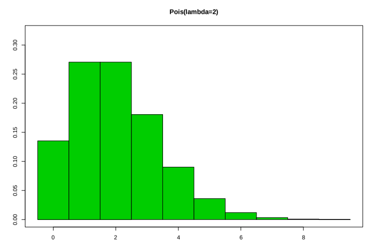
Kapittel 5: Diskrete sannsynlighetsfordelinger

TMA4240 Statistikk (F2 og E7)

5.4-5.6: Negativ binomisk, geometrisk og Poisson fordeling:
mandag 13.september 2004.



Repetisjon - Binomisk

Binomisk fordeling

$$b(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x}, \qquad x = 0,1,2,\ldots,n$$

Fenomén:

i) n uavhengige forsøk
ii) Suksess/fiasko i hvert av forsøkene
iii) $p = P(\text{suksess})$ er lik i alle forsøkene

Sannsynligheten: $p = \mathrm{P}(\text{suksess}) = 1 - \mathrm{P}(\text{fiasko}) = 1 - q$ er den samme i alle forsøkene.

Registrerer: X = antall suksesser i n repeterte forsøk under identiske forhold.

Forventningsverdi og varians:

$$\mu = \mathrm{E}(X) = np,$$
$$\sigma^2 = \mathrm{Var}(X) = np(1-p) = npq.$$

Eksempel:
Teller dager tog et er forsinket i løpet av et år, gitt uavhengig fra dag til dag.
Teller antall hvite kuler som trekkes fra urne med svarte og hvite **med tilbakelegging**.

Kommentar: Kanskje den vanligste og viktigste diskrete fordelingen.

Eksempel - Binomisk

Fallskjermhopp

- $n = 500$ fallskjermhopp
- $p = \frac{1}{500} = 0.002$ for at skjermen ikke virker.
- Anta $X = \sharp$ fiaskohopp $\sim \text{bin}(x; n, p)$
  (Antagelser OK?)
- Beregn

$$\begin{aligned} P(\text{minst en fiasko}) &= P(X \geqslant 1) \\ &= 1 - P(X = 0) \\ &= 1 - \binom{500}{0} p^0 (1-p)^{500-0} \\ &= 1 - 1 \cdot 1 \left(\frac{499}{500}\right)^{500} \quad = \quad 0.63. \end{aligned}$$

Repetisjon - Hypergeometrisk

Hypergeometrisk fordeling

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, 2, \ldots, \min(k, n).$$

Fenomén: Suksess/fiasko-eksperiment. Har en populasjon med størrelse N, og k av disse regnes som suksess hvis de blir trukket. Trekker n ganger **uten tilbakelegging**.

Sannsynligheten: er ikke konstant fra ett trekk til det neste siden vi jobber uten tilbakelegging.

Registrerer: $X =$ antall suksesser i n forsøk, her altså ikke identiske.

Forventningsverdi og varians:

$$\mu = \mathrm{E}(X) = \frac{nk}{N},$$
$$\sigma^2 = \mathrm{Var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N}\left(1 - \frac{k}{N}\right).$$

Eksempel: Spør $n = 10$ studenter i en klasse (på $N = 150$ hvor $k = 83$ er for EU) om de er for eller mot EU, teller antall for og mot. Spør ikke samme person mer enn én gang.

Kommentar: Tenk her på eksperimentets "natur". I praksis vil vi normalt ikke vite k på forhånd. (Det gjelder også parameterne(konstantene) i de andre fordelingene.) Seinere i pensum vil vi lære å *estimere* eller *anslå* k. I eksempel med trekking av kuler fra urne - uten tilbakelegging - er k kjent.
Hvis $N \gg n$, spiller det liten rolle om det er med eller uten tilbakelegging. Da er binomisk fordeling med $p = \frac{k}{N}$ en bra tilnærming.

Eksempel - Hypergeometrisk

Eksempel: Meningsmåling

- Ja/nei-spørsmål:
  Gjør regjeringen en bra jobb?
- Typisk at
  - $N = 4000000$.
    - $k = \#$ som ville svart JA.
    - $N - k = \#$ som ville svart NEI.
  - $n = 1000$ som faktisk blir spurt.
  - $X = \#$ av de $n = 1000$ personene som svarer JA.

Binomisk og negativ binomisk

- Forsøksrekke, registrerer A (suksess) eller A^* (fiasko) i hvert forsøk.
- $P(A) = p$ i hvert forsøk.
- Forsøkene er uavhengige.

Binomisk	**Negativ binomisk**
• Bestemmer totalt antall forsøk, n, på forhånd.	• Antall forsøk er ikke bestemt på forhånd, men eksperimentet avsluttes når k suksesser er oppnådd.
• X=antall suksesser på n forsøk	• X=antall forsøk til k suksesser er oppnådd.

Negativ binomisk fordeling

Negativ binomisk eksperiment: utføres som et binomisk eksperiment med den forskjell at forsøkene gjentas til et *fast* antall suksesser inntreffer. Dvs.

1. Eksperimentet består av et på forhånd ukjent antall forsøk.
2. Hvert forsøk: inntreffer hendelsen A (suksess) eller ikke (fiasko).
3. Sannsynligheten for hendelsen A (suksess), $P(A) = p$, er den samme fra forsøk til forsøk.
4. De gjentatte forsøkene er uavhengige av hverandre.
5. Eksperimentet avsluttet når et bestemt antall, k, av hendelsen A (suksesser) har inntruffet.

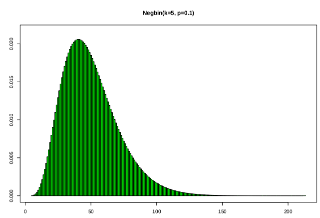
Negativ binomisk fordeling: Vi ser på gjentatte uavhengige forsøk som kan resultere i hendelsen A (suksess) med sannsynlighet p og komplementet til hendelsen A (A'=fiasko) med sannsynlighet $1 - p$.
La den stokastiske variabelen X angi antall forsøk som må gjøres for at hendelsen A (suksess) inntreffer k ganger. X har da en *negativ binomisk fordeling* med sannsynlighet

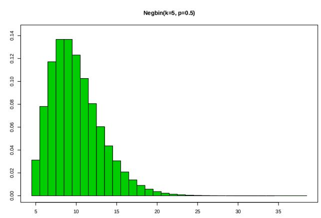
$$b^*(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{(x-k)}$$

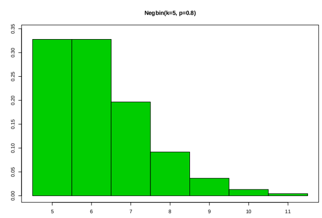
for $x = k, k+1, k+2, \ldots$.

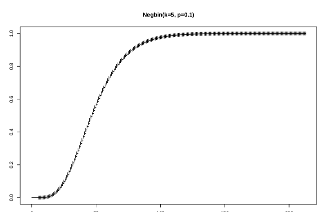
Negativ binomisk fordeling (forts.)

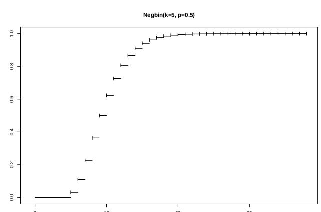
$k = 5, p = 0.1$ $\qquad$ $k = 5, p = 0.5$ $\qquad$ $k = 5, p = 0.8$

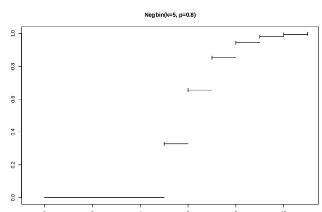












Sjokoladesalg

- Per skal selge sjokolader i nabolaget for å tjene penger for speidergruppen. Han har fått beskjed om å komme hjem etter at han har solgt k sjokolader. Sannsynligheten p for å få solgt en sjokolade i et hus er konstant.
- Antall hus X Per må innom er negativt binomisk fordelt.

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k (1-p)^{(x-k)}$$

for $x = k, k+1, k+2, \ldots.$

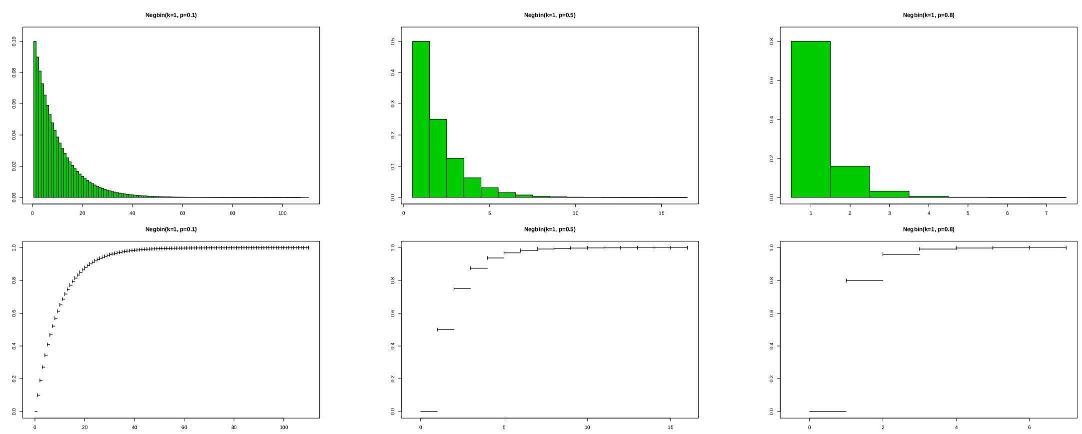
Geometrisk fordeling

Negativ binomisk med k=1: $g(x;p) = p(1-p)^{(x-1)}$ $\quad x = 1, 2, 3, \ldots$

TEO 5.4: Forventning og varians i den geometriske fordelingen $g(x;p)$ er

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

$p = 0.1$ $\qquad$ $p = 0.5$ $\qquad$ $p = 0.8$



5.6 Poisson prosess og fordeling

Poisson prosess: Vi ser på om en hendelse inntreffer eller ikke innenfor et intervall eller en region.

1. Antall hendelser som inntreffer i et intervall eller i en spesifisert region, er uavhengig av antall hendelser som inntreffer i ethvert annet disjunkt intervall eller region.
2. Sannsynligheten for at en enkelt hendelse inntreffer innenfor et lite intervall eller liten region, er proporsjonal med lengden av intervallet eller størrelsen på regionen, og er ikke avhengig av antallet hendelser som inntreffer utenfor intervallet eller regionen.
3. Sannsynligheten for at mer enn en hendelse skal inntreffe innenfor et kort intervall eller liten region er negliserbar.

Poisson fordeling: La den stokastiske variabelen X representere antallet hendelser i et gitt intervall eller region av størrelse t. Sannsynlighetsfordelingen til X er

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

hvor λ er gjennomsnittlig antall hendelser per enhet intervall eller region (og $e = 2.71828$).

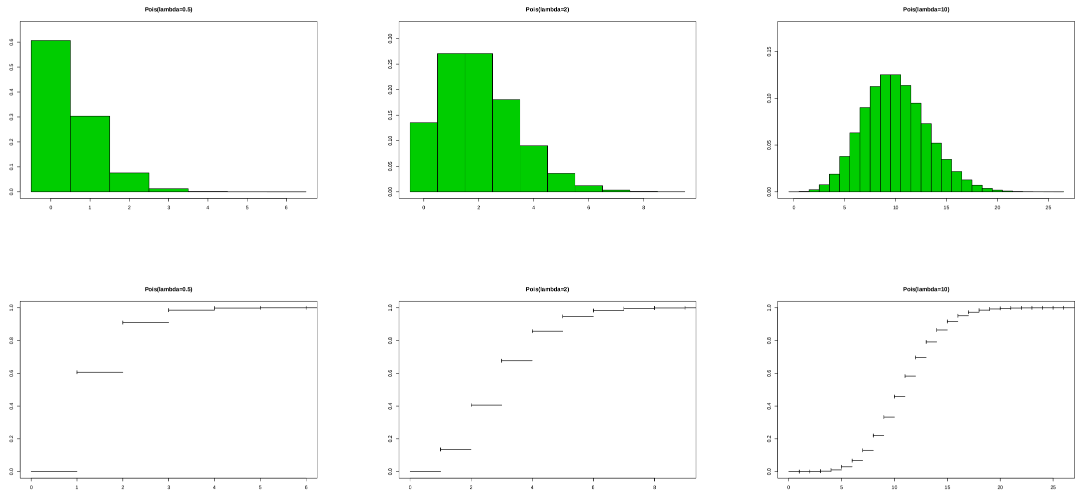
Binomisk- og Poissonfordeling

TEO 5.6 La X være en binomisk stokastisk variabel med sannsynlighetsfordeling $b(x; n, p)$.
Når $n \to \infty$, $p \to 0$, og $\mu = np$ holdes konstant, så er

$$b(x; n, p) \to p(x; \mu)$$

Poisson fordeling (forts.)

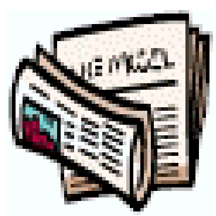
$\mu = \lambda t = 0.5$ $\quad$ $\mu = \lambda t = 2$ $\quad$ $\mu = \lambda t = 10$



TEO 5.45 Forventning og varians i Poissonfordelingen $p(x;\lambda t)$ er begge $\mu = \lambda t$.

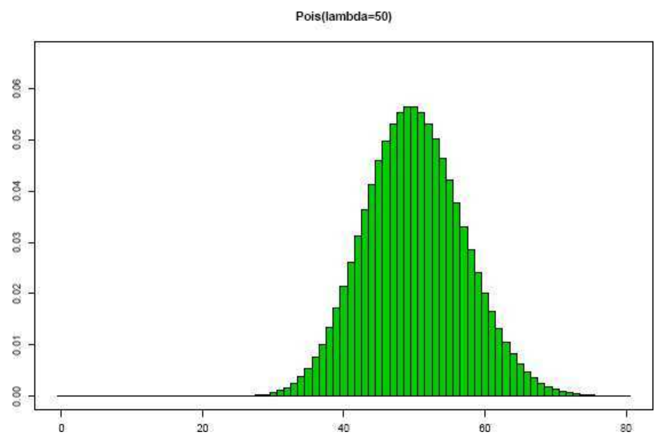
Optimal leveranse av Dagbladet

- Daglig selges rundt 220 000 eksemplarer av Dagbladet hos tilsammen 11 000 utsalgssteder.
- Dagbladet ønsker å bruke statistiske modeller for å betemme hvor mange eksemplarer som skal leveres til hvert utsalgssted hver salgsdag for at avisen skal tjene mest.
  - Leveres for mange eksemplarer blir noen ikke solgt og er en unødvendig kostnad.
  - Leveres for få eksemplarer gå utsalgsstedet utsolgt og avisen taper salgsinntekter.
  - Økonomer i avisen kan angi en kostnad eksemplar som ikke blir solgt og for eksemplarer som kunne vært solgt (tapt salg). Dette kan våre avhengig av ukedag, type utsalgssted og andre størrelser.
- Kan vi finne fordelingen til antall aviser som kan selges på hvert salgssted hver salgsdag kan vi optimalt bestemme hvor mange aviser som skal leveres til hvert salgssted hver salgssdag.
- Et slikt system er implementert ved Dagbladet!



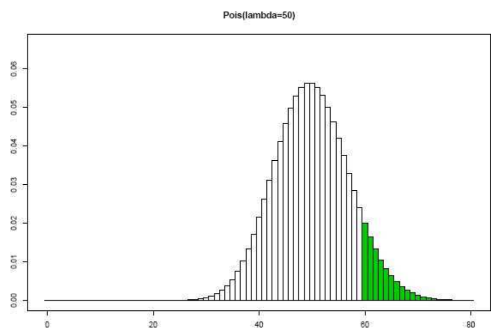
Fordelingen til avissalg

- Dagens salg av Dagbladet i en dagligvareforretning på City Syd (idealisert).
  1. Ser vi på salget i to disjunkte tidsintervall så er disse uavhengige. (Har mange aviser og går ikke utsolgt.)
  2. Kundene ankommer butikken fordelt over hele åpningstiden. Noen av kundene kjøper Dagbladet, og vi har en underliggende intensitet for kjøp på λ.
  3. To salg er ikke fullstendig sammenfallende på tidsaksen.
- Salget er Poisson-fordelt med forventing λt.



Fordelingen til avissalg (forts.)

- Forventet salg er avhengig av utsalgssted og salgsdag. Klar effekt av:
  - Ukedag
  - Sesong, helligdager, høytider, spesielle hendelser, trender over lengre perioder.
  - Type utsalgssted, geografi.
- Basert på data tilbake i tid (her 3.5 år) kan man anslå forventet salg for hver utsalgssted og hver salgsdag – frem i tid.
- Leveranse kan så bestemmes som en percentil i Poisson-fordelingen med denne forventningen.



- Metoden anbefaler leveringstall på en normaldag, og skaleres i forhold til dagens forside (totalopplaget).