last change: oct.13^{th} 2012@18.00

WS12/13
Multivariate Analysis
SolEx1

1) **multinomialKoeffizienten**

$$\binom{n}{n_1,...,n_q} = \binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}...\binom{n-n_1...-n_{q-1}}{n_q}$$

sol:

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}... = \frac{n!}{n_1!(n-n_1)!}\frac{(n-n_1)!}{n_2!(n-n_1-n_2)}... = \frac{n!}{n_1!n_2!...} = \binom{n}{n_1,...,n_q}$$

interpretation: having n balls,

the number of choices for: putting n_1 balls in box$_1$,n_2 balls in box$_2$,...,n_q balls in box$_q$ means:

the number of choices for: putting n_1 balls in box$_1$,n_2 (out of $n-n_1$) balls in box$_2$...,n_q balls out of n_q balls in box$_q$

2) schreiben sie ein trinomialModell und berechnen sie die **marginalen** und **conditionalen** verteilungen

$$\mathbf{N} \sim \text{Mnom}_q(\mathbf{p}, n) \Longleftrightarrow \begin{array}{ll} \mathbb{P}(n_1) = \binom{n}{n_1} p_1^{n_1} (1-p_1)^{n-n_1} & \text{for } q=2 \\ \mathbb{P}(n_1,n_2) = \binom{n}{n_1}\binom{n-n_1}{n_2} p_1^{n_1} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2} & \text{for } q=3 \end{array}$$

marginals: a binomial can be thought of a trinomial in which the 2^{nd} and 3^{rd} boxes are pooled[1] or, formally,

$$\mathbb{P}(n_1) = \sum_{n_2=0}^{n-n_1} \mathbb{P}(n_1,n_2) = \sum_{n_2}^{n-n_1} \binom{n}{n_1} p_1^{n_1} \binom{n-n_1}{n_2} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2}$$
$$= \binom{n}{n_1} p_1^{n_1} \sum_{n_2}^{n-n_1} \binom{n-n_1}{n_2} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2}$$

we show that

$$\sum_{n_2}^{n-n_1} \binom{n-n_1}{n_2} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2} = (1-p_1)^{n-n_1}$$

[1] see script on continTables

this is so because

$$\sum_{n_2}^{n-n_1} \binom{n-n_1}{n_2} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2}$$
$$= \frac{(1-p_1)^{n-n_1}}{(1-p_1)^{n-n_1+(n_2-n_2)}} \sum_{n_2}^{n-n_1} \binom{n-n_1}{n_2} p_2^{n_2} (1-p_1-p_2)^{n-n_1-n_2}$$
$$= (1-p_1)^{n-n_1} \sum_{n_2}^{n-n_1} \binom{n-n_1}{n_2} \left(\frac{p_2}{1-p_1}\right)^{n_2} \left(1-\frac{p_2}{1-p_1}\right)^{n-n_1-n_2} = (1-p_1)^{n-n_1}$$

next, compute **conditionals:**

$$\text{given } (N_1, N_2, N_3) \sim \text{Mnom}(p_1, p_2, p_3, n) \ \ \text{find } \mathbb{P}(\, n_2, n_3 | \ n_1) = \frac{\mathbb{P}(n_1, n_2, n_3)}{\mathbb{P}(n_1)}$$

know:

$$(n_1, n_2, n_3) \sim \text{Mnom}(p_1, p_2, p_3, n) \Longrightarrow \mathbb{P}\,(n_1) \sim \text{Bnom}(p_1, n)$$
$$\text{Bnom}(p_1, n) = \text{Mnom}(p_1, 1-p_1, n) = \text{Mnom}(p_1, p_2+p_3, n) \Longrightarrow$$

$\Longrightarrow$

$$\mathbb{P}(\, n_2, n_3 | \ n_1) = \frac{\mathbb{P}(n_1, n_2, n_3)}{\mathbb{P}(n_1, n_2+n_3)}$$
$$= \frac{\binom{n}{n_1}\binom{n-n_1}{n_2}\overbrace{\binom{n-n_1-n_2}{n_3}}^{=1} p_1^{n_1} p_2^{n_2} p_3^{n_3}}{\binom{n}{n_1}\binom{n-n_1}{n_2+n_3} p_1^{n_1} (p_2+p_3)^{n_2+n_3}}$$
$$= \frac{\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3} p_2^{n_2} p_3^{n_3}}{\underbrace{\binom{n-n_1}{n_2+n_3}}_{=1} (p_2+p_3)^{n_2+n_3}}$$
$$= \binom{n-n_1}{n_2} \overbrace{\binom{n-n_1-n_2}{n_3}}^{=n_3} \left(\frac{p_2}{p_2+p_3}\right)^{n_2} \left(\frac{p_3}{p_2+p_3}\right)^{n_3}$$
$$\Longrightarrow \text{Mnom}\left(\frac{p_2}{p_2+p_3}, \ \frac{p_3}{p_2+p_3}, n-n_1\right)$$

correlation between multinomial variables:

$$\mathbb{P}\,(X_1 = n_1, \ldots, X_k = n_k) = \binom{n}{n_1, \ldots, n_k} p_1^{n_1} \cdots p_k^{n_k}$$

indicator variables

$$I_{ij} = \begin{cases} 1 & \text{if } i^{th} \text{ ball hits } j^{th} \text{ box} \\ 0 & \text{else} \end{cases}$$

$\Longrightarrow$

$$\mathbb{E}X_j = \mathbb{E}\sum_i^n I_{ij}$$

also

$$\mathbb{E}I_{ij} = p_j$$

$$\mathbb{E}I_{i_1j_1}I_{i_2j_2} = \begin{cases} p_j & \text{if } i_1 = i_2 = i, \quad j_1 = j_2 = j \\ p_j^2 & \text{if } i_1 \neq i_2, \quad j_1 = j_2 = j \\ 0 & \text{if } i_1 = i_2 = i, \ j_1 \neq j_2 \\ p_{j_1}p_{j_2} & \text{if } i_1 \neq i_2, \ j_1 \neq j_2 \end{cases}$$

$\Longrightarrow$

$$\mathbb{C}\left(I_{i_1j_1}, I_{i_2j_2}\right) = \mathbb{E}I_{i_1j_1}I_{i_2j_2} - \mathbb{E}I_{i_1j_1}\mathbb{E}I_{i_2j_2} = \begin{cases} 0 & \text{if } i_1 \neq i_2 \\ p_j - p_j^2 & \text{if } i_1 = i_2 = i, \ j_1 = j_2 = j \\ -p_{j_1}p_{j_2} & \text{if } i_1 = i_2 = i, \ j_1 \neq j_2 \end{cases}$$

$$\begin{aligned} \mathbb{C}\left(X_{j_1}, X_{j_2}\right) &= \mathbb{C}\left(\sum_{i_1}^n I_{i_1j_1}, \sum_{i_2}^n I_{i_2j_2}\right) = \mathbb{C}\left(\sum_{i_1}^n \sum_{i_2}^n I_{i_1j_1}, I_{i_2j_2}\right) \\ &= \sum_{i_1}^n \sum_{i_2}^n \mathbb{C}\left(I_{i_1j_1}, I_{i_2j_2}\right) = \sum\sum_{i_1=i_2} + \underbrace{\sum\sum_{i_1\neq i_2}^n}_{=0} = \sum_i^n \mathbb{C}\left(I_{ij_1}, I_{ij_2}\right) \\ &= \begin{cases} np_j\left(1-p_j\right) & \text{if } \ j_1 = j_2 = j \\ -np_{j_1}p_{j_2} & \text{if } \ \ j_1 \neq j_2 \end{cases} \end{aligned}$$

note:

$$\sum = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ & \sigma_{22} & & \vdots \\ \vdots & & \ddots & \\ \sigma_{k1} & \ldots & & \sigma_{kk} \end{bmatrix}$$

has rowsums zero and is singular.