4. Stokastiska variabler

En *stokastisk variabel* (s.v.) är en funktion som definieras i utfallsrummet. Varje stokastisk variabel har en viss sannolikhetsstruktur.

Ex: Man kastar två tärningar. Låt

$$X = \text{summan av de två antal ögon.}$$

Då är X en stokastisk variabel, som kan anta värdena 2,3,4,5,6,7,8,9,10,11 och 12 med motsvarande sannolikheter $\frac{1}{36}$, $\frac{2}{36}$, $\frac{3}{36}$, $\frac{4}{36}$, $\frac{5}{36}$, $\frac{6}{36}$, $\frac{5}{36}$, $\frac{4}{36}$, $\frac{3}{36}$, $\frac{2}{36}$ och $\frac{1}{36}$.

Ex.1b: En urna innehåller 20 bollar numrerade från 1 till 20. Man drar 3 bollar på måfå ur urnan utan återläggning. Man slår vad om att minst en av bollarna har nummer 17 eller större. Vad är då sannolikheten att man vinner?

Ex.1c: Man kastar ett mynt så länge som man får en klave men inte fler än n gånger. Sannolikheten att få en klave är p. Man kan definiera en s.v.

$$X = \text{antal myntkast.}$$

Vilka värden och med hur stor sannolikhet kan X anta?

4.2 Diskreta stokastiska variabler

Definition: En s.v. är *diskret* om den kan anta ett ändligt eller uppräkneligt oändligt många olika värden.

Definition: Storheterna $p(a) = P(X = a) = p_X(a)$ kallas gemensamt för *frekvensfunktionen* eller *sannolikhetsfunktionen* (probability mass function) för den s.v. X.

Definition: En funktion F av en s.v. X

$$F(x) = P(X \leq x), \; -\infty < x < \infty$$

kallas den *(kumulativa) fördelningsfunktionen* (distribution function) av X.

4.3 Väntevärde

Definition: Om X är en diskret s.v. med frekvensfunktionen $p(x)$, då är *väntevärdet* (expected value) av X

$$\mathbf{E}[X] = \sum_{p(x)>0} x p(x).$$

Ex.3d: 120 elever delas i 3 bussar. En av bussarna har 36 elever, en 40 elever och den sista 44 elever. Man väljer en av de 120 eleverna på måfå. Låt X vara antalet elever i bussen, där eleven som man har valt, sitter. Vad är $\mathbf{E}[X]$?

4.4 Väntevärde av en funktion av en s.v.

Proposition 4.1: Om X är en diskret s.v. och x_i, $i \geq 1$, är dess möjliga värden med resp. sannolikheter $p(x_i)$, då för varje funktion g $(g(x) \in \mathbf{R})$

$$\mathbf{E}[g(X)] = \sum_i g(x_i)p(x_i).$$

Följdsats 4.1: Om a och b är konstanter, då är

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b.$$

4.5 Varians

Definition: Om X är en s.v. med väntevärde μ, då är *variansen* av X

$$\mathrm{Var}(X) = \mathbf{E}[(X - \mu)^2] = \mathbf{E}[X^2] - \mu^2.$$

Roten av variansen kallas *standardavvikelse* av X, dvs.

$$SD(X) = \sqrt{\mathrm{Var}(X)}.$$

Om a och b är konstanter, då är

$$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X).$$

4.6 Bernoulli- och binomialfördelning

Ett försök med två möjliga utfall: "success" och "failure". Man definierar en s.v.

$$X = \begin{cases} 1 & \text{om "success"} \\ 0 & \text{om "failure"} \end{cases}$$

med frekvensfunktionen

$$p(1) = P(X = 1) = p \text{ och } p(0) = P(X = 0) = 1 - p.$$

Den s.v. X har en *bernoullifördelning* med parameter $p \in (0, 1)$.

Låt oss utföra n oberoende försök ovan och definiera

$$X = \text{antalet succéer i } n \text{ försök}.$$

Då har X en *binomialfördelning* med parametrar (n, p), $X \sim Bin(n, p)$.

Frekvensfunktionen av X är

$$P(X = i) = \binom{n}{i} p^i(1-p)^{n-i}, \ i = 0, 1, ..., n.$$

Ex.6c: (Wheel-of-fortune) En spelare väljer ett tal mellan 1 och 6. Tre tärningar kastas. Om talet, som spelaren har valt, kommer i, $i = 1, 2, 3$, gånger, vinner hon/han \$i. Om talet inte kommer förlorar hon/han \$1. Är spelet ett "fair game" för spelaren?

Ex.6d: Anta att ögonfärget bestäms av ett visst par av gener: d är den dominerande genen (bruna ögon) och r den resessiva genen (blåa ögon). Då är en person med gener dd dominant, med gener rr resessiv och med gener rd (eller dr) hybrid. Dominant och hybrid är lika i utseendet (bruna ögon). Ett barn får en gen från sin mamma och en från sin pappa. Anta att både föräldrar är hybrida. Vad är då sannolikheten att 3 av 4 barn har bruna ögon?

4.7 Poissonfördelning

Låt X vara en diskret s.v. som kan anta värdena 0,1,2,... Den har en *Poissonfördelning* med parameter λ, om för någon $\lambda > 0$

$$p(i) = P(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}, \ i = 0, 1, 2, ...,$$

$X \sim Pois(\lambda)$.

Poissonfördelningen har många tillämpningar därför att den kan användas för att approximera binomialfördelningen med parametrar (n, p), när n är stor och p liten så att np inte är för stor eller för liten. Några exempel om variabler som kan antas ha en Poissonfördelning:

- antal skrivfel per sida i en bok
- antal personer, som lever mer än hundra år, i en kommun
- antal felknäppta telefonnummer per dag
- antal kunder som besöker ett postkontor på en viss dag

Ex.7a: Anta att antal skrivfel/sida i läroboken är Poisson(0.5)-fördelat. Vad är sannolikheten att det finns minst ett skrivfel på sidan 154?

Ex.7b: En produkt gjort av en viss maskin är felaktig m.s. 0.1. Hur stor är sannolikheten att ett stickprov av 10 produkter har högst en felaktig produkt?

Ex: (Födelsedagsexemplet) Var och en av n personer har sin födelsedag lika sannolikt på vilken av helst av de 365 dagarna av året. Man vill räkna sannolikheten att i en grupp av n (oberoende) personer alla har olika födelsedagar.

I Ex.5i i Kapitel 2 använde vi kombinatorik för att lösa problemet och fick att om $n \geq 23$, är sannolikheten ovan mindre än 0.5.

4.8. Andra diskreta fördelningar

4.8.1. Geometrisk fördelning

Betrakta oberoende försök sådana att en succé inträffar (på varje försök) med sannolikhet p. Man fortsätter att utföra försöken tills en succé inträffar för första gången. Låt X vara antalet försök som behövs. Då är

$$P(X = n) = (1 - p)^{n-1}p, \ n = 1, 2, ...$$

Man säger att X har en *geometrisk fördelning* med parameter p, $X \sim Geom(p)$.

Ex.8a: Man har N vita och M svarta bollar i en urna. Man tar bollar på måfå från urnan, en boll åt gången, tills man får en svart boll. Om man antar att bollen återläggs varje gång innan man tar den nästa, vad är sannolikheten att

a) exakt n bollar måste tas?

b) minst k bollar måste tas?

4.8.2. Negativ binomialfördelning

Anta att man utför oberoende försök, vart och ett av vilka har succésannolikheten p, $0 < p < 1$, tills r succéer har inträffat. Låt X vara antalet försök som måste utföras. Då är

$$P(X = n) = \binom{n-1}{r-1} p^r (1 - p)^{n-r}, \ n = r, r + 1, ...$$

Man säger att X har en *negativ binomialfördelning* med parametrar (r, p), $X \sim NegBin(r, p)$.

Ex.8e: (Banachs tändstickaproblem) En piprökare har alltid två askar av tändstickor med sig, en i sin vänster ficka och en i sin höger ficka. Varje gång han behöver en tändsticka, tar han asken från en av fickorna på måfå. Betrakta en situation att piprökaren har just märkt att en av askarna är tomm. Anta att var och en av askarna hade N tändstickor i början. Vad är sannolikheten att det finns exakt k tändstickor kvar i den andra asken, $k = 0, 1, ..., N$?

4.8.3. Hypergeometrisk fördelning

Anta att man tar n bollar på måfå utan återläggning från en urna som innehåller N bollar, m vita och $N-m$ svarta. Låt X vara antalet vita bollar som har dragits. Då är

$$P(X=i)=\frac{\begin{pmatrix} m \\ i \end{pmatrix}\begin{pmatrix} N-m \\ n-i \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}},\ i=0,1,...,n$$

och X har en *hypergeometrisk fördelning* med parametrar (n,N,m),
$X \sim HypGeom(n,N,m)$.

Obs! $P(X=i)=0$ om inte $n-(N-m)\leq i\leq min(n,m)$.

Ex.8h: (Fångst-återfångstmetoden; Capture-recapture method, Mark-release method) Ett okänt antal, säg N, djur bor på ett område. Ekologer vill skatta N. De fångar först några, säg m, av djuren, märker dem och återlägger till populationen. Efter att de märkta djuren har blandat sig med de andra, fånger egologerna igen några, säg n, djur. Låt X vara antalet märkta djur på andra fångsten. Man antar att antalet djur, N, förblir det samma mellan fångsterna och att varje djur har lika stor sannolikhet att bli fångat både gång. Då är $X \sim HypGeom(n,N,m)$.

4.9. Egenskaper av fördelningsfunktionen F

Definition: $F(b)=P(X\leq b)$, $-\infty<b<\infty$, kallas *fördelningsfunktionen* för den s.v. X.

Några egenskaper av fördelningsfunktionen F:

1) F är en icke-avtagande funktion: om $a<b$, då är $F(a)\leq F(b)$

2) $\lim\limits_{b\to\infty} F(b)=1$

3) $\lim\limits_{b\to-\infty} F(b)=0$

4) F är kontinuerlig till höger: för varje b och för var och en avtagande sekvens b_n, $n\geq 1$, som konvergerar mot b, $\lim\limits_{b\to\infty} F(b_n)=F(b)$

Kapitel 2.6:

- En följd av händelser $\{E_n,\, n \geq 1\}$ är *växande* om $E_1 \subset E_2 \subset ... \subset E_n \subset E_{n+1} \subset ...$ och den är *avtagande* om $E_1 \supset E_2 \supset ... \supset E_n \supset E_{n+1} \supset ...$
- Om $\{E_n,\, n \geq 1\}$ är en växande följd av händelser, definierar man
$$\lim_{n\to\infty} E_n = \cup_{i=1}^{\infty} E_n,$$
där $E_n \subset E_{n+1}$ för varje n. Om $\{E_n,\, n \geq 1\}$ är en avtagande följd av händelser, definierar man
$$\lim_{n\to\infty} E_n = \cap_{i=1}^{\infty} E_n,$$
där $E_n \supset E_{n+1}$ för varje n.
- **Proposition 6.1**: Om $\{E_n,\, n \geq 1\}$ är antingen växande eller avtagande följd av händelser, då är
$$\lim_{n\to\infty} P(E_n) = P(\lim_{n\to\infty} E_n).$$

Ex.9a: En s.v. X har fördelningsfunktionen

$$X = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \leq x < 1 \\ \frac{2}{3}, & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

Hitta

a) $P(X < 3)$

b) $P(X = 1)$

c) $P(X > \frac{1}{2})$

d) $P(2 < X \leq 4)$