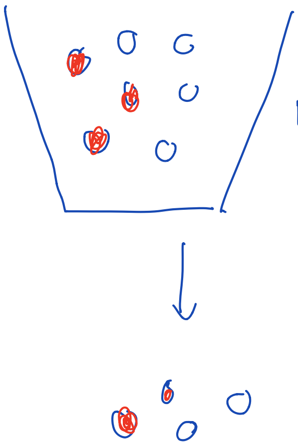
Forventning og varians, hypergeometrisk fordeling.



N kuler hvorav k røde

Utvalg: n kuler hvorav X røde

$X \sim$ hypergeometrisk (N, k, n)

Forventning:

$$EX = \sum_x x P(X=x) = \sum_x x \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \ldots = n\frac{k}{N} \qquad (1) \quad (9.2.\ 2.\ \text{time})$$

Alternativt bevis av (1)

Skriver X som sum av indikatorvariable

$$X = I_1 + I_2 + \ldots + I_n$$

hvor $I_i = 1$ hvis i'te kule vi trekker er rød.

Enhver permutasjon av $I_1, \ldots, I_n$, f.eks.

$$I_3, I_1, I_4, I_2 \quad \text{o.s.v.}$$

må ha samme fordeling (exchangeable random variables)

Dermed må alle $E(I_i)$, $Var(I_i)$ og $Cov(I_i, I_j)$, $i \neq j$ være like.

$$E(I_i) = E(I_1) = P(I_1 = 1) = P(\text{første kule rød}) = \frac{k}{N}.$$

slik at

$$EX = E(I_1 + \ldots + I_n) = EI_1 + EI_2 + \ldots + EI_n$$

$$= n\frac{k}{N} = np \quad \text{hvor } p = \frac{k}{N} \iff Np = k$$

Videre blir

$$Var\, X = Var(I_1 + I_2 + \ldots + I_n)$$

$$= \sum_{i=1}^{n} Var\, I_i + 2\sum_{i=1}^{n-1}\sum_{j=1}^{i} Cov(I_i, I_j) \qquad (2)$$

$$\operatorname{Var} I_i = E(I_i^2) - (E I_i)^2 = p - p^2 = p(1-p)$$

$$\operatorname{Cov}(I_i, I_j) = \operatorname{Cov}(I_1, I_2)$$

$$= E(I_1 I_2) - E I_1 \cdot E I_2$$

$$= P(I_1 = 1 \cap I_2 = 1) - p^2$$

$$= P(I_1 = 1) P(I_2 = 1 \mid I_1 = 1) - p^2$$

$$= p \frac{k-1}{N-1} - p^2 = p\left(\frac{k-1}{N-1} - p\right) < 0$$

Insatt i (2) får vi

$$\operatorname{Var} X = np(1-p) + \cancel{2} \frac{n(n-1)}{\cancel{2}} \cdot \left(p \frac{k-1}{N-1} - p^2\right)$$

$$= np(1-p) \underbrace{\left[1 + \frac{n-1}{1-p}\left(\frac{k-1}{N-1} - p\right)\right]}_{*}$$

$$= np(1-p) \frac{N-n}{N-1} \xrightarrow{N \to \infty} np(1-p)$$

$$\downarrow_{n \to N}$$

$$0$$

$p = \frac{k}{N} \qquad k = pN$

$$\psi = 1 + \frac{n-1}{1-p}\left(\frac{pN-1}{N-1} - p\right)$$

$$= \frac{(1-p)(N-1) + (n-1)(pN-1) - p(N-1)(n-1)}{(1-p)(N-1)}$$

$$= \frac{\cancel{(1-p)}(N-1) + (n-1)(\cancel{pN}-1 - \cancel{pN} + p)}{\cancel{(1-p)}(N-1)}$$

n

-1

$$= \frac{N-1-n+1}{N-1} = \frac{N-n}{N-1}$$

Geometrisk fordeling

$\underbrace{FFF}_{x=4}S \qquad P(S)=p$

$\qquad\qquad\qquad P(F)=1-p$

Punktsannsynlighetsfunksjon

$f(x)=P(X=x)=p(1-p)^{x-1}$ for $x=1,2,\ldots$

Kumulativ fordeling $F(x)=1-(1-p)^x$

Fordelingen er hukommelsesløs fordi

$$P(\underbrace{X-k=x}_{\text{ytterligere antall delforsøk}} \mid \underbrace{X>k}_{\text{ingen suksess i første k bernoulliforsøk}}) = \frac{P(X-k=x \cap X>k)}{P(X>k)}$$

$$= \frac{P(X=k+x)}{P(X>k)} = \frac{p(1-p)^{k+x-1}}{(1-p)^k} = p(1-p)^x$$

d.v.s. $X-k \mid X>k$ har samme fordeling som X.

Vis at $EX = \frac{1}{p}$

Lov om total forventning (ikke pensum)

$$EX = E(X \mid X=1)P(X=1) + E(X \mid X>1)P(X>1)$$

$$= 1 \cdot p + (EX+1)(1-p)$$

$$EX(1-1+p) = p + 1 - p$$

$$EX = \frac{1}{p}$$

$\mathrm{Var}\,X = \frac{1-p}{p^2}$ kan utledes tilsvarende v.h.a. lov om totalvarians.

Alternativ 1 : momentgenererende funksjoner

Alternativ 2 :

$$EX = \sum_{x=1}^{\infty} x f(x)$$

$$= p \sum_{x=1}^{\infty} x (1-p)^{x-1}$$

$$= p \sum_{x=1}^{\infty} \frac{d}{dp}\left(-(1-p)^x \right)$$

$$= -p \frac{d}{dp}\left(\sum_{x=0}^{\infty} (1-p)^x - 1 \right)$$

$$= -p \frac{d}{dp}\left(\frac{1}{1-(1-p)} - 1 \right)$$

$$= -p \frac{d}{dp}\left(\frac{1}{p} - 1 \right)$$

$$= +p \frac{1}{p^2} = \frac{1}{p}$$

Negativ binomisk fordeling

Antall delforsøk for k'te suksess $P(S) = p$.

Eks.: $k = 3$

$$\underbrace{\overbrace{F\,F\,F\,S}^{Y_1}\;\overbrace{F\,F\,F\,F\,S}^{Y_2}\;\overbrace{F\,S}^{Y_3}}_{X = 11}$$

Merk at

$$X = \sum_{i=1}^{k} Y_i$$

hvor $Y_i \sim \text{geom}(p)$

Dermed er

$$EX = E\left(\sum_{i=1}^{k} Y_i\right)$$

$$= \sum_{i=1}^{k} EY_i = k \cdot \frac{1}{p} = \frac{k}{p}$$

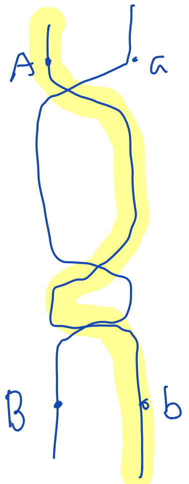
og

$$\text{Var}\, X = \text{Var}\left(\sum_{i=1}^{k} Y_i\right)$$

$$= \sum_{i=1}^{k} \text{Var}\, Y_i = k\,\frac{1-p}{p^2}$$

Poissonprosess

Eksempel fra genetikk: Antall overkrysninger X mellom to loci under meiosen



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

Sannsynlighet for at A rekombineres med b blir

$$P(X \text{ odde})$$

$$= P(X=1 \cup X=3 \cup \ldots)$$

$$= P(X=1) + P(X=3) + \ldots$$

$$= \sum_{x=1,3,5,\ldots} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=1,3,5,\ldots} \frac{\lambda^x}{x!}$$

Taylorutvikling av

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6} + \ldots$$

Taylorutvikling av

$$e^{-\lambda} = 1 - \lambda + \frac{\lambda^2}{2} - \frac{\lambda^3}{6} + \ldots$$

Dermed er

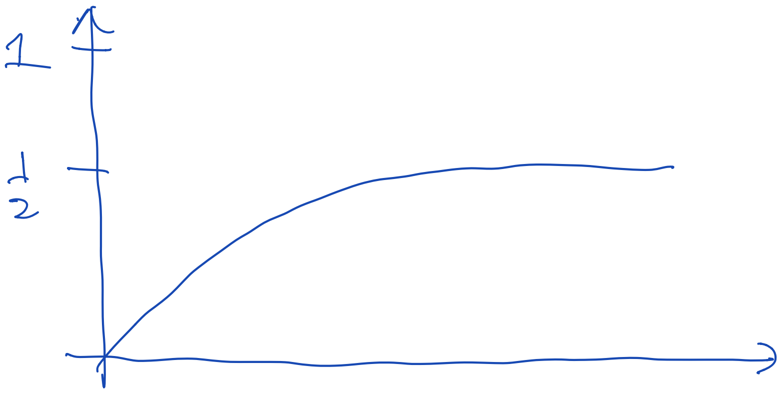
$$e^{\lambda} - e^{-\lambda} = 2\left(\lambda + \frac{\lambda^3}{6} + \frac{\lambda^5}{120} + \ldots\right) = 2 \sum_{x=1,3,5,\ldots} \frac{\lambda^x}{x!}$$

slik at

$$P(X \text{ odde}) = e^{-\lambda} \frac{1}{2}\left(e^{\lambda} - e^{-\lambda}\right)$$

$$= \frac{1}{2}\left(1 - e^{-2\lambda}\right)$$

$P(X \text{ odde}) = P(\text{rekomb})$



λ (avstand målt i forventet antall overkrysninger)