STK1100 våren 2018

Poissonfordelingen

Svarer til avsnitt 3.7 i læreboka

Ørnulf Borgan
Matematisk institutt
Universitetet i Oslo

Poissonfordelingen er oppkalt etter den franske fysikeren og matematikeren Siméon Denis Poisson (1781-1840)

$E(X)$ $X \sim Bin(n,p)$ $p = p_n$

Poisson viste at hvis $n \to \infty$ og $p \to 0$ på en slik måte at $np \to \lambda > 0$ så vil

$$\binom{n}{x} p^x (1-p)^{n-x} \to \frac{\lambda^x}{x!} e^{-\lambda}$$

Dette betyr at hvis n er stor og p er liten, så vil $\binom{n}{x} p^x (1-p)^{n-x} \approx \frac{\lambda^x}{x!} e^{-\lambda}$ $\lambda = np$

Poisson brukte dette til å finne en tilnærmelse til den binomiske fordelingen når n er stor og p er liten

Seinere fant en ut at en kunne bruke den grensen Poisson fant som punktsannsynligheten til en stokastisk variabel

Vi sier at en stokastisk variabel X er Poissonfordelt med parameter λ hvis den har punktsannsynlighet

$$p(x;\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad \text{for } x = 0, 1, 2, \ldots.$$

Merk at $\sum_{x=0}^{\infty} p(x;\lambda) = 1$ slik det skal være for en punktsannsynlighet

Poissonfordeling med MATLAB

Punktsannsynlighet:

$$p(x;\lambda) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

MATLAB: poisspdf(x,λ)

Kumulativ fordeling:

$$F(x;\lambda) = P(X \le x) = \sum_{y=0}^{x} \frac{\lambda^y}{y!} e^{-\lambda}$$

MATLAB: poisscdf(x,λ)

Tabell A.2 bak i boka gir $F(x;\lambda)$ for noen verdier av λ

Vi vil ikke bry oss om disse tabellene

Antar at $np = \lambda$

Har at x er gitt naturlig tall

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \frac{n(n-1)(n-2)\cdots(n-x+1)}{n \cdot n \cdot n \cdots n} \frac{(np)^x}{x!} (1-p)^n \frac{1}{(1-p)^x} \qquad p = \frac{\lambda}{n}$$

$$= \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n \cdot n \cdots n}}_{\substack{n \to \infty \\ 1}} \frac{\lambda^x}{x!} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{e^{-\lambda} \;\; n \to \infty} \frac{1}{(1-\lambda/n)^x}$$

$$\to \frac{\lambda^x}{x!} e^{-\lambda} \qquad n \to \infty$$

Sammenligning av binomisk og Poisson

Vi vil sammenligne de binomiske fordelinger med $n = 20,\ p = 0.10$ og $n = 200,\ p = 0.01$ med Poissonfordelingen med $\lambda = 2$

MATLAB:
```
x=(0:10)'
[x binopdf(x,20,0.1) binopdf(x,200,0.01) poisspdf(x,2)]
```

Resultatet er gitt på neste slide

x	$b(x;20,0.10)$	$b(x;200,0.01)$	$p(x;2)$
0	0.1216	0.1340	0.1353
1	0.2702	0.2707	0.2707
2	0.2852	0.2720	0.2707
3	0.1901	0.1814	0.1804
4	0.0898	0.0902	0.0902
5	0.0319	0.0357	0.0361
6	0.0089	0.0117	0.0120
7	0.0020	0.0033	0.0034
8	0.0004	0.0008	0.0009
9	0.0001	0.0002	0.0002

Vi kan ofte bruke Poissonfordelingen til å beskrive forekomsten av «sjeldne begivenheter» :

- Antall tvillingfødsler i løpet av ett år på et sykehus
- Antall krefttilfeller i løpet av ett år i en kommune
- Antall ulykker i løpet av én måned på en byggeplass

Vi vil se på et konkret eksempel om litt, men først vil vi finne forventning og varians for Poisson-fordelingen

Forventning og varians

Momentgenererende funksjon:

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{\infty} e^{tx} \cdot p(x) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} e^{-\lambda}$$

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

Kumulantgenererende funksjon

$$R_X(t) = \ln\{M_X(t)\} = \lambda e^t - \lambda$$

Forventning

$$\mu = E(X) = R_X'(0) = \lambda$$

Varians:

$$\sigma^2 = V(X) = R_X''(0) = \lambda$$

Eksempel:
Forekomst av anencefali i Edinburgh 1956-66

Anencefali er en alvorlig defekt hos fosteret som gjør at hjernen ikke utvikler seg som den skal. Antall barn født med anencefali i Edinburgh i de 132 månedene fra 1955 til 1956 er gitt i tabellen:

# anencefali	0	1	2	3	4	5	6	7	8	9+
# måneder	18	42	34	18	11	6	0	2	1	0

Kan dataene beskrives med Poissonfordelingen?

Har at gjennomsnittlig x-verdi er ca. lik E(X)
Antall tilfeller er 18·0+42·1+34·2+...+8·1 = 260 $\quad \frac{260}{132} = 1{,}97$

La X være antall tilfeller av anencefali i en måned

Hvis X er Poissonfordelt har vi at $\quad 132 \cdot p(x;1.97)$

$$P(X=x) = p(x;\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

Matlab: poisspdf(0:8, 1.97)

Vi ser på n = 132 måneder

Forventet antall måneder med x tilfeller er $n \cdot p(x;\lambda)$

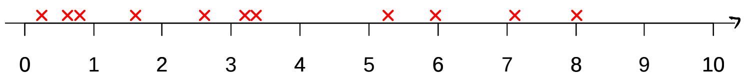
For λ kan vi bruke gjennomsnittlig antall tilfeller per måned som er 1.97

# tilfeller	0	1	2	3	4	5	6	7	8
# observert	18	42	34	18	11	6	0	2	1
# forventet	18.4	36.2	35.7	23.5	11.6	4.5	1.4	0.4	0.1

Poissonfordeling passer bra

Poisson prosessen

Vi observeret begivenheter (markert med x) som hender over tid:



Vi antar at:

- sannsynligheten for at det inntreffer én begivenhet i et intervall av lengde Δt er $\alpha\Delta t + o(\Delta t)$ $\quad \frac{o(\Delta t)}{\Delta t} \to 0$ så dette går mot 0 når Δt
- sannsynligheten for at det inntreffer mere enn én begivenhet i et intervall av lengde Δt er $o(\Delta t)$
- antall begivenheter i et intervall er uavhengig av hvor mange begivenheter som har skjedd tidligere

0 x antall begivenheter

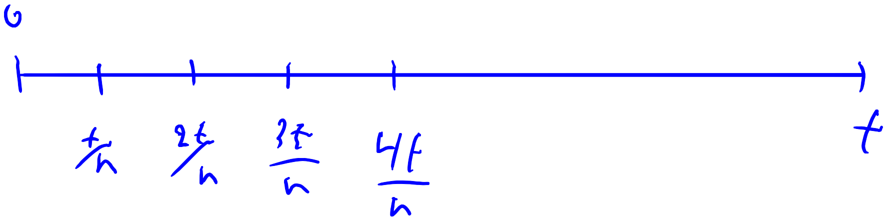
La $P_k(t)$ være sannsynligheten for at det inntreffer k begivenheter i et intervall av lengde t

Da har vi at $P_k(t) = \frac{(\alpha t)^k}{k!}e^{-\alpha t}$

Antall begivenheter i et intervall av lengde t er Poisson fordelt med parameter $\lambda = \alpha t$ $\quad \frac{E(x)}{t} = \alpha$ raten til prosessen

Merk at α er forventet antall begivenheter per tidsenhet

α er raten til Poissonprosessen



Deler i n deler

Hver delintervall har lengde $\Delta t = t/n$

Ser fra delintervall om lengdeintervallet

$$p = \alpha \Delta t = \alpha\, t/n$$

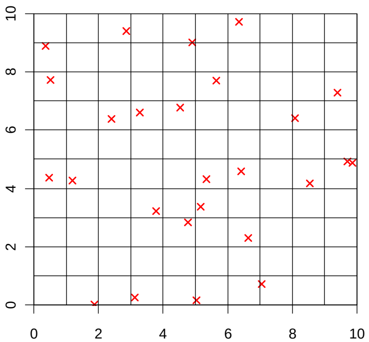
Uavhengighet fra intervall til intervall

Dermed (tilnærmet) X = antall begivenheter $\sim \mathrm{Bin}(n, p)$

n er stor og p er liten:

Dette (tilnærmet) $X \sim \mathrm{Poisson}(\lambda)$ der $\lambda = np = \alpha t$

Vi kan også se på punkter i planet:



Anta at:

- forventet antall punkter per arealenhet er α
- det er ingen sammenfallende punkter
- antall punkter i disjunkte områder er uavhengige

Da har vi en Poissonprosess i planet

Dette er en modell for punkter som er «tilfeldig fordelt»

La X være antall punkter i et område R med areal $a(R)$

Da er X Poissonfordelt med parameter $\lambda = \alpha \cdot a(R)$

Hvis tilfeldig er antall treff i et område Poisson fordelt med $\lambda = 0{,}93$

Eksempel:
Bombetreff av tyske V-1 raketter i Syd-London

Ser på antall treff av V-1 bomber i 576 små områder i Syd-London ($0.25\ km^2$ hver) fra juni 1944 til mars 1945

# treff	0	1	2	3	4	5+
# observert	229	221	93	35	7	1
# forventet	226.7	211.4	98.5	30.6	7.1	1.6

$(n \cdot P(X=x) = 576 \cdot \frac{\lambda^x}{x!} e^{-\lambda}$

Poissonfordelingen passer bra

Case studie: Krefttilfeller i Sømna


PINSEVENN I SEX-SAK

KVINNER KJØPER SEX

SØNDAG

Dagbladet

BYGD I KREFT-SJOKK

FIRE REDDET FRA DYNAMITT-SKIP

Dagbladet 10. januar 1993

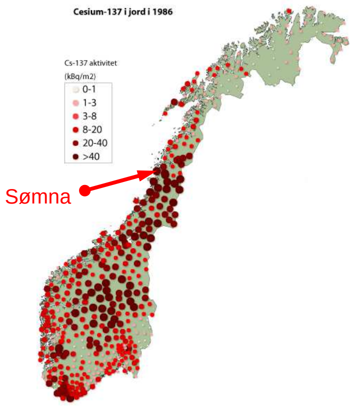

POLITIKERE ER SKREMT

Krever Stortings-orientering om kreftteoriene i Sømna

– Skremmende. Slik reagerer politikere Dagbladet har snakket med på reportasjen fra Sømna om frykten for kreftøkning som følge av Tsjernobyl.

Dagbladet 11. januar 1993

- Det ble observert tre tilfeller av hjernesvulst i Sømna kommune i 1992
- Det var uvanlig mange i en så liten kommune
- Saken vakte stor oppsikt i media og blant politikere.
- Ble sett i sammen-heng med Tsjernobyl ulykken 26. april 1986



KREFT registeret
INSTITUTT FOR POPULASJONSBASERT KREFTFORSKNING

Etter Kreftregisterets statistikk vil en for en kommune av Sømnas størrelse og befolknings-sammensetning i gjennomsnitt observere ett tilfelle av hjernesvulst hvert sjette år *hvis* kreftrisikoen er som i resten av landet

Mer presist: Hvis kreftrisikoen i Sømna var som ellers i landet, vil antall tilfeller av hjernesvulst i løpet av ett år være Poisson fordelt med $\lambda = 0.16$

Det gir: P(minst 3 tillfeller av hjernesvulst)

$$= \sum_{k=3}^{\infty} \frac{0.16^k}{k!} e^{-0.16} = 0.0006$$

Det som skjedde i Sømna var svært usannsynlig hvis kommunen hadde samme kreftrisko som landet forøvrig

---

Tror at det skyldes ren tilfeldighet

— Tilfellet Sømna ser foreløpig ut som en ren tilfeldighet, sier Frøydis Langmark, Kreftregisterets leder.

I 1992 ble det registrert tre tilfeller av hjernesvulst i Sømna kommune i på Helgeland. Det normale i tidligere år har vært null til ett tilfelle. Hittil i år er det ikke registrert noen tilfeller.

Leder av Kreftregisteret, Frøydis Langmark, bruker uttrykket «cluster», eller opphopning om det som er skjedd. Hun sier dette er et ikke uvanlig fenomen, og at man sjelden finner spesielle grunner til at slikt skjer. Men Kreftregisteret vil likevel fortsette med å undersøke saken, for å finne mulige årsaker.

Aftenposten 20. oktober 1993

- Hvordan kan vi forklare at krefttilfellene i Sømna skyldes en ren tilfeldighet?
- Må da være oppmerksom på hvorfor krefttilfellene i Sømna vakte oppsikt
- Det skjedde nettopp fordi det ble registrert uvanlig mange krefttilfeller i denne ene kommunen i dette ene året
- Alle de kommunene og alle de årene der det ikke skjer noe oppsiktsvekkende, er det ingen som bryr seg om

- Vi må derfor spørre: Hva er sannsynligheten for at vi en gang i blant i en eller annen kommune vil observere noe så påfallende som det en gjorde i Sømna i 1992 ved en ren tilfeldighet?
- Regneeksempel: Vi tenker oss at vi har 100 kommuner med samme størrelse og befolkningsstruktur som Sømna, og at vi observerer antall krefttilfeller i disse kommunene i en tiårs periode.
- Sannsynligheten for at vi i minst én av kommunene vil oppleve minst tre tilfeller av hjernesvulst i løpet av ett år ved en ren tilfeldighet er:

$$1-(1-0.0006)^{1000}=0.45$$

- Det er sannsynlig at noe så usannsynlig som det som skjedde i Sømna vil skje en gang i blant ved en ren tilfeldighet