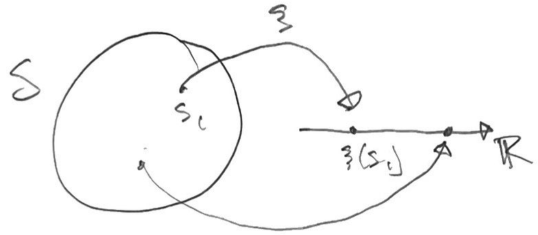
Stokastiska variabler

En stokastisk variabel ξ är en funktion från ett underliggande rum S till de reella talen, $\mathbb{R}$.



$\xi : S \to \mathbb{R}$

"slumpen" drar en punkt $s \in S$ och skickar den till $w = \xi(s) \in \mathbb{R}$, ett utfall.

Om värdemängden, $\Omega \subseteq \mathbb{R}$, dvs mängden av alla värden som ξ kan anta, är diskret, t.ex. $\Omega = \{1, \ldots, 6\}$, varvid vi betraktar ξ som ett försök med diskret utfallsrum, så kallas ξ för en diskret stokastisk variabel.

Om däremot Ω är ett kontinuerligt utfallsrum, t.ex. $\Omega = \mathbb{R}$ eller $\Omega = [0,1]$, så kallas ξ för en kontinuerlig stokastisk variabel.

Stokastiska variabler formaliserar slumpmässiga försök.

För diskreta stokastiska variabler så beskriver sannolikhetsfunktionen (i folkmun diskreta täthetsfunktionen) $f(x) = \mathbb{P}(\xi = x) = \mathbb{P}(\{\xi = x\})$, $x \in \Omega$, sannolikheten för händelsen $A = \{\xi = x\}$.

Vidare definieras <u>fördelnings-funktionen</u> för ξ som $F(x)=\mathbb{P}(\xi\leq x)$, $-\infty<x<\infty$, dvs sannolikheten för händelsen $A=\{\xi\leq x\}$.

Till skillnad från sannolikhetsfunkt. så kan man definiera fördelningsfunktioner för både diskreta och kontinuerliga stok. var.

<u>Def</u> <u>Fördelningen</u> för ξ är funktionen $P(A)=\mathbb{P}(\xi\in A)$, $\underset{\text{händelse}}{A\subseteq\Omega}$

<u>Sats</u> Fördeln. funkt. $F(x)$ är en växande funktion som är högerkontinuerlig för varje x, med $\lim_{x\to-\infty}F(x)=0$ och $\lim_{x\to\infty}F(x)=1$

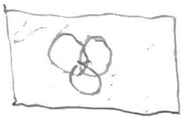
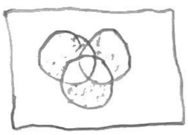
$\mathbb{P}(\xi\in[a,b])=\mathbb{P}(a\leq\xi\leq b)=F(b)-F(a)$ för $a\leq b$, samt $\mathbb{P}(\xi>x)=1-F(x)$

(<u>Växande</u>: Om $x\leq y$ så gäller att $\{\xi\leq x\}\subseteq\{\xi\leq y\}$
$\Rightarrow F(x)=\mathbb{P}(\xi\leq x)\leq\mathbb{P}(\xi\leq y)=F(y)$)

Låt $\Omega=\{x_1,\ldots,x_n\}$ (där $x_1<\ldots<x_n$) för en diskret stokastisk variabel. Då gäller $F(x)=F(x_k)=\sum_{i=1}^{k}f(x_i)=$
$=\sum_{i=1}^{k}\mathbb{P}(\xi=x_i)$, där $x_k=\max\{x_i\in\Omega: x_i\leq x\}$.

<u>Ex</u> Entreprenör lämnar lämnar offert på jobben A, B, C, $\mathbb{P}(\text{få A})=0.7$, $\mathbb{P}(\text{få B})=0.8$, $\mathbb{P}(\text{få C})=0.4$ (oberoende händelser). Bestäm $f(x)$ och $F(x)$ för den stokastiska variabel ξ som beskriver försöket, samt beräkna $\mathbb{P}(\text{få som mest 2,5 jobb})$.

$\{\xi = x_i\}$ $\quad \Omega = \{x_1, x_2, x_3, x_4\} = \{0, 1, 2, 3\}$

x_i	$f(x_i) = \mathbb{P}(\xi = x_i) =$	$F(x_i) = \mathbb{P}(\xi \leq x_i)$
0 	$\mathbb{P}(A^c \cap B^c \cap C^c) \overset{\text{ober.}}{=} \mathbb{P}(A^c)\mathbb{P}(B^c)\mathbb{P}(C^c)$ $= (1-\mathbb{P}(A))(1-\mathbb{P}(B))(1-\mathbb{P}(C)) = 0.3 \cdot 0.2 \cdot 0.6 = 0.036$	0.036
1 	$\mathbb{P}(A \cap (B \cup C)^c) + \mathbb{P}(B \cap (A \cup C)^c) + \mathbb{P}(C \cap (A \cup B)^c)$ $= \mathbb{P}(A)\mathbb{P}(B^c \cap C^c) + \mathbb{P}(B)\mathbb{P}(A^c \cap C^c) + \mathbb{P}(C)\mathbb{P}(A^c \cap B^c)$ $= \mathbb{P}(A)\mathbb{P}(B^c)\mathbb{P}(C^c) + \mathbb{P}(B)\mathbb{P}(A^c)\mathbb{P}(C^c) + \mathbb{P}(C)\mathbb{P}(A^c)\mathbb{P}(B^c)$ $= 0.7 \cdot 0.2 \cdot 0.6 + 0.8 \cdot 0.3 \cdot 0.6 + 0.4 \cdot 0.3 \cdot 0.2 =$ $= 0.084 + 0.144 + 0.024 = 0.252$	$0.036 + 0.252$ $= 0.288$
2 	$\mathbb{P}(A \cap B) + \mathbb{P}(B \cap C) + \mathbb{P}(A \cap C) - 3\mathbb{P}(A \cap B \cap C) =$ $\overset{\text{ob.}}{=} \mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(B)\mathbb{P}(C) + \mathbb{P}(A)\mathbb{P}(C) - 3\mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$ $= 0.7 \cdot 0.8 + 0.8 \cdot 0.4 + 0.4 \cdot 0.7 - 3 \cdot 0.7 \cdot 0.8 \cdot 0.4 = 0.488$	$0.288 + 0.488$ $= 0.776$
3 	$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = 0.7 \cdot 0.8 \cdot 0.4$ $= 0.224$	1

$$\mathbb{P}(\xi \leq 2{,}5) = \mathbb{P}(\xi \in \{0, 1, 2\}) = F(2) = 0.776$$

$\boxtimes$

Def:

För en diskret stokastisk variabel ξ med sannolikhetsfunktion $f(x) = \mathbb{P}(\xi = x)$, $x \in \Omega$, definieras väntevärdet som $\mathbb{E}[\xi] = \sum_{x \in \Omega} x f(x) = \sum_{x \in \Omega} x \mathbb{P}(\xi = x)$ och betecknas μ.

Om $\Omega = \{x_1, \ldots, x_n\}$ är ändligt: $\mathbb{E}[\xi] = \sum_{i=1}^{n} x_i f(x_i)$

Om $\Omega = \{x_1, x_2, \ldots\}$ är oändligt: $\mathbb{E}[\xi] = \sum_{i=1}^{\infty} x_i f(x_i)$.

Beskriver en "centralpunkt" för fördelningen.

Ett slags viktat medelvärde där $f(x)$ styr viktningen.

Ex Spel med två tärningskast, där person 1 vinner om tärningarna får samma poängtal och person 2 vinner om de får olika poängtal. Om pers 1 vinner får denne 14 kr av pers 2 och om pers 2 vinner får denne 4 kr av pers 1.

a) Förväntad vinst för pers 1?

b) Hur mycket skall pers 2 betala till pers 1 när pers 2 förlorar för att den förväntade vinsten för båda ska bli 0 kr?

Lösn: ξ = vinsten för pers 1 efter en omgång

$$\mathbb{P}(\text{pers 1 vinner}) = \frac{\#\text{ varianter med samma poängtal på båda}}{\#\text{ uppsättningar av med två tärn.}} = \frac{6}{6\cdot 6} = \frac{1}{6}$$

$$\Rightarrow \mathbb{P}(\text{pers 2 vinner}) = 1 - \frac{1}{6} = \frac{5}{6}$$

x_i	$f(x_i) = \mathbb{P}(\xi = x_i)$
-4	$5/6$
14	$1/6$

Förväntad vinst:

$$\mathbb{E}[\xi] = \sum_{i=1}^{2} x_i f(x_i) = -4\tfrac{5}{6} + 14\tfrac{1}{6} = -1$$

dvs en förväntad/genomsnittlig förlust.

Pers 2 betalar y kr $\Rightarrow \mathbb{E}[\xi] = -4\frac{5}{6} + y\frac{1}{6} = 0$

$\Rightarrow y = 20$ kr

Utfall/realiseringar av en stokastisk variabel ligger spridda runt $\mathbb{E}[\xi]$. Men hur mäter vi utfallens spridning runt $\mathbb{E}[\xi]$?

Def Variansen för en ~~stokastisk~~ stokastisk variabel ξ ges av $\mathrm{Var}(\xi) = \mathbb{E}[(\xi - \mathbb{E}[\xi])^2] = \mathbb{E}[(\xi - \mu)^2] = \mathbb{E}[\xi^2] - \mu^2$ och betecknas $\sigma^2 = \mathrm{Var}(\xi)$; $\sigma = \sqrt{\mathrm{Var}(\xi)}$ kallas standardavvikelsen.

försök ..., samt beräkna $\mathbb{P}$(få som mest 2,5 jobb).

För en diskret stokastisk variabel ξ gäller $\mathbb{E}[g(\xi)] = \sum_{x \in \Omega} g(x) \mathbb{P}(\xi = x)$ för godtyckliga funktioner $g(x)$.

$\Rightarrow Var(\xi) = \mathbb{E}[\xi^2] - \mathbb{E}[\xi]^2 = \sum_{x \in \Omega} x^2 \mathbb{P}(\xi = x) - \left(\sum_{x \in \Omega} x \mathbb{P}(\xi = x)\right)^2$

Ex Fortsättning tärningskastspelet

$\Omega = \{-4, 14\}$, $\mathbb{P}(\xi = -4) = 5/6$, $\mathbb{P}(\xi = 14) = \frac{1}{6}$,

$Var(\xi) = \sum_{i=1}^{2} x_i^2 \mathbb{P}(\xi = x_i) - \mu^2 = \left((-4)^2 \frac{5}{6} + 14^2 \frac{1}{6}\right) - (-1)^2 = \frac{270}{6}$

$\Rightarrow \sigma = S(\xi) = \sqrt{Var(\xi)} = \sqrt{\frac{270}{6}} \approx 6.708$

Notera att σ^2, till skillnad från σ, uttrycker spridningen på kvadratskala.

Härnäst tittar vi på ett antal diskreta fördelningar som är väldigt vanligt förekommande och som beskriver ett antal vanligt förekommande scenarion, vilket också är anledningen till att de har fått egna namn.

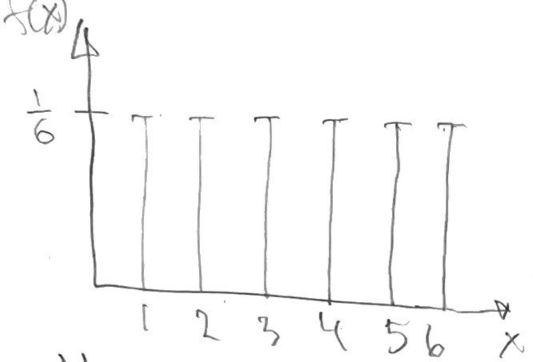
Likformig fördelning

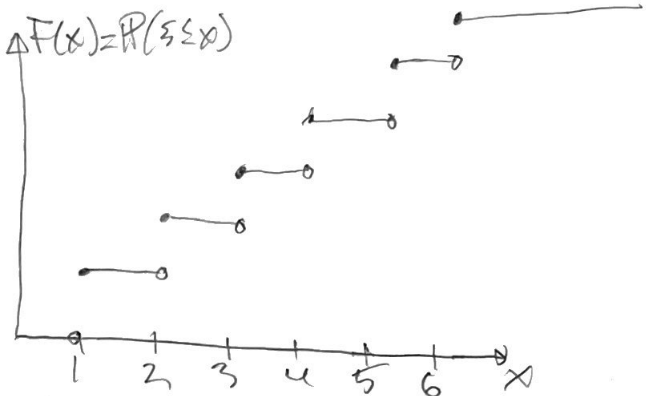
Denna beskriver det upplägg vi hittills har haft, där det är lika sannolikt att få varje utfall i Ω.

Def

Om ξ antar värden i $\Omega = \{1, \ldots, n\}$ med lika stor sannolikhet, dvs $f(x) = \mathbb{P}(\xi = x) = \frac{1}{n}$ för varje $x \in \Omega = \{1, \ldots, n\}$, så sägs ξ vara likformigt fördelad på Ω.

Ex Symmetrisk tärning: $\Omega = \{1, \ldots, 6\}$, $f(x) = \mathbb{P}(\xi = x) = \frac{1}{6}$ för alla $x \in \Omega$





Vi skriver $\xi \sim Uni(n)$ eller $\xi \in Uni(n)$, alternativt $\xi \sim Likf(n)$

Hypergeometrisk fördelning

Om ξ har sannolikhetsfunktionen

$$f(k) = \mathbb{P}(\xi = k) = \frac{G}{M} = \frac{\binom{N_p}{k}\binom{N-N_p}{n-k}}{\binom{N}{n}}, \quad 0 \le k \le N_p, \ 0 \le n \le N.$$

så följer ξ en hypergeometrisk fördelning med parametrarna

N = totalt antal objekt

n = urvalsstorlek

N_p = totalt antal objekt med speciell egenskap.

Vi skriver $\xi \sim Hyp(N, n, N_p)$

Från N objekt, där N_p av dem har en speciell egenskap, drar vi slumpmässigt n st objekt utan återläggning. Här ger $f(k)$ sannolikheten att k av de dragna objekten har egenskapen.

$f(k)$ är positiv om $\max(0, n+N_p-N) \le k \le \min(N_p, n)$.

Ex Urna med 17 kulor, där 5 är vita.
Vi plockar upp 9 kulor utan återläggning,
$\xi = \#$ vita kulor vi får upp
Finn fördelningen för ξ.

Lösn:
$\xi \sim Hyp(N, n, Np)$, $N = 17$, $n = 9$, $Np = 5$

$$\Rightarrow f(k) = \mathbb{P}(\xi = k) = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}} = \frac{\binom{5}{k}\binom{12}{9-k}}{\binom{17}{9}}$$

$$\text{t.ex. } f(2) = \frac{\binom{5}{2}\binom{12}{7}}{\binom{17}{9}} = \frac{10 \cdot 792}{24310} = 0.3258$$

$\boxminus$

Sats
För $\xi \sim Hyp(N, n, Np)$ gäller att
$\mathbb{E}[\xi] = n \cdot p$ och $Var(\xi) = \frac{N-n}{N-1} \cdot n \cdot p \cdot (1-p)$, där
$p = \frac{Np}{N}$.

$\boxminus$

Binomialfördelningen

En stokastisk variabel kallas Bernoulli-fördelad med parameter $p \in [0,1]$ om den beskriver ett binärt experiment, $\Omega = \{A^c, A\}$ (alt. $\Omega = \{0, 1\}$), där $\mathbb{P}(\xi = A) = p$ och $\mathbb{P}(\xi = A^c) = 1-p$ (alt. $\mathbb{P}(\xi = 1) = p$, $\mathbb{P}(\xi = 0) = 1-p$). Tänk slantsingling.

Låt $\xi_1, \dots, \xi_n \in \{0, 1\}$ vara $n \geq 1$ $Ber(p)$-fördelade och oberoende. Då är $\xi = \xi_1 + \dots + \xi_n$, som beskriver antalet "lyckade" binära försök Binomialfördelad.

Med andra ord, ξ är binomial fördelad med parametrarna (n,p) om

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

Vidare gäller att

$$\mathbb{E}[\xi]=np \quad \text{och} \quad \mathrm{Var}(\xi)=np(1-p)$$

Vi skriver $\xi\sim \mathrm{Bin}(n,p)$ eller $\xi\in \mathrm{Bin}(n,p)$.

Notera att detta är samma upplägg som för en hypergeometrisk fördelning. Fast här har vi återläggning.

Ex

Vi köper en fröpåse med 15 frön. På påsen står det att fröna har en 90% groddbarhet. $\mathbb{P}$(åtminstone 14 frön börjar gro).

Lösn:

$p=0.9$, $n=15$, $\xi=\#$ frön som börjar gro $\sim \mathrm{Bin}(n,p)$

$$\mathbb{P}(\xi\geq 14)=\mathbb{P}(\xi=14)+\mathbb{P}(\xi=15)=\binom{15}{14}0.1^{14}(1-0.1)^{15-14}+\binom{15}{15}0.1^{15}(1-0.1)^{15-15}$$

$$\approx 0.3432+0.2059=0.5491$$

Förväntat antal frön som gror:

$$\mathbb{E}[\xi]=np=15\cdot 0.9=13.5$$

Standardavvikelsen:

$$\sigma=\sqrt{\mathrm{Var}(\xi)}=\sqrt{np(1-p)}=\sqrt{1.35}\approx 1.162$$

Poissonfördelning

Denna fördelning beskriver typiskt
ξ = antal gånger som en viss sorts händelse inträffar (oberoende händelser inom ett tidsintervall)
när det inte finns något teoretiskt max-antal (jämför med tidigare fördeln.)

- $\underline{\text{Def}}$ Vi säger att ξ är Poissonfördelad med parameter $\lambda > 0$ om

$$f(k) = \mathbb{P}(\xi = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, \ldots \quad \left(\begin{matrix}\xi \sim Poi(\lambda) \text{ eller} \\ \xi \sim Po(\lambda)\end{matrix}\right)$$

Det gäller ~~att~~ $\mathbb{E}[\xi] = \lambda = Var(\xi)$, så λ kallas ofta för "intensiteten" eller "det genomsnittliga antalet händelser". ⊠

- Boken tar upp Poisson-processer i detta skede men eftersom de är kopplade till kontinuerliga fördelningar så tar vi upp dem senare.

$\underline{\text{Ex}}$ Antalet fel på en viss typ av lakansväv kan antas vara Poisson-fördelat med i genomsnitt 3 fel per 10 meter väv.

a) En kund köper fem vävbitar om vardera 2 meter. Finn $\mathbb{P}$(finns fler än två fel totalt på dessa).

b) Beräkna $\mathbb{P}$(slumpmässigt vald väv om 12 meter är felfri).

Lösn

Vi kan betrakta de 5 bitarna som en 10-meters väv pga att felen antas inträffa oberoende av varandra. Genomsn. antal fel per 10 m är $\lambda = 3$. ξ = antal fel på rullarna $\sim Poi(\lambda) = Poi(3)$

$$\mathbb{P}(\xi = k) = \frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-3}3^k}{k!} \Rightarrow \mathbb{P}(\xi = 0) = \frac{e^{-3}3^0}{0!} \approx 0.05, \mathbb{P}(\xi=1) \approx 0.149$$

$\mathbb{P}(\xi = 2) \approx 0.224$, $\mathbb{P}(\xi \geq 3) = 1 - \mathbb{P}(\xi < 3) = 1 - \mathbb{P}(\xi \leq 2) \approx 1 - (0.05 + 0.149 + 0.224) = 0.577$

b) Genomsnittl. antal fel per meter är $3/10 = 0.3 \Rightarrow$ Genomsnittl. # fel per 12 m är $\lambda = 12 \cdot 0.3 = 3.6$

Vi söker $\mathbb{P}(\xi = 0)$ för $\xi \sim Poi(3.6)$:

$$\mathbb{P}(\xi = 0) = \frac{e^{-3.6}\,3.6^0}{0!} = e^{-3.6} \approx 0.03$$

Här får vi även $\mathbb{E}[\xi] = \lambda = 3.6$ så $Var(\xi) = 3.6$

Några approximationer

Ibland är uträkningar för en viss fördelning jobbiga.

- Om $n > 10$ och $p < 0.1$ så gäller $\mathbb{P}(\xi = k) = \binom{n}{k}p^k(1-p)^{n-k} \approx \frac{e^{-np}(np)^k}{k!}$, dvs $Bin(n,p) \approx Poi(np)$
- Om $\frac{n}{N} < 0.1$ så gäller $\mathbb{P}(\xi = k) = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}} \approx \binom{n}{k}p^k(1-p)^{n-k}$, där $p = \frac{Np}{N}$, dvs $Hyp(N, n, Np) \approx Bin(n,p)$ ("återläggning eller ej spelar liten roll")
- Om $n > 10$ och $p + \frac{n}{N} < 0.1$, där $p = \frac{Np}{N}$, gäller $Hyp(N, n, Np) \approx Poi(np)$

Ex

$X \sim Bin(100, 0.02)$, dvs $n > 10$ och $p < 0.1 \Rightarrow X$ har ungefär samma fördeln. som $Y \sim Poi(2)$

$$\mathbb{P}(Y \leq 3) = \sum_{k=0}^{3} \frac{e^{-2}\lambda^k}{k!} \approx 0.857$$

$$\mathbb{P}(X \leq 3) = \sum_{k=0}^{3} \binom{100}{k} 0.02^k (1-0.02)^{100-k} \approx 0.860 \quad (\text{jobbigt med } \tbinom{100}{k})$$

Ex $X \sim Hyp(N, n, Np)$, $N = 200$, $n = 10$, $Np = 54$, så $\frac{n}{N} = \frac{10}{200} = 0.05 < 0.1$ varvid X är ungefär lika-fördelad med $Y \sim Bin(n, p)$, $n = 10$, $p = \frac{54}{200} = 0.27$

$$\mathbb{P}(Y > 5) = 1 - \sum_{k=0}^{5} \binom{10}{k} 0.27^k 0.73^{10-k} \approx 1 - 0.971278 = 0.0287$$

Dator ger $\mathbb{P}(X > 5) = 1 - \sum_{k=0}^{5} \frac{\binom{54}{k}\binom{200-54}{10-k}}{\binom{200}{10}} \approx 0.0225$

Ex

$X \sim Hyp(5000, 100, 150)$, så $p = \frac{150}{5000} = 0.03$ och $p + \frac{n}{N} = 0.05 < 0.1$ samt $n = 100 > 10 \Rightarrow X \approx Y$ i fördeln., där $Y \sim Poi(\lambda)$, $\lambda = np = 3$.

$$\mathbb{P}(Y > 3) = 1 - \sum_{k=0}^{3} \frac{e^{-3}3^k}{k!} \approx 0.3528$$

$$\mathbb{P}(X > 3) = 1 - \mathbb{P}(X \leq 3) = 1 - \sum_{k=0}^{\infty} \frac{\binom{150}{k}\binom{5000-150}{100-k}}{\binom{5000}{100}} \approx 0.3527$$

enl. dator.