Slumpvariabel (Stokastisk variabel)

En stokastisk variabel är en variabel som antar olika numeriska värden bestämda av slumpen.

Exempel: Diskret stokastisk variabel

- Antal flickor i en slumpmässigt vald trebarnsfamilj
- Antal telefonsamtal till en växel under en slumpmässigt vald minut
- Summan av antal prickar vid kast med två tärningar

Exempel: Kontinuerlig stokastisk variabel

- Vikten på ett slumpmässigt valt nyfött barn
- Livslängden på en slumpmässigt vald glödlampa

Diskret stokastisk variabel

$X =$ Variabeln

$x =$ Numeriska värdet

$\Pr(X = x) = p(x)$ betecknar sannolikheten att *variabeln* X antar värdet x.
$p(x)$ kallas slumpvariabelns sannolikhetsfördelning eller sannolikhetsfunktion.

$\Pr(X \leq x) = F(x)$ betecknar sannolikheten att *variabeln* X är mindre eller lika med värdet x.
$F(x)$ kallas slumpvariabelns fördelningsfunktion.

För sannolikheterna $p(x)$ gäller följande två villkor

1. $p(x) \geq 0$
2. $\sum_{alla\ x} p(x) = 1$

Exempel:

$X =$ Antalet rum i en slumpmässigt vald lägenhet?

X har följande sannolikhetsfunktion och fördelningsfunktion

x	$\Pr(X = x)$	$F(x) = \Pr(X \leq x)$	x^2	$2x + 3$
1	0.020	0.020	1	5
2	0.230	$0.020 + 0.230 = 0.250$	4	7
3	0.300	$0.250 + 0.300 = 0.550$	9	9
4	0.345	$0.550 + 0.345 = 0.895$	16	11
5	0.100	$0.895 + 0.100 = 0.995$	25	13
6	0.005	$0.995 + 0.005 = 1.000$	36	15

$$
\begin{aligned}
\Pr(X \leq 2) &= F(2) = 0.25 \\
\Pr(X > 2) &= 1 - \Pr(X \leq 2) \\
&= 1 - F(2) = 0.75 \\
\text{Alter. } \Pr(X > 2) &= p(3) + p(4) \\
&\quad + p(5) + p(6) \\
&= 0.3 + 0.345 + 0.1 \\
+0.005 &= 0.75 \\
\Pr(2 \leq X \leq 4) &= \Pr(X \leq 4) - \Pr(X < 2) \\
&= \Pr(X \leq 4) - \Pr(X \leq 1) \\
&= F(4) - F(1) \\
&= 0.895 - 0.02 = 0.875 \\
\text{Alter. } \Pr(2 \leq X \leq 4) &= p(2) + p(3) + p(4) \\
&= 0.23 + 0.3 + 0.345 = 0.875
\end{aligned}
$$

Väntevärde

Väntevärdet (eller förväntade värdet) för variabeln X definieras
$E(X) = \mu = \sum_{alla\ x} xp(x)$
$= x_1 p(x_1) + x_2 p(x_2) + ... + x_n p(x_n)$

Väntevärdet i exemplet ovan blir:

$$
\begin{aligned}
E(X) &= \mu = \sum_{alla\ x} xp(x) \\
&= x_1 p(x_1) + x_2 p(x_2) + ... + x_6 p(x_6) \\
&= 1 \cdot 0.02 + 2 \cdot 0.23 + ... + 6 \cdot 0.005 = 3.29
\end{aligned}
$$

Väntevärde för en funktion av en s.v. X

$$
\begin{aligned}
E[h(x)] &= \sum_{alla\ x} h(x)\, p(x) = h(x_1)\, p(x_1) \\
&\quad + h(x_2)\, p(x_2) + ... + h(x_n)\, p(x_n)
\end{aligned}
$$

Forts. exemplet: $h(x) = x^2$

$$
\begin{aligned}
E[h(x)] &= E\left(x^2\right) = \sum_{alla\ x} x^2 p(x) \\
&= x_1^2 p(x_1) + x_2^2 p(x_2) + ... + x_6^2 p(x_6) \\
&= 1^2 \cdot 0.02 + 2^2 \cdot 0.23 + ... + 6^2 \cdot 0.005 \\
&= 1 \cdot 0.02 + 4 \cdot 0.23 + ... + 36 \cdot 0.005 \\
&= 11.84
\end{aligned}
$$

Om $h(x) = x^3$ blir $E[h(x)] = E\left(x^3\right) = \sum_{alla\ x} x^3 p(x)$
Om $h(x) = x^8$ blir $E[h(x)] = E\left(x^8\right) = \sum_{alla\ x} x^8 p(x)$

Forts. exemplet: $h(x) = 2x + 3$

$$
\begin{aligned}
E[h(x)] &= E(2X+3) = \sum_{alla\ x} (2x+3)\, p(x) \\
&= (2x_1+3)\, p(x_1) + (2x_2+3)\, p(x_2) \\
&\quad +... + (2x_6+3)\, p(x_6) \\
&= 2x_1 p(x_1) + 2x_2 p(x_2) + ... + 2x_6 p(x_6) \\
&\quad + 3p(x_1) + 3p(x_2) + ... + 3p(x_6) \\
&= 2\left[x_1 p(x_1) + x_2 p(x_2) + ... + x_6 p(x_6)\right] \\
&\quad + 3\left[p(x_1) + p(x_2) + ... + p(x_6)\right] \\
&= 2 \sum_{alla\ x} xp(x) + 3 \sum_{alla\ x} p(x) \\
&= 2E(X) + 3 = 2 \cdot 3.29 + 3 = 9.58
\end{aligned}
$$

Väntevärdet för en linjär funktion $aX + b$ av en s.v. X

$E(aX + b) = aE(X) + b$ där a och b är konstanter.

Varians och standardavvikelse

Variansen för en slumpvariabel X definieras som väntevärdet för funktionen $(X - \mu)^2$, dvs
$V(X) = \sigma^2 = E(X-\mu)^2 = \sum_{alla\ x} (x-\mu)^2 p(x)$
$= (x_1 - \mu)^2 p(x_1) + (x_2 - \mu)^2 p(x_2)$
$+.... + (x_n - \mu)^2 p(x_n)$

Forts. exemplet:

$$
\begin{aligned}
V(X) &= E(X-\mu)^2 = \sum_{alla\ x} (x-\mu)^2 p(x) \\
&= (x_1-\mu)^2 p(x_1) + (x_2-\mu)^2 p(x_2) \\
&\quad +... + (x_6-\mu)^2 p(x_6) \\
&= (1-3.29)^2 \cdot 0.02 + (2-3.29)^2 \cdot 0.23 \\
&\quad +... + (6-3.29)^2 \cdot 0.005 \\
&= 1.016
\end{aligned}
$$

Alternativt

$$\begin{aligned} V(X) &= E(X-\mu)^2 = E\left(X^2 - 2X\mu + \mu^2\right) \\ &= E\left(X^2\right) - 2\mu E(X) + \mu^2 \\ &= E\left(X^2\right) - 2E(X)E(X) + [E(X)]^2 \\ &= E\left(X^2\right) - 2[E(X)]^2 + [E(X)]^2 \\ &= E\left(X^2\right) - [E(X)]^2 = 11.84 - 3.29^2 = 1.016 \end{aligned}$$

Standardavvikelsen (den genomsnittliga avvikelsen från väntevärdet) ges av

$$SD(X) = \sigma = sqrt(V(X))$$

Forts. exemplet:

$$\sigma = sqrt(V(X)) = sqrt(1.016) = 1.008$$

Forts. exemplet: $h(x) = 2x + 3$

Sen tidigare har vi att $E[h(x)] = 2E(X) + 3$

$$\begin{aligned} V[h(x)] &= E[h(x) - E[h(x)]]^2 \\ &= E[(2X+3) - (2E(X)+3)]^2 \\ &= E[2X + 3 - 2E(X) - 3]^2 \\ &= E[2X - 2E(X)]^2 \\ &= 2^2 E[X - E(X)]^2 = 2^2 V(X) \\ &= 4 \cdot 1.016 = 4.06 \end{aligned}$$

Variansen för en linjär funktion $aX + b$ av en s.v. X

$V(aX + b) = a^2 V(X)$ där a och b är konstanter.

Tvåpunktsfördelad (Bernoullifördelad) slumpvariabel

Slumpvariabeln kan anta två värden, t. ex. klave eller krona, med sannolikhetsfördelning

x	0	1
$p(x)$	$1-p=q$	p

Väntevärdet för denna $0-1$ variabel blir

$$E(X) = \sum xp(x) = 0(1-p) + 1p = p$$

$$E\left(X^2\right) = \sum x^2 p(x) = 0^2(1-p) + 1^2 p = p$$

Variansen blir då

$$V(X) = E\left(X^2\right) - [E(X)]^2 = p - p^2 = p(1-p)$$