

TMA4240/TMA4245 Statistikk

Oppsummering diskrete sannsynlighetsfordelinger

*Binomisk fordeling**

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

Fenomen:

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

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

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

Forventningsverdi og varians:

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

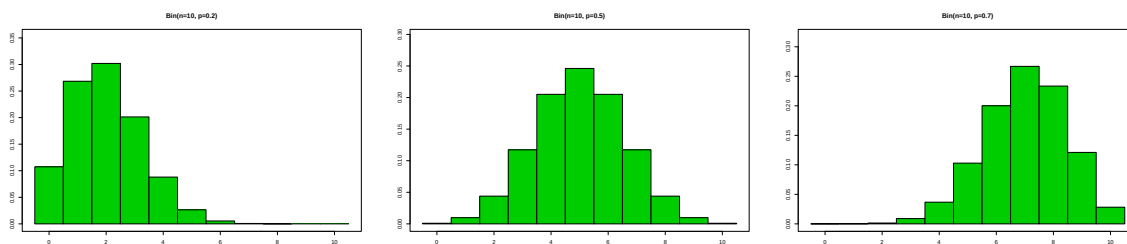
Eksempel:

Teller antall dager et tog er forsinket i løpet av n dager, 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. En sum av stokastiske variabler som er binomisk fordelt med samme p er også binomisk fordelt med p og summen av antall forsøk.

Sammenheng med andre fordelinger :

- Når n blir stor og $\mu = np$ holdes konstant, går binomisk fordeling mot Poisson fordeling.
- For store n kan binomisk fordeling tilnærmes med en normalfordeling (sjekk at $np > 5$ og $n(1-p) > 5$, og bruk kontinuitetskorreksjon).



Multinomisk fordeling

$$f(x_1, x_2, \dots, x_k; n) = \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \quad \sum_{i=1}^k x_i = n, \quad \sum_{i=1}^k p_i = 1.$$

Fenomen:

- i) n uavhengige forsøk
- ii) Hvert forsøk gir ett av k mulige utfall $E_1, E_2, \dots, E_k$
- iii) $p_1 = P(E_1), \dots, p_k = P(E_k)$ er de samme i alle forsøkene.

Sannsynligheten: $p_i = P(E_i)$ er den samme i alle forsøkene.

Registrerer: $X_1, X_2, \dots, X_k$: antall ganger hver av hendelsestypene opptrer når en gjør n uavhengige forsøk.

Forventningsverdi og varians:

$$\mu_i = E(X_i) = np_i, \\ \sigma_i^2 = \text{Var}(X) = np_i(1 - p_i).$$

Eksempel: Urne med kuler med forskjellig farge, 3 røde, 2 blå og 9 grønne, dvs $k = 3$ "typer", trekker $n = 14$ ganger **med tilbakelegging** og registrerer antall ganger en trakk hver farge.

Kommentar: Utvidelse av binomisk fordeling.

Hypergeometrisk fordeling

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

Fenomen: 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.

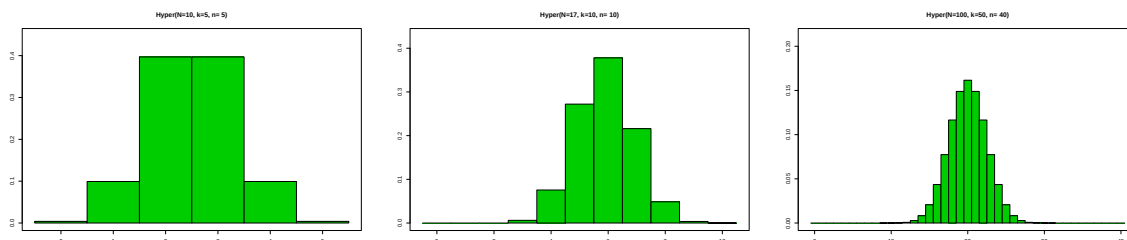
Forventningsverdi og varians:

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

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", og se om eksperimentet kan beskrives slik som listet opp under Fenomen.

Sammenheng med andre fordelinger: 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.



Multivariabel hypergeometrisk fordeling

$$f(x_1, x_2, \dots, x_k; a_1, a_2, \dots, a_k, N, n) = \frac{\binom{a_1}{x_1} \binom{a_2}{x_2} \dots \binom{a_k}{x_k}}{\binom{N}{n}}, \quad \sum_{i=1}^k x_i = n, \quad \sum_{i=1}^k a_i = N.$$

Fenomen: Det er k mulige (forskjellige) utfall/"typer" som kan nummereres. Total populasjon er N , hvorav a_1 er av type 1, a_2 er av type 2 osv. Trekker n ganger **uten tilbakelegging**.

Sannsynlighetene: er ikke de samme fra et forsøk til det neste, vi jobber **uten tilbakelegging**.

Registrerer: $X_1, X_2, \dots, X_k$: antall ganger hver av hendelsestypene opptrer med n forsøk.

Forventningsverdi og varians:

$$\mu_i = E(X_i) = \frac{na_i}{N}$$

Variansen til X_i har tilsvarende form som for hypergeometrisk fordeling.

Eksempel: Gallup i en liten kommune; hvilket parti vil du stemme på? En spør n av de N stemmeberettigede. Totalt i kommunen ville a_1 stemme bygdelista, a_2 Senterpartiet, $\dots$, og a_k ville ikke stemme. Samme person blir ikke spurt mer enn en gang.

Kommentar: Dette er en generalisering av hypergeometrisk fordeling, på samme vis som multinomisk fordeling er en generalisering av binomisk fordeling. I prinsippet kan dere i grunnkurset i statistikk klare dere med hypergeometrisk ved å se på en og en hendelse. F.eks. kan vi definere utfallet at personen som blir spurt vil stemme bygdelista som suksess, og alt annet fiasko.

Geometrisk fordeling

$$g(x; p) = p(1 - p)^{x-1}, \quad x = 1, 2, 3, \dots, (\infty).$$

Fenomen: Identiske, uavhengige forsøk som resulterer i suksess/fiasko

Sannsynligheten: $p = P(\text{suksess}) = 1 - P(\text{fiasko}) = 1 - q$ er konstant for alle forsøk.

Registrerer: X = antall forsøk som må gjøres før vi får første suksess.

Forventningsverdi og varians:

$$\mu = E(X) = \frac{1}{p},$$

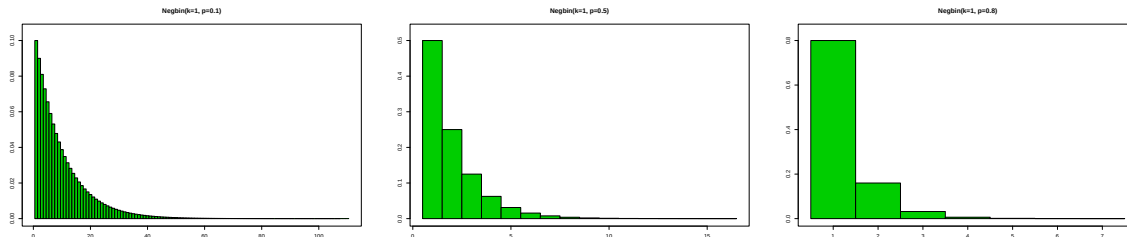
$$\sigma^2 = \text{Var}(X) = \frac{1-p}{p^2}.$$

Eksempel:

Antall ganger en kaster en terning før første sekser, f.eks. for å begynne et spill Ludo.

Antall dager før strømmen går hvis sannsynligheten for strømbrudd er den samme hver dag.

Kommentar: Dette er et spesialtilfelle av negativ binomisk fordeling med $k = 1$. Problemstillingen er så vanlig at den har et eget navn, i tillegg til at fordelingen har form som en geometrisk følge for $x = 1, 2, 3, \dots$

**Negativ binomisk fordeling**

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

Fenomen: Identiske, uavhengige forsøk som resulterer i suksess/fiasco

Sannsynligheten: $p = P(\text{suksess}) = 1 - P(\text{fiasko}) = 1 - q$ er konstant for alle forsøk.

Registrerer: X = antall forsøk som må gjøres før vi har oppnådd k suksesser.

Forventningsverdi og varians:

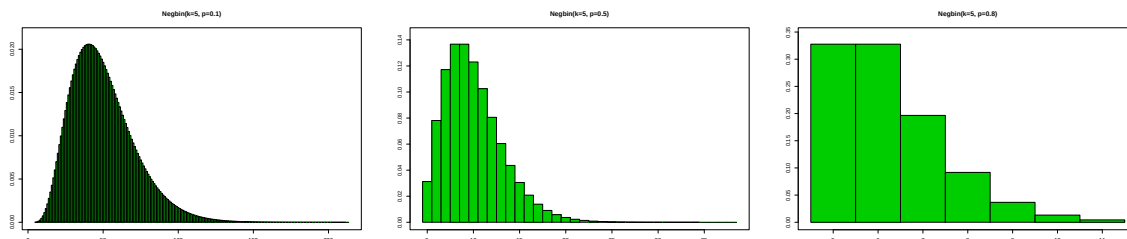
$$\mu = E(X) = \frac{k}{p},$$

$$\sigma^2 = \text{Var}(X) = k \frac{1-p}{p^2}.$$

Eksempel: Kaster et kronestykke helt til vi har fått 7 mynt og teller antall kast.

Antall dager en maskin kan brukes før den har feilet 11 ganger, forutsatt at feilene ikke skyldes alder eller tidligere feil, men opptrer tilfeldig. Hvor mange biler må passere før man må skifte film i "fotoboksen" som står plassert i ett spesielt veikryss?

Kommentar: Sammelikn negativ binomisk og binomisk fordeling. Denne har "negativ" som fornavn, fordi vi tenker motsatt. I stedet for et fast antall forsøk og telle antall suksesser, har vi her et fast antall suksesser og teller antall forsøk.



Poisson-fordeling*

$$p(x; \mu) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} = \frac{\mu^x}{x!} e^{-\mu}, \quad x = 0, 1, 2, 3, \dots, (\infty).$$

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

- 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.
- 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.
- Sannsynligheten for at mer enn en hendelse skal inntreffe innenfor et kort intervall eller liten region er neglisierbar.

Registrerer: X = antall hendelser innenfor et fast tidsintervall t .

Forventningsverdi og varians:

$$\begin{aligned}\mu &= E(X) = \lambda t, \\ \sigma^2 &= \text{Var}(X) = \lambda t = \mu.\end{aligned}$$

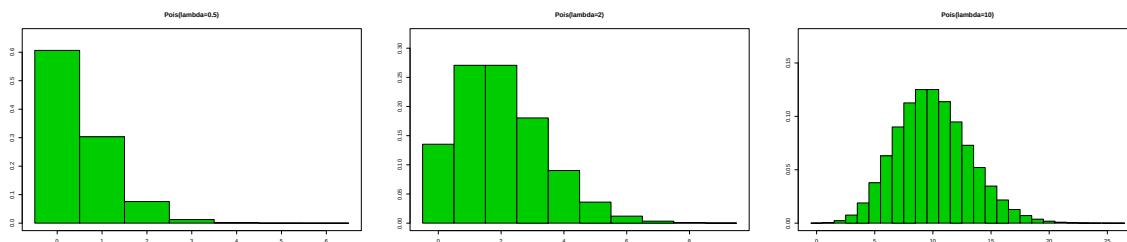
Eksempel:

Ankomst av kunder til en billettluke. Antall jordskjelv i løpet av 10 år i California. Antall ulykker i alpinbakke. Antall aviser som selges ved et utsalgssted. Antall kjempeflodbølger som treffer kysten av Japan de neste 50 årene.

Kommentar: Situasjoner hvor antakelsene i Poisson-fordelingen stemmer, kaller vi Poisson-prosesser. Poisson skrives med stor forbokstav siden fordelingen er oppkalt etter fransk matematiker.

Sammenheng med andre fordelinger :

- Summen av Poisson-fordelte variable er også Poisson-fordelt med summen av forventningsverdiene.
- Hvis n er stor og p er liten i en Binomisk fordeling, kan den tilnærmes med en Poisson-fordeling med $\mu = np$.
- Ventetiden mellom to hendelser i en Poisson-prosess er eksponensialfordelt. Ventetiden mellom mer enn to hendelser er gamma-fordelt.



Diskret uniform fordeling

$$f(x; k) = \frac{1}{k}, \quad x = x_1, x_2, \dots, x_k.$$

Fenoméén: Et sett med k disjunkte, nummererte utfall eller "typer".

Sannsynligheten: Er den samme for alle utfallene.

Registrerer: X = verdien på utfallet som inntreffer. (Ikke nummeret, dvs. x_i , ikke i .)

Forventningsverdi og varians:

$$\mu = E(X) = \frac{1}{k} \sum_{i=1}^k x_i,$$
$$\sigma^2 = \text{Var}(X) = \frac{1}{k} \sum_{i=1}^k (x_i - \mu)^2.$$

Eksempel: Ett terningkast, registrere antall øyne.

Kommentar: Merk "likheten" i formelen mellom empirisk gjennomsnitt og forventningsverdien, og **forskjellen** mellom empirisk varians og varians.