1 Föreläsning III

1.1 Diskret (Sannolikhets-)fördelning

Med *diskret* menas i matematik, att något antar ett ändligt antal värden eller *uppräkneligt* oändligt med värden ex.vis $\{1, 2, 3, \ldots\}$. Med fördelning menas ur sannolikheterna fördelas mellan olika utfall.

1.1.1 Likformig fördelning

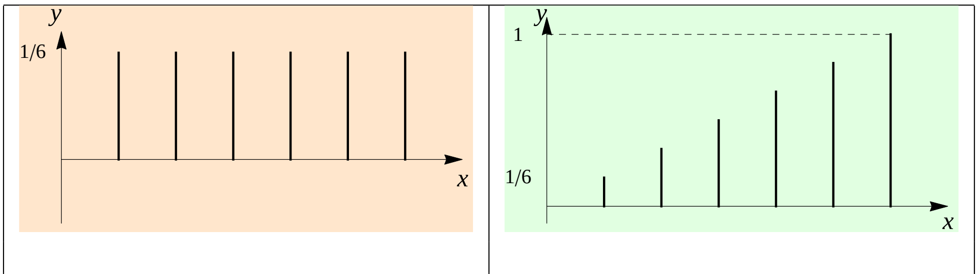
Ex 3.1 I exempel 1.1 har vi sannolikheterna $\frac{1}{6}$ för alla utfalll $1, 2, 3, 4, 5, 6$ (poäng vid ett tärningskast). Man talar då om en *likformig* fördelning, alla utfall har samma sannolikhet. Vi inför nu en *stokastisk variabel*, ξ, (läses "xi"). För att förskorta ned texten/förklaringen av en typ av utfall används en sådan variabel. Låt ξ =antal poäng vid ett kast med en tärning.
Händelsen A =antal poäng ≤ 4 kan då skrivas

$$A = \{\xi \leq 4\} \text{ med motsvarande sannolikhet } P(\xi \leq 4) = \frac{4}{6} = \frac{2}{3}.$$

Vi går ett steg till och inför en *frekvens-* eller *sannolikhetsfunktion* $f(x) := P(\xi = x)$. Det följer att

$$f(x) = \frac{1}{6}, \quad x = 1, 2, 3, 4, 5, 6.$$

Vi sätter $f(x) = 0$ för övriga (heltals)värden på x. Vi kan dessutom rita $y = f(x)$ tillsamman med *fördelningsfunktionen* $F(x) := P(\xi \leq x)$.



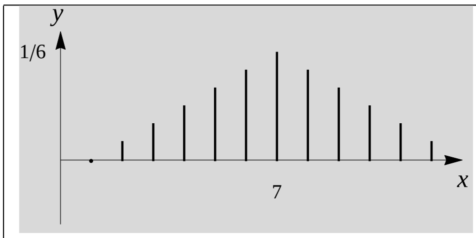
Frekvensfunktion | Fördelningsfunktion

Man skriver att $\xi \in$ Likf(6) "ξ tillhör klassen av likformig fördelningar med parameter 6".

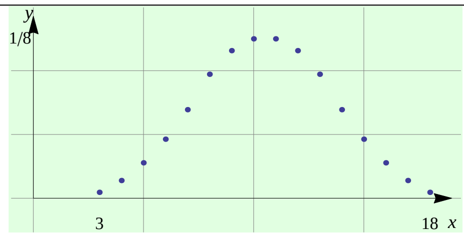
Ex 3.2 I exempel 1.2 har vi summann av två tärningskast poäng. Låt η ("eta") vara summans poäng. Det är klart att η antar värdena $2, 3, \ldots, 12$ med olika sannolikheter. Vi har att, med $P(\eta = x) =: f(x)$, *frekvensfunktionen* för η.

$$P(\eta = 2) = f(2) = \frac{1}{36}, \quad P(\eta = 3) = f(3) = \frac{2}{36} \text{ etc.}$$

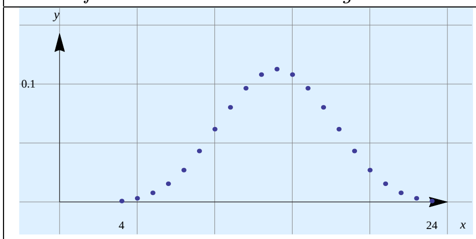
P.s.s. kan man låta η, betyda summan av tre tärningskasts poäng och rita motsvarande frekvensfunktion.



Frekvensfunktion
för summan av två tärningskast.



Frekvensfunktion
för summan av tre tärningskast.



Frekvensfunktion
för summan av fyra tärningskast.

Kommentarer

- Det är typiskt för en fördelningsfunktion, att den är växande mot 1: Mer exakt $\lim_{x\to\infty} f(x) = 1$.
- I exempel 1.2 är fördelning inte likformig, eftersom på utfallsrummet $\Omega_1 = \{2, 3, 4, \ldots, 12\}$ är sannolikheterna olika.

1.1.2 Binomialfördelning

Ex 3.3 (Ex 2.12) Vid fiske med kastspö är sannolikheten att få napp (fiskfångst) p vid varje kast. Karin kastar n gånger. Vad där sannolikheten att hon får exakt x napp (fiskar)? Antag att händelsen varje kast är oberoende.

Lösning:

Sökt sannolikhet $P(A)$, där A är händelsen att få napp exat x gånger av $n$är

$$P(A) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots n.$$

Med ξ =antal fångade fiskar vid n kast kan vi skriva

$$P(\xi = x) = f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots n.$$

Man säger att ξ är binomialfördelad med parametrar n och p. Detta skrivs kort

$$\xi \in \mathrm{Bin}(n, p).$$

Kommentarer

- Binomialfördelning föreligger då man gör n obeoende försök med samma sannolikhet p för "lyckat utfall" vid varje försök.
- Binomialfördelning handlar om dragning *med* återläggning och *utan* hänsyn till inbördes ordning.

Frekvensfunktionens graf där $P(\xi = x) = f(x) = \binom{4}{x} p^x (1-p)^{4-x}$
och $\xi \in Bin(4, 0.2)$, *d.v.s.* $p = 0.2$.
Väntevärdet $E(\xi) = \mu = n \cdot p = 4 \cdot 0.2 = 0.8$ *är utsatt.*

1.1.3 Hypergeometrisk fördelning

Ex 3.4 Vid en kvalitetskontroll av $N = 1000$ cykelramar, räknar man att det är fel/defekt på $r := 50$ ramar. Man gör kontrollen genom att välja ut och kontrollera $n = 30$ ramar. Vad är sannolikheten att exakt 2 ramar av dessa utvalda är defekta?

Lösning:

Vi får en relativ felfrekvens på $p := \dfrac{r}{N} = 5\%$. Den tolkar vi som sannolikheten att en slumpvist val ram är defekt. Vi kan inför den stokastisk variabel ξ =antal defekta ramar bland de n valda. Vi söker alltså sannolikheten $P(\xi = 2) =: f(2)$.

$$P(\xi = 2) = \frac{g}{m} = \frac{\binom{50}{2} \cdot \binom{1000-50}{30-2}}{\binom{1000}{30}} \approx 0.26 = 26\%.$$

Kommentarer

- Denna stokastiska variabel ξ är *Hypergeometrisk fördelad* med parametrar N, n och p (eller r).
  Detta skrivs $\xi \in \mathrm{Hyp}(N, n, p)$.
- Hypergeometrisk fördelning handar om dragning *utan* återläggning och *utan* hänsyn till inbördes ordning av de n slumpmässigt valda av N.
- Frekvensfunktionen i exemplet ovan är

$$P(\xi = x) = f(x) := \frac{\binom{50}{x} \cdot \binom{1000-50}{30-x}}{\binom{1000}{30}}.$$

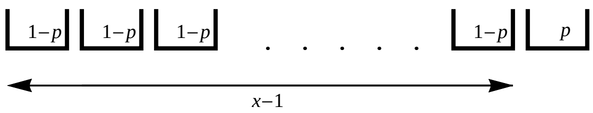
- Förutom dessa tre fördelningar finns ett antal andra, några fler som går att förklara och några som inte kan förklaras lätt.

1.1.4 Geometrisk fördelning

Ex 3.5 För fiskaren kan frågan vara: Vad är sannolikheten att jag får napp f.f.g. vid kast (nummer) x? Vi antar att fiskelyckan är oberoende kast för kast. V låter η_1 =antal kast tills fiskaren får napp f.f.g. Antag att det sker vid kast nr x. Vi får sannolikheten och frekvensfunktionen till

$$P(\eta_1 = x) = f(x) = (1-p)^{x-1} \cdot p.$$

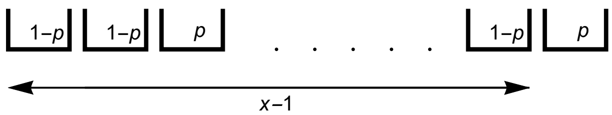
Fördelningen kallas *geometrisk fördelning.* $\eta_1 \in \text{Geo}(p)$.



Ex 3.6 För fiskaren kan frågan vara: Vad är sannolikheten att jag får napp f.a.g. (för andra gången) vid kast (nummer) x? Vi antar att fiskelyckan, som ovan, är oberoende kast för kast. V låter η_2 =antal kast tills fiskaren får napp f.f.g. Antag att det sker vid kast nr x. Det betyder att napp erhåll f.f.g. vid något av de $x-1$ första kasten. Det sker alltså på ett av de $x-1$ första kasten. Vi kan skriva $x-1 = \binom{x-1}{1}$. Vi får sannolikheten och frekvensfunktionen till

$$P(\eta_2 = x) = f(x) = (1-p)^{x-2} \cdot p.$$

Fördelningen kallas *negativ bimomialfördelning med parametrar* 1 *och* p. $\eta_2 \in \text{Neg}(1, p)$.



I figuren inträffar det första lyckade kastet (d.v.s. napp) vid det tredje kastet.

Övning: Hur ser frekvensfunktionen för att få det n :e lyckade kastet vid kast nummer x?

Ex 3.7 För de anförda fördelningarna måste gälla att summan över alla sannolikheter = 1.

$$\sum_{\text{alla utfall } x} P(\xi = x) = 1.$$

- Binomialfördelning: $\xi \in \text{Bin}(n, p)$:

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p + (1-p))^n = 1.$$

- Geometrisk fördelning: $\eta_1 \in \text{Geo}(p)$:

$$\sum_{x=1}^{\infty} (1-p)^{x-1} \cdot p = p \cdot \frac{1}{1-(1-p)} = 1.$$

1.1.5 Indikatorvariabel

Ex 3.8 I ett spel går det ut på att få flest poäng. Det går till så att två eller fler kastar tärning. Man får 1 poäng för tärningspoängen 1 eller 4 och noll poäng för övriga tärningspoäng (2,3,5,6).
Vi inför händelsen/mängden $A = \{1, 4\}$ och således $A^c = \{2, 3, 5, 6\}$. Som stokastisk variabel inför vi en *indikatorvariabel*, som $\mathcal{I} = \mathcal{I}_A$. Den skall vara sådan att den antar värdet 1, om utfallet hamnar i A och 0, om utfallet hamnar i A^c.

$$\begin{cases} P(\mathcal{I} = 1) = p = \dfrac{1}{3} \\ P(\mathcal{I} = 0) = \ (1 - p) = \dfrac{2}{3} \end{cases} \tag{1}$$

Definition 1 En speciell stokastisk variabel är indikatorvariabeln $\mathcal{I} = \mathcal{I}_A$. Denna variabel antar bara två värden: $\mathcal{I} = 1$ eller 0 med sannolikheten p för värdet 1, d.v.s. $P(\mathcal{I} = 1) = p$ och således måste

$$P(\mathcal{I} = 0) = 1 - p.$$

Kommentarer

- Man kan se indikatorvariabeln som en variabel som beskriver lyckat (1) med sannolikheten p och misslyckat (0) med sannolikheten $1 - p$.

1.1.6 ∗ Summor av indikatorvariabler

Låt $\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_n$ vara n oberonde likafördelade indikatorvariabler. Vi skall beskriva fördelningen

$$\xi = \sum_{k=1}^{n} \mathcal{I}_k.$$

Den kan anta alla heltalsvärden mellan 0 och n. Vad är sannolikheten att $\xi = k$ för ett $k = 0, 1, 2, 3, ..., n$?

Lösning:

Det betyder att k stycken antar värdet 1 och övriga $n - k$ antar värdet 0. *En* sådan sekvens har sannolikheten

$$P(\mathcal{I}_1 = 1 \cap \mathcal{I}_2 = 0 \cap .. \cap \mathcal{I}_n = 1)$$

Med precis k ettor och $n - k$ nollor. P.g.a. oberoende och att vi har snitt (och) får vi

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

för *en* sekvens med k ettor. Det finns $\binom{n}{k}$ sådana sekvenser (med k ettor).
Sannolikheten $P(\xi = k)$ är alltså

$$P(\xi = k) = \binom{n}{k} p^k (1 - p)^{n-k},\ k = 0, 1, 2, ..., n \tag{2}$$

d.v.s. ξ är binomialfördelad med parametrar (n, p).

Vi verifierar att totala sannolikheten är 1:

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1^n = 1.$$

1.1.7 Poissonfördelning

Typiska händelser som brukar vara *Poissonfördelade* är antal utfall som inträffar slumpmässigt i tiden under ett givet tidsintervall med en viss intensitet, λ =antal/tidsenhet .

Ex 3.9 Låt den stokastisk variabeln ξ var antal inkommande telefonsamtal som inkommer till en växel under ett 2 minuter långt tidsintervall. Det är en fördelning som oftast kan betraktas som en Poissonfördelning.

Definition 2 Om $\xi \in \text{Po}(\lambda)$ betyder, per definition, att

$$P(\xi = x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{3}$$

Ex 3.10

Låt den stokastisk variabeln ξ vara antal inkommande telefonsamtal till Chalmers växel (tel.nr 772 1000) med $\lambda = 3$ (=antal inkommande samtal under ett tvåminuters intervall)och att ξ är Poissonfördelad.

(a) Beräkna sannolikheten att det kommer in exakt ett samtal under ett tvåminutersintervall.

(b) Beräkna sannolikheten att det kommer in minst två samtal under ett tvåminutersintervall.

Lösning:

(a) Sannolikheten att det kommer in exakt ett samtal på en tvåminutersperiod är

$$P(\xi = 1) = e^{-3} \cdot \frac{3^1}{1!} = 0.149.$$

(b) Sannolikheten att det kommer in minst två samtal på en tvåminutersperiod är

$$P(\xi \geq 2) = e^{-3} \sum_{x=2}^{\infty} \frac{3^x}{x!}.$$

För att undvika beräkna summan, beräknar vi sannolikheten genom att betrakta komplementhändelsen.

$$P(\xi \geq 2) = 1 - P(\xi \leq 1) = 1 - e^{-3} \left(\frac{3^0}{0!} + \frac{3^1}{1!} \right) = 0.80$$

1.1.8 Väntevärde

Väntevärdet för en fördelning är ett *lägesmått*, dess tyngdpunkt på $x-$axeln.

Ex 3.11 Vilket medelvärde/genomsnitt får man när man kastar en tärning n gånger?

Lösning:

Vi får antingen utfallet $1, 2, 3, 4, 5$ eller 6 med sannolikheten för varje utfall $f(x) = \frac{1}{6}$. Väntevärdet av 1 kast med en tärning är medelvärdet

$$\frac{1+2+3+4+5+6}{6} = 3.5$$

Vi kan se $\frac{1}{6} = P(\xi = x)$ för $x = 1, 2, 3, 4, 5, 6$ och skriva medelvärdet som

$$\sum_{x=1}^{6} x \cdot P(\xi = x).$$

Medelvärdet, kallas i sannolikhetseorin *väntevärde* eller *eller förväntat värde* och definieras som nedan för en stokastisk variabel, ξ. Detta motsvarar medelvärde för ett *observerat stickprov.*

$$E(\xi) := \sum_{x \in \Omega} x \cdot P(\xi = x) = \mu. \tag{4}$$

Med beteckningen $P(\xi = x) = f(x)$, en *frekvensfunktion* kan väntevärdet skrivas

$$E(\xi) = \sum_{x \in \Omega} x \cdot f(x) = \mu. \tag{5}$$

Ex 3.12 För de presenterade fördelningarna ger vi nu deras väntevärden.

Fördelning	Väntevärde
Likf (N)	$\frac{N+1}{2}$
Bin (n, p)	$n \cdot p$
Hyp (N, n, p)	$n \cdot p$
Geo (p)	$\frac{1}{p}$
Po (λ)	λ

Väntevärdet för $\xi \in \text{Bin}\,(n, p)$ är alltså $E(\xi) = n\,p(= \mu)$

Ex 3.13 För 25 kast med en tärning fick man följande *observerade värden*:

$$\begin{array}{cccccccccc} 4 & 2 & 6 & 2 & 5 & 6 & 3 & 6 & 3 & 1 \\ 5 & 4 & 1 & 1 & 4 & 3 & 5 & 3 & 3 & 6 \\ 6 & 3 & 6 & 2 & 3 & 3 & 1 & 1 & 1 & 5 \\ 2 & 1 & 2 & 4 & 1 & 4 & 5 & 1 & 6 & 4 \\ 5 & 4 & 6 & 2 & 4 & 1 & 3 & 5 & 5 & 6 \end{array} \tag{6}$$

med frekvenser, respektive relativa frekvenser (undre raderna)

1	2	3	4	5	6
10	6	9	8	8	9
0.2	0.12	0.18	0.16	0.16	0.18

Vi ser att de relativa frekvenserna ligger nära värdet $\dfrac{1}{6} = 0.1666\ldots$, sannolikheten för att få utfallen $x = 1, 2, 3, 4, 5, 6$.
För medelvärdet av alla 50 utfallen behövs först hela poängsumman, som är

$$1 \cdot 10 + 6 \cdot 2 + 9 \cdot 3 + 8 \cdot 4 + 8 \cdot 5 + 9 \cdot 6 = 175$$

med medelvärde

$$\overline{x} = \frac{175}{50} = 3.5\,.$$

Det givetvis en tillfällighet att medelvärdet är lika med väntevärdet men medelvärdet skiljer sig i allmänhet mycket från väntevärdet.

Kommentarer

- I figuren på sidan 3 är väntevärdet $\mu = E(\xi) = 4 \cdot 0.2 = 0.8$ utsatt. $x-$värdet$= 0.8$ är alltså tyngdpunktens $x-$koordinat för grafen.

1.1.9 Varians och standardavvikelse

Dessa mått är de två *spridningsmått* som används inom matematisk statistik. Begreppen används i två betydelser. Dels har vi en "observerad" variant, där vi använder de observerade värdena. Den observerade variansen är summan av de kvadratiska avstånden från medelvärdet.
Motsvarende standardavvikelse s är roten ur variansen, d.v.s. varianen är s^2.

Definition 3 Varians $V(\xi)$ och standardavvikelse $\sigma > 0$ för en stokastisk variabel diskret fördelning med sannolikhetsfunktion $f(x) = P(\xi = x)$, och väntevärde $E(\xi) = \mu$ definieras som

$$V(\xi) = \sigma^2 = \sum_x (x - \mu)^2 f(x). \tag{7}$$

Kommentarer

- Dessa två spridningsmått är mast intressanta då man jämför två, säg standardavvikelser från två liknande fördelnngar, se nästa exempel.
- Standardavvikelse ser lite annorlunda ut när dens skall beräkna ur observerade värden

Ex 3.14 Det finns en form för variansen, som är lite enklare. Vi härleder den. Vi kommer ihåg att

$$\sum_x f(x) = 1, \quad \sum_x x \cdot f(x) = \mu \text{ samt } (x-\mu)^2 = x^2 - 2\mu x + \mu^2.$$

$$\begin{aligned} V(\xi) &= \sum_x x^2 f(x) - 2\mu \sum x\, f(x) + \mu^2 \sum_x f(x) = \\ &= \sum_x x^2 f(x) - 2\mu^2 + \mu^2 = \sum_x x^2 f(x) - \mu^2. \end{aligned}$$

Ex 3.15

(a) Beräkna väntevärde, varians och standardavvikelse för $\xi \in$ Likf(6).

(b) En tärning med lite ovanlig form beskrivs med den stok. var. ξ_1 och tabellen

x	1	2	3	4	5	6
$P(\xi_1 = x)$	1/8	1/8	1/4	1/4	1/8	1/8

Beräkna väntevärde, varians och standardavvikelse för ξ_1 och jämför med (a). Försök ge en förklaring! Undersök om det verkligen är en sannolikhetsfördelning!

Lösning:

(a) Väntevärde

$$\mu = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} = 3.5\,.$$

Varians

$$V(\xi) = \sigma^2 = \frac{1}{6}(1^2+2^2+\ldots 5^2+6^2) - \left(\frac{7}{2}\right)^2 = \ldots = \frac{35}{12} \text{ och } \sigma = \sqrt{\frac{35}{12}} \approx 1.7.$$

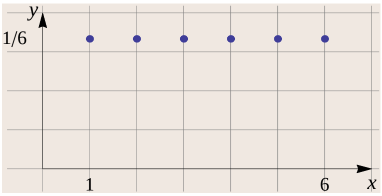
(b) För denna tärning är väntevärdet

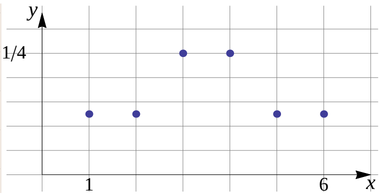
$$E(\xi_1) = \mu_1 = \frac{1}{8}(1+2+5+6) + \frac{1}{4}(3+4) = \frac{7}{2}$$

D.v.s. samma väntvärde som för en vanlig tärning.
Variansen är

$$V(\xi_1) = \sigma_1^2 = \frac{1}{8}(1^2+2^2+5^2+6^2) + \frac{1}{4}(3^2+4^2) - = \frac{29}{2} - \left(\frac{7}{2}\right)^2 = \frac{9}{4} \text{ och } \sigma_1 = 1.5\,.$$

Vi ser att båda fördelningarna har samma väntevärde men att ξ_1 har mindre standardavvikelse än ξ (vanlig tärning). Vad beror det på? För ξ_1 är sannolikheterna mer koncentrerade till väntevärdet, alltså en mindre spridning.





Slutligen adderar vi värdena av frekvensfunktionen:

$$\frac{1}{8} + \frac{1}{8} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} = \ldots = 1$$

d.v.s. det är en sannolikhetsfördelning.