1 Kombinasjoner(Uavhengig av Rekkefølge)

$nCk = \binom{n}{k}$ - uten tilbakelegging

$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$ - med tilbakelegging

2 Permutasjoner(Avhengig av Rekkefølge)

$nPk = k!\binom{n}{k}$ - uten tilbakelegging

n^k - med tilbakelegging

3 Sannsynligheter

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A|B) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$ - Bayes Regel

4 Forventning, Varians og Covarians

$\sigma^2 = E[X^2] - E[X]^2$

$Cov(X,Y) = E[XY] - \mu_x\mu_y \quad Cov(aX_1 + X_2, Y) = aCov(X_1, Y) + Cov(X_2, Y)$

$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

5 Diskrete Fordelinger

Binomisk - $\binom{n}{k}p^x(1-p)^{n-x}$ - Antall gunstige med tilbakelegging

1. n uavhengige forsøk
2. Enten intreffer A, eller ikke.
3. $P(A)$ er lik i alle forsøkene

Hypergeometrisk - $\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$ - Antall gunstige uten tilbakelegging

k = tot gunstige, N = tot mulige, n = antall "trukket", x = antall "vunnet"

Negativ Binomisk - $\binom{x-1}{k-1}p^k(1-p)^{x-k}$ - P for å få k'te gunstige i x'te forsøk

Merk: Geometrisk er Negativ Binomisk med x=1.

Poisson - $\frac{\mu^x}{x!}e^{-\mu}$ - Antall forekomster i tidsintervall

1. Ikke overlappende tidsinterval har uavhengig antall forekomster
2. Det finnes en konstant λ slik at $P(X = 1) \approx \lambda\, dt$ i [t,t+dt]
3. $P(X \geq 2) \approx 0$ i [t,t+dt]

6 Kontinuerlige Fordelinger

Normalfordeling - $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, $E[X] = \mu$, $Var[X] = \sigma^2$

1. Binomisk fordeling opptrer som $N(0,1)$ når $np, n(1-p) \geq 5$
2. t-fordelingen (og de fleste andre) opptrer som $N(0,1)$ når $n \geq 30$
3. Generelt opptrer alle fordelinger seg som $N(0,1)$ når $n \to \infty$

Eksponensialfordeling - $\lambda e^{-\lambda t}$, $\lambda, t > 0$ - "glemsk"

Gammafordeling - $\frac{1}{\beta^\alpha \Gamma(\alpha)}x^{\alpha-1}e^{-\frac{x}{\beta}}$, $x > 0$

1. $\alpha = 1$ er eksponsialfordelingen med $\lambda = \frac{1}{\beta}$
2. $\alpha = n$ er fordelingen til ventetiden til n-te hendelse i en Poisson-prosess

Kji-kvadrat-fordeling - $\frac{1}{2^{v/2}\Gamma(v/2)}x^{\frac{v}{2}-1}e^{-\frac{x}{2}}$, $x \geq 0$, $E[X] = v$, $Var[X] = 2v$

Fordelingen til summen av to kji-kvadrat-fordelte variable tilsvarer fordelingen til summen av frihetsgradene

t-fordeling - $\frac{Z}{\sqrt{\frac{V}{\nu}}}$, $Z \sim N(0,1)$, $V \sim \chi^2(\nu)$, Z og V uavhengige.

$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ er t-fordelt med n-1 frihetsgrader der;

S = empirisk standardavvik $\quad X \sim N(0,1)$

7 Funksjoner

Variabelskifte - $Y = u(X)$, $X = u^{-1}(Y) = w(Y)$

Y har fordeling: $g(y) = f(w(y))\, w'(y)$

Momentgenererende - $M_X(t) = E[e^{tX}]$, $E[X^k] = M_X^{(k)}(0)$

$M_{aX_1+bX_2}(t) = M_{X_1}(at) \cdot M_{X_2}(bt)$

Ordningsvariable(sorterte) - får $\{Y_1 \leq Y_2 \leq ... \leq Y_n\}$ fra $\{X_1, X_2, ..., X_n\}$

$f_{Y_1}(y) = n(1 - F_X(y))^{n-1} f_X(y)$

$f_{Y_k}(y) = \frac{n!}{(k-1)!(n-k)!} f_X(y) F_X(Y)^{k-1}(1 - F_X(y))^{n-k}$

8 Estimering

Forventningsrett: $E[\hat{\theta}] = \theta$, Minste $Var[\hat{\theta}]$ gir beste estimator

Middelverdi: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, Uavhengig av fordelingen til X_i

Normalfordeling - $\mu_{\bar{X}} = E[\bar{X}] = \frac{1}{n}E[\sum_{i=1}^{n} X_i] = E[X] = \mu$

$$\sigma^2_{\bar{X}} = \frac{1}{n^2}Var[\sum_{i=1}^{n} X_i] = \frac{n}{n^2}Var[X] = \frac{\sigma^2}{n}$$

Kjenner vi σ^2 :

Z=$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Kjenner vi ikke σ^2 bruker vi estimatoren S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{n\sum_{i=1}^{n}X_i^2 - (\sum_{i=1}^{n}X_i)^2}{n(n-1)}$$

$$V = \frac{n-1}{\sigma^2}S^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 \sim \chi^2_{n-1}$$

Dersom $n > 30$ kan vi bruke estimeringen: $Z = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim N(0,1)$

Dersom $n < 30$ må vi bruke $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim T_{n-1}$

Sannsynlighetsmaksimeringsestimator(SME):

$$L(X_1, X_2, ..., X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

$$ln(L(X_1, X_2, ..., X_n; \theta)) = \sum_{i=1}^{n} f(X_i; \theta)$$

$$\frac{\partial ln(L(X_1, X_2, ..., X_n; \theta))}{\partial \theta} = \frac{\partial}{\partial \theta}(\sum_{i=1}^{n} f(X_i; \theta)) = 0 \text{ løser for } \theta \text{ gir } \hat{\theta} = \text{SME}$$

9 Konfidensintervall

Normalfordeling - $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$

t-fordeling - $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$

χ^2-fordeling - $P(\chi^2_{1-\alpha/2,v} < V < \chi^2_{\alpha/2,v}) = 1 - \alpha$

Prediksjonsintervall (forutse verdien på X, med kjent σ, ukjent μ):

$$Y = X - \bar{X}, \quad E[Y] = 0, \quad Var[Y] = (1 + \frac{1}{n})\sigma^2, \quad Z = \frac{X - \bar{X}}{\sigma(1+\frac{1}{n})} \sim N(0,1)$$

Estimering av forskjellen mellom to middelverider($\bar{X}_1 - \bar{X}_2$):

$$\bar{X}_1 - \bar{X}_2 \sim N(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

Hvis σ_1 & σ_2 er ukjente, men like ($\sigma_1 = \sigma_2 = \sigma$), estimerer vi $\sigma^2 = S_P^2$:

$$S_P^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}, \quad \text{nå fordelt med } \bar{X}_1 - \bar{X}_2 \sim T_{\frac{\alpha}{2}, n_1+n_2-2}$$

Hvis σ_1 & σ_2 er ukjente og ulike ($\sigma_1 \neq \sigma_2$), må vi bruke $S_1 for \sigma_1$, $S_2 for \sigma_2$ og:

$$\nu = \frac{(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2})^2}{\frac{(\frac{s_1^2}{n_1})^2}{n_1-1} + \frac{(\frac{s_2^2}{n_2})^2}{n_2-1}} \text{ som antall frihetsgrader i t-fordelingen(rund av til heltall)}$$

Estimering av p i Binomial Fordeling(p^* som observator):

$$Z_{obs} = \frac{p^* - p}{\frac{p(1-p)}{n}} \quad n \text{ stor nok til å bruke } Z_{obs} \sim N(0,1)$$

Estimering av forskjell mellom to binomialfordelinger:

$$p_1^* - p_2^* \sim N(p_1 - p_2, \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}})$$

Estimering av Varians: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

10 Hypotesetesting

Type 1 feil: $\alpha = P(forkaste\, H_0 | H_0\, er\, rett)$ = signifikansnivå

Type 2 feil: $\beta = P(ikke\, forkaste\, H_0 | H_1\, er\, rett)$, (1-$\beta$)= styrken

P-verdi: α-verdi slik at vi akkurat forkaster H_0

Test av forventning(μ) til normalfordeling:

kjent σ, tosidig test $\sim H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$;

$$P(\mu_0 - \frac{\sigma}{\sqrt{n}} z_{\alpha/2} < \bar{X} < \mu_0 + \frac{\sigma}{\sqrt{n}} z_{\alpha/2}) = 1 - \alpha$$

kjent σ, énsidig test $\sim H_0 : \mu \leq \mu_0$, $H_1 : \mu > \mu_0$;

$$P(\bar{X} < \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha) = 1 - \alpha$$

ukjent σ, tosidig test $\sim H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$;

$$P(\mu_0 - \frac{S}{\sqrt{n}} t_{\alpha/2,n-1} < \bar{X} < \mu_0 + \frac{S}{\sqrt{n}} t_{\alpha/2,n-1}) = 1 - \alpha$$

To-utvalg; Test av to forventninger$\sim H_0 : \mu_1 - \mu_2 = d_0$, $H_1 : \mu_1 - \mu_2 \neq d_0$:

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad T = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Testing av Variansen til en normalfordeling:

$H_0 : \sigma^2 = \sigma_0^2$, $H_1 : \sigma^2 \neq \sigma_0^2$;

$$W = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

$$P(\frac{\chi^2_{1-\frac{\alpha}{2},n-1}}{n-1} \sigma_0^2 < S^2 < \frac{\chi^2_{\frac{\alpha}{2},n-1}}{n-1} \sigma_0^2) = 1 - \alpha$$

11 Lineær Regresjon

$$Y = \beta_0 + \beta_1 X + \epsilon, \quad \epsilon \sim N(0, \sigma^2), \quad Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$$

$$\hat{y} = b_0 + b_1 x$$

$$b_0 = \bar{y} - b\bar{x} \qquad E[b_0] = \beta_0 \qquad Var[b_0] = \frac{\sigma^2 \sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad E[b_1] = \beta_1 \qquad Var[b_1] = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - n\bar{X} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y})^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - (b_0 - b_1 x_i))^2$$

$$V = \frac{(n-2)s^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (Y_i - b_0 - b_1 x_i)^2 \sim \chi^2_{n-2}$$

Hypotese-testing: $T = \frac{b_1 - \beta_1}{S/\sqrt{S_{xx}}} \sim t_{n-2}$ $\quad$ $S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$

Prediksjonsintervall: y_0 = fremtidig verdi av Y_0

$$T = \frac{\hat{Y}_0 - Y_0}{s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})}}} \sim t_{n-2}$$