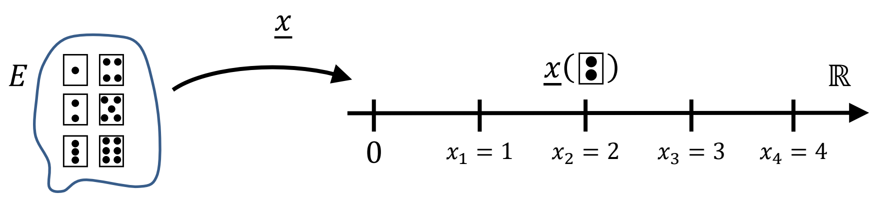
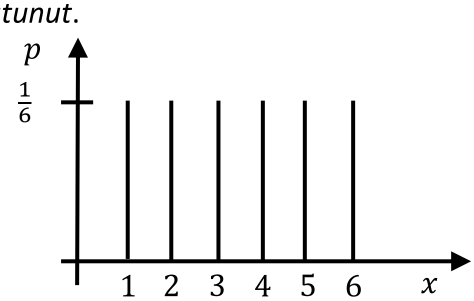
Diskreetin satunnaismuuttujan odotusarvo, keskihajonta ja varianssi

Kuten tilastojakaumia voitiin esittää tunnuslukujen (keskiarvo, moodi, mediaani, jne.) avulla, niin vastaavasti todennäköisyysjakaumia voidaan esittää/luonnehtia tunnusluvuilla.

Esimerkki Tarkastellaan nopanheittoa, kun satunnaismuuttujana $\underline{x}$ on nopan silmäluku (eli $\underline{x}$ on funktio, joka liittää nopan silmälukuun sen lukuarvon)



Nyt $p_1 = P(\underline{x} = 1) = \frac{1}{6}$ ja vastaavasti $p_2 = p_3 = p_4 = p_5 = p_6 = \frac{1}{6}$ eli satunnaismuuttuja $\underline{x}$ on *tasajakautunut*.



Saadaan keskiarvo:

$p_1 \cdot x_1 + p_2 \cdot x_2 + \ldots + p_6 \cdot x_6$

$= \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \ldots + \frac{1}{6} \cdot 6 = 3\frac{1}{2}$

kertomalla satunnaismuuttujan $\underline{x}$ arvot vastaavilla pistetodennäköisyyksillä ja laskemalla tulot yhteen. Tätä keskiarvoa sanotaan odotusarvoksi.

Määritelmä, *diskreetin satunnaismuuttujan odotusarvo*:
Olkoot diskreetin satunnaismuuttujan $\underline{x}$ arvot $x_1, x_2, \ldots, x_n$ ja vastaavat pistetodennäköisyydet $p_1, p_2, \ldots, p_n$. Tällöin tämän sat.muuttujan *odotusarvo* on

$$E(\underline{x}) = \mu = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n = \sum_{i=1}^{n} p_i x_i \,.$$

Merkki μ lausutaan: ”myy”. Odotusarvo on sopiva luku kuvaamaan jakauman keskikohtaa. Myös merkinnät $E\underline{x}$ tai $\mathbb{E}\underline{x}$ ovat käytössä. (Eräänä suosituksena olisi käyttää vahvennettua E-kirjainta $\mathbb{E}$ erotuksena otosavaruuden E symbolista.)

Määritelmä, *diskreetin satunnaismuuttujan keskihajonta ja varianssi*: Olkoot diskreetin satunnaismuuttujan $\underline{x}$ arvot $x_1, x_2, \dots, x_n$ ja vastaavat pistetodennäköisyydet $p_1, p_2, \dots, p_n$ sekä odotusarvo μ. Tällöin tämän satunnaismuuttujan *keskihajonta* on

$$D\underline{x} = \sigma = \sqrt{p_1(x_1-\mu)^2 + p_2(x_2-\mu)^2 + \ \dots + p_n(x_n-\mu)^2}$$
$$= \sqrt{\sum_{i=1}^{n} p_i(x_i-\mu)^2}$$

Keskihajonnan neliö

$$D^2\underline{x} = \sigma^2$$

on satunnaismuuttujan $\underline{x}$ *varianssi*. Myös merkintöjä

$$\mathbb{D}\underline{x} \quad \text{ja} \quad \mathbb{D}^2\underline{x}$$

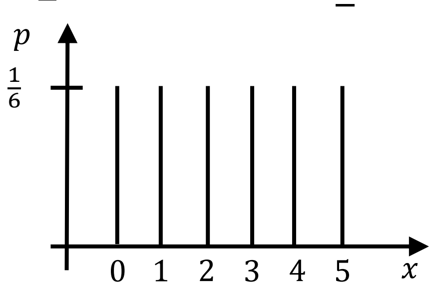
käytetään.

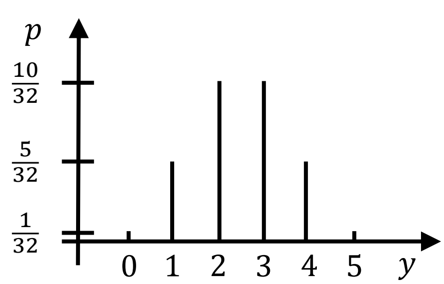
Esimerkki: Määritä/Laske satunnaismuuttujien $\underline{x}$ ja $\underline{y}$ keskihajonnat ja varianssit, kun näiden satunnaismuuttujien jakaumat ovat

x_k	0	1	2	3	4	5
p_k	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

y_k	0	1	2	3	4	5
p_k	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Nyt $\underline{x}$ on tasajakautunut ja $\underline{y}$ on binomijakautunut satunnaismuuttuja.





Havaitaan, että satunnaismuuttujien odotusarvot ovat samat 2,5, vaikka selvästi ne ovat eri tavoin jakautuneet.
Laskemalla saadaan (kannattaa laskea ensin varianssit, josta neliöjuuren otto antaa keskihajonnat):

$$D^2\underline{x} = \sigma^2 = \frac{1}{6}(0-2{,}5)^2 + \frac{1}{6}(1-2{,}5)^2 + \ \ldots + \frac{1}{6}(5-2{,}5)^2$$

$$= \frac{17{,}5}{6} \approx 2{,}92$$

$$\Rightarrow D\underline{x} = \sigma = \sqrt{D^2\underline{x}} = \sqrt{\frac{17{,}5}{6}} \approx 1{,}71$$

$$D^2\underline{y} = \sigma^2 = \frac{1}{32}(0-2{,}5)^2 + \frac{5}{32}(1-2{,}5)^2 + \ \ldots + \frac{1}{32}(5-2{,}5)^2$$

$$= \frac{40}{32} \approx 1{,}25$$

$$\Rightarrow D\underline{y} = \sigma = \sqrt{D^2\underline{y}} = \sqrt{\frac{40}{32}} \approx 1{,}12$$

Havaitaan, $D\underline{x} > D\underline{y}$, eli $\underline{x}$:n arvot ovat hajaantuneet enemmän, OK.

Edellä vilahti sana binomijakauma (…binomijakautunut…). Mitä se on?

Binomijakauma

Kertausta: Binomitodennäköisyyttä voitiin hyödyntää toistokokeissa.

Jos tapahtuman A todennäköisyys on p, eli $P(A) = p$, niin $P(\bar{A}) = (1-p) =: q$. Toistetaan tarkasteltavaa koetta n kertaa, jolloin tapahtuman A esiintymiskertojen lukumäärää voidaan tarkastella satunnaismuuttujana $\underline{x}$. Mahdollisia $\underline{x}$:n arvoja ovat siis $0,1,2,\ldots,n$. Näitä arvoja vastaa todennäköisyydet

$$P(\underline{x} = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k}$$

Saadaan binomijakautunut satunnaismuuttuja $\underline{x}$, merk. $\underline{x} \sim Bin(n,p)$.

Määritelmä, *binomijakauma ja sen odotusarvo/keskihajonta*:
Satunnaismuuttuja $\underline{x}$ noudattaa binomijakaumaa parametrein n, p, jos sen arvot ovat $x_0 = 0, x_1 = 1, \ldots, x_n = n$ ja vastaavat todennäköisyydet ovat

$$p_k = P(\underline{x} = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad \text{missä } k = 0,1,\ldots,n.$$

Binomijakautuneen satunnaismuuttujan $\underline{x}$ odotusarvolle ja keskihajonnalle voidaan johtaa lausekkeet

$$\mu = E(\underline{x}) = np \qquad \text{ja} \qquad \sigma = D\underline{x} = \sqrt{np(1-p)} = \sqrt{npq}$$

Esimerkki: Olkoot erään veikkauksen tulosvaihtoehdot $1, \mathrm{x}, 2$ yhtä todennäköisiä. Veikataan neljää kohdetta.

a) Muodosta satunnaismuuttujan $\underline{x}$="ristien lukumäärä" jakauma
b) Esitä jakauma histogrammina, viivadiagrammina
c) Laske satunnaismuuttujan $\underline{x}$ odotusarvo ja keskihajonta

a) Koska eri veikkauskohteilla ei ole vaikutusta toisiin, niin veikkausta voidaan pitää toistokokeena ja satunnaismuuttujaa $\underline{x}$ binomijakautuneena. Mahdolliset $\underline{x}$:n arvot (ristien lukumäärät) ovat

$$\mathcal{A}_{\underline{x}} = \{0,1,2,3,4\}.$$

Binomitodennäköisyyden nojalla

$$p_k = P(\underline{x} = k) = \binom{4}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{4-k}, \qquad k = 0,1,2,3,4$$

Joten

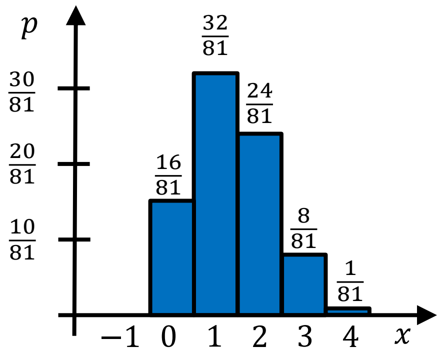
$$p_0 = P(\underline{x} = 0) = \binom{4}{0} \cdot \left(\frac{1}{3}\right)^0 \cdot \left(\frac{2}{3}\right)^4 = \frac{16}{81} \approx 0{,}198$$

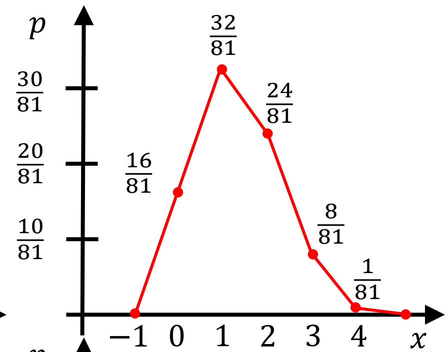
$$p_1 = P(\underline{x} = 1) = \binom{4}{1} \cdot \left(\frac{1}{3}\right)^1 \cdot \left(\frac{2}{3}\right)^3 = \frac{32}{81} \approx 0{,}395$$

$$p_2 = P(\underline{x} = 2) = \binom{4}{2} \cdot \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^2 = \frac{24}{81} \approx 0{,}296$$

$$p_3 = P(\underline{x} = 3) = \binom{4}{3} \cdot \left(\frac{1}{3}\right)^3 \cdot \left(\frac{2}{3}\right)^1 = \frac{8}{81} \approx 0{,}098$$

$$p_4 = P(\underline{x} = 4) = \binom{4}{4} \cdot \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^0 = \frac{1}{81} \approx 0{,}012$$





c) $\underline{x} \sim Bin\left(4, \frac{1}{3}\right)$, joten

$$\mu = E\left(\underline{x}\right) = np = 4 \cdot \frac{1}{3} = 1\frac{1}{3}$$

ja

$$\sigma = D\underline{x} = \sqrt{npq} = \sqrt{4 \cdot \tfrac{1}{3} \cdot \tfrac{2}{3}}$$

$$= \sqrt{\tfrac{8}{9}} = \frac{2\sqrt{2}}{3} \approx 0{,}943$$