F6 Diskreta stokastiska variabler.

Christian Tallberg

Avdelningen för Nationalekonomi/Statistik

Karlstads universitet

Stokastisk variabel (slumpvariabel)

Tidigare hade vi experiment definierade genom det diskreta utfallsrummet $\mathbf{\Omega} = \{w_1, w_2, ...w_n\}$, bestående av möjliga händelser $A_1, A_2, A_3, ...$.

Vi tilldelade varje möjlig händelse en viss sannolikhet $P(A_i)$ som hade vissa egenskaper (axiom).

- Exempel $\mathbf{\Omega} = \{$"regnar", "regnar inte"$\}$, $A =$ "regnar", $P(A) = 0.1, P(A^*) = 0.9$.

Nu koncentrerar vi oss på experiment och utfall som kan utttryckas, eller översättas till, *reella tal*.

Exempel:

- 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

Hur gör vi med experimentet "regnar"/"regnar inte".

- Givet ett experiment med utfallsrum $\mathbf{\Omega}$, kallar vi en funktion X, som tilldelar varje utfall (eller mängd av utfall) $w \in \mathbf{\Omega}$, ett och endast ett reellt tal $X(w) = x$, en *stokastisk variabel*. Dvs, den stokastiska variabeln X definierar en funktion från utfallsrummet till de reella talen. Den har definitionsmängden $\mathbf{\Omega} = \{w_1, w_2, ...w_n\}$ och värdemängden $X(w) = x$. Se exempel 3.1, sid 46 i Blom.

Slumpvariabler kan vara *diskreta* eller *kontinuerliga*.

Beteckningskonvention:

- Slumpvariabler betecknas med stora bokstäver: X, Y, Z osv.
- Ett numeriskt värde som antas av en slumpvariabel betecknas med motsvarande lilla bokstav: x, y, z osv.

Exempel: Låt $X =$ antal prickar vi kommer att få vid ett tärningskast. Nu kastar vi tärningen och får en trea. Då har slumpvariabeln X antagit värdet $x = 3$.

Sannolikhetsfördelning för en diskret stokastisk variabel

- En *diskret* slumpvariabel kan bara anta ett ändligt antal möjliga värden, oftast (men inte alltid) heltalsvär

- Eftersom slumpen bestämmer vilket värde X skall anta, så kan vi tala om *sannolikheten* att X antar olika värden.

- Vi låter $p_x(x)$ beteckna sannolikheten att slumpvariabeln X skall anta det numeriska värdet x, dvs
$$p_x(x) = P(X = x).$$

- Sannolikheterna, $p_x(x)$, för alla möjliga värden x, som den diskreta slumpvariabeln X kan anta, kallas slumpvariabelns sannolikhetsfördelning (eller sannolikhetsfunktion).

För sannolikheterna $p_x(x)$ gäller följande villkor

1. $0 \leq p_x(x) \leq 1$

2. $P(X \in A) = \sum_{k \in A} p_x(k)$

3. $\sum_{k \in \Omega} p_x(k) = 1$

- Sannolikhetsfördelningen för en diskret slumpvariabel kan presenteras i tabellform eller med stolpdiagram.

Fördelningsfunktion för en diskret stokast variabel

- Vi låter $F_x(x)$ beteckna sannolikheten att slumpvariabeln X skall anta ett värde mindre eller lika med det numeriska värdet x, dvs
$$F_x(x) = P(X \leq x), \quad -\infty < x < \infty.$$

och beräknas som
$$F_x(x) = \sum_{j \leq x} p_x(j), \quad j \leq x.$$

- Sannolikheterna, $F_x(x)$, för alla möjliga värden x, kallas slumpvariabelns fördelningsfunktion.

- Fördelningsfunktionen för en diskret slumpvariabel kan presenteras i tabellform eller med trappstegsdiagram

Exempel: Vi tänker oss en stor population av småföretag. Låt X = Antalet anställda i ett slumpmässigt valt småföretag.

Sannolikhetsfördelningen och fördelningsfunktionen för X ges av följande tabell:

x	$p_x(x) = P(X = x)$	$F_x(x) = P(X \leq x)$
1	0.020	0.020
2	0.230	$0.020 + 0.230 = 0.250$
3	0.300	$0.250 + 0.300 = 0.550$
4	0.345	$0.550 + 0.345 = 0.895$
5	0.100	$0.895 + 0.100 = 0.995$
6	0.005	$0.995 + 0.005 = 1.000$

Vi kan med hjälp av fördelningsfunktionen beräkna bland annat följande sannolikheter:

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

För en fördelningsfunktion $F_x(x)$ gäller att:

- $$F_x(x) \to \begin{cases} 0 & \text{då } x \to -\infty \\ 1 & \text{då } x \to \infty \end{cases}$$
- $F_x(x)$ är en icke-avtagande funktion av x.
- $F_x(x)$ är kontinuerlig till höger för varje x.

Bernoullifördelningen

Vi skall se på några olika sannolikhetsfördelningar som brukar användas som sannolikhetsmodeller i vissa typiska situationer. Den första är Bernoullifördelningen.

- En slumpvariabel är *Bernoullifördelad* om den antar endast värdena 0 och 1 med sannolikheten $(1 - p)$ och p respektive. $(0 \leq p \leq 1)$. Sannolikhetsfördelningen ser alltså ut så här:

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

En typisk användning av Bernoullifördelningen är följande:

- Vi har ett slumpförsök där vi bara är intresserade av ifall en viss händelse A inträffar eller ej.
- Låt $P(A) = p$ och $P(A^*) = 1 - p$.
- Låt X vara en indikatorvariabel för händelsen A. Dvs om A inträffar, så blir $X = 1$, och om A inte inträffar, så blir $X = 0$. Alltså är X en slumpvariabel som anger om händelsen A inträffar eller ej.
- Då är X Bernoullifördelad med
$$P(X = 1) = p.$$

Exempel: Slumpförsök = Ett kast med en tärning. Låt X vara indikatorvariabel för händelsen att få sexa. Dvs,

$$X = \begin{cases} 0 \text{ om vi inte får en sexa} \\ 1 \text{ om vi får en sexa} \end{cases}.$$

Då är X Bernoullifördelad med $\pi = 1/6$

Binomialfördelningen

Används som modell i situation av följande slag:

- Ett slumpförsök upprepas n gånger (*oberoende* upprepningar). Varje gång två möjliga resultat (Bernoulliförsök): A och icke-A (A^*).
- Sannolikheten för A är konstant vid varje upprepning av försöket, $P(A) = p$.
- X = *Antalet gånger* som A inträffar totalt.
- Då är X en binomialfördelad variabel med parametrarna n och p.

- Formell definition: X är en binomialfördelad variabel om den antar värdena $x = 0, 1, 2, ..., n$ med sannolikheterna:

$$p_x(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- Kortfattat betecknas det oftast med något av följande tre alternativ:

$X \sim Bin(n;\pi)$ X är $Bin(n;\pi)$ $X \in Bin(n;\pi)$.

- Varför ser sannolikheterna ut som de gör?
- Notera, att för fallet $n = 1$ har vi Bernoullifördelningen

Exempel: Två identiska oberoende kast med ett symmetriskt mynt. Låt X = Antal krona. Ge sannolikhetsfördelning, väntevärde och varians för X.

Lösningsalternativ 1: De möjliga utfallen:

$$\mathbf{S} = \{(kr, kr), (kr, kl)(kl, kr)(kl, kl)\}$$

ger följande sannolikhetsfördelning

x	0	1	2
$p_x(k)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Lösningsalternativ 2: Identifiera fördelningen för X :

$$X \sim Bin\,(2; 0.5)$$

Då kan sannolikhetsfördelningen, väntevärde och varians beräknas på följande sätt:

$$\begin{aligned} p_x\,(0) &= \binom{2}{0} 0.5^0 0.5^2 = 0.25 \\ p_x\,(1) &= \binom{2}{1} 0.5^1 0.5^1 = 0.5 \\ p_x\,(2) &= \binom{2}{2} 0.5^2 0.5^0 = 0.25 \end{aligned}$$

- *Tabell 2* i tabellsamlingen ger de kumulativa sannolikheterna, dvs fördelningsfunktionens sannolikheter

  $$F_x\,(x) = \sum_{j \le x} p_x\,(j) = p\,(0) + p\,(1) + ... + p\,(x)$$

  för $n = 2, 3, ..., 25, 30, 35, 40, 45, 50$ och $\pi = 0.05, 0.10, 0.15, ..., 0.95$.

- Binomialsannolikheter kan även enkelt erhållas med t ex SPSS (för många fler värden på n och π).

Exempel: Vi gör tio kast med ett asymetriskt mynt där sannolikheten för krona är $p = 0.2$. Oberoende antas.

a) Vilken fördelning har $X =$ Antalet krona?

b) Bestäm $P\,(X \le 3)$.

c) Bestäm $P\,(1 \le X \le 3)$.

d) Bestäm $P\,(X \ge 7)$.

e) Bestäm $P\,(X = 5)$.

Svar:

a)

$$X \sim Bin\,(10; 0.2)\,.$$

b)

$$P\,(X \le 3) = 0.8791$$

c)

$$\begin{aligned} P\,(1 \le X \le 3) &= P\,(X \le 3) - P\,(X < 1) \\ &= P\,(X \le 3) - P\,(X \le 0) \\ &= 0.8791 - 0.1074 = 0.7717 \end{aligned}$$

d)

$$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0.9991 = 0.0009$$

e)

$$\begin{aligned} P(X = 5) &= P(X \leq 5) - P(X \leq 4) \\ &= 0.9936 - 0.9672 = 0.0264 \end{aligned}$$

Hypergeometrisk fördelning (jfr med binomialfördelningen)

Typisk situation:

- Population med N individer, varav Np har en viss egenskap, medan de övriga $N - Np$ saknar denna egenskap.
- Från populationen väljs slumpmässigt *utan återläggnin* ett stickprov med n individer.
- Låt $X =$ Antal individer i stickprovet, som har den aktuella egenskapen.
- Då är X en *hypergeometriskt fördelad* variabel.

- Formell definition: X är en hypergeometriskt fördelad variabel om sannolikhetsfördelningen ges av:
$$p_x(k) = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}},$$
för heltalsvärden k, sådana att $0 \leq k \leq Np$ och $0 \leq n - k \leq N - Np$.
- Kortfattat skriver man
$$X \in Hyp(N; n; p)$$

Exempel:

I en låda ligger tio lampor, varav 30 % av lamporna är defekta. Man drar fem stycken slumpmässigt *utan återläggning*.

a) Vad är sannolikheten att högst en av de dragna lamporna är defekta?

b) Vad är sannolikheten att minst en av de dragna lamporna är defekta?

c) Beräkna väntevärde och varians.

Svar:

a) Låt X =Antalet defekta lampor i urvalet. $X \sim Hyp\,(N = 10; n = 5; p = 0.3)$. Då gäller att

$$\begin{aligned} P\,(X \leq 1) &= P\,(X = 0) + P\,(X = 1) \\ &= \frac{\binom{3}{0}\binom{10-3}{5-0}}{\binom{10}{5}} + \frac{\binom{3}{1}\binom{10-3}{5-1}}{\binom{10}{5}} \\ &= \frac{21}{252} + \frac{105}{252} = \frac{126}{252} = 0.5. \end{aligned}$$

b)

$$\begin{aligned} P\,(X \geq 1) &= 1 - P\,(X = 0) \\ &= 1 - \frac{21}{252} = 0.92. \end{aligned}$$

Poissonfördelningen

Ibland är man intresserad av att studera antal gånger en speciell händelse inträffar inom en given tidsperiod eller på ett givet fysiskt objekt.

- Händelserna inträffar oberoende av varandra med en viss intensitet μ, som hela tiden är samma.

- Låt X = Antal gånger som händelsen inträffar under ett tidsintervall av given längd.

- Då är X en *Poissonfördelad* variabel (egentligen under en mer noggrann formulering av förutsättningarna).

Exempel:

- X = Antal telefonsamtal till en växel mellan 9.00 och 9.10 en vardagsmorgon.

- X = Antal tryckfel på en slumpmässigt vald sida i en bok.

- X = Antal kunder som kommer till en butik mellan 12.00 och 14.00.

- Formell definition: X är en Poissonfördelad variabel om den antar värdena $k = 0, 1, 2, \ldots$ med sannolikheterna:

$$p_x\,(k) = \frac{\mu^k e^{-\mu}}{k!}, \quad \mu > 0$$

  där e är den matematiska konstanten 2.718....

Kortfattat skriver man

$$X \in Po\,(\mu)\,,$$

där konstanten μ är lika med intensiteten, dvs det genomsnittliga antalet gånger som händelsen skall inträffa under en tidsperiod av given längd.

Exempel:

$X =$ Antalet flygkrascher i civilflyget i Sverige. Antag att X är en Poissonfördelad variabel. Vi vet att det inträffar i genomsnitt 2.1 flygkrascher i civilflyget per år i Sverige, så vi sätter därför $\mu = 2.1$.

a) Vad är sannolikheten för högst en olycka under ett år?

Alternativ 1:

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) \\ &= \frac{2.1^0 e^{-2.1}}{0!} + \frac{2.1^1 e^{-2.1}}{1!} \\ &= 0.122 + 0.257 = 0.38 \end{aligned}$$

- *Tabell 3* i tabellsamlingen ger de kumulativa sannolikheterna, dvs fördelningsfunktionens sannolikheter

$$F_x(k) = P(X \leq k)$$

för $\mu = 0.1, 0.2, 0.3, ..., 28.0$.

- Poissonsannolikheter kan även enkelt erhållas med t ex SPSS (för många fler värden på μ).

Alternativ 2:

$$P(X \leq 1) = 0.380 \text{ enligt tabell 3.}$$

b) Vad är sannolikheten för exakt tre olyckor under ett år?

Alternativ 1:

$$P(X = 3) = \frac{2.1^3 e^{-2.1}}{3!} = 0.189.$$

Alternativ 2:

$$\begin{aligned} P(X = 3) &= P(X \leq 3) - P(X \leq 2) \\ &= 0.839 - 0.650 \\ &= 0.189 \text{ enligt tabell 3.} \end{aligned}$$