Kapittel 5: Spesielle diskrete sannsynlighetsfordelinger

Hittil: Sett på diskrete fordelinger definert direkte som:

y	0	1	2
$f(y)$	0.6	0.3	0.1

Eventuelt med et matematisk uttrykk for sannsynlighetsfordelinga:

$$f(x) = \frac{1}{30}(x^2 + 4) \quad x = 0, 1, 2, 3.$$

I praksis: Noen spesielle fordelingstyper opptrer oftere enn andre.

Dette kapitlet: Viktige **diskrete** fordelinger.

Fordelingsklasser/familier

Vi skal se på disse kjente klassene av **diskrete** sannsynlighetsfordelinger

- Binomisk fordeling.
- Hypergeometrisk fordeling.
- Geometrisk fordeling.
- Negativ-binomisk fordeling.
- Poisson-fordeling.

Egentlig klasser / familier av fordelingen.

Disse er så vanlige at vi vil gjøre det enklere å bruke dem.

Hovedmål:

Gi generelle formler for punktsannsynlighet.

Bruke tabeller for kumulativ fordelingsfunksjon.

Gi generelle formler for forventning og varians.

5.2 Binomisk fordeling

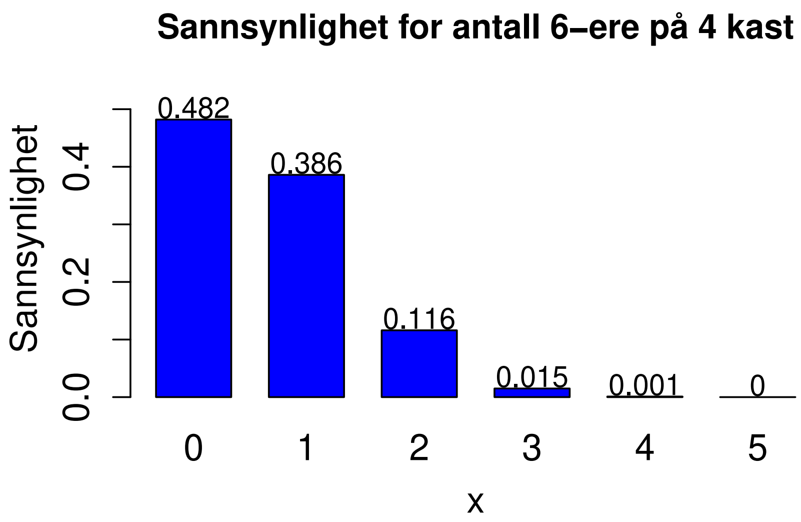
Vi har allerede sett eksempler der vi faktisk har en (underliggende) binomisk fordeling:

Eksempel: Terningspill.

- Ville finne sannsynligheten for minst én sekser i fire kast med en terning.
- Vi kan i staden innføre en stokastisk variabel:

$$X = \text{antall 6-ere på 4 kast}$$

Punktsannsynligheten for X:



(Ser òg fra fordelinga at

$$P(X \geq 1) = 1 - P(X=0) = 1 - 0{,}482 = 0{,}518)$$

Binomisk forsøk - binomisk fordeling

La X være hvor mange ganger en hendelse A inntreffer på n delforsøk.
X er **binomisk fordelt** dersom.

Terning - eksemplet:

$n = 4, \quad p = 1/6$

- Hvert delforsøk har bare to mulige utfall:
  - Suksess (hendelsen A **inntreffer**).
  - Fiasko (hendelsen A **inntreffer ikke**).
- Resultata i de n delforsøka er **uavhengige**.
- **Samme sannsynlighet** for suksses/fiasko i alle n delforsøk

$$P(A) = p \qquad P(A') = 1 - p$$

suksess

Merk: Slike delforsøk kalles **Bernoulli**-delforsøk.

Regne sannsynlighet i binomisk fordeling?

Eksempel: Terningspill.

X = antall 6-ere på $n = 4$ kast.
Definerer A_i = 6-er i kast nr i. Har da at $P(A_i) = p = 1/6$. Antar uavhengighet.

$$
\begin{aligned}
P(X=0) &= P(A_1' \cap A_2' \cap A_3' \cap A_4') = (1-p)^4 = \binom{4}{0} p^0 (1-p)^{4-0} = \left(\frac{5}{6}\right)^4 = 0.482 \\
P(X=1) &= P(A_1 \cap A_2' \cap A_3' \cap A_4') + \cdots + P(A_1' \cap A_2' \cap A_3' \cap A_4) \\
&= 4 \cdot p^1 (1-p)^3 = \binom{4}{1} p^1 (1-p)^{4-1} = 4 \cdot \frac{1}{6} \cdot \left(\frac{5}{6}\right)^3 = 0,386 \\
P(X=2) &= P(A_1 \cap A_2 \cap A_3' \cap A_4') + \cdots + P(A_1' \cap A_2' \cap A_3 \cap A_4) \\
&= 6 \cdot p^2 (1-p)^2 = \binom{4}{2} p^2 (1-p)^{4-2} = 6 \cdot \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = 0,116 \\
P(X=3) &= P(A_1 \cap A_2 \cap A_3 \cap A_4') + \cdots + P(A_1' \cap A_2 \cap A_3 \cap A_4) \\
&= 4 \cdot p^3 (1-p)^1 = \binom{4}{3} p^3 (1-p)^{4-3} = 4 \cdot \left(\frac{1}{6}\right)^3 \cdot \frac{5}{6} = 0,015 \\
P(X=4) &= P(A_1 \cap A_2 \cap A_3 \cap A_4) = p^4 = \binom{4}{4} p^4 (1-p)^{4-4} = 1 \cdot \left(\frac{1}{6}\right)^4 \cdot \left(\frac{5}{6}\right)^0 = 0,001
\end{aligned}
$$

$$P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) = 1$$

Punktsannsynlighet, binomisk fordeling:

Sannsynlighetsfordelinga (punktsannsynligheten) til en **binomisk fordelt** stokastisk variabel, X, med parametrer $\boldsymbol{n}$ og $\boldsymbol{p}$ er

skrivemåte i boka

$$f(x) = b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n$$

Mye brukt (kortform-)notasjon

$$X \sim \text{bin}(n, p)$$

- Merk at n og p **kan varieres**, slik at uttrykket for punktsannsynligheten gjelder **uansett** antall delforsøk $\boldsymbol{n}$ og sannsynlighet $\boldsymbol{p}$. Gir opphav til ulike binomiske fordelinger.
- n og p kalles da **parametrene** i fordelinga.
- **Kumulativ** fordelingsfunksjon finnes ved summering. Er (heldigvis) tabulert i tabell- og formelsamlinga for noen vanlige valg av n og p.
- Både punkts.s. og kumulativ fordelingfunksjon finnes i mange kalkulatorer, finnes i *R*.

Regning av binomiske sannsynligheter, praksis

- Punktsannsynlighet kan regnes fra formel:

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

- Kumulativ fordelingsfunksjon

$$F(x) = P(X \le x) = \sum_{k=0}^{x} P(X = k) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}, \quad X = 0, 1, \ldots, n$$

Finnes i formelsamling for $n = 2, \ldots, 20$, og noen valg av p.

- Punksannsynlighet og kumulativ fordelingsfunksjon finnes på noen kalkulatorer.
- Punktsannsynlighet og kumulativ fordelingsfunksjon i *R*: *dbinom(x,n,p)*, *pbinom(x,n,p)*.

```
> dbinom(2,size=4,prob=1/6)
[1] 0.1157407
> pbinom(2,size=4,prob=1/6)
[1] 0.9837963
> 1 - dbinom(0,size=4,prob=1/6)
[1] 0.5177469
```

$X \sim bin(n = 4, p = 1/6)$

$P(X = 2) = 0{,}115$

$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$= 0{,}984$

$P(X \ge 1) = 1 - P(X=0) = 0{,}518$

urealistisk?

Eksempel: Skiskyting.

La X være antall treff på 5 stående skudd. Trur du X er binomisk? Hva er (omtrent) p?

Bruk direkte regning/formelsamling:

Antar resultata av skudda er uavhengige.

Antar $p = 0{,}8$. Da er $X \sim bin(n=5, p=0{,}8)$

$$P(\text{4 treff}) = P(X=4) = \binom{5}{4} \cdot 0{,}8^4 \cdot (1-0{,}8)^1 = \underline{\underline{0{,}410}}$$

$$P(\text{minst 4 treff}) = P(X \geq 4)$$

$$= \begin{cases} P(X=4) + P(X=5) = 0{,}410 + \binom{5}{5} \cdot 0{,}8^5 \cdot 0{,}2^0 = \underline{\underline{0{,}737}} \\ 1 - P(X \leq 3) \underset{\text{tabell}}{=} 1 - 0{,}263 = \underline{\underline{0{,}737}} \end{cases}$$

$$P(\text{minst 2, færre enn 5}) = P(2 \leq X < 5)$$

$$= \begin{cases} P(X=2) + P(X=3) + P(X=4) = \underline{\underline{0{,}665}} \\ P(X \leq 4) - P(X \leq 1) \underset{\text{tab}}{=} 0{,}672 - 0{,}007 = \underline{\underline{0{,}665}} \end{cases}$$

Forventning og varians, binomisk variabel

Det finnes en generell formel for forventninga og variansen till en binomisk fordelt stokastisk variabel, bare basert på verdiene av parametrene n og p:

Formler for forventning og varians (T 5.1):

For en **binomisk** fordelt stokastisk variabel med n delforsøk og sannsynlighet p har vi at

$$\mu = E(X) = np \qquad \text{og} \qquad \sigma^2 = Var(X) = np(1-p)$$

<u>Bevis:</u> Innfør n uavhengige (indikator) Bernoulli-variabler $I_1, \ldots, I_n$ (notat kapittel 4) slik at

$$I_k = \begin{cases} 1, \text{suksess i delforsøk k} \\ 0, \text{fiasko i delforsøk k} \end{cases}$$

i	0	1
$P(I_k = i)$	$1-p$	p

$$X = \sum_{k=1}^{n} I_k$$

$$E(I_k) = 0 \cdot (1-p) + 1 \cdot p = p \qquad Var(I_k) = E(I_k^2) - [E(I_k)]^2 = 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p(1-p)$$

$$E(X) = E\left(\sum_{k=1}^{n} I_k\right) = \sum_{k=1}^{n} E(I_k) = \sum_{k=1}^{n} p = \underline{\underline{np}}$$

$$Var(X) = Var\left(\sum_{k=1}^{n} I_k\right) \underset{\text{uavh.}}{=} \sum_{k=1}^{n} Var(I_k) = \sum_{k=1}^{n} p(1-p) = \underline{\underline{np(1-p)}}$$

Eksempel: Skiskyting.

Finn forventning og varians til antall treff på en stående-serie? Anta $p = 0.8$ og uavhengighet mellom skudd.

$$X \sim bin(n = 5, p = 0{,}8)$$

$$E(X) = np = 5 \cdot 0{,}8 = \underline{\underline{4}}$$

$$Var(X) = np(1-p) = 5 \cdot 0{,}8 \cdot 0{,}2 = \underline{\underline{0{,}8}}$$

$$SD(X) = \sqrt{0{,}8} = \underline{0{,}894}$$

Eksempel: Skiskyting.

På "sprint" skytes én liggende og én stående 5-skuddserie. Hva blir forventning og varians til antall treff i et løp om $p_1 = 0.8$ for stående, og $p_2 = 0.9$ for liggende.

$Y \sim bin(n=5, p=0{,}9)$

$E(Y) = 5 \cdot 0{,}9 = \underline{4{,}5}$ $\quad Var(Y) = 5 \cdot 0{,}9 \cdot 0{,}1 = \underline{0{,}45}$

Treff totalt: $T = X + Y$

$E(T) = E(X+Y) = E(X) + E(Y) = \underline{\underline{8{,}5}}$

$Var(T) = Var(X+Y) = Var(X) + Var(Y)$

antar uavh.

$= 0{,}8 + 0{,}45 = \underline{\underline{1{,}25}}$ $\quad SD(T) = \underline{1{,}12}$

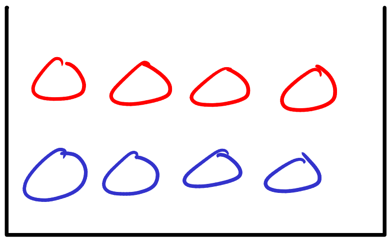
Spm: Er T binomisk fordelt? Nei, p ikke konstant.

Kommentarer

- Det blir binomisk fordeling når du trekker fra ei urne med N enheter der $100 \cdot p\%$ er av type A, og trekkinga skjer **med** tibakelegging.
  Eks. 8 kuler, 4 røde. Trekker 3.
  $X =$ antall røde som trekkes. $X \sim \text{bin}(n = 3, p = 4/8)$.

- Om N er stor (og mye større enn n) har vi tilnærma binomisk fordeling sjøl **uten** tilbakelegging. (Eks. meiningsmåling, $N = 4 \cdot 10^6$, $n = 1000$)

- Om $N >> n$ (og n stor nok) kan vi **tilnærme** binomisk fordeling med normalfordeling (meir om dette i kapittel 6).



3 m/tilbakelegg.

$X =$ # røde

$X \sim \text{bin}(n=3, p=4/8)$

Suksess / fiasko

Multinomisk fordeling

Generalisering til tilfellet hvor du har k mulige utfall (i binomisk er $k = 2$) i hvert delforsøk.

- Vi har n **uavhengige** delforsøk.
- I hvert delforsøk har vi $\boldsymbol{k}$ mulig utfall: $E_1, E_2, \ldots, E_k$.
- Sannsynlighetene for utfalla er **de samme** i alle delforsøka:

$$P(E_1) = p_1, \quad P(E_2) = p_2, \ldots, \quad P(E_K) = p_k$$

La $X_1, X_2, \ldots, X_k$ være antall ganger utfalla $E_1, E_2, \ldots, E_k$ inntraff totalt.
Da vil $(X_1, X_2, \ldots, X_k)$ være **multinomisk** fordelt. Simultanfordeling:

$$f(x_1, \ldots, x_k;\, p_1, \ldots, p_k,\, n) = \binom{n}{x_1, \ldots, x_k} p_1^{x_1} \cdots p_k^{x_k} = \frac{n!}{x_1! \cdots x_k!}\, p_1^{x_1} \cdots p_k^{x_k}$$

Merk at vi alltid har at $\sum_{i=1}^{k} x_i = n$ og $\sum_{i=1}^{k} p_i = 1$.

multinomial-koeffisient

Eksempel: Blodtyper.

Blodtypefordeling i Norge: A: 48%, 0: 40%, B: 8%, AB: 4%.

5 antatt tilfeldige pasienter kommer inn på legevakta.
Hva er sannyligheten for at det er én med blodtype A, to med blodtype 0, to med blodtype B og ingen med blodtype AB?

$$P(X_1 = 1, X_2 = 2, X_3 = 2, X_4 = 0) = \frac{5!}{1! \cdot 2! \cdot 2! \cdot 0!} \cdot 0{,}48^1 \cdot 0{,}4^2 \cdot 0{,}08^2 \cdot 0{,}04^0$$

$$= 30 \cdot 0{,}00049$$

$$= \underline{\underline{0{,}0147}}$$

(OBS: Stort sett alle slike utfall vil ha små s.s.
Men det er veldig mange mulige utfall.)

5.3 Hypergeometrisk fordeling

Likner på binomisk fordeling, men sannsynligheten vil **avhenge** av resultatet i tidligere trekninger.
Vi har allerede sett fleire eksempler/oppgaver hvor vi har hypergeometrisk fordeling.

Eksempler: Hypergeometrisk

1. Urne med 10 røde og 10 blå kuler, trekker 5 kuler tilfeldig **uten** tilbakelegging.
   Hva blir fordelinga for antall blå av de uttrukne?

2. Område med 50 hjort, 20 av dem er merka. Feller 10 (tilfeldig) i jakta.
   Hva blir fordelinga for antall merka av disse?

3. 7 jenter og 5 gutter i en barnehage, velger tilfeldig 3 som får være med til butikken.
   Hva blir fordelinga for antall gutter av disse?

Hypergeometrisk forsøk - hypergeometrisk fordeling

Hypergeometrisk forsøk

1. Har en populasjon med $\boldsymbol{N}$ enheter. Av disse blir $\boldsymbol{k}$ klassifisert som "suksess" og $\boldsymbol{N-k}$ som "fiasko".
2. Trekker tilfeldig ut $\boldsymbol{n}$ enheter **uten** tilbakelegging.

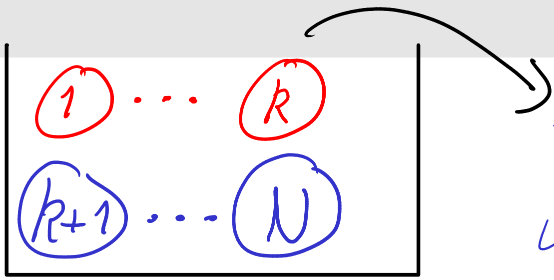
Ville blitt binomisk om det var med tilbakelegging.

Punktsannsynlighet:

La X være antall "suksesser" i et slikt forsøk. Da er X **hypergeometrisk** fordelt med parametrer N, k og n. Sannsynlighetsfordelinga (punktsannsynligheten) til X:

$$f(x) = h(x;\, N, n, k) \quad = \quad \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad \max[0, n - \underbrace{(N-k)}_{\#\text{ fiaskoer i urna}}] \le x \le \min[n, k]$$

\# suksess i urna



trekker n u/tilbakelegg.

Punktsannsynligheten er basert på velkjent regnemetode fra kapittel 2

Eksempel: Jakt.

Område med 40 hjort, 12 av dem er merka. Feller 3 (tilfeldig) i jakta.

Hva blir sannsynligheten for at 2 av disse var merka?

Direkte regning:

$$P(X=2) = \frac{\# g}{\# m} = \frac{\binom{12}{2}\binom{28}{1}}{\binom{40}{3}} = \frac{66 \cdot 28}{9880} = \underline{\underline{0,187}}$$

Hypergeometrisk: $N = 40$, $k = 12$, $n = 3$.

$$P(X=2) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{12}{2}\binom{40-12}{3-2}}{\binom{40}{3}} = \underline{\underline{0,187}}$$

Hypergeometrisk: Bruk av tabell for kumulativ fordelingsfunksjon.

$$P(X=2) = P(X \le 2) - P(X \le 1) \underset{\text{tabell}}{=} 0,9777 - 0,7907 = \underline{\underline{0,1870}}$$

Forventning og varians, hypergeometrisk variabel

Formler for forventning og varians (T 5.2):

Også for en **hypergeometrisk** fordelt stokastisk variabel kan en finne forventning og varians fra parameterverdiene:

$$\mu = E(X) = np \qquad \text{og} \qquad \sigma^2 = Var(X) = np(1-p)\underbrace{\frac{N-n}{N-1}}_{\leq 1}$$

der $p =$ total suksessannsynlighet $= \frac{k}{N}$. (Tar ikke med bevis.)

mindre varians enn binomisk

Eksempel: Jakt.

Område med 40 hjort, 12 av dem er merka. Feller 3 (tilfeldig) i jakta.

La X være antall merka av de som blei felt. Hva blir forventning og varians til X?

$E(X) = np = 3 \cdot \frac{12}{40} = \frac{9}{10} = \underline{\underline{0{,}9}}$

$Var(X) = np(1-p) \cdot \frac{N-n}{N-1} = 3 \cdot \frac{12}{40} \cdot \frac{28}{40} \cdot \frac{37}{39} = \underline{\underline{0{,}598}}$, $SD(X) = \underline{0{,}773}$

Hypergeometrisk og binomisk

Dersom $\boldsymbol{n}$ (størrelsen av utvalget) er **lite** i forhold til $\boldsymbol{N}$ (populasjonen):

- Sammensetninga av enhetene endrer seg lite slik at sannsynligheten for suksess endrer seg lite (tilnærma uavhengighet mellom trekningene).
- Hypergeometrisk kan da **tilnærmes** med binomisk med $p = \frac{k}{N}$.
  (Kan vises at hypergeometrisk sannsynlighet går asymptotisk (når $N \to \infty$) mot binomisk sannsynlighet. Se også på utrykka for forventning og varians.)
- **Tommelfingerregel** for når tilnærminga er god: $\boldsymbol{\frac{n}{N} \leq 0.05}$.
- Tilsvarende multinomisk finnes det og en **multivariat hypergeometrisk** fordeling. (Se boka.)

Eksempel: Jakt.

Rekn ut sannsynlighet $P(X = 2)$ ved tilnærming til binomisk fordeling.

$\frac{n}{N} = \frac{3}{40} = 0{,}075$ (Så litt stor etter tommelfingerregelen.)

$P(X=2) \approx \binom{3}{2}\left(\frac{12}{40}\right)^2\left(1-\frac{12}{40}\right)^{3-2} = \underline{\underline{0{,}189}}$ (Likevel god tilnærming.)

5.4 Negativ-binomisk og geometrisk fordeling

Mulige eksempler: Negativ-binomisk

1. Et forsvarsverk kan tåle 5 bombetreff før det svikter. Et fly slipper bomber med en treffprosent på 20.
   Hva blir fordelinga til antall bomber som må slippes for å slå ut forsvarsverket?

2. En dørselger må selge 6 bokverk før han gir seg for dagen. Han veit at sannsynligheten for at han får til et salg er 25% per salgsforsøk.
   Hva blir fordelinga til antall salgsforsøk han må gjøre per dag?

3. Urne med røde og svarte kuler. p er andel røde kuler. Trekker **med tilbakelegging** inntil k røde kuler er trukket ut.
   Hva blir fordelinga til antall trekninger som må gjøres?

Negativ-binomisk forsøk og fordeling

Negativ-binomisk forsøk:

Tilsvarende et binomisk forsøk (rekke av uavhengige Bernoulli-delforsøk med samme suksess-sannsynlighet) med en viktig forskjell:

- Gjentar delforsøka til vi oppnår $\boldsymbol{k}$ "suksesser" (ikke et på forhand bestemt antall delforsøk).

Punktsannsynlighet:

- La $\boldsymbol{X}$ være **antall delforsøk** du må gjøre for å oppnå at hendelsen $\boldsymbol{A}$ ("suksess") skal inntreffe $\boldsymbol{k}$ ganger i ei Bernoulli-forsøksrekke.

- $\boldsymbol{X}$ blir da **negativ-binomisk** fordelt med sannsynlighetsfordeling (punktsannsynlighet):

$$f(x) = b^*(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, k+2, \ldots$$

må bruke minst k del-
forsøk for k suksesser.

Forklaring av punktsannsynligheten

$X = \#$ delforsøk for å få k suksesser.

For eks. kunne vi fått denne rekka av delforsøk:

$$\underbrace{s\,f\,s\,s\,f\,f\,f\,s \cdots s}\,s \quad \text{siste er alltid suksess}$$

$k-1$ suksesser, $x-k$ fiaskoer på $x-1$ delforsøk

$$P(X=x) = P(k-1 \text{ suksesser på } x-1 \text{ delforsøk} \cap \text{suksess siste delforsøk})$$

uavh. delforsøk

$$= P(\text{———— " ————}) \cdot P(\text{———— " ————})$$

$$= \binom{x-1}{k-1} \cdot p^{k-1} \cdot (1-p)^{x-1-(k-1)} \cdot p$$

$$= \binom{x-1}{k-1} \cdot p^{k} \cdot (1-p)^{x-k}$$

Spesialtilfelle: Geometrisk fordeling

Eksempel: Spillrekke. Hver gang du spiller et spill har du sannsynlighet p for å vinne. Resultat av ulike spill er uavhengige.

- La X være antall ganger du må spille for å vinne **første** gang.
- X vil da være **geometrisk** fordelt.
- Spesialtilfelle av negativ-binomisk, teller antall delforsøk til du har **første** suksess.

Punktsannsynlighet:

La X være antall delforsøk du må gjøre for å oppnå at hendelsen A ("suksess") skal inntreffe **første** gang i ei rekke Bernoulli-delforsøk. X er da **geometrisk** fordelt med sannsynlighetsfordeling (punktsannsynlighet):

$$f(x) = g(x;\, p) = p\,(1-p)^{x-1}, \quad x = 1, 2, \ldots$$

Geometrisk fordeling forts.

- Forklaring av punktsannsynligheten.

Negativ binom med $k=1$: $P(X=x) = \binom{x-1}{0} p^1 (1-p)^{x-1} = p(1-p)^{x-1}$

Alternativ forklaring: $P(X=x) = P(\underbrace{f \cap f \cap \cdots \cap f}_{x-1 \text{ fiaskoer}} \cap s) \underset{\text{uavh}}{=} (1-p)^{x-1} \cdot p$

- Vise at dette blir ei ekte fordeling.

$0 < 1-p < 1$

$$\sum_x f(x) = \sum_{x=1}^{\infty} p(1-p)^{x-1} = p \sum_{x=0}^{\infty} (1-p)^x = p \frac{1-(1-p)^{\infty}}{1-(1-p)} = \frac{p}{p} = \underline{\underline{1}}$$

(Se „noen rekker" i tabell- og formelsamlinga.)

- Kan du finne et enkelt uttrykk for den kumulative fordelingsfunksjonen $F(x)$?

Ja, faktisk:

$$F(x) = P(X \le x) = \sum_{k=1}^{x} p(1-p)^{k-1} \underset{\text{Se øvingsoppgave}}{=} \underline{\underline{1-(1-p)^x}}, \quad x = 1, 2, \ldots$$

Forventning og varians, geometrisk og negativ-binomisk

Forventning og varians geometrisk (T 5.3):

For en geometrisk fordelt stokastisk variabel, X, har vi at

$$\mu = E(X) = \frac{1}{p} \qquad \text{og} \qquad \sigma^2 = Var(X) = \frac{1-p}{p^2}$$

Se obligoppgave.

Forventning og varians negativ-binomisk:

For en negativ-binomisk fordelt stokastisk variabel, Y, har vi at

$$\mu = E(Y) = \frac{k}{p} \qquad \text{og} \qquad \sigma^2 = Var(Y) = k\,\frac{1-p}{p^2}$$

Merk at et neg. binomisk forsøk er et geometrisk forsøk gjentatt k ganger: $Y = X_1 + \cdots + X_k$

Så: $E(Y) = E\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} E(X_i) = \sum_{i=1}^{k} \frac{1}{p} = \underline{\underline{\frac{k}{p}}}$ Tilsvarende for varians.

Eksempel: Bombetokt

Et forsvarsverk kan tåle 5 bombetreff før det svikter.

Et fly slipper bomber med en treffprosent på 20.

- Hva er sannsynligheten for at det første treffet skjer på slipp 3?
- Hva er sannsynligheten for at det første treffet skjer før slipp 3?
- Hva er forventa antall slipp til forsvarsverket svikter?

X = # slipp til første treff. $X \sim \text{geom}(p = 0{,}2)$

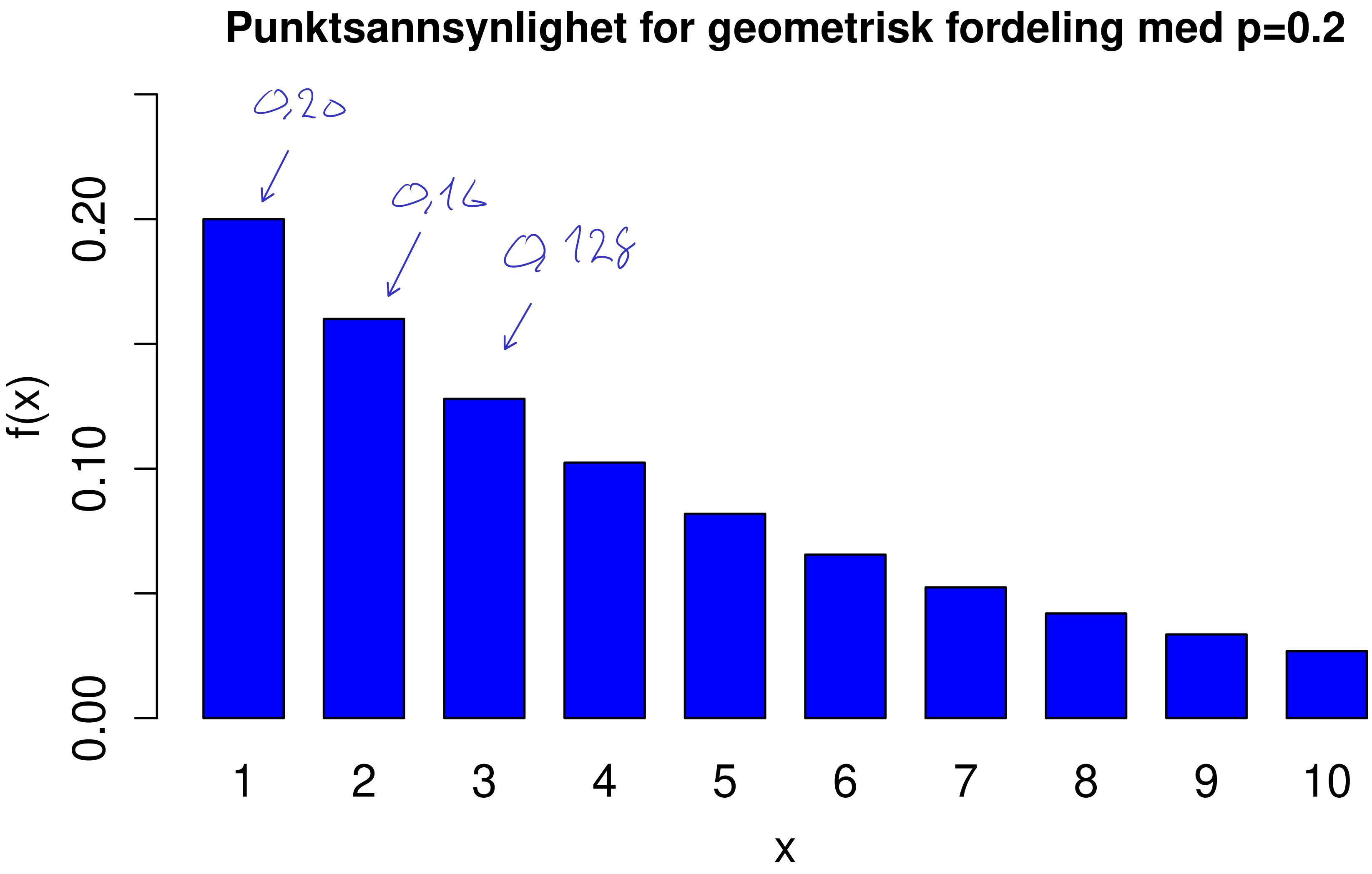
$P(X=3) = p(1-p)^2 = 0{,}2 \cdot 0{,}8^2 = \underline{\underline{0{,}128}}$

$$P(X<3) = \begin{cases} P(X \leq 2) = 1-(1-p)^2 = 1-0{,}8^2 = \underline{\underline{0{,}36}} \\ P(X=1) + P(X=2) = p(1-p)^0 + p(1-p)^1 = 0{,}2 + 0{,}16 = \underline{\underline{0{,}36}} \end{cases}$$

Y = # slipp til femte treff. $Y \sim \text{negbin}(k=5, p=0{,}2)$

$E(Y) = k \cdot \frac{1}{p} = 5 \cdot \frac{1}{0{,}2} = 5 \cdot 5 = \underline{\underline{25}}$

$1/p$ er forventa slipp til første treff.


Punktsannsynlighet for geometrisk fordeling med p=0.2
0,20
0,16
0,128
0.00
0.10
0.20
f(x)
1
2
3
4
5
6
7
8
9
10
x

5.5 Poisson-fordeling

Poisson-fordelinga opptrer som oftest når vi er interessert i antall forekomster av en hendelse over ei viss tid eller rom (lengde/areal/volum).

Mulige eksempler: Poisson

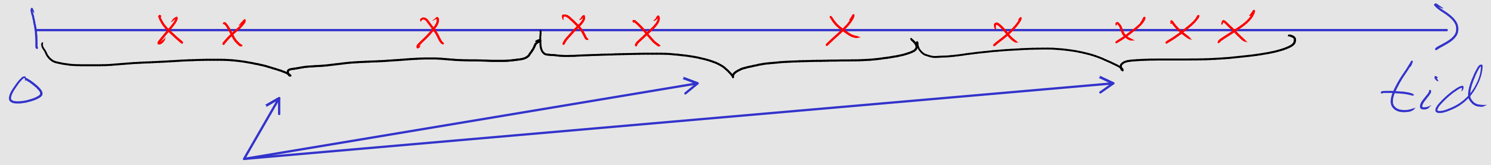
1. Antall vulkanutbrudd på Island siste 1000 år.
2. Antall skrivefeil i løpet av 10 sider du har skrevet.
3. Antall mygglarver i ei vannprøve av et visst volum.

Det er ikke sikkert det er riktig å bruke Poisson-fordeling i disse tilfellene, visse **krav** må være oppfylte.

Poisson-prosess

Kan bruke Poisson-fordeling for antall forekomster av en hendelse A om følgende er oppfylt:

Hendelsestidspunkt på tidslinje:



Krav til Poisson:

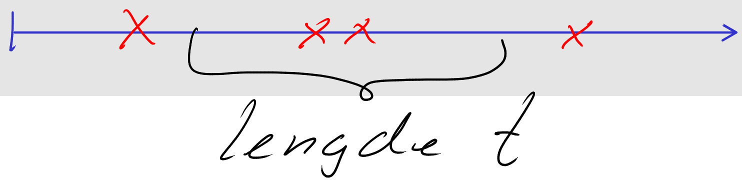
- Antall forekomster av A i disjunkte tidsintervall er **uavhengige**.
- Forventa antall forekomster av A per tidsenhet er **konstant** (kalles raten $\boldsymbol{\lambda}$).
- A kan **ikke** opptre 2 ganger på eksakt samme tidspunkt.

Obs: Bruker **tid** som samlebetegnelse på tid/rom/area/lengde etc.

Om disse krava er oppfylt sier vi at vi har en **Poisson-prosess**.

- La X være antall ganger A inntreffer i en Poisson-prosess i løpet av tid t.
  Da er X **Poisson-fordelt** med parameter λt. Skriver:

$$X \sim \text{Poisson}(\lambda t)$$



Punktsannsynlighet, Poisson-fordeling:

Sannsynlighetsfordelinga (punktsannsynligheten) til en **Poisson**-fordelt stokastisk variabel, X, er

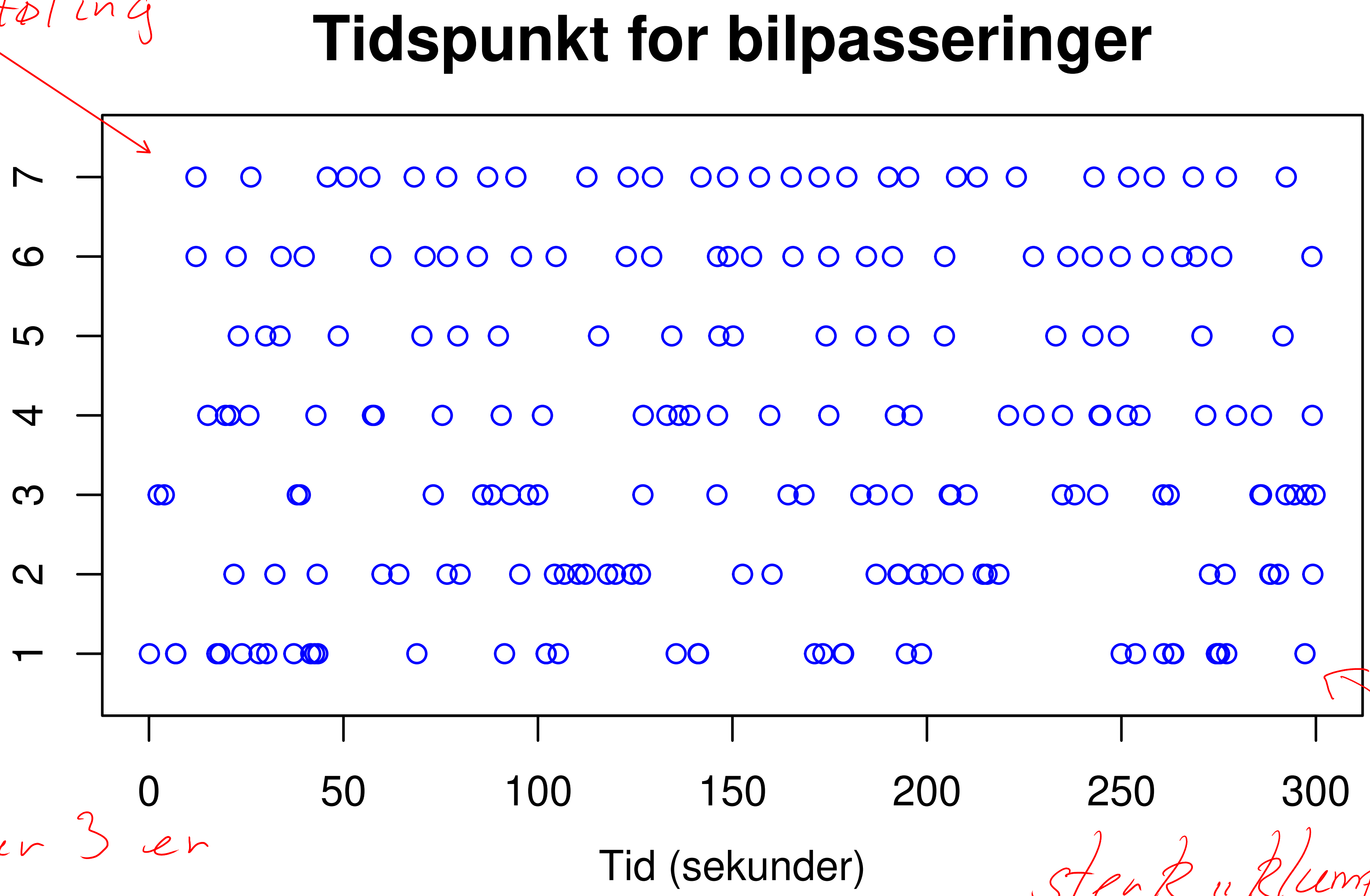
$$f(x) = p(x;\, \lambda t) = \frac{e^{-\lambda t} \cdot (\lambda t)^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

- Dette er sannsynligheten for å observere x forekomster innen en region av størrelse t. Dette kan være ei viss tid, viss lengde, visst areal, visst volum etc.
- Parameteren λ er gjennomsnittlig antall forekomster per “tidsenhet” (hendelsesraten).
- Merk at i en Poisson-prosess er hendelsesraten konstant uansett tidspunkt, og tidspunktet for én hendelse påvirker ikke tidspunktet for andre.
  Dette betyr at om du ser på tidspunkta for hendelser er det ingen tendens ”opphoping” eller ”frastøting” for hendelsene, heilt tilfeldig.

Eksempel: Er du god til å gjenkjenne "tilfeldighet"?

- La oss tenke oss at en trafikkteller registerer tidspunkt for bilpassering mellom 0 og 300 sekund.
- På neste side er 7 realisasjoner som alle har rate på 1/10, det vil si 1 bil per 10 sekund i gjennomsnitt, omtrent 30 biler på 300 sekund.
- **Bare 1** av disse realisasjonene som er fra en Poisson-prosess, de andre har tendenser til opphoping eller frastøting.
- Den nederste har mest tendens til opphoping: Bilene "klomper seg sammen".
- Den øverste har mest tendens til frastøting: Bilene "holder avstand".
- Hvor er Poisson-prosessen?

sterk
„frastøting"



Nummer 3 er
Poisson-prosessen.

sterk „klumping"

Forventning og varians, Poisson-fordeling

Forventning og varians Poisson (T 5.4):

forventa antall forekomster per tidsenhet * lengde tidsrom

For en **Poisson**-fordelt stokastisk variabel har vi at

$$\mu = E(X) = \lambda t \qquad \text{og} \qquad \sigma^2 = Var(X) = \lambda t$$

Skriver dermed også ofte sannsynlighetsfordelinga med μ som parameter:

$$f(x) = p(x;\, \mu) = \frac{e^{-\mu} \cdot \mu^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

Denne siste er den som er oppgitt i tabell- og formelsamlinga.

Bevis: Forventning og varians.

Viser først at $p(x; \lambda t)$ er ei ekte fordeling:

$$\sum_x p(x; \lambda t) = \sum_{x=0}^{\infty} \frac{e^{-\lambda t} \cdot (\lambda t)^x}{x!} = e^{-\lambda t} \sum_{x=0}^{\infty} \frac{(\lambda t)^x}{x!} \overset{\text{formelhefte}}{=} e^{-\lambda t} \cdot e^{\lambda t} = 1$$

$$E(X) = \sum_x x \cdot p(x; \lambda t) = \sum_{x=0}^{\infty} x \cdot \frac{e^{-\lambda t} \cdot (\lambda t)^x}{x!} = \sum_{x=1}^{\infty} x \cdot \frac{e^{-\lambda t} \cdot (\lambda t)^x}{x!}$$

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

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \cdot \frac{e^{-\lambda t} \cdot (\lambda t)^x}{x!} = \sum_{x=1}^{\infty} \frac{x^2 e^{-\lambda t} (\lambda t)^x}{x!} = \lambda t \sum_{x=1}^{\infty} \frac{x e^{-\lambda t} (\lambda t)^{x-1}}{(x-1)!} = \lambda t \sum_{x=0}^{\infty} \frac{(x+1) e^{-\lambda t} (\lambda t)^x}{x!}$$

$$= \lambda t \left(\sum_{x=0}^{\infty} \frac{x e^{-\lambda t} (\lambda t)^x}{x!} + \sum_{x=0}^{\infty} \frac{e^{-\lambda t} (\lambda t)^x}{x!} \right) = \lambda t \cdot (E(X) + 1) \qquad = \lambda t(\lambda t + 1)$$

$$Var(X) = E(X^2) - [E(X)]^2 = \lambda t(\lambda t + 1) - (\lambda t)^2 = \underline{\underline{\lambda t}}$$

Regning av poissonsannsynligheter, praksis

- Punktsannsynlighet kan regnes fra formel:

$$f(x) = P(X = x) = \frac{e^{-\mu} \cdot (\mu)^x}{x!}$$

- Kumulativ fordelingsfunksjon

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} P(X = k) = \sum_{k=0}^{x} \frac{e^{-\mu} \cdot (\mu)^k}{k!}$$

  Finnes i formelsamling for $n = 2, \ldots, 20$ og noen valg av $\mu = \lambda t$.

- Punktsannsynlighet og kumulativ fordelingsfunksjon finnes i noen kalkulatorer.

- Funksjoner for Poisson-fordeling i *R*: *dpois(x,μ)*, *ppois(x,μ)*.

```
> dpois(3,2)
[1] 0.180447
> ppois(3,2)
[1] 0.8571235
```

$X \sim \text{Poisson}(\mu = 2)$

$P(X = 3) = \frac{e^{-2} \cdot 2^3}{3!} = \underline{\underline{0{,}180}}$

$P(X \leq 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = \underline{\underline{0{,}857}}$

Eksempel: Vulkanutbrudd.

Vi antar at antall vulkanutbrudd på Island følger en Poisson-fordeling med rate $\lambda = 0.1$ per år. La X være antall vulkanutbrudd neste 5 år.

- Hva blir forventa antall utbrudd?
- Hva er sannsynligheten for ett utbrudd?

$X \sim \text{Poisson}(\lambda t) = \text{Poisson}(0{,}5)$

– $E(X) = \lambda t = 0{,}1 \cdot 5 = \underline{\underline{0{,}5}}$

– $P(X = 1) = \dfrac{e^{-\lambda t} \cdot (\lambda t)^1}{1!} = \dfrac{e^{-0{,}5} \cdot 0{,}5^1}{1} = \underline{\underline{0{,}303}}$

Binomisk og Poisson

Tilnærme binomisk med Poisson (T 5.5):

Når antall delforsøk n er **stort** og sannsynligheten p er **liten** gjelder:

$$b(x; n, p) \approx p(x; \lambda t)$$

Binomisk punkts.s. kan tilnærmes med Poisson.

Her er $\mu = np = \lambda t$.

Kan bruke Poisson-formler (enklere) som tilnærmingsformler for binomiske sannsynligheter. Merk dette gjelder **bare** om antall forsøk er stort og sannsynligheten liten i hvert delforsøk.

Forklaring: Ved å dele opp tidsaksen i en Poisson-prosess i svært små intervall av lengde t/n kan vi i hvert intervall neglisjere s.s. for fleire enn 1 forekomst. Altså omtrent Bernoulli-delforsøk med s.s. p. Da blir antall forekomster tilnærma $bin(n,p)$.

Eksempel: Vulkanutbrudd.

- Antar antall vulkanutbrudd på Island følger en Poisson-fordeling med rate $\lambda = 0.1$ per år.
- La X være antall vulkanutbrudd neste 5 år. Hva er $P(X = 1)$?
- Vi vil gå "motsatt veg" av vanlig, og tilnærme Poisson-sannsynligheten ved å dele opp 5-års-intervallet i små deler som vi antar er tilnærma Bernoulli-forsøk.
- Prøv med intervall på 1 år (5 intervall) og 1/10 år (50 intervall).

– $P(X=1) = \dfrac{e^{-\lambda t} \cdot (\lambda t)^1}{1!} = \dfrac{e^{-0,5} \cdot 0,5^1}{1!} = \underline{\underline{0,303}}$

– Delen opp $t=5$ års-intervallet i $n=5$ intervall av lengde 1 år.

$np = \lambda t \Rightarrow 5p = \lambda \cdot 5 \Rightarrow p = \lambda = 0,1$

$$P(X=1) \approx b(1; n=5, p=0{,}1) = \binom{5}{1} \cdot 0{,}1^1 \cdot (1-0{,}1)^{5-1}$$

$$= \underline{\underline{0{,}328}}$$

- Deler opp i $n=50$ intervall av lengde $0{,}1$ år.

$$np = \lambda t \Rightarrow 50 \cdot p = \lambda \cdot 5 \Rightarrow p = \frac{\lambda}{10} = 0{,}01$$

$$P(X=1) \approx b(1, n=50, p=0{,}01) = \binom{50}{1} \cdot 0{,}01^1 \cdot (1-0{,}01)^{50-1}$$

$$= \underline{\underline{0{,}304}}$$

Andre funksjoner i *R*

- Funksjoner for hypergeometrisk fordeling i *R*: *dhyper(x,k,N-k,n)*, *phyper(x,k,N-k,n)*.

```
> dhyper(2,4,6,4)
[1] 0.4285714
> phyper(2,4,6,4)
[1] 0.8809524
```

$P(X=2)$

$P(X\leq 2)$

$X \sim hyper(N=10, k=4, n=4)$

- Funksjoner for geometrisk fordeling i *R*: *dgeom(x,p)*, *pgeom(x,p)*.

```
> dgeom(3,1/3)
[1] 0.09876543
> pgeom(3,1/3)
[1] 0.8024691
```

$P(X=3)$

$P(X\leq 3)$

$X \sim geom(p=1/3)$

- Funksjoner for negativ-binomisk fordeling i *R*: *dnbinom(x,k,p)*, *pnbinom(x,k,p)*.

```
> dnbinom(7,3,0.5)
[1] 0.03515625
> pnbinom(7,3,0.5)
[1] 0.9453125
```

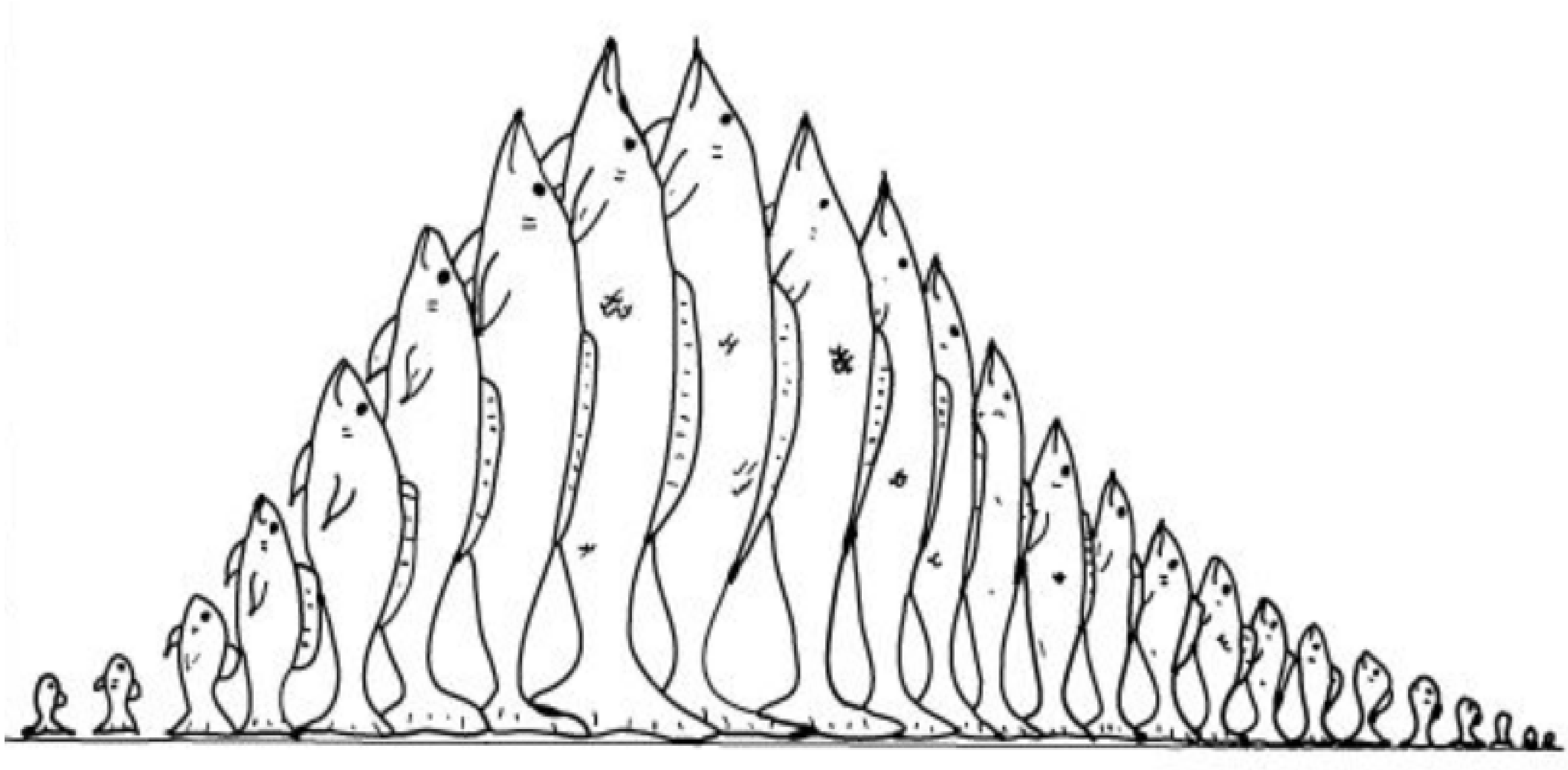
$P(X=7)$

$P(X\leq 7)$

$X \sim negbin(k=3, p=0{,}5)$

Merk i geometrisk og negativ-binomisk at i *R* er den stokastiske variabelen antall fiaskoer før suksess nummer k, ikke antallet delforsøk du trenger for å få suksess nummer k.

Litt på sida (fra 2fliplearning.blogspot.com):



Poisson-fordeling med $\mu = 8$.