-\infty}^{\infty} u^2 f(u)\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2\, e^{-\frac{1}{2}u^2}\, du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u\, (u\, e^{-\frac{1}{2}u^2})\, du$$

Felhasználva, hogy

$$\int u\, e^{-\frac{1}{2}u^2}\, du = -\, e^{-\frac{1}{2}u^2}$$

és parciálisan integrálva, kapjuk:

$$\int_{-\infty}^{\infty} u \left(u\, e^{-\frac{1}{2}u^2} \right) du = -\left[u\, e^{-\frac{1}{2}u^2} \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{1}{2}u^2}\, du \cdot$$

Az első tag kiesik, a második pedig $\sqrt{2\pi}$-vel egyenlő, így Var [u] = 1.

A normalizált Gauss-eloszláshoz tartozó **valószínűségi eloszlásfüggvény**:

$$F(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2}\, dt = \phi(u) \quad (\textbf{hibaintegrál}), \quad F(\infty)=1. \qquad 2.25$$

A normalizált Gauss-függvény, a $\phi(u)$ hibaintegrál vagy az

$$erf(u) = \frac{2}{\sqrt{\pi}} \int_{0}^{u} e^{-t^2}\, dt \qquad \textbf{hibafüggvény} \qquad 2.26$$

értékeit matematikai kézikönyvekben táblázatosan megtaláljuk[1]. A két függvény összefüggése

$$\Phi(u) = 0{,}5\, (1 + erf\, (u/\sqrt{2})). \qquad 2.27$$

A Gauss-eloszlás alkalmazása

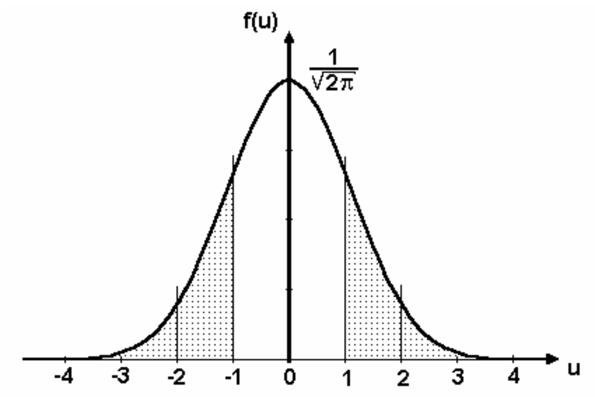
Tételezzük fel, hogy ismerjük egy adott sokaságban egy bizonyos valószínűségi változó eloszlását, és ez normális eloszlás, adott μ várható értékkel és σ szórással. Milyen valószínűséggel esik a valószínűségi változó értéke a várható érték körüli, adott sugarú intervallumba, tehát $x_1 = \mu - \Delta x$ és $x_2 = \mu + \Delta x$ közé?

Ilyen feladatoknál az aktuális változót 2.23 alapján úgy transzformáljuk, hogy normalizált Gauss-eloszlást kapjunk. Az intervallum végpontjai:

$$u_1 = -\,\Delta x/\sigma \quad \text{és} \quad u_2 = \Delta x/\sigma = v. \qquad 2.28$$

2.11-et alkalmazva $P(u_1 \le u \le u_2) = F(u_2) - F(u_1) = \phi(v) - \phi(-v)$. A táblázatokban viszont csak pozitív argumentumra találjuk meg a ϕ hibaintegrált. Az ábrán a két pöttyözött terület a függvény szimmetriája miatt egyenlő, azaz

$$\phi(-u) = 1 - \phi(u) \qquad P(-v \le u \le v) = 2\,\phi(v) - 1 \qquad 2.29$$



Annak a valószínűsége pedig, hogy a változó értéke kiessen az adott szimmetrikus intervallumból, tehát egy adott tűrésnél jobban eltérjen a várható értéktől:

$$P(u \leq -v \cup u \geq v) = 1- (2\phi(v)-1) = 2(1-\phi(v)). \qquad 2.30$$

Példa

Legyen a másodéves lányok magasságeloszlásának valószínűségi sűrűségfüggvénye
(a magasságot h-val jelölve és cm-ben mérve):

$$f(h) = \frac{1}{\sqrt{2\pi}5} e^{-0,5\left(\frac{h-165}{5}\right)^2}$$

a/ Mi annak a valószínűsége, hogy egy véletlenül kiválasztott lány magassága 160 és 170 cm között legyen?
b/ Ha az évfolyamon 80 lány van, várhatóan hány magasabb 160 cm-nél, hánynak a magassága 160 cm és 170 cm közötti, illetve 170 és 175 cm közötti érték?
c/ Van-e 175 cm-nél magasabb lány? És 150 cm-nél alacsonyabb?

Megoldás:
Térjünk át a normalizált eloszlásra (*A*) és keressük ki az 1. Táblázatból a $\phi(u)$ függvény értékeit (*B*). Az eloszlás várható értéke μ = 165 cm, szórása σ = 5 cm.

A

h (cm)	u
150	-3
160	-1
170	1
175	2

B

v	$\phi(v)$	$2\phi(v)-1$
0	0,5000	0
1	0,8413	0,6826
2	0,9772	0,9544
3	0,9986	0,9972

a/ A 160 - 170 cm intervallum a várható érték körüli σ sugarú intervallumnak felel meg (v=1), a táblázat szerint ebbe az intervallumba 68,26% valószínűséggel esnek a magasságok.
b/ A 160 cm a normalizált változóban u = -1-nek felel meg, annak valószínűsége, hogy $u \leq -1$ legyen, $P(u\leq-1) = \phi(-1) = 1-\phi(1)$ = 15,27 %, 1-P = 84,13%. Ennyi annak a valószínűsége, hogy valakinek a magassága nagyobb legyen, mint 160 cm. Mivel $P = n_{kedvező}/N_{összes}$, az évfolyamon 80 lány közül n = 80·0,8413 = 67 lánynak a magassága 160 cm-nél nagyobb. Hasonlóan, a 160 és 170 cm közötti magasságú lányok száma n = 80·0,6826 = 54,6, azaz 54-55 lány magassága esik 160 és 170 cm közé. A [170 cm, 175 cm] intervallum a normalizált változóban az [1,2] intervallumnak felel meg. $P(1\leq u\leq2)=\phi(2)-\phi(1)=0,1359$. Ez 11 személyt jelent.
c/ A 175 cm-es magasság u = 2-nek felel meg. $P(u>2)=1-P(u\leq2)=1-\phi(2)=0,0228$. Ez azt jelenti, hogy várhatóan két 175 cm-nél magasabb lány van.
A 150 cm u = -3-nak felel meg, $P(u<-3)=\phi(-3)=1-\phi(3)=0,0014$. Ez 0,1 személynek felel meg, tehát az évfolyamon valószínűleg nincs 150 cm-nél alacsonyabb lány.

Célszerű megjegyezni, hogy a normális eloszlásnál a várható érték körüli σ sugarú intervallumba (v=1) a valószínűségi változó értéke 68,3 % valószínűséggel esik, a 2σ sugarú intervallumba pedig (v=2) kb. 95,4 % valószínűséggel.

A középérték eloszlásának tulajdonságai

Mérjük egy sokaságon a ξ sajátosságot n-szer. Képzeljünk el egy n mérésből álló mintát, ahol az egyes mérések (egyelőre elképzelt) eredményei $x_1,...,x_n$. Ezek valószínűségi változók, még nem tudjuk, milyen értéket kapnak a mérésnél. Az $x_1,...,x_n$ valószínűségi változó számtani közepe

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

szintén valószínűségi változó, tehát tartozik hozzá egy $f(x_1,...,x_n)$ valószínűségi sűrűségfüggvény, és kérdezhetjük, mi ennek a várható értéke és varianciája. Az egyes mérési eredmények függetlenek egymástól, tehát

$$f(x_1,...,x_n) = f(x_1)...f(x_n).$$

Mivel ugyanazt a mérést ismételjük, az egyes mérési eredmények várható értéke $E[x_i] = \mu$ és varianciája $Var[x_i] = \sigma^2$ azonos minden egyes mérésre. Az összeg és konstansszoros várható értékére és varianciájára kapott 2.15, 2.17, 2.20, 2.21 formulákat alkalmazva kapjuk, hogy

$$E\,[\overline{x}] = 1/n \sum_{i=1}^{n} E\,[x_i] = \mu \qquad \text{és}$$

$$Var\,[\overline{x}] = 1/n^2 \sum_{i=1}^{n} Var\,[x_i] = \frac{1}{n} Var[x] = \frac{1}{n}\sigma^2 \qquad 2.31$$

Azaz a középérték várható értéke megegyezik az egyes mérések várható értékével, varianciája viszont n-ed része az egyes mérésének.

A centrális határeloszlás tétele szerint bármilyen eloszlású sokaság esetén az n elemű minta számtani középértékének eloszlása a minta elemszámának növekedésével egy olyan normális eloszláshoz tart, melynek várható értéke megegyezik az eredeti eloszlás várható értékével.
Ez azt jelenti, hogy ha már egyetlen mérési eredmény is átlagnak, pl. időátlagnak tekinthető, akkor várható, hogy az Gauss-eloszlású lesz. A mérési eredmények viszont nagyon gyakran ilyen átlagértékek. A mutató tehetetlensége miatt egy átlagértéknek megfelelő helyzetbe áll be. Ha elektronikusan gyűjtünk adatot, azt is egy bizonyos ideig tesszük, és az átlagjelet dolgozzuk tovább fel. Így a gyakorlatban legtöbbször normális eloszlású mérési eredményekkel találkozunk.

Az eloszlásparaméterek becslése

A mintavétel és mérés célja, hogy információt kapjunk a sokaságon az adott tulajdonság eloszlásáról, azaz meg tudjuk becsülni az eloszlásparamétereket a sokaság elemszámánál sokkal kisebb minta alapján. A becsült paramétereket hullámvonallal fogjuk jelölni. Egy becslés **torzítatlan** a θ paraméterre nézve, ha a becsült és valóságos várható értékek megegyeznek, azaz

$$E\,[\tilde{\theta}] = \theta \qquad \text{vagy} \qquad E\,[\tilde{\theta} - \theta] = 0, \qquad \text{a hiba várható értéke 0.}$$

A várható érték és a variancia becslése

A várható értéket úgy vezettük be véges elemű, diszkrét sokaságra, mint a sokaságra vett átlagát az adott tulajdonságnak (2.13). Ha most nem az egész sokaságot vesszük, csak egy mintát belőle, becsülhetjük úgy az egész sokaságra vonatkozó átlagot, hogy csak a mintára átlagolunk, azaz a várható értéket, μ-t a következőképp becsüljük:

$$\tilde{\mu} = 1/n \sum_{i=1}^{n} x_i \qquad 2.32$$

Amíg a mérésről csak beszélünk, $\tilde{\mu}$ valószínűségi változó, az x_i valószínűségi változók számtani közepe, melynek várható értéke megegyezik az egyes mérés várható értékével. Tehát $\tilde{\mu} = \mu$, a becslés torzítatlan.

Hasonlóan okoskodva, a varianciát becsülhetjük az egyes mérések hibanégyzetének átlagával:

$$\tilde{\sigma}^2 = 1/n \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Meg akarjuk határozni ennek a becslésnek a várható értékét; de egyszerűbb, ha $n\tilde{\sigma}^2$ várható értékét számítjuk ki.

$$E\,[n\tilde{\sigma}^2] = E\left[\sum_i (x_i - \bar{x})^2\right] = E\left[\sum_i \left((x_i - \mu) - (\bar{x} - \mu)\right)^2\right] =$$

$$= \sum_i E\left[\left((x_i - \mu) - \frac{1}{n}\left(\sum_j x_j - \mu\right)\right)^2\right] = \sum_i E\left[\left(\left(1 - \frac{1}{n}\right)(x_i - \mu) - \frac{1}{n}\sum_{j \neq i}(x_j - \mu)\right)^2\right] =$$

$$= \sum_i \left(\left(1 - \frac{1}{n}\right)^2 E\left[(x_i - \mu)^2\right] - \frac{2}{n}\left(1 - \frac{1}{n}\right) E\left[\sum_{j \neq i}(x_i - \mu)(x_j - \mu)\right] + \frac{1}{n^2} E\left[\left(\sum_{j \neq i}(x_j - \mu)\right)^2\right]\right)$$

Az első tag éppen az egyes mérések varianciájának (1-1/n) -szerese. A második tagban a j=i tag kimarad, így tehát az összeg egyes tagjaiban a tényezők függetlenek, szorzatuk várható értéke a tényezők várható értékének szorzata, ami viszont 0. A harmadik tagot kifejtve $(x_j-\mu)^2$-es tagok fognak szerepelni és vegyes szorzatok. Az utóbbiak várható értéke, az előző okfejtés értelmében 0. Így végül a fenti összeg a következőképp alakítható:

$$E\,[n\tilde{\sigma}^2] = \left(1 - \frac{1}{n}\right)^2 n\sigma^2 + \frac{1}{n^2}\sum_i \sum_{j \neq i} E\left[(x_j - \mu)^2\right] =$$

$$= \frac{(n-1)^2 \sigma^2}{n} + \frac{1}{n^2}(n-1)n\sigma^2 = \frac{n-1}{n}(n-1+1)\sigma^2 = \mathbf{(n\text{-}1)\ \sigma^2}\ ,$$

azaz a variancia n-szeresének becsült értéke a valóságos variancia (n-1)-szerese,

$$\tilde{\sigma}^2 = (n-1) / n\ \sigma^2\ . \qquad 2.33$$

Ez a becslés torzított, így célszerűbb a varianciát a mérési eredményekből a következőképp becsülni:

$$\tilde{\sigma}^2 = s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad 2.34$$

s_x -et az egyes mérési eredmények **korrigált tapasztalati szórásának** nevezzük.

Mivel 2.31 szerint a középérték varianciája az egyes mérések varianciájának n-ed része,

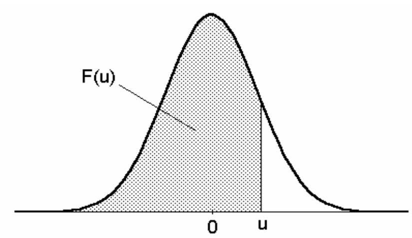
$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}}\ , \qquad 2.35$$

a **középérték korrigált tapasztalati szórása (standard deviációja):**

$$\boxed{s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} \qquad \text{és} \qquad s_{\bar{x}} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{(n-1)n}}} \qquad 2.36$$

I. táblázat

A normális eloszlás F(u) eloszlásfüggvénye



$$F(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-1/2\,t^2}\,dt = \phi(u)$$

u	0	1	2	3	4	5	6	7	8	9
0,0	,50000	,50399	,50798	,51197	,51595	,51994	,52392	,52790	,53188	,53586
0,1	,53983	,54380	,54776	,55172	,55567	,55962	,56356	,56749	,57142	,57535
0,2	,57926	,58317	,58706	,59095	,59483	,59871	,60257	,60642	,61026	,61409
0,3	,61791	,62172	,62552	,62930	,63307	,63683	,64058	,64431	,64803	,65173
0,4	,65542	,65910	,66276	,66640	,67003	,67364	,67724	,68082	,68439	,68793
0,5	,69146	,69497	,69847	,70194	,70540	,70884	,71226	,71566	,71904	,72240
0,6	,72575	,72907	,73237	,73565	,73891	,74215	,74537	,74857	,75175	,75490
0,7	,75804	,76115	,76424	,76730	,77035	,77337	,77637	,77935	,78230	,78524
0,8	,78814	,79103	,79389	,79673	,79955	,80234	,80511	,80785	,81057	,81327
0,9	,81594	,81589	,82121	,82381	,82639	,82894	,83147	,83398	,83646	,83891
1,0	,84134	,84375	,84614	,84850	,85083	,85314	,85543	,85769	,85993	,86214
1,1	,86433	,86650	,86864	,87076	,87286	,87493	,87698	,87900	,88100	,88298
1,2	,88493	,88686	,88877	,89065	,89251	,89435	,89617	,89796	,89973	,90147
1,3	,90320	,90490	,90658	,90824	,90988	,91149	,91309	,91466	,91621	,91774
1,4	,91924	,92073	,92220	,92364	,92507	,92647	,92786	,92922	,93056	,93180
1,5	,93319	,93448	,93574	,93699	,93822	,93943	,94062	,94179	,94295	,94408
1,6	,94520	,94630	,94738	,94845	,94950	,95053	,95154	,95254	,95352	,95449
1,7	,95543	,95637	,95728	,95818	,95907	,95994	,96080	,96164	,96246	,96327
1,8	,96407	,96485	,96562	,96638	,96712	,96784	,96856	,96926	,96995	,97062
1,9	,97128	,97193	,97257	,97320	,97381	,97441	,97500	,97558	,97615	,97670
2,0	,97725	,97778	,97831	,97882	,97932	,97982	,98030	,98077	,98124	,98169
2,1	,98214	,98257	,98300	,98341	,98382	,98422	,98461	,98500	,98537	,98574
2,2	,98610	,98645	,98679	,98713	,98745	,98778	,98809	,98840	,98870	,98899
2,3	,98928	,98956	,98983	,99001	,99036	,99061	,99086	,99111	,99134	,99158
2,4	,99180	,99202	,99224	,99245	,99266	,99286	,99305	,99324	,99343	,99361
2,5	,99379	,99369	,99413	,99430	,99446	,99461	,99477	99492,	,99506	,99520
2,6	,99534	,99547	,99560	,99573	,99586	,99598	,99609	,99621	,99632	,99643
2,7	,99653	,99664	,99674	,99683	,99693	,99702	,99711	,99720	,99728	,99736
2,8	,99744	,99752	,99760	,99767	,99774	,99781	,99788	,99795	,99801	,99807
2,9	,99813	,99819	,99825	,99830	,99836	,99841	,99846	,99851	,99856	,99860
3,0	,99865	,99869	,99874	,99878	,99882	,99886	,99889	,99893	,99896	,99900