Stokastiske variable og sannsynlighetsfordelinger

Dekkes av kapittel 3 i læreboka.

Definisjon: Stokastisk variabel er en variabel som får sin numeriske verdi bestemt av utfallet av et stokastisk forsøk. □

Eksempler: Oljeinnhold i boreprøve, funksjonstid til elektronisk komponent, temperatur i morgen kl 12, antall feil i ny programkode, antall ulykker på en veistrekning, resultatet av et terningkast, osv (de fleste variable kan sees på som stokastiske). □

Stokastiske variable er enten diskrete (f.eks. antall feil, antall ulykker, terningkast) eller kontinuerlige (f.eks. oljeinnold, funksjonstid, temperatur). En diskret variabel kan kun anta et tellbart antall verdier, mens en kontinuerlig variabel kan anta enhver verdi i et intervall.

Notasjon: Stokastiske variable betegnes med store bokstaver, X, Y, Z, osv ("før forsøket er utført"), mens numerisk verdi til den stokastiske variabelen ("etter forsøket er utført") betegnes med små bokstaver, x, y, z, osv. $\square$

Diskrete og kontinuerlige stok. var. må behandles noe forskjellig. For diskrete variable gir det mening å snakke om $P(X = a)$, mens for kontinuerlige variable vil alltid $P(X = a) = 0$. For kontinuerlige variable ser vi bare på sannsynligheter for å få verdier i et intervall, f.eks. $P(a \leq X \leq b)$.

Diskrete sannsynlighets-fordelinger (3.2)

Definisjon: $f(x)$ er punktsannsynligheten til den diskrete stok. var. X dersom for alle mulige utfall x:

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $f(x) = P(X = x)$ $\square$

Eksempel: Myntkast, kaster en mynt 3 ganger. La X="antall kron".

Utfallsrom	x	Utfallsrom	x
KKK	3	KMM	1
KKM	2	MKM	1
KMK	2	MMK	1
MKK	2	MMM	0

Vi ser at:

$$\begin{aligned} f(0) &= P(X=0)=1/8 \\ f(1) &= P(X=1)=3/8 \\ f(2) &= P(X=2)=3/8 \\ f(3) &= P(X=3)=1/8 \end{aligned}$$

Dette er punktsannsynligheten til X. $\square$

Definisjon: Den <u>kumulative fordelingen</u> til en diskret stok. var. X er:

$$F(x)=P(X\leq x)=\sum_{t\leq x} f(t) \quad \square$$

Kontinuerlige sannsynlighetsfordelinger (3.3)

Definisjon: $f(x)$ er sannsynlighetstettheten til den kontinuerlige stok. var. X dersom:

1. $f(x) \geq 0$ for alle $x \in \mathcal{R}$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$ $\square$

Merk:

$$\text{Diskret:} \quad P(a \leq X \leq b) = \sum_{a \leq x \leq b} f(x)$$

$$\text{Kontinuerlig:} \quad P(a \leq X \leq b) = \int_a^b f(x)dx$$

Merk også at for diskrete stok. var. er $P(X \leq a) \neq P(X < a)$, mens for kontinuerlige stok. var. er $P(X \leq a) = P(X < a) + P(X = a) = P(X < a)$.

Eksempel: La Y="funksjonstid til elektronisk komponent". Anta at det er kjent at Y har sannsynlighetstetthet:

$$f(y) = \begin{cases} \frac{1}{5}e^{-y/5} & ,y \geq 0 \\ 0 & ,\text{ellers} \end{cases}$$

(Sjekk selv at $\int_{-\infty}^{\infty} f(y)dy = 1$). Vi kan nå regne ut sannsynligheter, f.eks:

$$\begin{aligned} P(Y \leq 5) &= \int_{-\infty}^{5} f(y)dy = \int_{0}^{5} \frac{1}{5}e^{-y/5} \\ &= -e^{-1} + e^{0} = \underline{\underline{0.63}} \\ P(4 < Y < 5) &= \int_{4}^{5} f(y)dy = \int_{4}^{5} \frac{1}{5}e^{-y/5} \\ &= -e^{-1} + e^{-4/5} = \underline{\underline{0.08}} \quad \square \end{aligned}$$

Definisjon: Den <u>kumulative fordelingen</u> til en kontinuerlig stok. var. X er:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \quad \square$$

Funksjonstideksemplet:

$$\begin{aligned} F(y) &= \int_{-\infty}^{y} f(t)dt = \int_{0}^{y} \frac{1}{5} e^{-t/5} dt = 1 - e^{-y/5} \\ F(5) &= 1 - e^{-5/5} = \underline{\underline{0.63}} \quad \square \end{aligned}$$

Simultanfordeling (3.4)

Simultanfordeling = samtidig fordeling til to eller flere stok. var.

Definisjon: Dersom X og Y er to *diskrete* stok. var. så er $f(x,y)$ den simultane punktsannsynligheten til X og Y dersom

1. $f(x,y) \geq 0$
2. $\sum_x \sum_y f(x,y) = 1$
3. $f(x,y) = P(X = x, Y = y)$ $\square$

Definisjon: Dersom X og Y er to *kontinuerlige* stok. var. så er $f(x,y)$ simultantettheten til X og Y dersom

1. $f(x,y) \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dx dy = 1$
3. $P((X,Y) \in A) = \int \int_A f(x,y) dx dy$ $\square$

Eksempel: Vann fra vannverk. La

$$\begin{aligned} X &= \text{relativt vannforbruk} \\ Y &= \text{relativt humusinnhold} \end{aligned}$$

Anta at X og Y har simultan sannsynlighetstetthet

$$f(x,y) = \begin{cases} x - \frac{1}{2}y + \frac{3}{4} & ,0 \le x \le 1,\ \ 0 \le y \le 1 \\ 0 & \text{,ellers} \end{cases}$$

(Sjekk selv at $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$).
Gir f.eks. at:

$$\begin{aligned} &P(X < 0.5, Y > 0.75) \\ &= \int_{0.75}^{1}\int_{0}^{0.5}(x - \frac{1}{2}y + \frac{3}{4})dxdy = \underline{\underline{\frac{9}{128}}} \quad \square \end{aligned}$$

Definisjon: To stok. var. X og Y er uavhengige (dvs informasjon om verdien til den ene variabelen påvirker ikke fordelingen til den andre) hvis og bare hvis

$$f(x, y) = g(x) \cdot h(y) \quad \text{,for alle } x, y$$

der $g(x)$ er punktsanns./tettheten til X og $h(y)$ er punktsanns./tettheten til Y. $\square$

Eksempel: I vanneksemplet ser vi at $f(x, y)$ ikke kan skrives på formen $g(x) \cdot h(y)$, dvs X og Y er ikke uavhengige (de er med andre ord *avhengige*). $\square$