1. Mathematatik, Folgen und Reihen, Potenzreihen

2. Mathematik, Differential- und Integralrechnung

3. Binomial- und Hypergeometrische Verteilung, Approximationen

4. Normalverteilung, Berechnen von Wahrscheinlichkeiten

5. Chi-Quadrat-Verteilung, Erwartungswert, Varianz

6. Nullhypothese, Alternativhypothese, Tests, p-Werte

7. Lineare Algebra, Matrizenrechnung

8. t-Test und Varianzanalyse
Produkt von Matrizen

9. Eigenwerte, Eigenvektoren

3a. Bsp. $X \sim Bi(n,p), \quad p = 0.1, n = 25$

exakt (Binomial):

$$\begin{aligned} P(X \geq 5) &= \sum_{k=5}^{25} \binom{n}{k} p^k (1-p)^{n-k} \\ &= 1 - P(X \geq 4) = 1 - \text{CDF('Binomial',4,p,n)} \\ &= 0.097994 \end{aligned}$$

approximativ (ZGWS, normal)

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - P\Big(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{4 - np}{\sqrt{np(1-p)}}\Big) \\ &\approx 1 - \Phi\Big(\frac{1.5}{5 \cdot 0.3}\Big) = 1 - \Phi(1) = \Phi(-1) \\ &= \text{CDF('Normal',-1,0,1)} = 0.15866 \end{aligned}$$

$$\begin{aligned} P(X \geq 5) &= 1 - P(X < 5) \\ &\approx 1 - P\Big(\frac{X - np}{\sqrt{np(1-p)}} < \frac{5 - np}{\sqrt{np(1-p)}}\Big) \\ &= 1 - \Phi\Big(\frac{5 - 2.5}{5 \cdot 0.3}\Big) \\ &= 1 - \Phi\Big(\frac{5}{3}\Big) = \Phi\Big(-\frac{5}{3}\Big) = 0.04779 \end{aligned}$$

approximativ (ZGWS, normal, Stetigkeitskorrektur):

$$\begin{aligned} P(X > 5) &\approx 1 - P\Big(\frac{X - np - 0.5}{\sqrt{np(1-p)}} \le \frac{5 - np - 0.5}{\sqrt{np(1-p)}}\Big) \\ &= 1 - \Phi\Big(\frac{4.5 - 2.5}{5 \cdot 0.3}\Big) = 1 - \Phi\Big(\frac{4}{3}\Big) \\ &= \Phi\Big(-\frac{4}{3}\Big) = 0.09121 \end{aligned}$$

Die Approximation der Binomial- durch eine Normalverteilung ist hier nicht so gut (vor allem ist $p = 0.1$ klein).

3b. Hypergeometrische und Binomial Verteilung

Seien

$$f(k|Bi(m,p)) = \binom{m}{k} p^k (1-p)^{m-k} \quad \text{Binomialwkt.}$$

und

$$f(k|H_{N,n,m}) = \frac{\binom{n}{k} \cdot \binom{N-n}{m-k}}{\binom{N}{m}} \quad \text{Hypergeom.wkt.}$$

Satz: Es gilt Für $N \to \infty$, $n \to \infty$, $\frac{n}{N} \to p$ gilt:

$$f(k|H_{N,n,m}) \to \binom{m}{k} p^k (1-p)^{m-k} = f(k|Bi(m,p))$$

Beweis: Es gilt

$$f(k|H_{N,n,m}) = \frac{\binom{n}{k}\cdot\binom{N-n}{m-k}}{\binom{N}{m}}$$

$$= \frac{n!\cdot(N-n)!\cdot m!(N-m)!}{k!(n-k)!\cdot(N-n-m+k)!(m-k)!\cdot N!}$$

$$= \binom{m}{k}\frac{(N-n)!}{(N-n-m+k)!}\frac{n!}{(n-k)!}\frac{(N-m)!}{N!}$$

$$= \binom{m}{k}(N-n-m+k+1)\cdots(N-n)\cdot$$

$$\frac{(n-k+1)\cdots n}{(N-m+1)\cdots N}$$

$$= \binom{m}{k}\frac{(N-n-m+k+1)\cdots(N-n)}{N^{m-k}}\cdot$$

$$\frac{(n-k+1)\cdots n}{N^k}\frac{N^m}{(N-m+1)\cdots N}$$

$$= \binom{m}{k}(1-p-\frac{m-k-1}{N})(1-p-\frac{m-k-2}{N})\cdots(1-$$

$$\frac{(p-\frac{k-1}{N})\cdots p}{(1-\frac{m-1}{N})(1-\frac{m-2}{N})\cdots(1-\frac{1}{N})\cdot 1}$$

$$\to \binom{m}{k}(1-p)^{m-k}p^k = f(k|Bi(m,p))$$

5a. χ^2-Verteilung

Def.: Seien $X_i \sim \mathcal{N}(0,1), i = 1,\ldots,n$, und unabhängig. Dann ist

$$Y = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$$

χ^2-verteilt mit n Freiheitsgraden.

Erwartungswert und Varianz lassen sich leicht ausrechnen:

$$\begin{aligned}
\mathbf{E}(Y) &= n\mathbf{E}(X)^2 = var(X) + (\mathbf{E}(X))^2 = n(1+0) = n \\
\mathrm{var}(Y) &= n \cdot \mathrm{var}(X^2) = n(\mathbf{E}X^4 - (\mathbf{E}(X^2))^2) \\
&= n(3 - 1^2) = 2n
\end{aligned}$$

Anmerkung: Die Dichte von Y ist gegeben durch $(y > 0)$

$$f_{\chi_n^2}(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} e^{-\frac{y}{2}} y^{\frac{n}{2}-1}.$$

5b. Sei $X \sim Exp(\lambda)$. Es gilt:

$$\mathbf{E}X = \lambda \qquad \text{und} \qquad \text{var}X = \lambda^2.$$

Beweis:

$$\begin{aligned}
\mathbf{E}X &= \int_0^\infty x \cdot \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\, dx = \\
&= x(-e^{-\frac{x}{\lambda}})|_0^\infty - \int_0^\infty (-e^{-\frac{x}{\lambda}})\, dx \\
&= \lambda.
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}X^2 &= \int_0^\infty x^2 \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\, dx = \\
&= x^2(-e^{-\frac{x}{\lambda}})|_0^\infty - \int_0^\infty 2x(-e^{-\frac{x}{\lambda}})\, dx \\
&= 2\lambda^2 \\
\text{var}X &= 2\lambda^2 - \lambda^2 = \lambda^2.
\end{aligned}$$

6. Im zweiseitigen Fall ist $H_0 : \mu = \mu_0$. $H_A : \mu \neq \mu_0$,
d.h. der Parameterraum ist $\Theta = \mathbb{R}$.

Im einseitigen Fall gibt es folgende Varianten
$H_0 : \mu \leq \mu_0$ und $H_A : \mu > \mu_0$,
d.h. $\Theta = \mathbb{R}$ (V) oder
$H_0 : \mu = \mu_0$ und $H_A : \mu > \mu_0$,
d.h. $\Theta = \{\mu : \mu \geq \mu_0\}$ (ÜA 12a)

Beide Fälle werden gleich behandelt.
Wenn t die Realisierung von T ist und $t < 0$ dann wird H_0 ohnehin nicht abgelehnt, der p-Wert ist größer als 0.5.

Der Fall
$H_0 : \mu \geq \mu_0$ und $H_A : \mu < \mu_0$ oder
$H_0 : \mu = \mu_0$ und $H_A : \mu < \mu_0$
ist analog.

7. Matrizenmultiplikation ist assoziativ, nicht kommutativ.

Sei $\mathbf{X} = (X_1, \ldots, n)$ der Beobachtungsvektor. Sei $\mathbf{I}$ die Einheitsmatrix und $\mathbf{1}$ die Matrix, die nur aus Einsen besteht. Es gilt (bitte nachrechnen)

$$(n-1)s^2 = \sum_{i=1}^{n} (X_i - \overline{X})^2 = \mathbf{X}^T A \mathbf{X},$$

wobei

$$A = \mathbf{I} - \frac{1}{n}\mathbf{1}$$

Weiterhin gilt: A ist symmetrisch und (bitte nachrechnen)

$$A^2 = A$$

Bem.: Matrizen mit dieser Eigenschaft heißen <u>idempotent</u>. Seien die $X_i \sim \mathcal{N}$ und die Matrix A idempotent. Dann sind Quadratische Formen $\mathbf{X}^T A \mathbf{X} \sim \chi^2_{fg}$ mit $fg = rg(A)$ Freiheitsgraden.

8. Betrachten wir die F-Statistik aus der einfaktoriellen Varianzanalyse

$$F = \frac{MSB}{MSE} = \frac{N-k}{k-1}\frac{SSB}{SSW} = \frac{N-k}{k-1}\frac{\sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{Y})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2}$$

und setzen die Anzahl der Behandlungen auf $k = 2$. Dann wird mit $\overline{Y} = \frac{1}{N}(n_1\overline{Y}_1 + n_2\overline{Y}_2)$, $N = n_1 + n_2$

$$\begin{aligned} F &= \frac{N-2}{2-1}\Big(\frac{n_1(\overline{Y}_1 - \frac{1}{N}(n_1\overline{Y}_1 + n_2\overline{Y}_2))^2}{(n_1-1)s_1^2 + (n_2-1)s_2^2} + \\ &\quad \frac{n_2(\overline{Y}_2 - \frac{1}{N}(n_1\overline{Y}_1 + n_2\overline{Y}_2))^2}{(n_1-1)s_1^2 + (n_2-1)s_2^2}\Big) \end{aligned}$$

Die Zähler in den rechten Brüchen sind zusammen gleich

$$\begin{aligned} &= n_1((1 - \frac{n_1}{N})\overline{Y}_1 - \frac{n_2}{N}\overline{Y}_2)^2 + n_2(-\frac{n_1}{N}\overline{Y}_1 + (1 - \frac{n_2}{N})\overline{Y}_2)^2) \\ &= \frac{n_1 n_2^2}{N^2}(\overline{Y}_1 - \overline{Y}_2)^2 + \frac{n_1^2 n_2}{N^2}(\overline{Y}_1 - \overline{Y}_2)^2 \\ &= \frac{n_1 n_2}{N}(\overline{Y}_1 - \overline{Y}_2)^2 \end{aligned}$$

Also:

$$F = \frac{(N-2)n_1n_2}{N}\frac{(\overline{Y}_1-\overline{Y}_2)^2}{(n_1-1)s_1^2+(n_2-1)s_2^2}$$
$$= \frac{1}{\frac{1}{n_1}+\frac{1}{n_2}}\frac{(\overline{Y}_1-\overline{Y}_2)^2}{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{N-2}}$$

Vergleichen Sie das mit der t-Statistik (gemeinsame Varianzen)

$$T = \frac{\overline{X}_1-\overline{X}_2}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}}$$

und das ist dassselbe wie das Quadrat von T.

Bem.: Wir wissen

$$t = \frac{\mathcal{N}(0,1)}{\sqrt{\frac{1}{n+m-2}\chi^2_{n+m-2}}}$$

(so ist die t-Verteilung definiert)

und es ist

$$t^2 = \frac{\mathcal{N}(0,1)^2}{\frac{1}{n+m-2}\chi^2_{n+m-2}} = \frac{\frac{1}{1}\chi^2_1}{\frac{1}{n+m-2}\chi^2_{n+m-2}} = F_{1,n+m-2}$$

(so ist die F-Verteilung definiert).