

Innføring i medisinsk statistikk

KLH3004 - Høst 2009

Kapittel 4. Stokastisk variabel og Diskret sannsynlighetsfordeling

Harald Johnsen, aug. 2009

1

Felles for a, b, c og d:

I hvert enkelt tilfelle kan vi ikke forutsi hva utfallet av ”eksperimentet” blir. Men det kan angis en *mengde* (utfallsrommet) av mulige enkelutfall slik at hvert enkelt forsøk gir som resultat ett og bare ett av resultatene i mengden. For eksempel i kast med én terning vet vi på forhånd at utfallet må bli ett og bare ett av tallene 1, 2, 3,...,6 der tallet betegner antall øyne opp.

I hvert tilfelle kan en, i alle fall tenke seg, at et eksperimentet kan gjentas under samme betingelser så mange ganger en ønsker, og det er i tillegg underlagt *statistisk regelmessighet*.

Statistisk regelmessighet

Anta et stokastisk forsøk, for eksempel kast med myntstykke, der A er hendelsen ”kron” opp. Myntstykket kastes n ganger og antall ganger som resulterer i ”kron”, kalles n_A . Det viser seg erfaringsmessig at den relative hyppighet av A , n_A/n , har en tendens til å stabilisere seg i nærheten av et bestemt tall p når n gjøres større, og i en ny forsøksserie vil erfaringsmessig n_A/n igjen stabilisere seg rundt p . Den relative hyppigheten av A er et anslag (*estimat*) for *sannsynligheten* for ”kron” i et kast, og det er opplagt at $0 \leq \frac{n_A}{n} \leq 1$.

3

Stokastisk forsøk el. eksperiment

Et *rent teknisk* begrep for en prosess der hensikten er å framskaffe data om hendelser der utfallet kan variere og ikke forutsies med sikkerhet. Hvilke muligheter som foreligger, er imidlertid kjent.

Noen eksempler på stokastiske forsøk:

- a) Kast med pengestykke. Registrerer resultat (utfall) som ”Kroneside opp” eller ”Myntside”opp.
- b) Kast med en trening. Registrerer resultat (utfall) som antall øyne som terningen viser.
- c) Kast med et pengestykke inntil resultatet ”Krone” opptrer første gang. Noterer antall kast (utfall) i kastserien.
- d) Sykdom. Registrerer resultat (utfall) som ”Frisk”, ”Kronisk syk” eller ”Død”.

2

Utfallsrom, enkeltutfall og hendelse

Mengden av alle *enkelutfall* (eng.: elementary outcomes) i et (stokastisk) forsøk kalles *utfallsrommet* (eng.: sample space) og betegnes vanligvis med S . Enkelutfallene blir derfor *elementer* utfallsrommet. I eksemplene ovenfor blir utfallsrommene:

- a) $S = \{\text{Kron, Mynt}\}$
- b) $S = \{1, 2, 3, 4, 5, 6\}$
- c) $S = \{1, 2, 3, 4, 5, 6, 7, 8, \dots\}$
- d) $S = \{\text{Frisk, Kronisk syk, Død}\}$

Merk at elementer i et utfallsrom kan være en målbar størrelse uttrykk som tall, eller være kvalitative som for eksempel utfallet av en sykdom.

4

Utfallsrom – diskret eller kontinuert

I eksemplene a, b og d består utfallsrommet av endelig mange elementer. I eksempel c består utfallsrommet av uendelig, men nummererbart antall elementer. Alle situasjonene ovenfor er eksempler på *diskrete utfallsrom*.

Dersom utfallsrommet består av uendelige mange elementer som ikke lar seg nummerere, sies det å være *kontinuert*. Dette er tilfelle når vi har å gjøre med en kontinuert måleskala som f. eks. høyde, vekt osv.

Hendelse

En *hendelse* (eng.: *event*) består av ett eller flere av elementene i utfallsrommet og tilfredsstiller et eller annet kriterium. F. eks. kan det være av interesse å registrere hvor vidt minst 4 øyne opptrer i et terningkast. Her blir hendelsen mengden $A = \{4, 5, 6\}$, m.a.o. en undermengde av utfallsrommet S . Hendelsen A inntreffer når terningen viser 4, 5 eller 6 øyne.

I tilfeller der utfallsrommet er endelig, vil enhver undermengde av utfallsrommet kunne utgjøre en hendelse.

5

Kontinuert stokastisk variabel (mer om dette i Kap. 5)

En stokastisk variabel med mulige verdier som ikke kan nummereres, kalles en kontinuert stokastisk variabel.

Synonym: Målevariabel, skalavariabel

Noen egenskaper til kontinuert stokastisk variabel:

Sannsynligheten for å innta en bestemt verdi er null.

Sannsynligheter uttrykkes over intervall av enkeltverdier ved hjelp av en *sannsynlighetstetthet*.

7

Stokastisk variabel

Forenklet sagt er en stokastisk variabel, (stor) X , en *funksjon* av utfallet e i et eksperiment. For hvert utfall har $X(e)$ en bestemt, numerisk verdi. dvs. et tall. X kan godt ha samme verdi for ulike utfall, men kun én verdi for hvert enkelt utfall.

Eksempel: Kast med mynt to ganger. Eksempelvis vil enkeltutfallet KM bety ”Kron” i 1. kast og ”Mynt” i andre kast osv.

Utfallsrommet blir da $S = \{KK, KM, MK, MM\}$ med enkeltutfallene $e_1 = KK$, $e_2 = KM$, $e_3 = MK$ og $e_4 = MM$. Dersom antall ”Kron” er hendelsen av interesse, får vi:

$$X(e_1)=2, X(e_2)=1, X(e_3)=1, X(e_4)=0$$

Vanligvis sløyfes argumentet e i uttrykket for en stokastisk variabel. I eksemplet overfor kan X da defineres slik:

$$X = \begin{cases} 0, & e = MM \\ 1, & e = MK \text{ eller } KM \\ 2, & e = KK \end{cases}$$

6

Diskret stokastisk variabel

En stokastisk variabel som bare kan anta et endelig antall verdier, eller høyst et nummerert antall verdier, kalles en diskret stokastisk variabel.

Synonym: Tellevariabel

Noen egenskaper:

Variabelen kan eventuelt ha uendelig mange verdier, men de må kunne ordnes i en sekvens og telles.

Sannsynligheten for en enkelte verdier er gitt av en *punktsannsynlighet*

8

Sannsynlighetsfordeling for diskret stokastisk variabel - punktsannsynlighet

Verdiene som en diskret variabel kan anta og tilhørende sannsynlighet kan uttrykkes ved hjelp av en

Punktsannsynlighet (probability mass function)

Sannsynlighetsfordelingen uttrykkes som $P(X=x)$, der x betegner de mulige verdier for X . Kast med rettferdig terning, $x = 1, 2, 3, \dots, 6$. Sannsynligheten er den samme for alle utfall (uniform fordeling)

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

9

Sannsynlighetsfordeling – eksempel

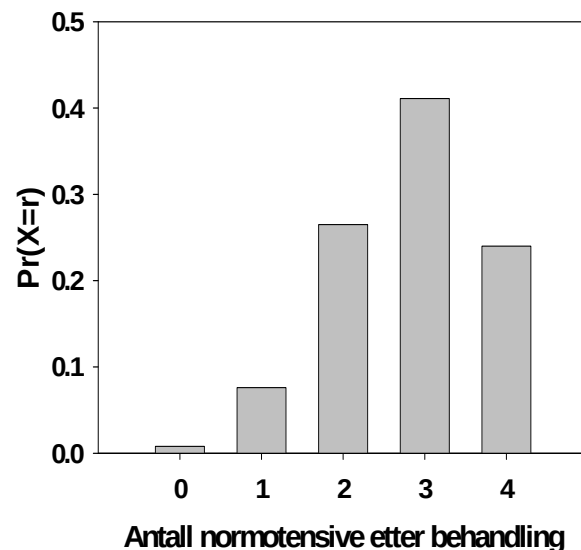
Eks 4.6: Sannsynlighetsfordeling for antall som blir normotensive av 4 hypertonikere som har fått blodtrykkssenkende medikament

x	0	1	2	3	4
$P(X=x)$	0.008	0.076	0.265	0.411	0.240

Merk at sannsynligheten for en bestemt verdi alltid ligger mellom 0 og 1 og at summen av sannsynlighetene for alle mulige verdier av X alltid er lik 1.

10

Sannsynlighetsfordeling



11

Forventningsverdi for diskret stokastisk variabel

Et mål for “tyngdepunktet” i sannsynlighetsfordelingen

Beregnes som et veid gjennomsnitt av hver enkelt verdi x_i av den stokastiske variabelen. Vektene er sannsynligheten for de enkelte verdier den stokastiske variabelen kan ha.

$$\mu = E[X] = x_1 P(X = x_1) + x_2 P(X = x_2) + \dots + x_n P(X = x_n) = \sum_{i=1}^n x_i P(X = x_i)$$

Spesialtilfelle: Ved *uniform* sannsynlighetsfordeling, dvs. sannsynligheten er den samme for alle verdier av X , er forventningen lik den aritmetiske middelvei av de mulige verdier av X .

NB: På engelsk kan “mean” bety både middelvei (mean value) og forventning (expected value), alt etter meningssammenhengen. På norsk skjeller vi mellom disse to begrepene, og de er ikke identiske.

12

Forventningsverdi - eksempel

Kast med terning, X er mulige antall øyne, x er realiseringen av X .

x	1	2	3	4	5	6
$P(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

$$E[X] = \sum_{x=1}^6 x P(X=x)$$

$$= 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot P(X=3)$$

$$+ 4 \cdot P(X=4) + 5 \cdot P(X=5) + 6 \cdot P(X=6)$$

$$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5$$

Forventningen tilsvarende gjennomsnittet av øyne ved (uendelig) mange terningkast. Merk som her at forventningsverdien ikke nødvendigvis tilhører utfallsrommet for det enkelte forsøk.

13

Forventningsverdi - eksempel

EKS 4.9; Forventningsverdien for X = antall normotensive etter blodtrykksbehandling av 4 hypertonicere

x	0	1	2	3	4
$P(X=x)$	0.008	0.076	0.265	0.411	0.240

$$\mu = E[X] = \sum_{x=0}^4 x P(X=x)$$

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

$$= 0 \cdot 0.008 + 1 \cdot 0.076 + 2 \cdot 0.265 + 3 \cdot 0.411 + 4 \cdot 0.240 = 2.80$$

Merk at dette er *populasjonsforventningen*. Den er regnet ut uten å gjøre noe forsøk og observere noen ting som helst.

14

Varians for diskret stokastisk variabel

Variansen $Var(X)$ er et mål på spredningen av de underliggende data rundt forventningsverdien

$$\sigma^2 = Var(X) = \sum_{i=1}^n (x_i - \mu)^2 \cdot P(X = x_i)$$

Merk at variansen er forventningen til $(X - \mu)^2$ og kan derfor også uttrykkes som

$$Var(X) = E[(X - \mu)^2] = E(X^2 - 2\mu X + \mu^2)$$

$$= E(X^2) - 2\mu \cdot E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2$$

$$= E(X^2) - \mu^2 = E(X^2) - (E(X))^2$$

Fordi variansen alltid er ikke-negativ, følger også at

$$E(X^2) \geq (E(X))^2$$

15

Varians - eksempel

EKS 4.9: Variansen til X = antall normotensive etter blodtrykksbehandling av 4 hypertonicere

x	0	1	2	3	4
$P(X=x)$	0.008	0.076	0.265	0.411	0.240

$$E(X) = \mu = 2.8$$

$$Var(X) = \sum_{x=0}^4 (x - \mu)^2 \cdot P(X=x) = E(X^2) - \mu^2$$

$$= 0^2 \cdot P(X=0) + 1^2 \cdot P(X=1) + 2^2 \cdot P(X=2)$$

$$+ 3^2 \cdot P(X=3) + 4^2 \cdot P(X=4) - 2.8^2$$

$$= 1^2 \cdot 0.0076 + 2^2 \cdot 0.265 + 3^2 \cdot 0.411 + 4^2 \cdot 0.240 - 2.8^2 = 0.835$$

16

Standardavvik

Forventningen har samme måleskala (dimensjon) som den stokastiske variable har, f.eks. m (meter). Variansen er imidlertid et kvadratisk uttrykk og får i dette tilfellet dimensjonen m^2 . Den positive roten til variansen får imidlertid samme dimensjon som forventningen og kalles standardavviket:

$$SD(X) \equiv +\sqrt{Var(X)} = +\sqrt{\sigma^2} = \sigma$$

Eksempel 4.9 (hypertonicere):

$$Var(X) = 0.835$$
$$SD(X) = \sqrt{0.835} = 0.914$$

Merk at i likhet med varians er også standardavvik en ikke-negativ størrelse. Det er derfor meningsløst å oppgi standardavviket som for eksempel ± 0.914 .

Sannsynlighetsfordeling og frekvensfordeling

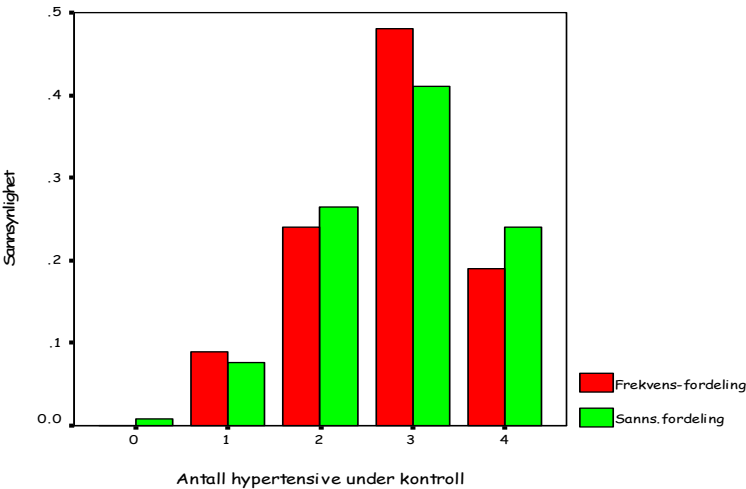
Eksempel 4.8: Sammenlikning av teoretisk sannsynlighetsfordeling med frekvensfordeling fra utvalgsforsøk. Utfall er antall hypertonicere som blir normotensive. Forsøket er basert på 100 leger som hver behandler 4 hypertonicere

Antall hypertonicere under kontroll, x	Sannsynlighetsfordeling P(X=x)	Frekvensfordeling (observert)
0	0.008	0.00 = 0/100
1	0.076	0.09 = 9/100
2	0.265	0.24 = 24/100
3	0.411	0.48 = 48/100
4	0.240	0.19 = 19/100

Her blir forventningen i utvalget (estimert forventning) basert på observasjoner:

$$\sum_{x=0}^4 x P(X = x)$$
$$= 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + 4 \cdot P(X = 4)$$
$$= 0 + 1 \cdot 0.09 + 2 \cdot 0.24 + 3 \cdot 0.48 + 4 \cdot 0.190 = 2.77$$

Sammenlikning av teoretisk sannsynlighetsfordeling og observert frekvensfordeling



Litt kombinatorikk

Har to grupper (mengder), I og II med henholdsvis n_1 og n_2 elementer:

$$I: \{a_1, a_2, a_3, \dots, a_{n_1}\}$$

$$II: \{b_1, b_2, b_3, \dots, b_{n_2}\}$$

Hvor mange ulike par kan dannes med ett element fra hver gruppe? Svar: $n_1 \cdot n_2$

Innfører er tredje gruppe

$$III: \{c_1, c_2, c_3, \dots, c_{n_3}\}$$

Kan da danne $n_1 \cdot n_2 \cdot n_3$ tripler med ett element fra hver gruppe.

Kombinatorikksetning: Ved kombinasjon av ett element fra hver av r grupper kan det dannes $n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_r$ r -tupler der n_i er antall elementer i i -te gruppe.

To eksempler:

- 1) En kafeteria tilbyr en 3 rettes middag med adgang til å velge én forrett blant 3, én hovedrett blant 6, samt én dessert blant 4. Antall mulige middager blir da $3 \cdot 6 \cdot 4 = 72$.
- 2) Tipping i fotball. 12 kamper, hver med 3 mulige utfall. Antall mulige måter å fylle ut en tipperekke på blir $3 \cdot 3 \cdot \dots \cdot 3 = 3^{12} = 531441$.

Ordnet utvalg

I det følgende skal vi alltid foreta utvelgning fra en endelig mengde (populasjon) av elementer som vi betegner {a₁, a₂, a₃,...,a_n}.

Definisjon:
En hvilken som helst ordnet sekvens av *k* elementer fra populasjonen kalles et *ordnet utvalg av størrelse k*.

Et utvalg av størrelse 3 kan for eksempel være {a₁, a₅, a₇} (forutsatt n≥7).

I prinsipp kan vi tenke oss at ett og ett element velges i gangen. For å kunne avgjøre hvor mange ordnede utvalg som er mulige av totalt *n* elementer må vi først presisere om utvelgelsen skjer *med eller uten tilbakelegging* av elementet etter hver trekning.

Noen regler fra kombinatorikken

Vi tenker oss en urne (boks) med *n* kuler nummerert fra 1 til *n*. Så trekkes *s* kuler én og én ad gangen.

Ordnet utvalg med tilbakelegging

Regel 1:
Når k kuler trekkes med tilbakelegging fra en boks med n kuler, vil antall ordnede utvalg være n^k.

Ved første trekning er det *n* muligheter. Kulen legges tilbake, og i neste trekning er det igjen *n* muligheter osv. Dette gjøres *k* ganger. Spesielt kan samme kule trekkes flere ganger. Det totale antall kombinasjoner av *k* merkede kuler blir da *n · n · n · ... · n = n^k*.

Eksempel: I det internasjonale alfabet (A, B, C,...,Z) er det 26 bokstaver. Til et bilnummer med 2 bokstaver kan det dannes 26² = 676 bokstavpar dersom en tillater samme bokstav å opptre begge gangene.

Ordnet utvalg uten tilbakelegging

Regel 2:
Når k kuler trekkes uten tilbakelegging fra en boks med n kuler, vil antall ordnede utvalg være n(n-1)(n-2)·...·(n-k+1).

Ved første trekning er det *n* muligheter, ved neste (*n-1*) muligheter og deretter (*n-2*) osv. Etter nest siste trekning ligger (*n-(k-1)*) kuler igjen i boksen, og antall muligheter i siste (den *k*-te) trekning blir (*n-(k-1)*)=*(n-k+1)*. Det totale antall ordnede kombinasjoner av *k* merkede kuler blir da

$${}_n P_k \equiv n(n-1) \cdot \dots \cdot (n-k+1).$$

Eksempel: Fra det internasjonale alfabet kan det dannes en bokstavkode bestående av 4 ulike bokstaver på ₂₆*P*₄=26(26-1)·...·(26-4+1)=26·25·24·23=358800 måter.

En konsekvens av Regel 2 er at dersom alle *n* kuler, dvs. *k=n*, trekkes ut uten tilbakelegging, kan disse kulene ordnes (*permuteres*) på

$${}_n P_n = n(n-1)(n-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$$

forskjellige måter.

Uttrykket *n(n-1)(n-2)·...·3·2·1* skrives *n!* og leses ”*n* faktultet”. (Vi definerer 0!=1). Dette gir:

Regel 3: Antall permutasjoner av n kuler er n!

Ikke-ordnet utvalg
Dersom *k* kuler trekkes fra *n* kuler *uten* tilbakelegging, fins det _{*n*}*P*_{*k*}=*n(n-1)(n-2)·...·(n-k+1)* ordnede utvalg på *k* kuler. Disse *k* kulene kan etterpå permuteres *k!* ganger, men kulene er de samme. Hvis vi ikke er interesserte i ordningen av kulene, men bare *hvilke* kuler som trekkes ut (slik som i Lotto), blir problemstillingen hvor mange ikke-ordnede utvalg som kan forekomme. Vi kaller dette antallet _{*n*}*C*_{*k*}. Hvert ikke-ordnet utvalg kan permuteres *k!* ganger, og vi får likheten

$${}_n P_k = {}_n C_k \cdot k !$$

Dette gir

$${}_n C_k = \frac{{}_n P_k}{k !} = \frac{n(n-1)(n-2) \cdots (n-k+1)}{k !}$$

$$= \frac{n(n-1)(n-2) \cdots (n-k+1) \cdot (n-k)!}{k! \cdot (n-k)!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

Uttrykket $\binom{n}{k}$, *binomialkoeffisienten*, leses ”n over k” og må ikke forveksles med $\frac{n}{k}$.

Regel 4: Når k kuler trekkes uten tilbakelegging fra en boks med n kuler, vil antall ikke-ordnede utvalg være

$${}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Eksempel: Lottotrekning. Trekker uten tilbakelegging 7 av 34 nummererte kuler. Antall muligheter

$${}_{34}C_7 = \binom{34}{7} = \frac{34!}{7!(34-7)!} = \frac{34!}{7! \cdot 27!} = \frac{34 \cdot 33 \cdot \dots \cdot 28}{7!} = 5379616$$

25

Binomisk modell

Et forsøk som er karakterisert ved

- i) Forsøket er (kan tenkes) sammensatt av n uavhengige enkeltforsøk
- ii) I hvert enkeltforsøk registreres hvorvidt én bestemt hendelse A opptrer eller ikke
- iii) $P(A) = p$ i alle enkeltforsøk (innebærer at $P(A^*) = 1 - p = q$)

kalles en *binomisk forsøksrekke*.

Eksempler på situasjoner som kan beskrives ved binomisk forsøksrekke:

- i) En kaster n ganger med et pengestykke og registrerer antall kast som resulterer i ”Kron”
- ii) n frø såes og en registrerer hvor mange som spirer etter en viss tid
- iii) En undersøker hvor mange enkeltfødslar som resulterer i pike

26

Binomisk fordeling

Vi lar X (stokastisk) være det antall av de n enkeltforsøk som resulterer i hendelsen A og ser på fordelingen til X når n er f. eks. 3. Da vil utfallsrommet S bestå av $2^3 = 8$ enkeltutfall

Utfallsrom (S)	Enkeltutfall	P(e _i)	X
A*A*A*	e ₁	(1-p) ³	0
A*A*A	e ₂	p (1-p) ²	1
A*A A*	e ₃	p (1-p) ²	1
A A*A*	e ₄	p (1-p) ²	1
A A A*	e ₅	p ² (1-p)	2
A A*A	e ₆	p ² (1-p)	2
A *A A	e ₇	p ² (1-p)	2
A A A	e ₈	p ³	3

Sannsynlighetsfordelingen blir nå:

$$P(X=0) = (1-p)^3, P(X=1) = 3(1-p)^2 p, P(X=2) = 3(1-p)p^2, P(X=3) = p^3$$

27

Dette kan sammenfattes som

$$P(X = k) = \binom{3}{k} p^k (1-p)^{3-k}, \quad k=0, 1, 2, 3$$

I det generelle tilfelle med n forsøk blir sannsynligheten for å få k enkeltutfall A av n mulige

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Dette er den *binomiske fordelingen*, og den har en sentral rolle sannsynlighetsregning og statistikk.

I praksis er det sjelden av interesse å beregne sannsynligheter av typen $P(X=k)$, men snarere av typen $P(X \leq k)$ eller $P(X > k)$.

Vi får formelen

$$P(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \quad i=0, 1, 2, \dots, k$$

28

Eksempel: Antar at sannsynligheten for pike i enkeltfødsel er $p = 0.5$. Hva er sannsynligheten for 9 piker i 18 fødsler?
 X =antall piker i $n=18$ fødsler
 $p = P(\text{"pike i enkeltfødsel"})$

Binomisk modell:

$n=18$ uavhengige forsøk

$A = \text{"pike"}$

$P(A)=p=0.5$ i alle forsøk (fødsler)

$X = \text{antall utfall } A \text{ i } n=18 \text{ forsøk}$

$$P(X=9) = \binom{18}{9} 0.5^9 (1-0.5)^{18-9} = \binom{18}{9} 0.5^9 0.5^9 = \binom{18}{9} 0.5^{18} = 0.186$$

Eksempel, forts: En spåkone påstår hun er synsk og at kan forutsi kjønnet i en fødsel. Ser på 18 fødsler og observerer 14 korrekte forutsigelser. Gjetter hun bare?

Spørsmål: Hva er sannsynligheten til å utsi minst 14 korrekte "spådommer" gitt at hun bare gjetter, dvs. $p=0.5$?

$$P(X \geq 14) = P(X=14) + P(X=15) + P(X=16) + P(X=17) + P(X=18)$$

$$= 0.0117 + 0.0031 + 0.0006 + 0.0001 + 0.0000 = 0.0155 \quad (\text{fra tabell})$$

Det er "liten" sannsynlighet for å forutsi korrekt minst 14 av 18 ganger ved ren gjetting, så observasjonene kan tyde på overnaturlige evner.

Forventning og varians til indikatorvariabel

Ser på et forsøk med bare to mulige utfall

$$\text{Indikatorvariabel: } I = \begin{cases} 1 \text{ ved utfall } A, & P(A) = p \\ 0 \text{ ved utfall } A^*, & P(A^*) = 1 - p = q \end{cases}$$

$$E(I) = 1 \cdot P(I=1) + 0 \cdot P(I=0) = p$$

$$E(I^2) = 1^2 \cdot P(I=1) + 0 \cdot P(I=0) = p$$

$$\text{Var}(I) = E(I^2) - (E(I))^2 = p - p^2 = p(1-p) = pq$$

Forventning og varians i binomisk fordeling

$$\text{Antall } A \text{ i } n \text{ forsøk: } X = I_1 + I_2 + \dots + I_n = \sum_{i=1}^n I_i$$

Forventning:

$$E(X) = E(I_1) + E(I_2) + \dots + E(I_n) = p + p + \dots + p = np$$

Uavhengige enkeltforsøk (ingen kovarians mellom I-ene) medfører:

$$\text{Var}(X) = \text{Var}(I_1) + \text{Var}(I_2) + \dots + \text{Var}(I_n) = pq + pq + \dots + pq = npq = np(1-p)$$

Poissonfordeling – ekstrastoff (bonus, ikke pensum)

Noen ganger har man å gjøre med en binomisk forsøksrekke der n er stor og p er svært liten. Forventet antall enkeltutfall er som alltid $\mu = np$. Under betingelsene ovenfor kan det da vises at

$$P(X=k) = \frac{\mu^k}{k!} e^{-\mu}$$

Dette er *poissonfordelingen*, og det kan videre vises at $E(X) = \text{Var}(X) = \mu$

Eksempler på poissonfordelte hendelser:

- i) Utsending partikler fra en radioaktiv kilde over et visst tidsrom
- ii) Antall kollisjoner i en sterkt trafikkert veikryss over et visst tidsrom
- iii) Antall sjeldne celler i et synsfelt under mikroskopet
- iv) Antall tilfeller av en sjelden sykdom i en stor populasjon over en viss tidsperiode