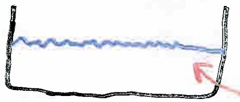


pH $= \mu$

$\bar{X}$ = målt pH-verdi

$E(\bar{X}) = \mu$, $Var(\bar{X}) = 0.01$

Eksempel: Vi skal bruke et måleinstrument til å måle pH-verdien til en løsning. Vi vet at målinger med dette instrumentet gir resultater som har en fordeling med forventningsverdi μ (sann pH-verdi) og varians $\sigma^2 = 0.01$. I stedet for å bare gjøre en måling bestemmer vi oss for å gjøre 10 målinger og bruke gjennomsnittet av disse 10 målingene som vårt resultat. Hva blir forventningsverdi og varians til dette gjennomsnittet? La $\bar{X} = \frac{1}{10}(X_1 + X_2 + \cdots + X_{10}) = \frac{1}{10}X_1 + \cdots + \frac{1}{10}X_{10}$ være gjennomsnittet av de 10 målingene. Vi antar at de 10 målingene er uavhengige.

$a = \frac{1}{10}$

$b = 0$

$Y = \bar{X} = \frac{1}{10} \cdot X_1 + \frac{1}{10} X_2 + \cdots + \frac{1}{10} X_{10}$

$$\begin{aligned}
\mathrm{E}(\bar{X}) &= \frac{1}{10}\mathrm{E}(X_1) + \cdots + \frac{1}{10}\mathrm{E}(X_{10}) \\
&= \frac{1}{10}\mu + \cdots + \frac{1}{10}\mu = \mu \\
\mathrm{Var}(\bar{X}) &= (\frac{1}{10})^2\mathrm{Var}(X_1) + \cdots + (\frac{1}{10})^2\mathrm{Var}(X_{10}) \\
&= (\frac{1}{10})^2\sigma^2 + \cdots + (\frac{1}{10})^2\sigma^2 \\
&= \frac{\sigma^2}{10} = \frac{0.01}{10} = 0.001
\end{aligned}$$

Dvs, gjennomsnittet har samme forventningsverdi som enkeltmålingene men lavere varians. Bruker vi tommelfingerregelen om at omtrent 95% av resultatene vil havne i intervallet $\mu \pm 2\sigma$ får vi at en enkeltmåling har ca 95% sannsynlighet for å havne innenfor $[\mu - 2 \cdot \sqrt{0.01}, \mu + 2 \cdot \sqrt{0.01}] = [\mu - 0.2, \mu + 0.2]$, mens gjennomsnittet av 10 målinger har ca 95% sannsynlighet for å havne innenfor $[\mu - 2 \cdot \sqrt{0.001}, \mu + 2 \cdot \sqrt{0.001}] = [\mu - 0.06, \mu + 0.06]$.

Enkeltmåling

Gj. sn. av 10 målinger

Generelt: Usikkerhet avtar med $\sqrt{n}$.
Ekv. informasjon/presisjon øker med $\sqrt{n}$.

Generelt: Når $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ er $E(\bar{X}) = \mu$

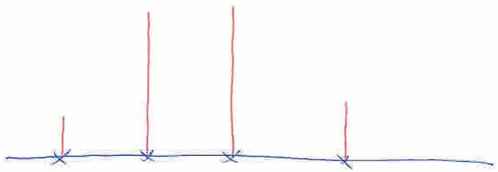
$Var(\bar{X}) = \frac{\sigma^2}{n}$

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).

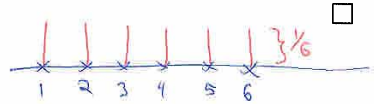
Diskret uniform fordeling (5.2)

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.

$$\mathrm{E}(X) = \frac{1}{6} \sum_{x=1}^{6} x = 3.5$$

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

Binomisk fordeling (5.3)

I situasjoner karakterisert ved

NB!

- n uavhengige enkeltforsøk. → Bestemt antall
- 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)

"fordelt"

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

Det kan videre vises at:

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

$E(X) = \sum_x x \cdot f(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = \ldots = np$

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. □

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

E-eksempel: EU → for, → mot, → vet ikke   3 mulige utfall 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

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

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 ($n=6$) 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

NB!

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

k | $N-k$

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. 13 | 39
- Antall defekte varer man avdekker i en stikkprøve. □

"mye større enn"

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 $p = k/N$.

OK når $N > 10 \cdot n$

Negativt binomisk fordeling (5.5)

Dersom vi har:

NB!

- 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

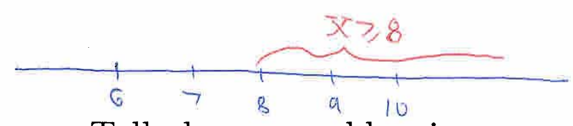
$$\mathrm{E}(X) = \frac{k}{p}\ , \quad \mathrm{Var}(X) = k\frac{1-p}{p^2}$$

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. □

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

$\hookrightarrow p = 0.91$



klassifiseres til A-vare. Tallerkenene pakkes i esker med 6 stk.

$\hookrightarrow k = 6$

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?

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

$$\mathrm{E}(X) = \frac{k}{p} = \frac{6}{0.91} = \underline{\underline{6.6}} \quad \square$$

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.

$\binom{n}{r} = \frac{n!}{r!(n-r)!} = {}_nC_r$ (HP30S: Under PRB)