Sannsynlighetsfordelinger:

Binomisk Fordeling:

1. n uavhengige delforsøk
2. Suksess eller ikke
3. P(A)=p i alle forsøk

- X = Antall ganger A intreffer på n forsøk.
- $X \sim binom(n, p)$
- $f(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x},\ x = 0, 1, 2, \ldots, n$
- $P(X \le x) = \sum_{k=0}^{x} P(X = k)$
- $E(X) = np \quad Var(X) = np(1-p)$

Hypergeometrisk:

1. Populasjon med N elementer.
2. k av disse regnes som "Suksess", $N - k$ som fiasko
3. Trekker n elementer uten tilbakelegging

- X, antallet suksesser.
- $f(x) = \frac{\binom{k}{x} \cdot \binom{N-k}{n-x}}{\binom{N}{n}}$
- $E(X) = np \quad Var(X) = np(1-p)\frac{N-n}{N-1},\ p = k/N$

Negativ-Binomisk:

X er antall forsøk en må gjøre for at en hendelse A skal intreffe k ganger

- $f(x) = \binom{x-1}{k-1} \cdot p^x (1-p)^{x-k},\ x = k, k+1, k+2, \ldots$
- $E(X) = k/p \quad Var(x) = k \cdot \frac{1-p}{p^2}$

Geometrisk:

X er antall forsøk en må gjøre for at hendelsen A intreffer første gang.

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

Geometrisk fordeling er minneløs!

Poisson:

Antall forekomster av hendelsen A er Poisson-fordelt hvis:

1. Antallet av A i disjunkte tidsintervall er uavhengige
2. Forventa antall av A er konstant lik λ(raten) per tidsenhet
3. Kan ikke få to forekomster samtidig

- X = antall forekomster av A i et tidsrom t
- $f(x) = \frac{(\lambda t)^x \cdot e^{-\lambda t}}{x!},\ x = 0, 1, 2, \ldots$
- $E(X) = \lambda t \quad Var(X) = \lambda t$
- $P(X \le x) = \sum_{k=0}^{x} P(X = k)$
- Ventetida til hendelse k er gammafordelt med $\alpha = k$ og $\beta = 1/\lambda$
- Ventetida til første hendelse,og mellom etterfølgende hendelser, er eksponensialfordelt

Uniform fordeling:

En kontinuerlig uniformt fordelt variabel, har samme sannsynlighet for alle verdier innen et intervall. Generelt har vi tetthetsfunksjonen: $f(x) = \begin{cases} \frac{1}{B-A},\ A \le x \le B \\ 0,\ \text{ellers} \end{cases}$

- $E(X) = \frac{A+B}{2} \quad Var(X) = \frac{(A-B)^2}{12}$

Gammafordeling:

En kontinuerlig variabel X er gammfordelt med parameter $\alpha > 0$ og $\beta > 0$ dersom tetthetsfunksjonen er gitt ved: $f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}},\ x > 0 \\ 0,\ \text{ellers} \end{cases}$

- $E(X) = \alpha\beta \quad Var(X) = \alpha\beta^2$

Eksponensialfordeling:

- $f(x; \beta) = \begin{cases} \frac{1}{\beta} e^{-\frac{x}{\beta}},\ x > 0 \\ 0,\ ellers \end{cases}$
- $E(X) = \beta \quad Var(X) = \beta^2$

Eksponensialfordelinga er minneløs!

Normalfordeling:

- $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- $P(a \le X \le b) = \int_a^b f(x)dx$

Standard normalfordeling:

- Alle normalfordelinger kan skrives som Standard normalfordeling
- $Z = \frac{X-\mu}{\sigma}$
- $F(x) = F(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Anta at $X_1, X_2, \ldots, X_n$ er uavhengige og normalfordelt. Da er: $Y = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n$ Være normalfordelt med:

- $E(Y) = \sum_{i=1}^{n} \alpha_i \mu_i \quad Var(Y) = \sum_{i=1}^{n} \alpha_i^2 \sigma_i^2$

Inferens:

QQ-Plot:

- Plotter observasjoner mot teoretiske("ideelle") observasjoner fra en aktuell fordeling.
- Teoretiske observasjoner er gitt ved invers kumulativ fordeling av jevnt spredte Sannsynlighetsfordelinger mellom 0 og 1.
- Om antatt fordeling stemmer skal plottet gi tilnermet rett linje.

Estimering:

Viktige estimatoregenskaper:

- En punktestimator Θ for en paramaeter θ er forventningsrett hvis $E(\Theta) = \theta$
- Variansen $Var(\Theta)$ burde synke med økende antall observasjoner
- Om en har to ulike estimatorer, så er den estimatoren med minst varians den mest effektive estimatoren.

Vanlige estimatorer:

Alle estimatorene vist til her er forventningsrett.

- μ: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad E(\overline{X}) = \mu \quad Var(\overline{X}) = \frac{\sigma^2}{n}$
- σ^2: $S^2 = \frac{1}{1-n}\sum_{i=1}^{n}(X_i - \overline{X})^2\ E(S^2) = \sigma^2 \quad Var(S^2) = \frac{2\sigma^4}{n-1}$
- p: $\hat{p} = \frac{X}{n} \quad E(\hat{p}) = p \quad Var(\hat{p}) = \frac{p(1-p)}{p}$ Binomisk
- $\mu_1 - \mu_2$: $\overline{X_1} - \overline{X_2} \quad Var(\overline{X_1} - \overline{X_2}) = \frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}$
- $\frac{\sigma_1^2}{\sigma_2^2}$: $\frac{S_1^2}{S_2^2}$
- $p_1 - p_2$: $\hat{p_1} - \hat{p_2}$, Binomisk
- $\mu_D : \overline{D}$

Utvalgsfordelinger:

$\overline{X}$, Z:

$\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)\ Z = \frac{\overline{X} - E(\overline{X})}{\sqrt{Var\overline{X}}} = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ Selv om populasjonen ikke er normalfordelt gjelder dette når $n \to \infty$.Regner vanligvis tilnærmina for god når $n > 30$

T

Hvis ukjent varians: $T = \frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ Dette gjelder tilnærmet andre fordelinger som har klokkeliknende form.

$\mathbf{S^2}$:

Forutsatt normalfordeling: $\frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2_{n-1}$

$\hat{\mathbf{p}}$:

Binomisk forsøk med sannsynlighet p, gitt at n er stor nok: $Z = \frac{\hat{p}-E(\hat{p})}{\sqrt{Var(\hat{p})}} = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$

$\hat{X}_1 - \hat{X}_2$:

Kjent varians:

- $\hat{X}_1 - \hat{X}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$
- $Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

Ukjent varians:

- $\sigma_1^2 = \sigma_2^2$: $T = \frac{(\overline{X_1} - \overline{X_2} - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
- $\sigma_1^2 \neq \sigma_2^2$: $T' = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_v$, $v = \frac{(S_1^2/n_1 + s_2^2/n_2)^2}{\frac{(S_1^2 n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2-1}}$

F

Fra to uavhengige NF utvalg: $F = \frac{S_1^2\sigma_2^2}{S_2^2\sigma_1^2} \sim F_{n_1-1, n_2-1}$

$\hat{\mathbf{p_1}} - \hat{\mathbf{p_2}}$

$Z = \frac{(\hat{p_1} - \hat{p_2}) - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \sim N(0,1)$

$\overline{\mathbf{D}}$

Gitt normalfordeling: $T = \frac{\overline{D} - \mu_D}{S_D/\sqrt{n}} \sim t_{n-1}$

Fra utvalgsfordelinger kan en utlede testobservatorer og konfidensintervaller!

Utlede Konfidensintervall:

Anta at vi har $X_1, X_2, \ldots, X_n$ stokastiske variabler, hvor sannsynlighetsfordelingen til disse inneholder en ukjent parameter θ. Anta også at vi har observasjoner $x_1, x_2, \ldots, x_n$. Har lyst å bruke disse for å finne et $100(1-\alpha)\%$ konfidensintervall:

1. Bestem en stokastiske variabel $Z = h(X_1, X_2, \ldots, X_n, \theta)$ som følger en kjent fordeling. Altså finn utvalgsfordelingen for paramaeteren θ
2. Finn kvantilene $Z_{\frac{\alpha}{2}}$ og $Z_{1-\frac{\alpha}{2}}$. Da har en at: $P(Z_{1-\frac{\alpha}{2}} \leq h(X_1, X_2, \ldots, X_n, \theta) \leq Z_{\frac{\alpha}{2}})$
3. Da er løsningen på ulikhetene $Z_{1-\frac{\alpha}{2}} \leq h(X_1, X_2, \ldots, X_n, \theta)$ og $Z_{1-\frac{\alpha}{2}} \geq h(X_1, X_2, \ldots, X_n, \theta)$ konfidensintervallet.