

Kom ihåg: X diskret s.v., $f_X(x)$ t.f.h.n.
och $F_X(x) := P(X \leq x)$ är en t.f.h.n.

OBS: $F_X(x) = P(X \leq x) = \sum_{y: y \leq x} P(X=y) = \sum_{y: y \leq x} f_X(y)$

så kännedom om f_X ger F_X .

P.S.S.



$$P(X=x_3) = F_X\left(\frac{x_3+x_4}{2}\right) - F_X\left(\frac{x_2+x_3}{2}\right)$$

Standardfördelningar

Tal:

Antag att vi gör n oberoende försök där varje försök lyckas m.s. p . Låt $X = \#$ lyckade försök. Bestäm t.f.h.n. för X .

L: Vi skall bestämma $P(X=k) \forall k$.

Vi har att

$$\begin{aligned} P(X=k) &= (\# \text{ försökssekvenser med exakt } k \text{ lyckade}) \\ &\quad \times (\text{sann. för en fix sek, med " " "}) \\ &= \underline{I} \times \underline{II}, \end{aligned}$$

Försök nr:	1	2	3	4	...	n-1	n
Status (l/m):	l	m	m	l	...	l	m
Sann:	p	(1-p)	(1-p)	p	...	p	(1-p)

se kvensen har sann = $p^{\# \text{lyckade}} (1-p)^{\# \text{missl.}}$

Vi får att $\Pi = p^k (1-p)^{n-k}$,

$I = \# \text{sätt att välja } k \text{ ur } n = \binom{n}{k}$

$$\Rightarrow P(X=k) = I \cdot \Pi = \binom{n}{k} p^k (1-p)^{n-k} \text{ för } k=0,1,\dots,n$$

Def: En s.v. X är binomialfördelad med parametrar $n \in \mathbb{N}^+$ och $0 < p < 1$

$(X \sim \text{Bin}(n, p))$ om

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

Ann: 1/ Binomialfördelningen uppkommer

när n oberoende försök genomförs.

Dessa försök kan antingen lyckas eller misslyckas.

2/ Fördelningen av en s.v. X bestäms av dess täthetsfkn och/eller ffn.

3/ Fallet $n=1$ kallas ibland för

Bernoulli-fördelningen ($X \sim \text{Ber}(p)$).

Def: Väntevärdet av en diskret s.v. $\underline{X}$ betecknas $E[\underline{X}]$ och ges av

$$E[\underline{X}] = \sum_{\text{alla } x} x P(\underline{X}=x).$$

Om $h: \mathbb{R} \rightarrow \mathbb{R}$ definierar vi

$$E[h(\underline{X})] = \sum_{\text{alla } x} h(x) P(\underline{X}=x) = \sum_{\text{alla } x} h(x) f_{\underline{X}}(x).$$

Ann: $\sum$

betyder $\sum$
alla värden
på x som
 $\underline{X}$ kan anta

skrivs ibland $\sum_x$.

2/ $E[h(\underline{X})]$ är bara def. för de fner

h och s.v. $\underline{X}$ s.a. $\sum_x |h(x)| f_{\underline{X}}(x) < \infty$.

" Andra konventioner existerar, dvs

$E[\underline{X}] = \pm \infty$ tillåts",

Tal: Låt $\underline{X}$ = resultatet av ett tärningskast. Beräkna $E[\underline{X}]$.

L: Vi har att $\underline{X} \in \{1, 2, \dots, 6\}$ och att

$P(\underline{X}=k) = \frac{1}{6}$ för $k \in \{1, 2, \dots, 6\}$. Vi får att

$$E[\underline{X}] = \sum_{k=1}^6 k P(\underline{X}=k) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5 //$$

Sats (räkne regler för E):

Låt $\underline{X}, \underline{Y}$ vara (diskreta) s.v. och

låt $c \in \mathbb{R}$. Vi har att

$$i) E[c] = c$$

$$ii) E[c\underline{X}] = c E[\underline{X}]$$

$$iii) E[\underline{X} + \underline{Y}] = E[\underline{X}] + E[\underline{Y}]$$

B: i) $E[c] = \{ \underline{X} \text{ är en s.v. } = c \text{ m.s. 1} \}$

$$= \sum_x x f_{\underline{X}}(x) = c \cdot f_{\underline{X}}(c) = c \cdot 1 = c$$

ii) Låt $h(x) = cx$. Vi har att

$$E[c\underline{X}] = E[h(\underline{X})] = \sum_x h(x) P(\underline{X} = x)$$

$$= \sum_x cx P(\underline{X} = x) = c \sum_x x P(\underline{X} = x)$$

$$= c E[\underline{X}]$$

iii) Vi väntar till kap 5. $\square$

Ann: Om $h, g: \mathbb{R} \rightarrow \mathbb{R}$ och $a, b \in \mathbb{R}$

så får vi att

$$E[a h(\underline{X}) + b g(\underline{X})] = a E[h(\underline{X})] + b E[g(\underline{X})]$$

Tal: Låt $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_{100}$ vara resultaten av 100 tärningskast.

Beräkna $E\left[\sum_{k=1}^{100} \underline{X}_k\right]$.

L: Metod 1: Låt $Z = \underline{X}_1 + \underline{X}_2 + \dots + \underline{X}_{100}$.

Hitta $P(Z=100), P(Z=101), \dots, P(Z=600)$ och beräkna sedan

$$E[Z] = \sum_{k=100}^{600} k P(Z=k)$$

Metod 2: Enl. satsen ovan har vi

$$\begin{aligned} \text{att } E[Z] &= E\left[\sum_{k=1}^{100} \underline{X}_k\right] = E\left[\underline{X}_1 + \sum_{k=2}^{100} \underline{X}_k\right] \\ &= E[\underline{X}_1] + E\left[\sum_{k=2}^{100} \underline{X}_k\right] = \dots = \sum_{k=1}^{100} E[\underline{X}_k] \end{aligned}$$

$$= 100 E[\underline{X}_1] = 100 \cdot 3.5 = 350 //$$

Tal: Låt $\underline{X} \sim \text{Bin}(n, p)$. Beräkna $E[\underline{X}]$.

L: Metod 1:

$$E[\underline{X}] = \sum_{k=0}^n k P(\underline{X}=k) = \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k}$$

= ... räkna ... = ?

Metod 2: Låt $\underline{X}_k = \begin{cases} 1 & \text{om exper. } k \text{ lyckat} \\ 0 & \text{annars.} \end{cases}$

$$\Rightarrow \underline{X} = \underline{X}_1 + \dots + \underline{X}_n \sim \text{Bin}(n, p)$$

$$\Rightarrow E[\underline{X}] = E\left[\sum_{k=1}^n \underline{X}_k\right] = \sum_{k=1}^n E[\underline{X}_k] = \sum_{k=1}^n p = np //$$

Def: Variansen $\text{Var}(X)$ av en (diskret) s.v. X definieras av

$$\text{Var}(X) = E[(X - E[X])^2].$$

Standardavvikelsen av X definieras

som $\sqrt{\text{Var}(X)}$.

Anm: i) $\text{Var}(X)$ betecknas ofta σ^2 eller σ_X^2
 ii) $\sqrt{\text{Var}(X)}$ " " σ " σ_X
 iii) $E[X]$ " " μ " μ_X

Sats (räkne regler för $\text{Var}(X)$):

Låt X vara en (diskret) s.v. Vi har att

- i) $\text{Var}(X) = \sum_x (x - E[X])^2$
- ii) $\text{Var}(X) = E[X^2] - E[X]^2$
- iii) $\text{Var}(X) = 0$ om $P(X=c)=1$ för något $c \in \mathbb{R}$.
- iv) $\text{Var}(cX) = c^2 \text{Var}(X) \quad \forall c \in \mathbb{R}$.
- v) Om X, Y är oberoende diskreta s.v.
 $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y),$

Ann: $v)$ är pre-matur. Vi återkommer till denna längre fram.

Intuitivt innebär oberoende att kännedom om $\underline{X}$ inte ger ngn info om $\underline{Y}$ och vice versa. Tänk, $\underline{X}, \underline{Y}$

resultaten av två olika tärningskast.

Lom i kåg!

$$\mathbb{E}[h(X)] = \sum_x h(x) \mathbb{P}(X=x)$$

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Sats (räkner regler för $\text{Var}(X)$):Låt X vara en (diskret) s.v. Vi har att

$$i) \quad \text{Var}(X) = \sum_x (x - \mathbb{E}[X])^2 \mathbb{P}(X=x)$$

$$ii) \quad \text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$iii) \quad \text{Var}(X) = 0 \quad \text{om} \quad \mathbb{P}(X=c)=1 \quad \text{för något } c \in \mathbb{R},$$

$$iv) \quad \text{Var}(cX) = c^2 \text{Var}(X) \quad \forall c \in \mathbb{R},$$

v) Om X, Y är oberoende diskreta s.v.

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y).$$

B:

i) Enligt ovan har vi att

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \underbrace{\mathbb{E}[X]}_{=\mu_X})^2] = \mathbb{E}\{h(X) = (X - \mu_X)^2 = (X - \mathbb{E}[X])^2\} \\ &= \sum_x h(x) \mathbb{P}(X=x) = \sum_x (x - \mu_X)^2 \mathbb{P}(X=x) \end{aligned}$$

ii) Vi har att

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mu_X)^2] = \mathbb{E}[X^2 + \mu_X^2 - 2\mu_X X] \\ &= \mathbb{E}[X^2] + \mathbb{E}[\mu_X^2] + \mathbb{E}[-2\mu_X X] \end{aligned}$$

$$= E[X^2] + \mu_X^2 - 2\mu_X E[X]$$

$$= E[X^2] + \mu_X^2 - 2\mu_X^2 = E[X^2] - \mu_X^2$$

där vi använt räkneregler för väntevärde,

iii) Om $P(X=c)=1 \Rightarrow (X-\mu_X)^2 = (c-c)^2 = 0$

$\Rightarrow \text{Var}(X) = E[(X-\mu_X)^2] = E[0] = 0$

iv) $\text{Var}(cX) = E[(cX - E[cX])^2]$

$$= E[(cX - cE[X])^2] = E[c^2(X - E[X])^2]$$

$$= c^2 E[(X - E[X])^2] = c^2 \text{Var}(X)$$

v) Tar vi senare (kap. 5) $\square$

Fler fördelningar

Tal: Emilia skjuter luftgevär mot

en liten tavla. Varje skott träffar

tavlan oberoende och med sannolikhet

p . Låt $X = \#$ försök som Emilia

behöver t.o.m. första träffen,

hitta t.f.n för X .

2: Vi skall beräkna $P(X=k)$
 för $k=1, 2, \dots$ ("ingen övre gräns").
 För att händelsen $\{X=k\}$ skall
 inträffa måste Emilia ha missat
 $(k-1)$ ggr och sedan träffat försök
 nr k . Dvs

$$\begin{aligned}
 P(X=k) &= P(m, m, m, \dots, m, t) \\
 &= \underbrace{P(m) \cdot P(m) \cdots P(m)}_{\substack{\uparrow \\ \text{ober.} \\ k-1 \text{ ggr}}} \cdot P(t) \\
 &= (1-p)^{k-1} p //
 \end{aligned}$$

Def: X är geometriskt fördelad
 med parameter $0 < p < 1$ ($X \sim \text{Geom}(p)$)

om

$$P(X=k) = (1-p)^{k-1} p \quad \text{för } k=1, 2, \dots$$

Tal: Samma som föregående tal,
 men nu låter vi $X = \# \text{ försök som}$
 Emilia behöver t.o.m. träff nr. x .

2: Om händelsen $\{X=k\}$ inträffar

- måste i) skott nr k vara en träff
 ii) A de $k-1$ skotten innan mest exakt
 $r-1$ vara en träff.

Betrakta en fix sekvens som uppfyller
 i) och ii):

Skott nr:	1	2	3	...	k-1	k
Händelse:	T	M	M		T	T

En sekvens med $r-1$ T och $(k-1)-(r-1)$
 $= k-r$ missar har sann.

$$p^{r-1} (1-p)^{k-r}$$

Vi ser att

$IP(X=k) = (\# \text{ sekr. som ovan}) \cdot (\text{sann. för en}$
 fix sådan sekr.) $\cdot$ sann att sista är träff

$$= \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} p = \binom{k-1}{r-1} p^r (1-p)^{k-r}$$

Def: X är negativt binomialfördelad
 med parametrar r, p ($X \sim \text{Neg Bin}(r, p)$)

om

$$IP(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \text{ för } k=r, r+1, \dots$$

Ann: Om $X \sim \text{geom}(p)$ kan det vara bra xigt att beräkna $E[X]$:

$$E[X] = \sum_{k=1}^{\infty} k P(X=k) = \sum_{k=1}^{\infty} k (1-p)^{k-1} p$$

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

men det finns ett hjälpverktyg!

Momentgenererande funktioner (mgf):

Def:

Låt X vara en s.v. Då kallas

$E[X^n]$ för det n :te momentet för X .

Def: Låt X vara en s.v. med t.f.h.n

$f_X(x)$. Mgf: en för X betecknas $m_X(t)$

och definieras av

$$m_X(t) := E[e^{tX}]$$

Om $m_X(t)$ är ändlig för t i ett

öppet intervall runt 0,

Vad är vitsen med mgf?

Sats: Vi har att

$$m_{\underline{X}}^{(n)}(0) = \left. \frac{d^n}{dt^n} m_{\underline{X}}(t) \right|_{t=0} = \mathbb{E}[\underline{X}^n]$$

för $n \in \mathbb{N}$.

B: (boken sid 63-64 är nogt noggrannare)

$$m_{\underline{X}}^{(n)}(0) = \left. \frac{d^n}{dt^n} \mathbb{E}[e^{t\underline{X}}] \right|_{t=0} \stackrel{(!?) }{=} \mathbb{E} \left[\left. \frac{d^n}{dt^n} e^{t\underline{X}} \right|_{t=0} \right]$$

$$= \mathbb{E}[\underline{X}^n e^{t\underline{X}}] \Big|_{t=0} = \mathbb{E}[\underline{X}^n] \quad \square$$

Tal: Låt $\underline{X} \sim \text{geon}(p)$. Bestäm $m_{\underline{X}}(t)$ och använd denna för att beräkna $\mathbb{E}[\underline{X}]$ och $\text{Var}(\underline{X})$.

L: Vi har att

$$m_{\underline{X}}(t) = \mathbb{E}[e^{t\underline{X}}] = \sum_{k=1}^{\infty} e^{tk} \mathbb{P}(\underline{X}=k) = \sum_{k=1}^{\infty} e^{tk} (1-p)^{k-1} p$$

$$= p e^t \sum_{k=1}^{\infty} e^{t(k-1)} (1-p)^{k-1} = p e^t \sum_{k=1}^{\infty} (e^t (1-p))^{k-1}$$

$$= p e^t \sum_{k=0}^{\infty} (e^t (1-p))^k = \begin{cases} \sum_{k=0}^{\infty} x^k = \frac{1}{1-x} & \text{för } |x| < 1 \end{cases}$$

$$= p e^t \frac{1}{1 - e^t (1-p)}.$$

"kan slippas, ersätt med resultat".

Vi får att

$$m'_X(t) = \frac{p e^t (1 - (1-p)e^t) - p e^t (-(1-p)e^t)}{(1 - (1-p)e^t)^2}$$

$$= \frac{p e^t}{(1 - (1-p)e^t)^2}$$

$$\Rightarrow E[X] = m'_X(0) = \frac{p}{p^2} = \frac{1}{p}.$$

på liknande sätt får vi att

$$E[X^2] = m''_X(0) = \dots = \frac{1 + (1-p)}{p^2}$$

$$\text{s.a. } \text{Var}(X) = E[X^2] - E[X]^2 = \frac{1 + (1-p)}{p^2} - \frac{1}{p^2}$$

$$= \frac{1-p}{p^2} //$$

Anm: Enkelt med dator men inte för hand. Man behöver dock inga "tricks".

Def: S.v. X är Poisson-fördelad med parameter $\lambda > 0$ ($X \sim \text{Poi}(\lambda)$) om

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{för } k=0, 1, 2, \dots$$

Anm: 1)
$$e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \Rightarrow \sum_{k=0}^{\infty} P(X=k)$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^\lambda = 1.$$

2) Poisson-fördelningen uppkommer som ett gränsvärde av binomial-fördelningen.

3) $\text{Poi}(\lambda)$ är vanligt förekommande.

T.ex. har # sönderfall _{tidsenhet} i ett stycke radioaktivt material denna fördelning.

Poissonprocessen (kort och informellt):

Poissonprocessen uppkommer när vi observerar händelser (sönderfall, jordbävningar, vita blodceller) i en mängd (tidsintervall, blodsdroppe).

Om en Poissonprocess har intensitet λ (händelser / tids eller rymdenhet)

så är antalet händelser i mängden
 $S \subset \mathbb{R}^d$ poisson-fördelat med parameter
 $\lambda | S$.

Anm: Boken använder $Poi(k)$ och
 intensitet λ . Alla andra använder
 $Poi(\lambda)$ ("men intensitetsbeteckningen
 varierar").

Tal1: Ett radioaktivt material har intensit.
 1,73 sönderfall/sekund. Vilken fördelning
 har antalet sönderfall under 35 sekunder;

L:

$$Poi(1,73 \cdot 35) = Poi(60,55) //$$

Tal2: Antalet vita blodkroppar är i
 genomsnitt 637/ml blod. Vilken fördelning
 har # vita blodkroppar i 1 cl blod?

L:

$$1 \text{ cl} = 10 \text{ ml.}$$

$$Poi(637 \cdot 10) = Poi(6370) //$$

Anm: ^{1/} Varför är Poisson lämpligt
 här?

^{2/} I tal 2, varför står det i genomsnitt
 (och inte intensitet)?

Tal (69 i boken): Låt $X \sim \text{Poi}(\lambda)$.

Bestäm $m_X(t)$ och använd detta för att bestämma $E[X]$ och $E[X^2]$.

L: Vi har att

$$m_X(t) = E[e^{tX}] = \sum_{k=0}^{\infty} e^{tk} P(X=k)$$

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

$$= e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}$$

Vi ser att $m_X'(t) = \lambda e^t e^{\lambda(e^t - 1)}$

$$\Rightarrow E[X] = m_X'(0) = \lambda e^0 \cdot e^{\lambda(e^0 - 1)} = \lambda \cdot 1 \cdot e^{\lambda(1-1)} \\ = \lambda e^{\lambda \cdot 0} = \lambda //$$

Vidare är

$$m_X''(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda e^t (\lambda e^t e^{\lambda(e^t - 1)})$$

$$\Rightarrow E[X^2] = m_X''(0) = \lambda + \lambda^2 //$$

Anm: $\text{Var}(X) = E[X^2] - E[X]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda //$

Hypergeometrisk fördelning

Betrakta en population med N individer,

Låt r vara antalet i populationen med en viss egenskap (andel med viss gen, andel sjuk, andel barn etc).

Vi väljer ut n individer slumpmässigt och låter $X = \#$ valda med egenskapen,

Vilken t.f.kn har X ?

Vi får att

$$P(X=k)$$



$$= \frac{(\# \text{ sätt välja } k \text{ av } r) \times (\# \text{ sätt välja } n-k \text{ av } N-r)}{\# \text{ sätt välja } n \text{ av } N}$$

$$i) \# \text{ sätt välja } k \text{ ur } Np = \binom{r}{k}$$

$$ii) \quad \text{---} \quad \text{---} \quad n-k \text{ ur } N(1-p) = \binom{N-r}{n-k}$$

$$iii) \quad \text{---} \quad \text{---} \quad n \text{ ur } N = \binom{N}{n}$$

$$\Rightarrow P(X=k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}$$

Def: X är hypergeometriskt fördelad med parametrar N, n, r ($X \sim Hg(N, n, r)$)

$$\text{om } P(X=k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}$$

där $k \leq n$, $k \leq Np$, $k \geq 0$, $n-k \leq N-r$
och $n \leq N$.

Anm: 1/ Detta skrivs ibland $H_g(N, n, p)$

där $p = \frac{r}{N}$ är proportionen som har
egenskapen.

2/ Om r och $N-r$ är stort och n
är litet så blir valen approximativt
oberoende: Säg att $N = 2 \cdot 10^7$ och
 $r = 10^7$. Om vi valt en person som
hade egenskapen är det nu $r-1 = 10^7 - 1 \approx 10^7$
som har egenskapen. D.v.s. om $X \sim H_g(N, n, r)$
med N, n, r som ovan $\Rightarrow X \approx \text{Bin}(n, p)$
där $p = \frac{r}{N}$.

Ex: En medicinsk studie syftar till
att se om aspirin hjälper mot
myokardinfarkt. Man gav 11034 placebo
och 11037 aspirin. Studien bestod av
 $N = 11034 + 11037 = 22071$ individer varav
 $r = 11034$ fick placebo.

Bevisbördan ligger på den som påstår något. Vår utgångspunkt måste vara att aspirin $\hat{e}$ påverkar förekomsten av hjärtattack (H_0). Efter 10 år hade $n=293$ fått hjärtattack varav 189 kom från placebogruppen. Vilken slutsats kan vi dra om H_0 ?

Låt $\bar{X} = \#$ från placebogruppen med h.-a.

Under H_0 är $\bar{X} \sim H_g(22071, 293, 11034)$

Vi har att $P(\bar{X} = 189) = \dots = 0.00000015$.

Är H_0 rimligt?

Tal: Kalle spelar bridge. Vad är sann. att han får en hand med högst en hjärter?

L: Låt $\bar{X} = \#$ hjärter. Vilken fördelning har $\bar{X}$? $N = \#$ kort = 52, $n = \#$ kort som Kalle får = 13 och $r = 13$ är antalet hjärter.

Vi ser att $X \sim H_g(N, n, r) = H_g(52, 13, 13)$.

Vi söker $P(X \leq 1) = P(X=0) + P(X=1)$

$$= \frac{\binom{r}{0} \binom{N-r}{n}}{\binom{N}{n}} + \frac{\binom{r}{1} \binom{N-r}{n-1}}{\binom{N}{n}}$$

$$= \frac{\binom{13}{0} \binom{39}{13}}{\binom{52}{13}} + \frac{\binom{13}{1} \binom{39}{12}}{\binom{52}{13}} //$$
