

Sandsynlighedsregning

Definition: Sandsynlighedsfelt

Et sandsynlighedsfelt (U, p) beskriver et sandsynlighedseksperiment E og består af to dele:

- Udfaldsrummet U , som er en liste over alle mulige udfald der kan ske, når E udføres:
 $U: u_1, u_2, \dots, u_n$
- Sandsynlighedsfordeling p , som er en liste over sandsynligheden af hvert udfald i U .
 $p(u_1), p(u_2), \dots, p(u_n)$
 - Der skal gælde: $p(u_1) + p(u_2) + \dots + p(u_n) = 1$
 - Der skal for hvert i fra 1 til n gælde: $p(u_i) \geq 0$

Sætning: Kombineret sandsynlighed

Lad s og t være udfald i et sandsynlighedsfelt.

Additionsprincippet: Sandsynligheden for at enten s eller t sker, når E udføres én gang er:

$$p(s) + p(t)$$

Multiplikationsprincippet: Sandsynligheden for at både s og t sker, når E udføres to gange, er:

$$p(s) \cdot p(t)$$

Definition: Hændelse

En hændelse H i et sandsynlighedsfelt, er en delmængde af udfaldsrummet. Hvis et af udfaldene i delmængden sker ved udførelse af E , så siger vi at hændelsen H skete.

Sætning: Sandsynlighed for hændelse

En hændelse H sker, hvis ét af udfaldene i H sker. Af additionsprincippet følger det, at sandsynligheden for at H sker, er summen af sandsynligheder for hvert udfald i H .

Definition: Symmetrisk sandsynlighedsfelt

Et sandsynlighedsfelt kaldes symmetrisk, hvis alle udfald i U har samme sandsynlighed:

$$p(u_1) = p(u_2) = \dots = p(u_n)$$

Sætning: Symmetrisk sandsynlighed

Hvis et symmetrisk sandsynlighedsfelt har n udfald i udfaldsrummet, vil sandsynligheden for hver af dem være $\frac{1}{n}$.

Bevis:

Lad $p(u_1) = p(u_2) = \dots = p(u_n) = p$. Da fås:

$$p(u_1) + p(u_2) + \dots + p(u_n) = \underbrace{p + p + \dots + p}_{n \text{ gange}} = n \cdot p = 1$$

$$p = \frac{1}{n}$$

Definition: Stokastisk variabel

En stokastisk variabel X beskriver et sandsynlighedseksperiment, hvor udfaldene i U alene består af tal. X henter altså sine værdier fra en liste af tal $x_1, x_2, \dots, x_n$. Sandsynligheden for udfaldet x_i betegnes p_i - det vil sige: $P(X = x_i) = p_i$. På tabelform fås:

t	x_1	x_2	$\dots$	x_n
$P(X = t)$	p_1	p_2	$\dots$	p_n

Definition: Middelværdi og spredning

Middelværdien for X bestemmes: $\mu = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n$

Spredningen for X bestemmes: $\sigma = \sqrt{(x_1 - \mu)^2 \cdot p_1 + (x_2 - \mu)^2 \cdot p_2 + \dots + (x_n - \mu)^2 \cdot p_n}$

Binomialfordeling

Definition: Binomialfordeling

En stokastisk variabel X , der angiver antallet af "succeser" ved uafhængige gentagelser af identiske sandsynlighedseksperimenter, siges at være *binomialfordelt*.

Sandsynligheden for "succes" ved én udførelse af eksperimentet kaldes sandsynlighedsparameteren, og betegnes p . Antallet af gange eksperimentet udføres kaldes antalsparameteren, og betegnes n .

At en stokastisk variabel X er binomialfordelt med sandsynlighedsparameter p og antalsparameter n betegnes med: $X \sim (n, p)$.

Sætning: Binomialfordelingens sandsynlighedsfunktion

For en binomialfordelt stokastisk variabel $X \sim b(n, p)$ gælder:

$$P(X = r) = K(n, r) \cdot p^r \cdot (1 - p)^{n-r}$$

Ræsonnement:

Da sandsynligheden for udfaldet "succes" er p , er sandsynligheden for den modsatte hændelse "fiasko" nødvendigvis $1 - p$, da de to sandsynligheder til sammen skal give 1.

Ved udførelse af n gentagelser af sandsynlighedseksperimentet fås en række af udfald, som er enten "succes" (S) eller "fiasko" (F). Et eksempel på en række kunne være:

S S F F F S F S ... S F

De tilhørende sandsynligheder bliver da:

$$p \ p \ (1 - p) \ (1 - p) \ (1 - p) \ p \ (1 - p) \ p \ \cdots \ p \ (1 - p)$$

Da rækken er en "både-og"-situation, skal sandsynligheden for at alle udfaldene falder præcist i denne række følge, beregnes ved at de enkelte sandsynligheder ganges sammen.

Hvis antallet af S sættes til r , vil antallet af F være $n - r$:

$$p \cdot p \cdot (1 - p) \cdot (1 - p) \cdot (1 - p) \cdot p \cdot (1 - p) \cdot p \cdot \cdots \cdot p \cdot (1 - p) = p^r \cdot (1 - p)^{n-r}$$

Det ses heraf, at sandsynligheden for enhver anden række af r ”succeser” og $n - r$ ”fiaskoer” vil have samme sandsynlighed, i det sandsynligheden af en række kun afhænger af antallet af ”succeser”. Dermed vides at sandsynligheden for enhver række af n udfald, med netop r succeser vil være:

$$p^r \cdot (1 - p)^{n-r}$$

Antallet af rækker med n udfald, som netop indeholder r ”succeser”, svarer til antallet af måder man kan vælge r blandt de n udfald til at være succeserne, hvilket vil svare til:

$$K(n, r)$$

Da hændelsen $X = r$ fås, hvis én af de rækker af udfald der indeholder netop r succeser realiseres, vil sandsynligheden for at hændelsen indtræffer være summen af sandsynligheder for disse rækker. Da sandsynligheden for hver enkelt er den samme, er summen blot antallet af rækker ganget med sandsynligheden for den enkelte række, hvilket giver:

$$P(X = r) = K(n, r) \cdot p^r \cdot (1 - p)^{n-r}$$

Sætning: Normalfordelingsapproximation

For en binomialfordelt stokastisk variabel $X \sim b(n, p)$ gælder, når $n \cdot p < 5$ og $n \cdot (1 - p) < 5$ at:

Sandsynligheden for et normalt udfald er ca. 95%:

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \approx 95\%$$

Sandsynligheden for et exceptionelt udfald er ca. 0,3%:

$$P(X < \mu - 3\sigma) + P(X > \mu + 3\sigma) = 0,3\%$$

Bevis:

Sætningen bevises ikke her.

Statistiske test

Definition: Populationsandel

I en population med N individer opfylder M af individerne at have egenskaben E .

Dermed bliver andelen af populationen med egenskaben E bestemt ved: $p = \frac{M}{N}$.

Definition: Stikprøve

En stikprøve er en delmængde af en population, som er udtaget ved at man vælger ét individ ad gangen på en helt tilfældig måde. Det betyder at ved hvert valg af et individ, har alle individer præcis samme sandsynlighed for at blive udvalgt til stikprøven.

Antallet af individer i stikprøven kaldes *stikprøvestørrelsen* og betegnes n . Det antages at stikprøven er meget mindre end populationen, så sandsynligheder ved valg af ét individ ikke ændrer sig målbart efter at et individ er udtrukket.

Definition: Stikprøveandel

Hvis der i en stikprøve med størrelse n , er m individer med egenskaben E , da kan andelen af individer i stikprøven med egenskaben bestemmes ved:

$$\hat{p} = \frac{m}{n}$$

Vi kalder denne størrelse for *stikprøveandelen*, og udtaler symbolet for den ”p hat”.

Sætning:

Hvis en stikprøve med størrelse n udtages af en population, og stikprøveandelen af individer med egenskaben E opgøres til $\hat{p}$, da vil vi med 95% sikkerhed kunne sige, at andelen p af individer i populationen med egenskaben E ligger i intervallet:

$$\left[\hat{p} - 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}; \hat{p} + 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right]$$

Vi kalder dette interval for et 95%-konfidensinterval for populationsandelen med egenskab E .

Størrelsen $s = 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$ kaldes den ”statistiske usikkerhed”, og vi siger at andelen med egenskab E i populationen med 95% sikkerhed er $\hat{p} \pm s$.

Bevis:

Vores bedste bud på p er $\hat{p}$. Hvis vi lader X være en stokastisk variabel der angiver antallet af valgte individer til stikprøven, som har egenskaben E , da er X binomialfordelt: $X \sim b(n, \hat{p})$. Dermed fås følgende

$$\mu = n \cdot \hat{p}$$

$$\sigma = \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}$$

Vi ved da at vi med 95% sandsynlighed skal have udfald i intervallet $[\mu - 2\sigma; \mu + 2\sigma]$. I dette interval vil μ svare til antallet af individer i stikprøven med egenskab n , mens $\mu \pm 2\sigma$ vil være grænserne for antal individer i stikprøven. Ved at dividere med n omregnes fra antal til andel:

$$\mu \pm 2 \cdot \sigma = n \cdot \hat{p} \pm 2 \cdot \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}$$

$$\frac{\mu \pm 2 \cdot \sigma}{n} = \frac{n \cdot \hat{p} \pm 2 \cdot \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}}{n} = \frac{n \cdot \hat{p}}{n} \pm \frac{2 \cdot \sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}}{n} = \hat{p} \pm 2 \cdot \frac{\sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}}{n}$$

Da $n = \sqrt{n^2}$ kan den sidste brøk omskrives til: $\frac{\sqrt{n \cdot \hat{p} \cdot (1 - \hat{p})}}{n} = \sqrt{\frac{n \cdot \hat{p} \cdot (1 - \hat{p})}{n^2}} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$. Dermed fås:

$$\frac{\mu \pm 2 \cdot \sigma}{n} = \hat{p} \pm 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Dermed er det vist, at 95%-konfidensintervallet er: $\left[\hat{p} - 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}; \hat{p} + 2 \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \right]$

Slag med 1, 2, 3 og n terninger

Definition:

En terning er en symmetrisk fordelt stokastisk variabel X med udfaldsrummet $U = \{1, 2, 3, \dots, m\}$.

Vi siger at terningen har m sider. Sandsynligheden for et udfald t er altså $P(X = t) = \frac{1}{m}$.

Almindeligvis taler vi mest om terninger med 6 sider, dvs. vi lader $m = 6$.

En 6-siders terning har således sandsynlighedsfeltet:

t	1	2	3	4	5	6
$P(X = t)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Sætning: Middelværdi for 6-siders terning

Middelværdien $E(X)$ for X der angiver udfald ved slag med en 6 siders terning er 3,5.

Bevis:

$$\begin{aligned} E(X) &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} \\ &= \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = \frac{7}{2} = 3,5 \end{aligned}$$

Ræsonnement: Slag med flere terninger

Ved slag med flere terninger bliver udfaldsrummet flertydigt. Det er f.eks. muligt at beskrive slag med n terninger hver hjælp af *øjensummen*, dvs. sum af udfald.

$$U = \{n, n + 1, \dots, m \cdot n\}$$

Mindste udfald er $\underbrace{1 + 1 + \dots + 1}_{n \text{ gange}} = n$ og største er $\underbrace{m + m + \dots + m}_{n \text{ gange}} = m \cdot n$

Principielt vil vi for terningerne $X_1, X_2, \dots, X_n$ kalde $X = X_1 + X_2 + \dots + X_n$ for øjensummen.

For 2 terninger med 6 sider (dvs. $n = 2$ og $m = 6$) fås altså udfaldsrummet:

$$U = \{2, 3, 4, \dots, 11, 12\}$$

Dette udfaldsrum er imidlertid ikke symmetrisk. $P(X = 2)$ er noget andet end $P(X = 6)$.

Vi kan i stedet lave udfaldet af slag med n terninger vi blot ser på de enkelte udald i en rækkefølge, altså kalde $X = (X_1, X_2, \dots, X_n)$. Da der på hver plads er m forskellige udfald, kan der altså dannes $\underbrace{m \cdot m \cdot \dots \cdot m}_{n \text{ gange}} = m^n$ forskellige udfald. Disse er lige sandsynlige:

$$P(X = (x_1, x_2, \dots, x_n)) = \frac{1}{m^n}$$

For et slag med 2 terninger med 6 sider fås altså: $X = (X_1, X_2)$ og $P(X = (x_1, x_2)) = \frac{1}{6^2} = \frac{1}{36}$

De mulige udfald ved at skelne mellem terningerne kan samles i en tabel:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Det gælder at alle udfald har samme sandsynlighed:

$$P(X = (1,1)) = P(X = (4,2)) = P(X = (1,6)) = \frac{1}{36}$$

Vi kan nu opfatte f.eks. ”øjensummen” X som en hændelse i dette udfaldsrum.

F.eks. $X = 6$:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

$$\text{Heraf fås } P(X = 6) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{5}{36}$$

Vi kan også udregne det med princippet: $\frac{\text{"Antal gunstige"}}{\text{"Antal mulige"}}$

Ved optælling ses at der er 5 gunstige udfald ud af 36 mulige, for hændelsen ”øjensum 6”.

Dermed opstilles for øjensum for 2 terninger med 6 sider, sandsynlighedsfeltet:

t	2	3	4	5	6	7	8	9	10	11	12
$P(X = t)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Som kan forkortes til:

t	2	3	4	5	6	7	8	9	10	11	12
$P(X = t)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Middelværdien for øjensum med to terninger X_1 og X_2 bliver da:

$$\begin{aligned}
 E(X) &= E(X_1 + X_2) \\
 &= 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + 5 \cdot \frac{4}{36} + 6 \cdot \frac{5}{36} + 7 \cdot \frac{6}{36} + 8 \cdot \frac{5}{36} + 9 \cdot \frac{4}{36} + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} \\
 &= \frac{2 + 6 + 12 + 20 + 30 + 42 + 40 + 36 + 30 + 22 + 12}{36} \\
 &= \frac{20 + 20 + 30 + 42 + 40 + 66 + 34}{36} \\
 &= \frac{70 + 42 + 40 + 100}{36} \\
 &= \frac{252}{36} \\
 &= \frac{240 + 12}{6 \cdot 6} = \frac{40 + 2}{6} = \frac{42}{6} \\
 &= 7
 \end{aligned}$$

Dette passer med:

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3,5 + 3,5 = 7$$

Dette er et mere generelt princip der kan formuleres:

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

For slag med tre terninger med 6 sider fås udfaldsrummet:

$$U = \{3, 4, 5, \dots, 17, 18\}$$

Nu er det svært at tegne de unikke udfald i en tabel, men den indeholder $m^n = 6^3 = 216$ forskellige og lige sandsynlige udfald. De har altså hver sandsynligheden $\frac{1}{216}$.

Øjensummen 3 realiseres kun af udfaldet (1,1,1), hvorfor $P(X = 3) = \frac{1}{216}$.

Øjensummen 4 realiseres af udfaldene (2,1,1), (1,2,1) og (1,1,2), hvor $P(X = 4) = \frac{3}{216} = \frac{1}{72}$

Øjensummen 5 realiseres af udfaldene: (2,2,1), (2,1,2), (1,2,2), (3,1,1), (1,3,1) og (1,1,3)

Dermed fås altså $P(X = 5) = \frac{6}{216} = \frac{1}{36}$

Og så videre...

Middelværdien bliver:

$$E(X) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 3,5 + 3,5 + 3,5 = 10,5$$

For slag med n terninger med 6 sider fås udfaldsrummet:

$$U = \{n, n + 1, \dots, 6n\}$$

Der bliver 6^n forskellige udfald, hver med sandsynlighed $\frac{1}{6^n}$.

Hvis X angiver øjensummen for terningerne $X_1, X_2, \dots, X_n$, da fås:

For $X = n$ er der kun udfaldet (1,1,...,1) der er gunstigt. Derfor fås: $P(X = n) = \frac{1}{6^n}$.

For $X = n + 1$ er der udfaldene (2,1,...,1), (1,2,...,1)...(1,1,...,2). Da der er n terninger, hvoraf vi kan vælge 2'eren på n måder, bliver sandsynligheden: $P(X = n + 1) = \frac{n}{6^n}$.

Herefter bliver det hurtigt *meget* kompliceret at regne på.

$X = n + 2$ kan realiseres ved (3,1,1,...,1), (1,3,1,...,1),..., (1,1,1,...,3), men også ved at få to 2'ere, hvilket kan ske på $K(n, 2)$ måder, hvor $K(n, 2)$ jo er antal måder at vælge 2 blandt n på.

Middelværdien bliver: $E(X) = E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n) = 3,5 \cdot n$