Diskrete sannsynlighets-fordelinger.

Dekkes av kapittel 5 i læreboka.

Husk: $f(x)$ er punktsannsynligheten til en diskret X dersom:

1. $f(x) \geq 0$
2. $\sum_x f(x) = 1$
3. $f(x) = P(X = x)$

Vi skal nå se på situasjoner der vi har en bestemt form på $f(x)$ ($f(x)$ gitt ved en kjent parametrisk funksjon).

Diskret uniform fordeling (5.2)

Dersom X har k mulige verdier $x_1, \cdots, x_k$ og alle er like sannsynlige sier vi at X er *diskret uniform fordelt*. Har da at

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

Eksempel: Terningkast, $k = 6$, $x_1 = 1, x_2 = 2, \cdots, x_6 = 6$. $\square$

Forventning og varians:

$$\mu = \mathrm{E}(X) = \sum_x x f(x) = \sum_x x \frac{1}{k} = \frac{1}{k} \sum_x x$$

$$\sigma^2 = \mathrm{Var}(X) = \sum_x (x-\mu)^2 f(x) = \frac{1}{k} \sum_x (x-\mu)^2$$

Eksempel: Terningkast.

$$\begin{aligned} \mathrm{E}(X) &= \frac{1}{6} \sum_{x=1}^{6} x = 3.5 \\ \mathrm{Var}(X) &= \frac{1}{6} \sum_{x=1}^{6} (x-3.5)^2 = 2.92 \quad \square \end{aligned}$$

Binomisk fordeling (5.3)

I situasjoner karakterisert ved

- n uavhengige enkeltforsøk.
- Sjekker "suksess" eller ikke "suksess" i hvert enkeltforsøk (om en bestemt hendelse inntreffer eller ikke).
- Sannsynligheten for "suksess" er den samme i alle enkeltforsøk, p.

har vi at $X =$ "antall suksesser i de n forsøkene" er *binomisk fordelt* med parametre n og p. Vi skriver $X \sim B(n, p)$, og har at (se boka for detaljer)

$$f(x) = b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Det kan videre vises at:

$$\mathrm{E}(X) = np \quad \text{og} \quad \mathrm{Var}(X) = np(1-p)$$

Eksempler på ting som typisk er binomisk fordelt:

- Antall komponenter i en test som fungerer etter tid t.
- Antall personer for EU i et tilfeldig utvalg.
- Antall pasienter som får en infeksjon på en sykehusavdeling.
- Antall brønner hvor man finner olje.
- Antall 6ere i n terningkast. $\square$

Multinomisk fordeling (5.3)

Generalisering av binomisk fordeling til situasjoner der man har mer enn to mulige utfall.

- n uavhengige enkeltforsøk.
- Hvert enkeltforsøk har k mulige utfall.
- Sannsynlighetene, $p_1, \ldots, p_k$ for hver av de k utfallene er de samme i alle enkeltforsøk.

Da er simultanfordelingen til $X_1, \ldots, X_k$, der X_j

representerer antall ganger utfall j inntreffer, $j = 1, \ldots, k$, en *multinomisk fordeling* med parametre n og $p_1, \ldots, p_k$. Har da (kan vises):

$$f(x_1, \ldots, x_k; p_1, \ldots, p_k, n) = \frac{n!}{x_1! \cdots x_k!}\, p_1^{x_1} \cdots p_k^{x_k}$$

der $\sum_{i=1}^k x_i = n$ og $\sum_{i=1}^k p_i = 1$. Det kan også vises at

$$\begin{aligned} &\mathrm{E}(X_i) = np_i\ , \quad \mathrm{Var}(X_i) = np_i(1 - p_i)\ , \\ &\mathrm{Cov}(X_i, X_j) = -np_ip_j \text{ for } i \neq j. \end{aligned}$$

Eksempel: En kafe selger tre sorter kaffe, vanlig, cappucino og espresso. De har registrert at 44% av kundene som bestiller kaffe bestiller vanlig kaffe, 32% cappucino og 24% bestiller espresso. 6 personer kommer til kafeen for å ta en kaffe - hva er sannsynligheten for at tre av dem bestiller vanlig kaffe, to av dem cappucino og en av dem espresso? Vi har $p_1 = 0.44$, $p_2 = 0.32$ og $p_3 = 0.24$, og får:

$$f(3, 2, 1) = \frac{6!}{3!2!1!} \cdot 0.44^3 \cdot 0.32^2 \cdot 0.24^1 = \underline{\underline{0.13}} \quad \square$$

Hypergeometrisk fordeling (5.4)

Brukes i situasjoner karakterisert ved

- Totalt N enheter.
- k av de N enhetene er "suksesser" (og dermed $N-k$ "ikke suksesser").
- Trekker ut n enheter uten tilbakelegging.

Har da at $X =$ "antall suksesser blant de n" er *hypergeometrisk fordelt* med parametre N, k og n.

$$f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0, 1, \ldots, \min(k, n),$$

og $\mathrm{E}(X) = \frac{nk}{N}$ og $\mathrm{Var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{k}{N} \cdot (1 - \frac{k}{N})$.

Eksempler på ting som er hypergeometrisk fordelt:

- Antall vinnerlodd man trekker ut i en loddtrekning.
- Antall spar i en korthånd.
- Antall defekte varer man avdekker i en stikkprøve. □

Merk: Trekker man *med tilbakelegging* blir X binomisk fordelt. Dersom $N >> n$ er det liten forskjell på trekning med og uten tilbakelegging - kan da tilnærme hypergeometrisk med binomisk med parametre $n = n$ og $p = k/N$.

Negativt binomisk fordeling (5.5)

Dersom vi har:

- Uavhengige enkeltforsøk inntil k "suksesser" er observert.
- Sjekker "suksess" eller ikke "suksess" i hvert enkeltforsøk.
- Sannsynligheten for "suksess" er den samme i alle enkeltforsøk, p.

og X = "antall forsøk til $k$te suksess" er X *negativt binomisk fordelt* med parametre k og p.

Merk: Ved binomisk ser vi på antall suksesser i et gitt antall forsøk (antall suksesser stokastisk), mens ved negativt binomisk ser vi på antall forsøk som må gjøres for å få et gitt antall suksesser (antall forsøk stokastisk).

Det kan vises ved et tilsvarende resonnement som for binomisk fordeling (se detaljer i boka) at:

$$f(x) = b^*(x;k,p) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \,, \quad x = k, k+1, \ldots$$

og at

$$\mathrm{E}(X) = \frac{k}{p} \,, \quad \mathrm{Var}(X) = k\frac{1-p}{p^2}$$

Eksempler på ting som typisk er negativt binomisk fordelt:

- Antall komponenter man må produsere før man har k feilfrie komponenter.
- Antall hus en støvsugerselger må besøke før han har solgt k støvsugere.
- Antall personer man må spørre for å få k underskrifter til en underskriftskampanje.
- Antall ganger man må kaste en terning før man har fått 6er k ganger. $\square$

Eksempel: Ved en porselensfabrikk gjør man en kvalitetssjekk på alle ferdige produkt. Kun produkt klassifisert som A-vare sendes til vanlig salg. For en type tallerkener har man erfart at 91% av de ferdig produserte tallerkenene klassifiseres til A-vare. Tallerkenene pakkes i esker med 6 stk.

Hva er sannsynligheten for at man må sjekkes 8 eller flere tallerker før man har nok tallerker klassifisert til A-vare til å fylle en eske? Og hva er forventet antall tallerker man må sjekke?

$$
\begin{aligned}
P(X \geq 8) &= 1 - [P(X = 6) + P(X = 7)] \\
&= 1 - \binom{6-1}{6-1} 0.91^6 (1 - 0.91)^{6-6} \\
&\quad - \binom{7-1}{6-1} 0.91^6 (1 - 0.91)^{7-6} \\
&= 1 - 0.5678 - 0.3066 = \underline{\underline{0.126}} \\
\mathrm{E}(X) &= \frac{k}{p} = \frac{6}{0.91} = \underline{\underline{6.6}} \quad \square
\end{aligned}
$$

Geometrisk fordeling (5.5)

Spesialtilfelle av negativt binomisk der man ser på antall forsøk til *første* suksess, dvs $\underline{k = 1}$.

F.eks. antall prøveboringer man må gjøre før første gang man finner olje.

Poisson-fordeling (5.6)

Typisk for Poisson-forsøk er at vi ser på antall forekomster av en hendelse innenfor et tidsintervall eller et spesifisert område.

Eksempler på ting som typisk er Poisson-fordelt:

- Antall telefoner til et kontor i løpet av en time.
- Antall visninger av en webside per time.
- Antall trafikkulykker på en veistrekning i løpet av ett år.
- Antall utrykninger fra en brannstasjon per uke.
- Antall trykkfeil per side i en bok.
- Antall trær per mål i en skog. $\square$

Forekomsten av hendelser følger en *Poisson-prosess* dersom:

1. Antallene hendelser i ikke-overlappende intervaller er uavhengige.

2. For ethvert lite intervall $[s, s + \Delta s]$ er $P(\text{en hendelse i } [s, s + \Delta s]) \approx \lambda \Delta s$.

3. $P(\text{to eller flere hendelser i } [s, s + \Delta s]) \approx 0$.

Dersom dette er oppfylt er X="antall hendelser i $[0, t]$" Poisson-fordelt med parameter $\mu = \lambda t$. Det kan da vises at:

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

og at $\mathrm{E}(X) = \lambda t$ og $\mathrm{Var}(X) = \lambda t$.

Parameteren λ representerer forventet antall hendelser per enhet og kalles *intensiteten* til prosessen.

Vi skal lære mer om Poisson-prosesser senere i kurset.

Poisson-fordeling som tilnærming til binomisk fordeling.

Når n er stor og p nær 0 (eller 1) kan binomisk fordeling tilnærmes med Poisson-fordeling med $\mu = np$. (Nyttig fordi det gir enklere regning).

Eksempel: Antall vinnerrekker i lotto. Sanns. for at en rekke gir 7 rette er $p = 1/\binom{34}{7} =$ 1/5379616. Antall innleverte rekker en uke er ca $n = 23500000$. La X="antall rekker som har 7 rette". Vi har da at $X \sim B(n,p) \approx$Poisson(μ), der $\mu = np = 23500000/5379616 = 4.37$. Dvs:

$$\begin{aligned} f(0) = P(X=0) &\approx \frac{4.37^0}{0!}e^{-4.37} = 0.013 \\ f(1) = P(X=1) &\approx \frac{4.37^1}{1!}e^{-4.37} = 0.055 \end{aligned}$$

x	0	1	2	3	4
$f(x)$	0.013	0.055	0.121	0.176	0.192

x	5	6	7	8	9
$f(x)$	0.168	0.122	0.076	0.042	0.020