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.

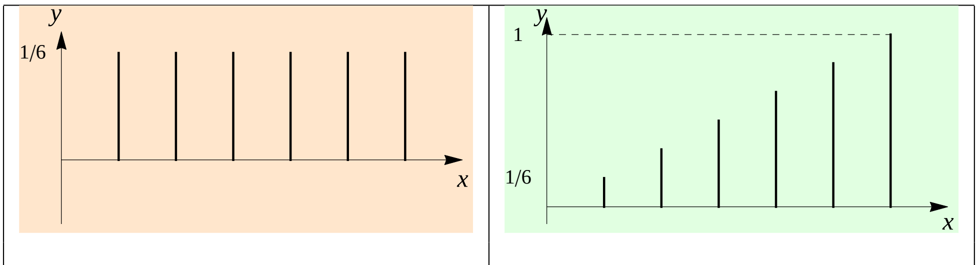
Ex 3.1 I exempel 1.1 har vi sannolikhterna $\frac{1}{6}$ för alla utfalll $1, 2, 3, 4, 5, 6$ (poäng vid ett tärningskast). Man talar då om en *likformig* fördelning, alla utfall har samma sannolikhet. Vi inför nu en *stokastisk variabel*, ξ, (läses "xi"). För att förskorta ned texten/förklaringen av 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).

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.

Ex 3.2 (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 varje kasts utfall 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

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

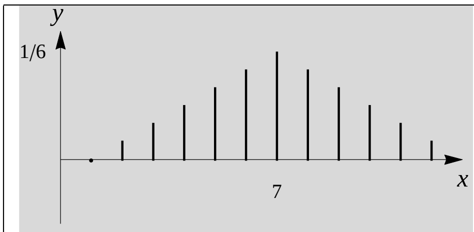
Kommentarer

- 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.

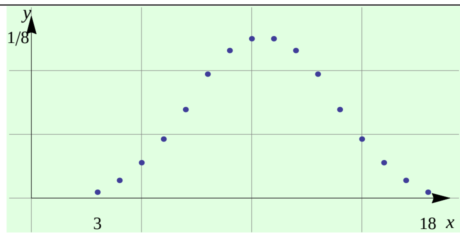
Ex 3.3 I exempel 1.2 har vi summan av två tärningskast poäng. Låt ξ vara summans poäng. Det är klart att ξ antar värdena $2, 3, \ldots, 12$ med olika sannolikheter. Vi har att, med $P(\xi = x) =: f(x)$

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

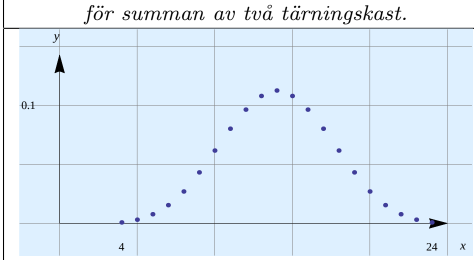
P.s.s. kan vi ta ξ, som summan av tre tärsningskasts poäng och rita den frekvensfunktionen.



Frekvensfunktion
för summan av två tärningskast.



Frekvensfunktion
för summan av tre tärningskast.



Frekvensfunktion
för summan av fyra tärningskast.

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

Lösning:

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

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

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 dragnng *utan* återläggning och *utan* hänsyn till inbördes ordning av de n valda av N.
- Förutom dessa tre fördelningar finns ett antal andra, några som går att förklara och några som inte kan förklaras lätt.

Ex 3.5 För fiskaren kan frågan vara: Vad är sannolikheten att jag får napp f.f.g. vid kast (nummer) x? Vi ana att fickelykan är oberoende kast för kast. V låter η =antal kast tills fiskaren får napp f.f.g. Antag att det sker vid kast nr x. Vi söker sannolikheten

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

Fördelningen kallas *geometrisk fördelning.* $\eta \in \mathrm{Geo}(p)$.

Ex 3.6 För de anförda fördelningarna måste gälla att summan över alla sanolikheter = 1.

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

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

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

- Geometrisk fördelning: $\eta \in \mathrm{Geo}(p)$:

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

Ex 3.7 Fiskaren bestämmer sig att fiska till dess att två fiskar fångas. Det betyder alltså napp f.a.g. (med sannolikheten för napp $= p = 0.2$, vid ett givet kast.) Antag att antal kast är $x = 4$, tills dess fångst fås för andra gången. Vad är sannolikheten $P(\eta = 4)$?

Lösning:

En sekvens kan då vara

$$\underbrace{(1-p)\cdot(1-p)\cdot p\cdot\ldots\cdot(1-p)\cdot(1-p)\cdot p}_{x \text{ kast}} = p^2(1-p)^{x-2}.$$

Det första nappet sker en gång av $x-1$ kast och alltså på $x - 1 = \binom{x-1}{1}$ sätt med $x = 4$. Alltså är sannolikheten

$$3(1-0.2)^2 \cdot 0.2^2 = 0.0768 \approx 8\%.$$

Vad är sannolikheten att få napp för n:e gången vid kast nummer x? Det är med ett utvidgat resonemang av exemplet ovan,

$$P(\eta = n) = \binom{x-1}{x-n}(1-p)^{x-n}p^n \tag{1}$$

Man säger att η är *negativt binomialfördelad* med parametrar n och p.

Kommentarer

- Vi kan skriva om binomialkoefficienten som

$$\binom{x-1}{x-n} = \{(x-1)-(x-n) = n-1\} = \binom{x-1}{n-1}.$$

Ex 3.8 Antal långtradare som per halvtimme som passerar Tingstadstunneln är i snitt $8.5 =: \lambda$. Man vet att detta antal, ζ, är poissonfördelat. Det betyder att $\zeta \in \text{Po}(\lambda)$ och sannolikhet- eller frekvensdfunktionen är

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

(a) Vad är sannolikheten att det kommer exakt 8 långtradare under en halvtimme?

(b) P.s.s. men minst 5.

(c) Visa att $\sum_{x=0}^{\infty} f(x) = 1$.

Lösning:

(a) Sannolikheten att det kommer exakt 8 långtradare under en halvtimme är

$$f(8) = \frac{e^{-8.5} \cdot 8.5^8}{8!} = 0.137508\ldots = 0.14\,.$$

(b) P.s.s. men minst 5:

$$P(\zeta \geq 5) = 1 - P(\zeta \leq 4) = 1 - F(4) = 1 - e^{-8.5}\left(1 + 8.5 + \frac{8.5^2}{2} + \frac{8.5^3}{6} + \frac{8.5^4}{24}\right) = 0.925636 \approx 0.93.$$

(c) Visa att...

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Ex 3.9 Bland ett parti brädor om 1000 stycken, väljs 20 ut för kontroll. Om dessa är korrekta, godkänns hela partiet. Om fler än en är defekt kasseras partiet. Om exat en av de utvalda är defekt går kontrollen över i en *andra* fas. Då väljs ytterligare 10 ut och partiet godkänns bara om alla dessa är korrekta. Vad är sannolikheten att partet godkänns, om det innehåller 30 defekta?

Lösning:

En sannolikhet som erhålls är $p = \dfrac{30}{1000} = 0.03$. Låt ξ =antal defekta bland de 20 först valda och η =antal defekta av de senare 10 valda. Sökt sannolikhet kan då skrivas

$$P(\xi = 0 \cup (\xi = 1 \cap \eta = 0))$$

som kan skrivas

$$P(\xi = 0 \cup (\xi = 1 \cap \eta = 0)) = \{\text{disjunkta}\} = P(\xi = 0) + P(\xi = 1 \cap \eta = 0) =$$

$$P(\xi = 0) + P(\eta = 0|\xi = 1) \cdot P(\xi = 1).$$

De stokastiska variablerna är hypergeometriskt fördelade och vi observerar att för η är $N_1 = 980$, antal defekta 29 och att vi drar 10 korrekta av $980 - 29 = 951$ brädor.

$$\begin{array}{rcl} \xi & \in & \text{Hyp}(1000, 20, 0.03) \\ \text{och} & & \\ \eta & \in & \text{Hyp}\left(980, 19, \frac{29}{980}\right). \end{array}$$

De inblandade sannolikheterna är

$$\begin{aligned} P(\xi = 0) &= \frac{\binom{970}{20} \cdot \binom{30}{0}}{\binom{1000}{20}}, \\ P(\xi = 1) &= \frac{\binom{970}{19} \cdot \binom{30}{1}}{\binom{1000}{20}} \text{ och} \\ P(\eta = 0|\xi = 1) &= \frac{\binom{951}{10} \cdot \binom{29}{0}}{\binom{980}{10}}. \end{aligned}$$

Sökt sannolikhet, d.v.s. sannolikhteten att partiet godkänns är alltså

$$\frac{\binom{970}{20}\cdot\binom{30}{0}}{\binom{1000}{20}} + \frac{\binom{951}{10}\cdot\binom{29}{0}}{\binom{980}{10}} \cdot \frac{\binom{970}{19}\cdot\binom{30}{1}}{\binom{1000}{20}} \approx 0.793 \approx 0.8.$$

Kommentarer

- På miniräknare brukar det finnas knappar för binomialkoefficient $\binom{n}{x}$ och för permutation $P(n,x)$.
- Den största binomialkoefficienten ligger innanför miniräknarens talområde och är
$$\binom{1000}{20} = 3.4 \cdot 10^{41}.$$

1.2 Varians och standardavvikelse

Emedan väntevärdet $\mu =: E(\xi) = \sum_x P(\xi = x)$ är ett *lägesmått*, är variansen $V(\xi) = \sigma^2$ (och standardavvikelsen σ) ett *spridningsmått.* Det definieras om

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

Ex 3.10 Beräkna variansen och standardavvikelsen för $\xi \in \text{Likf}(N)$ med $N = 6$.

Lösning:

Det är samma fördelning (tärningskast), som i ex. 1.1 och ex 3.1. Vi fick

$$E(\xi) = \mu = \frac{7}{2}(= 3.5).$$

Vi förenklar först (2).

$$V(\xi) = \sum_x (x^2 - 2x\mu + \mu^2) f(x) = \sum_x x^2 f(x) - 2\mu \sum_x x f(x) + \mu^2 \cdot 1 = \sum_x x^2 f(x) - \mu^2.$$

I fallet med ex 1.1 (och ex 3.1) får vi

$$V(\xi) = \sum_{x=1}^{6} x^2 \frac{1}{6} - \frac{49}{4} = \ldots = \frac{35}{12} \approx 2.9 \text{ och } \sigma \approx 1.7$$

För varians för de andra fördelningar hänvisas till boken eller formelsamlingen.

1.3 Summor av stok. variabler

Ex 3.11 En diskret fördelning med stokastisk variabel η ges av

$$\begin{cases} P(\xi = 1) &= 0.2 \\ P(\xi = 2) &= 0.5 \\ P(\xi = 3) &= 0.3 \end{cases}$$

(a) Beräkna sannolikheten $P(\xi \leq 2)$.

(b) Beräkna väntevärdet $E(\xi)$ och sätt ut det på en talaxel, tillsammans med utfallen.

(c) Beräkna väntevärdet av $a\xi + b$.

1.3.1 Indikatorvariabel

Ex 3.12 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.

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

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 $1 - p$.

1.3.2 ∗ 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

$$P(\xi = k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, 2, ..., n \tag{4}$$

Vi verifierar att totala sannolikheten är 1:

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

1.4 Approximation mellan fördelningar

Betrakta en hypergeometrisk fördelning Hyp(N, n, p). Om N är mycket större än n, d.v.s. om $n \ll N$ blir det ingen större skillnad mellan dragning *utan* återläggning och dragning *med* återläggning, alltså ingen större skillnad mellan hypergeometrisk och binomialfördelning. Då gäller approximationen (5). Man kan alltså approximera en hypergeometrisk fördelning med en binomialfördelning utan att göra alltför stort fel:

$$\frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x}. \tag{5}$$

Tumregel: $n/N < 0.1$.

Ex 3.13 I exempel 3.9 har vi $n/N = 30/1000 = 0.03 < 0.1$. Vi beräknar $P(\xi = 1)$ med approximationen (5). $p = 0.03$ och $n = 20$, som ger att

$$P(\xi = 1) \approx \binom{20}{1} 0.03^1 \cdot 0.97^{19} \approx 0.336$$

att jämföra med 0.341, som erhålls i exempel 3.9.

Nu kan hypergeometrisk fördelning approximeras med Poissonfördelning, om

$$n > 10, \quad p + n/N < 0.1 \tag{6}$$

Ex 3.14 Beräkna $P(\xi = 1)$ m.h.a. Poissonapproximation.

Lösning:

$$n = 20 > 10, \quad p + n/N = 0.03 + 0.02 < 0.1$$

Alltså går Poissonapproximation bra.

$$E(\xi) = \lambda = n \cdot p = 0.6 \Longrightarrow P(\xi = 1) \approx \frac{e^{-0.6} \cdot 0.6^1}{1!} = 0.329 \ldots \approx 0.33$$